NOTICE

AVIS

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

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

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

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

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

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

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

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

Canada

A Yes/No Vocabulary Test in A University Placement Setting

Tamara Loring

A Thesis

in

The TESL Centre

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Arts at
Concordia University
Montreal, Quebec, Canada

March 1995

THE AUTHOR HAS GRANTED AN
IRREVOCABLE NON-EXCLUSIVE
LICENCE ALLOWING THE NATIONAL
LIBRARY OF CANADA TO
REPRODUCE, LOAN, DISTRIBUTE OR
SELL COPIES OF HIS/HER THESIS BY
ANY MEANS AND IN ANY FORM OR
FORMAT, MAKING THIS THESIS
AVAILABLE TO INTERESTED
PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE
IRREVOCABLE ET NON EXCLUSIVE
PERMETTANT A LA BIBLIOTHEQUE
NATIONALE DU CANADA DE
REPRODUIRE, PRETER, DISTRIBUER
OU VENDRE DES COPIES DE SA
THESE DE QUELQUE MANIERE ET
SOUS QUELQUE FORME QUE CE SOIT
POUR METTRE DES EXEMPLAIRES DE
CETTE THESE A LA DISPOSITION DES
PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP
OF THE COPYRIGHT IN HIS/HER
THESIS. NEITHER THE THESIS NOR
SUBSTANTIAL EXTRACTS FROM IT
MAY BE PRINTED OR OTHERWISE
REPRODUCED WITHOUT HIS/HER
PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE
DU DROIT D'AUTEUR QUI PROTEGE
SA THESE. NI LA THESE NI DES
EXTRAITS SUBSTANTIELS DE CELLE-
CI NE DOIVENT ETRE IMPRIMES OU
AUTREMENT REPRODUITS SANS SON
AUTORISATION.

ISBN   0-612-01289-1

Canada

Name _**TAMARA LORING**_

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

**LINGUISTICS**   SUBJECT TERM   `0 2 9 0` **U·M·I**

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS
Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION
General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0688
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0680
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS
Language
    General 0679
    Ancient 0289
    Linguistics 0290
    Modern 0291
Literature
    General 0401
    Classical 0294
    Comparative 0295
    Medieval 0297
    Modern 0298
    African 0316
    American 0591
    Asian 0305
    Canadian (English) 0352
    Canadian (French) 0355
    English 0593
    Germanic 0311
    Latin American 0312
    Middle Eastern 0315
    Romance 0313
    Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY
Philosophy 0422
Religion
    General 0318
    Biblical Studies 0321
    Clergy 0319
    History of 0320
    Philosophy of 0322
Theology 0469

SOCIAL SCIENCES
American Studies 0323
Anthropology
    Archaeology 0324
    Cultural 0326
    Physical 0327
Business Administration
    General 0310
    Accounting 0272
    Banking 0770
    Management 0454
    Marketing 0338
Canadian Studies 0385
Economics
    General 0501
    Agricultural 0503
    Commerce-Business 0505
    Finance 0508
    History 0509
    Labor 0510
    Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
    General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0333
United States 0337
History of Science 0585
Law 0398
Political Science
    General 0615
    International Law and
        Relations 0616
    Public Administration 0617
Recreation 0814
Social Work 0452
Sociology
    General 0626
    Criminology and Penology ... 0627
    Demography 0938
    Ethnic and Racial Studies 0631
    Individual and Family
        Studies 0628
    Industrial and Labor
        Relations 0629
    Public and Social Welfare 0630
    Social Structure and
        Development 0700
    Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES
Agriculture
    General 0473
    Agronomy 0285
    Animal Culture and
        Nutrition 0475
    Animal Pathology 0476
    Food Science and
        Technology 0359
    Forestry and Wildlife 0478
    Plant Culture 0479
    Plant Pathology 0480
    Plant Physiology 0817
    Range Management 0777
    Wood Technology 0746
Biology
    General 0306
    Anatomy 0287
    Biostatistics 0308
    Botany 0309
    Cell 0379
    Ecology 0329
    Entomology 0353
    Genetics 0369
    Limnology 0793
    Microbiology 0410
    Molecular 0307
    Neuroscience 0317
    Oceanography 0416
    Physiology 0433
    Radiation 0821
    Veterinary Science 0778
    Zoology 0472
Biophysics
    General 0786
    Medical 0760

EARTH SCIENCES
Biogeochemistry 0425
Geochemistry 0996

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleoecology 0426
Paleontology 0418
Paleozoology 0985
Palynology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES
Environmental Sciences 0768
Health Sciences
    General 0566
    Audiology 0300
    Chemotherapy 0992
    Dentistry 0567
    Education 0350
    Hospital Management 0769
    Human Development 0758
    Immunology 0982
    Medicine and Surgery 0564
    Mental Health 0347
    Nursing 0569
    Nutrition 0570
    Obstetrics and Gynecology .. 0380
    Occupational Health and
        Therapy 0354
    Ophthalmology 0381
    Pathology 0571
    Pharmacology 0419
    Pharmacy 0572
    Physical Therapy 0382
    Public Health 0573
    Radiology 0574
    Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES
Pure Sciences
Chemistry
    General 0485
    Agricultural 0749
    Analytical 0486
    Biochemistry 0487
    Inorganic 0488
    Nuclear 0738
    Organic 0490
    Pharmaceutical 0491
    Physical 0494
    Polymer 0495
    Radiation 0754
Mathematics 0405
Physics
    General 0605
    Acoustics 0986
    Astronomy and
        Astrophysics 0606
    Atmospheric Science 0608
    Atomic 0748
    Electronics and Electricity 0607
    Elementary Particles and
        High Energy 0798
    Fluid and Plasma 0759
    Molecular 0609
    Nuclear 0610
    Optics 0752
    Radiation 0756
    Solid State 0611
Statistics 0463

Applied Sciences
Applied Mechanics 0346
Computer Science 0984

Engineering
    General 0537
    Aerospace 0538
    Agricultural 0539
    Automotive 0540
    Biomedical 0541
    Chemical 0542
    Civil 0543
    Electronics and Electrical 0544
    Heat and Thermodynamics ...0348
    Hydraulic 0545
    Industrial 0546
    Marine 0547
    Materials Science 0794
    Mechanical 0548
    Metallurgy 0743
    Mining 0551
    Nuclear 0552
    Packaging 0549
    Petroleum 0765
    Sanitary and Municipal 0554
    System Science 0790
Geotechnology 0428
Operations Research 0796
Plastics Technology 0795
Textile Technology 0994

PSYCHOLOGY
General 0621
Behavioral 0384
Clinical 0622
Developmental 0620
Experimental 0623
Industrial 0624
Personality 0625
Physiological 0989
Psychobiology 0349
Psychometrics 0632
Social 0451

♲

ABSTRACT

A Yes/No Vocabulary Test in a University Placement Setting

Tamara Loring

Receptive vocabulary knowledge appears to reflect broader receptive L1 and L2

language skills such as reading and listening comprehension. The Yes/No vocabulary

test, developed to provide a quick estimate of total receptive vocabulary size, requires

subjects to check off known words, leaving unknown words blank. Scoring formulae

based on signal detection theory offset the effects of guessing. Yes/No type tests, already

administered to ESL learners from several language backgrounds, are still in need of

validation. This study used a Yes/No test based on a general word list and a Yes/No test

based on a university-entrance level word list. The results of these two instruments were

measured against the results of an established North American university ESL placement

test (The Michigan Test of English Language Proficiency) and its vocabulary subtest.

While for most of this study's subjects the results of the more "academic" Yes/No

test correlated more closely with the overall Michigan Test and its vocabulary subtest,

two language subgroups produced odd results. Most of the Arabic-speaking subjects

obtained unusable scores, perhaps due to a high rate of guessing. The French-speaking

subjects, on the other hand, appeared to approach both Yes/No tests with caution, and

their results on the overall Michigan Test and its vocabulary subtest correlated more

closely with the less "academic" of the two Yes/No tests.

Acknowledgments

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

Theoretical Background

Vocabulary Knowledge and First Language Learning

Researchers during this century have increasingly recognized the importance of vocabulary knowledge to intelligence in general and to the learning of one's first language (L1). Anderson & Freebody (1981) cite various studies from Terman (1918) to the Stanford Achievement Tests (1973) showing a strong relationship between vocabulary knowledge and general intelligence. They also discuss R.L. Thorndike's 1973 fifteen-nation study revealing a very strong link between vocabulary knowledge and reading comprehension ability. Recently Meara (1994) found that one type of test of vocabulary size, the Yes/No test, correlated moderately well not only with reading comprehension tests but also with tests of listening comprehension and grammatical accuracy.

In their 1981 paper, Anderson & Freebody reviewed three different theories which have been used to explain the importance of vocabulary to linguistic ability. The instrumentalist position holds that simply knowing more individual words is the direct cause of better text comprehension. The aptitude school credits a learner's high vocabulary ability to mental agility. In other words, of two learners having the same amount of exposure to the culture, the more mentally agile learner will pick up more word meanings than his/her less able partner. The third, or knowledge position maintains that high scores on vocabulary tests are the result of deeper and broader knowledge of

1

the target culture. The knowledge position emphasizes conceptual frameworks, "icebergs" whose "tips" consist of individual words (Anderson & Freebody 1981, p. 82).

It can be seen that although several points of view exist on the nature of vocabulary knowledge, there appears to be agreement on its importance to L1 linguistic ability. Anderson and Freebody (1981) move on to suggest that if better ways of assessing L1 vocabulary size can be found, a "critical first step" will be taken towards understanding how word knowledge grows throughout the life span (p. 110).

Vocabulary Knowledge and Second Language (L2) Learning

Vocabulary knowledge is also seen as a key to language learning success for non-native speakers of English. Goulden, Nation and Read (1990) have explored ways to narrow the gap between language learners' L2 English vocabulary knowledge and that of their native-speaking peers. To do this, it seems especially important to help learners make the shift from a simpler, everyday vocabulary to the more complex and cognitively more demanding language of academic study, as Cummins (1981) has suggested:

> ...some aspects of language proficiency, such as reading skills, are strongly related to cognitive and academic development, whereas others involving such basic interpersonal communicative skills as oral fluency and phonology, are less related to cognitive and academic development (Cummins 1981, p.132).

Cummins has named the latter basic interpersonal communicative skills (BICS) and the former cognitive/academic language proficiency (CALP), emphasizing the different demands made by BICS and CALP on mental processing space. Corson (1983) has

described a gap between two types of language ability from a different perspective. He identifies a "lexical bar" separating the ability to use "very frequent, largely monosyllabic, mainly Anglo-Saxon vocabulary" from the ability to handle the "Graeco-Latin vocabulary of the English of academic study" (in Goulden, Nation & Read 1990, p. 342). Corson is concerned that speakers of certain peripheral dialect groups are prevented by this semantic barrier from attaining academic goals more readily available to speakers of more central dialects (Corson 1983).

A growing recognition that vocabulary knowledge is a key to broader areas of first and second language competence has led to calls for further exploration of a "functionally important personal lexicon" and empirical work on the feasibility of large-scale vocabulary learning among adults (Zechmeister, D'Anna, Hall, Paus & Smith 1992, p. 205). Goulden, Nation and Read (1990) estimate that adult L1 speakers acquire new vocabulary at an average rate of only two to three words per day. They suggest that direct teaching of vocabulary using a variety of techniques, "combined with large amounts of extensive reading accompanied by practice in guessing words from context, would allow second language learners to develop their vocabulary at a rate well above that of most native speakers" (p.356). The result would be, of course, a narrowing of the L2-L1 vocabulary knowledge gap.

Describing L2 Vocabulary Knowledge

Dichotomies: Receptive/productive, Active/passive

In L2 vocabulary research, it is common to find the terms *receptive* and

productive vocabulary used to approximate the meanings of comprehension and

production (Ringbom 1987, Palmberg 1987 & 1988, Nation 1990). For example, Nation

(1990) feels that native speakers possess larger receptive than productive vocabulary

knowledge. He describes L2 productive learning as not a subset of receptive knowledge

but an extension of it. It "involves what is needed for receptive learning plus the anbility

to speak or write needed vocabulary at the appropriate time" (Nation 1990, p. 5). He is

concerned with narrowing the resulting gap in vocabulary knowledge between L2 learners

and native speakers. For Nation, receptive knowledge of a word covers aural or visual

recognition, an expectation of the grammatical pattern in which it will occur, its

collocation, frequency, meanings and associations.

There are authors who refer to *passive* and *active* vocabulary/knowledge as

synonymous with receptive and productive vocabulary/knowledge (for L1, see

Zechmeister, D'Anna, Hall, Paus & Smith 1993). Levenston (1979) compares L2

passive/receptive vocabulary (words one knows) to active/productive vocabulary (words

one is able to use). Two other categories enter the picture. There is a set of words which

are not in active service because: a) a learner cannot call them to mind without actually

seeing or hearing them. Meara (1990) calls this *passive* vocabulary; or b) they are words

in reserve, waiting to be accessed by the learner when motivated by the right context.

Melka Teichroew (1982) has called this *possible use* knowledge; Palmberg (1988) calls

it *potential* vocabulary. The term "passive", then, can be confusing. Palmberg (1988)

credits Berman, Buchbinder and Beznedeznych (1968) with coining the terms *potential*

and *real* to distinguish between words that lie dormant for whatever reason and words

that are part of a learner's normal discourse. Real vocabulary may then be subdivided

into *passive real* and *active real*. One more term, *threshold*, will be discussed below.

A New Dimension: the Continuum

Vocabulary knowledge and/or control is often seen as an open-ended continuum

(Levenston 1979, Melka Teichroew 1982, Palmberg 1988) or series of continua

(Ringbom 1987). The idea that a learner gradually progresses from zero vocabulary

knowledge to complex, sophisticated vocabulary knowledge and the ability to retrieve it

has intuitive appeal. Continua are usually described as ranging from passive to active

(Levenston 1979) or from receptive to productive (Palmberg 1988). Some researchers see

the beginning of the continuum as the point where a learner first recognizes a word

(Melka-Teichroew (1982) calls this point the threshold). For Meara (1990), passive

vocabulary is an associational network that is qualitatively different from the active

continuum which begins where the learner starts accessing words without the aid of

external stimuli. The vocabulary continuum is considered to be open-ended because

humans generally add to their store of vocabulary knowledge (size and depth) all their

lives. Somewhere along the continuum each word will cross over from receptive

knowledge to productive use. Levenston (1979) adds a category he calls *threshold*

vocabulary, containing words that are "sometimes available, sometimes not" (p. 154).

Ringbom (1987), writing of lexical knowledge as varying through a number of different dimensions, prefers to see the knowledge component not as one progression but as a series of them: "Various continua can be drawn up, ranging from no knowledge at all to (theoretical) full knowledge" (p. 36). Figure 1 shows Ringbom's six continua for describing the basics of vocabulary knowledge.

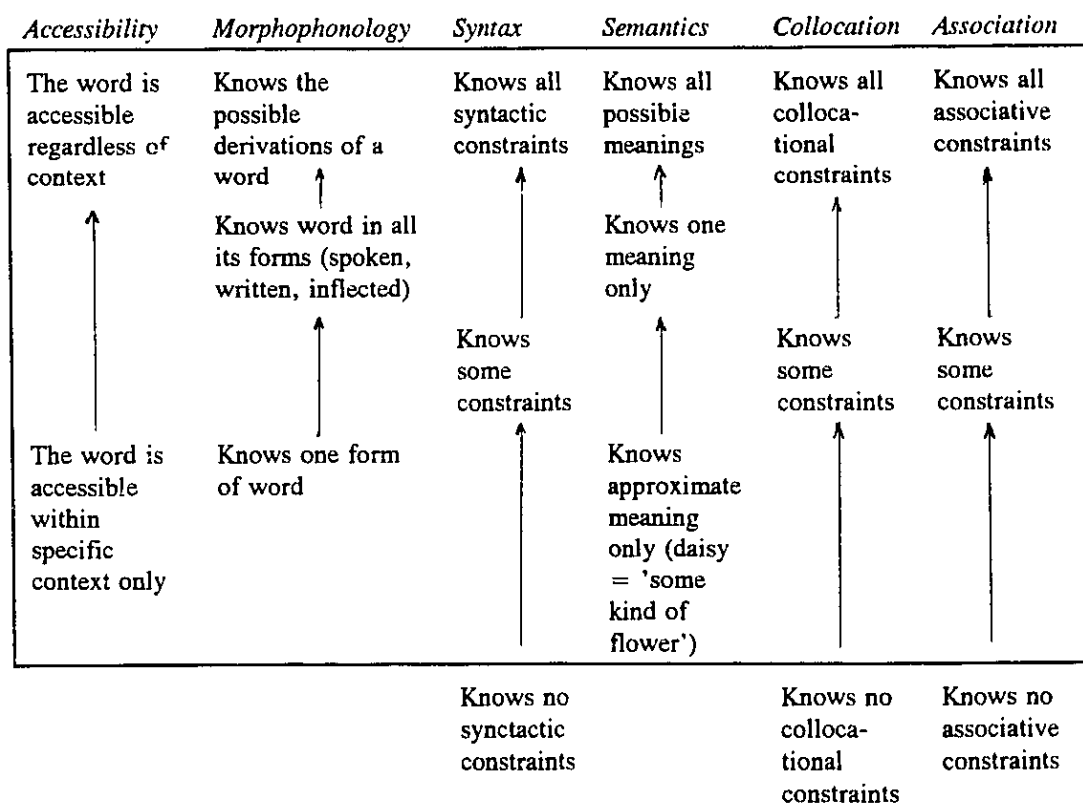| *Accessibility*	*Morphophonology*	*Syntax*	*Semantics*	*Collocation*	*Association*
The word is accessible regardless of context	Knows the possible derivations of a word ↑	Knows all syntactic constraints ↑	Knows all possible meanings ↑	Knows all collocational constraints ↑	Knows all associative constraints ↑
↑	Knows word in all its forms (spoken, written, inflected) ↑		Knows one meaning only ↑		
		Knows some constraints ↑		Knows some constraints ↑	Knows some constraints ↑
The word is accessible within specific context only	Knows one form of word		Knows approximate meaning only (daisy = 'some kind of flower')		
		Knows no synctactic constraints		Knows no collocational constraints	Knows no associative constraints

Figure 1. Lexical knowledge continua (adapted from Ringbom 1987, p. 37).

Although Ringbom is careful to specify that his continua refer only to lexical knowledge, the concept of lexical control (access and retrieval) is conceptually very close: each time a learner either acquires/forgets a meaning for a word, that word will be likely to shift to a different category, and this may well affect its accessibility.

Figure 2 is this researcher's attempt to schematize the various views put forward by those favoring the idea of a continuum. Although not represented in Figure 2, the dissenting viewpoint of Meara (1990) should be kept in mind: although he agrees that active vocabulary may be seen as a continuum, he suspects that passive vocabulary may be very different, and not part of a continuum in any sense.

passive   active

```
|            |  receptive————————————————————  ——————>
|            |  (increasing knowledge →)
|            |                        |-------------------------------------|
|            |                        |  productive—————————>
|            |                        |  (increasing control →)
```
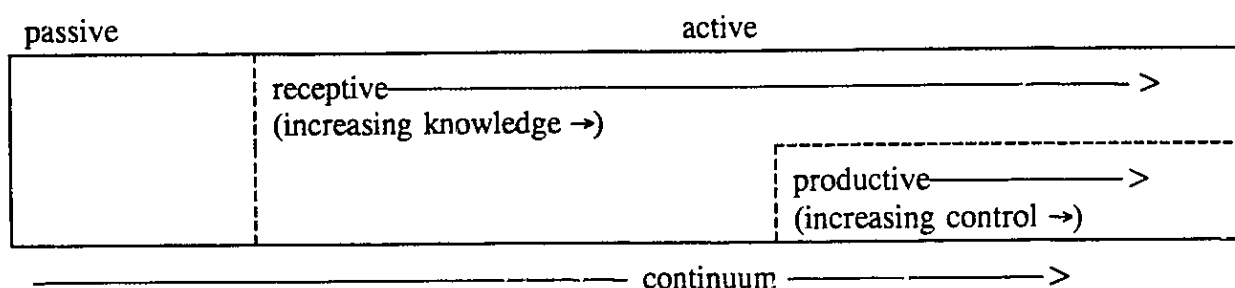
———————————————————————— continuum ——————————————>

Figure 2. L2 vocabulary knowledge seen as a continuum.

Note. 1) Passive vocabulary is inaccessible by the learner without external stimulus. It is separated from the active continuum by a bidirectional permeable threshold, which the model represents as a broken line. Passive vocabulary has also been called potential, or possible use vocabulary. Words vacillating between active and passive knowledge would belong to a learner's threshold vocabulary.

2) A second bidirectional permeable threshold (also shown here by a broken line) separates receptive from productive vocabulary.

3) Words begin their active careers receptively. Producing a word usually presumes receptive knowledge on the part of the learner.

4) A learner's receptive knowledge of a word may reach a highly sophisticated level yet may remain receptive. Productive ability (control) is qualitatively different from receptive knowledge (knowledge).

7

From Theory to Practice

What Aspects of L2 Vocabulary Knowledge Should We Test?

It would be ideal to be able to evaluate all the aspects of first or second language vocabulary knowledge, or at least some of the most important of them, such as:

- the ability to recognize and understand words (referred to in this study as *receptive* knowledge);

- the ability to retrieve words for active use (referred to in this paper as *productive* knowledge);

- the number of words a learner knows (vocabulary *size*);

- the extent to which each word is known (*depth* of vocabulary knowledge).

Realistically, however, researchers have not yet succeeded in delimiting these categories accurately, and still less, deciding how to test for them. The preceding discussion has demonstrated some of the ways in which researchers view receptive and productive knowledge. They agree that receptive vocabulary is larger than its subset, productive vocabulary. If we wish to evaluate vocabulary knowledge, it seems advisable to follow Melka-Teichroew's advice and to assess the largest set of words a learner knows: his or her receptive vocabulary. To do this, we may attempt to measure vocabulary size and the depth of knowledge of each word. Before reviewing the pitfalls and advantages inherent in assessing either size or depth, some further terminological clarification may be useful.

Most vocabulary research, whether in L1 or L2, investigates the number of words a learner knows and the quality of that knowledge. Most researchers refer to the *size* of a learner's vocabulary (some examples are Anderson & Freebody 1981, 1983, and D'Anna, Zechmeister & Hall 1991 for L1, Meara & Buxton 1987, Goulden, Nation & Read 1990 for L2). The quality of knowledge for each word is usually referred to as vocabulary *depth* (some examples are Anderson & Freebody 1981 for L1, Read 1993, Wesche & Paribakht 1993 for L2). A third term, vocabulary *breadth*, has been used to signify either size (Anderson & Freebody 1981 and Read 1993) or depth (Nation 1990). In this study the term "breadth" has been avoided. Reference is instead made to vocabulary size and depth of vocabulary knowledge.

Evaluating Vocabulary Depth

Vocabulary depth will be difficult to evaluate until agreement is reached on what it means to "know" a word. Anderson & Freebody (1981) describe vocabulary depth as knowing the necessary or essential features of words as opposed to features which are "merely characterisic" (p. 90). They use the following example: *having a back* is seen as a necessary feature of the word "chair". Without a back, this piece of furniture is considered not a chair, but a stool. A feature which is characteristic, but not essential, is *ability to fly* for "bird", since not all birds can fly (the penguin is given as an example). Extent of knowledge of a word, or familiarity, has also been described by Melka Teichroew (1982) as a continuum starting with simple recognition and moving through a "possible use" stage, then progressing by degrees towards full productive

9

knowledge. A three-stage version of the continuum is posited by Palmberg (1988) for the lexicon of second language learners: words move from potential vocabulary through receptive vocabulary to productive vocabulary. All boundaries between these three parts of the continuum are constantly changing as new words enter the lexicon. Anderson & Freebody (1981) cite Clark's (1973) assertion that depth of L1 vocabulary knowledge may be different for younger versus older learners, a position similar to that taken by Melka Teichroew (1982) and Wesche & Paribakht (personal communication, 1993).

Evaluating Vocabulary Size

Anderson & Freebody (1981) define vocabulary size as "the number of words for which the person knows at least some of the significant aspects of meaning"(p. 93). The fact that this definition embraces word recognition, Melka-Teichroew's starting point on the word familiarity continuum, is evident in their 1983 study, where they developed and administered a checklist test on which L1 subjects checked off words they felt they knew. In his 1985 study of Finnish EFL learners' active and passive vocabulary knowledge, Takala affirmed the importance of knowing even part of an L2 word's meaning:

> It can be conjectured that partial knowledge of a fair amount of basic words combined with some knowledge of basic morphological rules and the availability of an adequate context can lead to an adequate comprehension of text passages and to provide [sic] a good opportunity for more word learning" (Takala 1985, p. 162).

Takala found important differences in number of words known (both actively and passively) by EFL students who had been placed by their schools into classes (or "sets") of "fast, average, and slow" learners (p. 161). He called for further research into "when the observed large differences in vocabulary size in L2 emerge, and whether setting/streaming (and using different textbooks with different input) tends to increase or decrease such differences" (p. 163).

Research has shown that the size of an ESL learner's receptive vocabulary may predict some of that learner's broader receptive L2 skills. Once the decision has been made to test receptive vocabulary size, however, there are important factors to consider before the evaluation can begin. The following chapter will discuss some of these issues, such as how to decide what to count as words, how to choose which words to test, and how to test them.

CHAPTER 2: REVIEW OF THE LITERATURE

The Measurement of Vocabulary Size

Measuring vocabulary size is no simple matter. Estimates of vocabulary size have varied widely, mainly because individual researchers have not agreed on some basic issues, such as: How do we decide what to count as words? How do we choose which words to test? How do we test the chosen words? (Goulden, Nation & Read 1990).

How Do We Decide What to Count as Words?

Goulden, Nation and Read (1990) reviewed several attempts made since the early part of this century to evaluate receptive vocabulary size and attempted to set up a rational system for testing it. Using the largest non-technical dictionary available in English, Webster's Third New International Dictionary, they used criteria developed by Nagy & Anderson in 1984 to count "base words". These were defined as all unrelated dictionary entries except derived and proper words, compound words, and "others" (entries such as alternative spellings, archaic words and dialect words). Of a group of related words, they treated as base word the least inflected form of the group. Using these criteria, Goulden, Nation and Read administered a refined version of a test previously administered by Diack (1975) to a small group of English native speakers. The results led them to conclude that "well-educated adult native speakers of English" have vocabularies of approximately 17,000 base words. Zechmeister, D'Anna, Hall, Paus and Smith (1993) agree with Goulden, Nation & Read 1990 that estimates such as this,

based on wo.d families rather than on related individual words, were likely to be more trustworthy underpinnings for tests of Ll English vocabulary.

How Do We Choose Which Words to Test?

There are at least three factors to consider in choosing words to include in a test of receptive L2 vocabulary size: word frequency, cognacy, and end use. Traditionally, tests of vocabulary size have been constructed using words taken from frequency word lists. Until fairly recently, frequency was seen as sufficient for ESL purposes, as in Harlech-Jones' 1983 administration of a test based on the General Service List (West 1953). The Ll-L2 cognates present in a vocabulary size test may also affect results. (Al-Hazemi 1993, Meara, Lightbown & Halter 1994). The effects of such bias can be attenuated, however, by ensuring that the number of cognate items a test contains corresponds to the percentage of cognates occurring naturally in the language being tested. (Chen & Henning 1985, Meara, Lightbown & Halter 1994). Finally, it is essential to consider end use: in what context will the subject need to use the L2 vocabulary later? For example, a person entering university will need, in addition to a good basic grasp of core vocabulary, knowledge of terms that are qualitatively different from those of a person mainly interested in improving everyday interactive social skills.

How Do We Test the Chosen Words?

Several tests have aimed at measuring vocabulary size. Some succeed in measuring a degree of depth as well. The most frequently used vocabulary test has been the multiple choice format. Others to be discussed below include spew tests, lexical recognition tasks, and checklist tests, also known as Yes/No tests. A form of Yes/No test which incorporates safeguards against indiscriminate guessing is the focus of this study.

Multiple-choice tests

A typical item on this widely used type of test of receptive vocabulary knowledge consists of a phrase or sentence lacking one word and three to five options, of which one is correct and the others are distractors. The subject chooses the correct option and checks it off. Multiple-choice tests have been comparatively well tested for reliability, and are relatively easy to score. They have several disadvantages, however, for assessment of vocabulary size. A multiple-choice item measures more than its targeted word. A subject may obtain a correct answer on an item by guessing or by knowing the meaning of one or several of the words contained in the stem and/or the distractors and answering by process of elimination. To keep multiple-choice tests manageable in terms of time and length, only relatively small samples of a subject's total vocabulary may be tested at any one time. This may make multiple-choice tests useful predictors of overall vocabulary size only for extremely low proficiency levels, where vocabulary size is still minimal (Anderson & Freebody 1983).

Spew tests

A spew test is by nature a measure of the size of a subject's productive vocabulary. Within a specified time limit (usually one to two minutes), the subject, given a specific point of departure, must produce (orally or in writing) as many words as possible. It may be stipulated, for example, that all of the words produced begin with a particular letter. When the time is up, all words produced are tallied.

Although not designed to assess receptive vocabulary, the spew test has the advantage of allowing for a wide range of appropriate responses, and thus, according to Meara (1994) does not obviously penalize subjects who have specialized vocabularies. "Unlike most tests, (...) the spew test makes very few assumptions about the kinds of words people ought to know." (Meara 1994, p. 16). Spew tests appear to be useful in exploring particular aspects of vocabulary knowledge, such as lexical knowledge (the way in which lexical structures are represented or stored in the mental lexicon) and lexical control (learners' ability to access and retrieve individual words) (Sharwood-Smith 1984).

The spew test has disadvantages. Meara and colleagues found very low correlations between spew test totals and other measures of vocabulary size (Meara 1994). Both Meara (1994) and Palmberg (1987) found that the number of words produced was affected by individual differences such as competitiveness vs. minimal effort. It was difficult to establish target performance criteria, at least for tests giving specific letters as points of departure, since the number of words per initial letter varies from letter to letter and from language to language. It was also very difficult to score the responses. Meara (1994) gives these examples: should *be* count as much as *bicycle* or

bitumen? Also, what should be done with *become, became, becomes, becoming*? Confusion between letters (for example, *p* and *b*) occurred in both Meara's and Palmberg's tests, resulting in the rejection of data. The spew test may also be measuring something other than vocabulary growth manifested via spot checks on vocabulary size. At least in its written form, it may be evaluating "writing speed, lateral thinking or associative looseness" (Palmberg 1987, p. 206). He was also concerned about a possible ceiling effect, as some of his subjects ran out of words before the one-minute time limit was up. And, although native speakers produced more words than non-native speakers on Palmberg's test, the number of words they produced could not simply be multiplied in order to estimate overall vocabulary size. Palmberg felt that "little, if anything, can be concluded about the actual size of [the subjects'] productive English vocabulary" (Palmberg 1987, p. 215).

Tests of Word Recognition Speed (Lexical Decision Tasks)

Meara (1994) describes the administration of a lexical decision task in which subjects were asked to decide whether a given set of letters was or was not a word they knew. The time it took them to do this was measured. This test had the disadvantage of requiring highly sophisticated measuring equipment to distinguish the tiny differences between native-speaker and non-native speaker performance. In addition, Meara felt that the test did not really measure what it set out to measure: Meara and his student researchers "consistently found, for instance, that English children learning Spanish very often recognised Spanish words considerably faster than native speakers of Spanish did"

16

(Meara 1994, p. 17). This finding is surprising, and perhaps in need of verification, in light of previous research indicating that even fluent L2 learners tend to read more slowly in their L2 than in their L1, perhaps reflecting slow retrieval of L2 word meaning (Favreau & Segalowitz 1982).

In summary, multiple-choice tests can only adequately measure small vocabulary size. In addition, they involve testwise behavior and knowledge of more than the targeted words on the part of the test-taker. Spew tests are designed to test only the "productive" subset of a person's total word store. As measures of productive vocabulary, they may benefit from a change in orientation. Meara (1994) suggests that assessing the rate of word production rather than the number of words produced might yield more satisfying results. The lexical decision task, though apparently straightforward and quick, requires sophisticated equipment to detect differences in subjects' response times. One type of receptive vocabulary test, the checklist or Yes/No test, is easy to administer and simple for subjects to take. How it works, and its principle advantages and disadvantages, are described below.

Yes/No tests

What is a Yes/No Test?

The Yes/No test has been described as a variant of the lexical decision task (Anderson & Freebody 1983). Basically, it consists of a checklist on which subjects are asked to identify, usually by means of a check mark, those words they feel they know.

Early attempts at this type of testing, done with native speakers, simply presented lists of authentic English words to learners and asked them to tick off the ones for which they knew at least one meaning (Diack 1975, in Goulden, Nation & Read 1990). The validity of Diack's results is questionable. It has since been established that Diack's built-in verification for knowledge of word meaning did not focus on words which were neither easily known nor totally unknown, and about which the subjects may have had the most doubts (Goulden, Nation & Read 1990). It is therefore difficult to know to what extent Diack's subjects were overestimating their own abilities.

A number of more recent attempts at using the checklist, or "yes/no" format have resulted in refinements. It has been found, for example, that although individual Yes/No checklist tests at any one level are not particularly reliable, test reliability can be considerably enhanced by administering a minimum of two, and if possible, three or more individual tests at any one level, and then treating the average of the scores as one test result (Meara 1992). Another refinement has focused on ways to control for guessing on the part of the subjects. Signal detection theory has so far provided the most robust method for minimizing the effects of guessing. Its general principles and its relevance to the Yes/No test will be described in Chapter 3.

The Development of Yes/No Tests with Control for Guessing

Zimmerman, Broder, Shaughnessy & Underwood (1977) made the first attempt to control for guessing on a Yes/No test they administered to English native speaking adults. On a 60-word test, 40 of the words were authentic and 20 "words" were actually

18

plausible non-words, or distractors. Subjects were asked to rate each word according to the degree to which they felt they knew the word. Each distractor a subject claimed to "know" was scored in such a way as to lower the overall score obtained by that subject. Although this method (based on signal detection theory, described in Chapter 3) seemed to deal fairly successfully with guessing, the authors remained dissatisfied with their version of the Yes/No test. Meara (personal communication, 1995) has speculated that their lack of enthusiasm for the Yes/No test stems from their decision to split their original 6-point rating scale in two and treat each of the resulting 3-answer sets as either a "yes" or a "no".

Anderson and Freebody (1983) conducted a study in which they compared the performance of 120 native English speaking fifth graders on a multiple choice vocabulary test (195 items) and a Yes/No checklist test instrument containing the same 195 items which was administered one week later. Although the Yes/No test was scored using controls for guessing inspired by the study by Zimmerman et al. (1977), Anderson & Freebody retained the checklist format rather than a rating scale. A strong correlation (.84) was obtained between the two tests. Follow-up interviews with each subject yielded even more impressive results: yes/no answers and actual word knowledge were much more closely correlated (.35 on the strictest of three word knowledge criteria) than were multiple choice answers and actual word knowledge (.45 on the strictest of three word knowledge criteria). The authors concluded that this type of Yes/No instrument, with controls for guessing, was an "excellent indicator of the number of words a person truly knows" (p. 239). They stressed three caveats for the Yes/No, however:

1) It cannot determine which of a word's several meanings a person knows; 2) It cannot evaluate the effects of direct vocabulary instruction, and 3) it may have lower "reliability" and "predictive validity" (quotation marks are Anderson & Freebody's) because success on a multiple-choice test depends not only on knowledge of word meaning but also on "reasoning, planful use of working memory to hold response options in mind, and sensitivity to the subtle nuances of language use in cultured, mainstream circles" (p. 239), qualities which may discriminate in favor of students of higher ability or higher socioeconomic status.

The encouraging results obtained by Anderson & Freebody with English native speakers led to new research, this time with non-native speakers of English. Meara and Buxton (1987) were interested in finding a quick, practical tool for assessing the language abilities of people who act as subjects in empirical research. They developed and administered a new Yes/No test to 100 adult ESL students from several language backgrounds, along with a multiple choice vocabulary instrument which they developed based on the Cambridge First Certificate Examination. The two tests were administered on the same day (the multiple choice, a short break, then the Yes/No). The resulting .703 correlation between the two measures led the authors to conclude that the multiple choice and Yes/No tests were measuring "largely the same sort of thing" (p.147). The authors noted that had they been able, like Anderson and Freebody 1983, to use identical items in the two tests, even higher correlations might have been obtained. Interesting results were obtained for one of the language subgroups. The results for the 18 French speakers yielded a surprisingly high multiple-choice/Yes-No correlation: .829. It was speculated

20

that "imaginary words", or distractors, (which had been constructed by decomposing and reassembling Romance words) may have had some kind of effect on speakers of Romance languages. For this study, security considerations had made it impossible to use the actual Cambridge First Certificate Examination scoring procedure. The logical next step was to seek out a yardstick with which a Yes/No test instrument could be directly compared.

Meara and Jones (1988) arranged to compare the same Yes/No test used by Meara and Buxton (1987) against the Joint Entrance Test (JET), a placement instrument used by the Eurocentres Language Schools. This time, however, Meara and Jones developed a computerized version of their Yes/No test and ran three versions of it with approximately 250 subjects from a variety of language backgrounds. The subjects were students or prospective students at two of the Eurocentres Schools. The JET consisted of multiple-choice subtests (listening comprehension, grammar, and reading comprehension) followed by an oral interview. Comparison of the results of the multiple-choice portion of the JET against the Yes/No test yielded correlations of .664 (for 109 subjects in Cambridge, England) and .717 (for 159 subjects in London). Interestingly, for all but one language subgroup the correlations were even higher (.723 to .807). The exception was the French group, which this time yielded a much lower correlation than the group as a whole (.549 in Cambridge, .556 in London). These clearly divergent results for French L1 speakers caused the authors to speculate on the possibility that either their Yes/No test was systematically biased against this language group - or that it was the JET which was biased. It is also possible, of course, that the limited number

of French speakers tested (18 in Meara and Buxton, and an unspecified (and perhaps quite small) number in Meara and Jones) made the correlations unreliable. What emerged clearly from these early studies was the need to pay close attention to the performance of individual language subgroups on Yes/No tests.

More recent administrations of Yes/No instruments have been carried out with specific language groups. Extremely low-level Arabic learners of English were tested by Al-Hazemi in 1993. In eight experiments, Al-Hazemi administered computerized and pencil-and-paper versions of Meara and Jones' 1988 Yes/No vocabulary test, varying the types of distractors used and comparing the Yes/No to other types of proficiency tests such as Cloze tests and multiple choice. Although the Yes/No test type produced generally encouraging results with his subjects, Al-Hazemi found no relation between the Yes/No results and scores on the other language ability tests used. He also discovered that for some of these learners, Yes/No scores could be influenced by similar-sounding distractors. In addition, it appeared that for some subjects, "knowing" a word may have meant different things to different people. When asked to translate their apparently correct "yes" responses, a number of subjects showed that they had systematically misread the stimulus words. In one example, the target word "chicken" had been read as "kitchen". For these subjects, "yes" scores did not necessarily reflect knowledge of the targeted words. Al-Hazemi postulated that the Arabic language's visual representation as well as the English words' spelling and pronounceability may have contributed to this problem. Al-Hazemi concluded that the Yes/No is a useful tool for measuring vocabulary

knowledge, but that care must be taken in the selection of test items, especially distractors (Al-Hazemi 1993).

Meara, Lightbown & Halter (1994) compared the scores of Yes/No tests taken by low to intermediate level native speakers of Quebec French to the scores of these same speakers on an English placement instrument, the CELT (Comprehensive English Language Test: Harris & Palmer 1970, in Meara, Lightbown & Halter 1994). Their aim was to determine to what extent the presence of French-English cognates in the Yes/No test would affect the results obtained by francophones. Meara, Lightbown & Halter developed two new Yes/No tests using the word lists used by Meara and Jones. One test (NCOG) was composed entirely of items (both real words and distractors) which were not cognate with French. On the other test (COG) half the real words and half the distractors were cognate with French.

For 89 out of 107 subjects, both the NCOG and COG forms of the Yes/No test correlated fairly well with CELT scores (an estimate of overall vocabulary size based on both subparts of the vocabulary tests showed correlations of .746 and .757 with the listening and structure components of the CELT). It should be noted that the vocabulary subtest provided in the CELT battery was not administered by the school participating in this study. Thus, no direct comparison between the Yes/No and the CELT's vocabulary measure could be made.

Although the test without cognates correlated better with the CELT, the difference was slight. The authors feel that the presence of cognates on a Yes/No vocabulary test should not prevent valid results as long as the percentage of cognates on the test does not

23

exceed the number naturally occurring in the target language. For English, this naturally occurring percentage has been estimated at 42% for the first 3000 words (Roberts 1965, in Meara, Lightbown & Halter 1994), but the true percentage of "close cognate forms" may actually be perceived as lower (Meara, Lightbown & Halter 1994, p. 305). This is because significant differences in pronunciation or spelling may make some words difficult to recognize as cognate with French. A good example, taken from this study's Yes/No tests, is the English word *scorn*, which can be traced to the Old French *escarn*, *escharn*, *escar*, and *eschar* (Gove 1981).

An interesting phenomenon occurred with the remaining 18 subjects, who checked off so many false alarms that their scores were "negative", that is, unusable. The authors suspected that the high number of negative scores with these French speakers may have something to do with the high visual impact of English in Quebec daily life: Quebec residents whose level of active English may be low may nonetheless have "an unusually well-developed recognition vocabulary, even when they do not know what all the words mean" (Meara, Lightbown & Halter 1994, p. 306). The authors speculated that this trait may have caused these 18 subjects to overestimate their ability to understand English words, leading them to tick so many distractors that their scores, penalized for the numerous false alarms, lay outside the scoring grid.

Al-Hazemi (1993) and Meara, Lightbown and Halter (1994) studied the performance of learners whose English level was low to intermediate. The present study emphasizes evaluation of the ESL proficiency of learners whose level is expected to be advanced enough to enable them to undertake university studies.

Some Advantages of Yes/No Tests

The Yes/No test is simple to construct and easy to score. Unlike a multiple-choice test, a Yes/No test instrument can measure huge numbers of words within reasonable testing conditions, providing a fair sampling from a learner's whole vocabulary (Meara & Buxton 1987, Read 1988).

Word sampling can either be based on a frequency criterion, or taken from specialized texts in specific subject areas or more general academic or technical word lists for "special purpose vocabulary testing" (Meara 1991a).

Because Yes/No test results have tended to correlate well with multiple-choice-type vocabulary tests as well as tests of reading comprehension, grammar and listening comprehension, it may be possible to use the Yes/No test for diagnostic or placement purposes (Meara 1991a). They might also prove useful as an assessment tool for evaluating subjects in empirical research (Meara & Jones 1988).

An important possible use of the Yes/No format is to produce individual learner vocabulary profiles (Meara 1991a). This becomes especially interesting in cases where a learner has greater knowledge of academic or specialized vocabulary than vocabulary for everyday use, or vice versa. The resulting asymmetry in the learner profile can pinpoint areas where vocabulary work is needed most. In addition, vocabulary profiles can be useful in monitoring a learner's vocabulary growth over time (Meara 1991a).

Yes/No tests appear not to require knowledge of non-target items by the learner. In order to evaluate an item, a learner is not required to infer the meaning of or identify non-target words or distractors (Anderson & Freebody 1983). Each item is pared down

to a minimum: on an English Yes/No test, the subjects must check off an item only when sure they recognize/know it as an English word. When in doubt, they are urged to leave the test item blank.

Some Yes/No Test Limitations

The nature of vocabulary knowledge, in English at least, may become qualitatively different once a learner reaches higher levels of proficiency. In English, vocabulary size may be of greatest importance for learners whose lexicons are under 6,000 words. Beyond this level the L2 lexicon may begin to specialize, taking on more depth and, for this reason, a more sophisticated structure (Meara 1994a). Yes/No tests can be compiled from specialized lists corresponding to the subject areas needing exploration. This has been suggested by Meara (1991a) as a means of screening subjects for knowledge of specific target areas. The problem of tapping vocabulary depth remains, however. A specialized test may give a rough idea of a learner's familiarity with a set of words in a specific subject area, but it cannot indicate which meaning(s) of any word the learner claims to know.

Yes/No tests appear to yield unexpected results for certain language groups. Perhaps the most remarkable of these discrepancies occurs with native speakers of French. European French speakers obtained Yes/No scores which correlated highly (.829) with a "Cambridge Proficiency - style" vocabulary test (Meara & Buxton 1987). In the Eurocentres study, another set of European French speakers produced Yes/No test scores which correlated poorly with the JET placement measure (Meara & Jones 1988).

One group of Canadian French speakers showed a tendency to guess fairly indiscriminately on ESL Yes/No tests (Meara, Lightbown & Halter 1994).

As the examples concerning French and Arabic native speakers have shown, a significant roadblock to the Yes/No test's success appears to lie in the instrument's hypersensitivity to a subject's tendency to guess. Some subjects check off so many nonwords (known as "false alarms") that they obtain overall negative scores, as if they knew no words at all in English. This phenomen.. is poorly understood. Perhaps some cultures are less or more prone to risk-taking. It may be a manifestation of the test's importance to the subject. For example, subjects may feel they have more at stake if they are being evaluated for university admission, while those who have already been accepted and take the test for placement purposes only may come to the test task with a different attitude. An overreliance on guessing may also result from the way in which test instructions are handled.

The Yes/No test has been criticized because it cannot measure depth of vocabulary knowledge. There is no way of determining which meaning(s) of a word, if any, the subject thinks he/she knows. In fact, the Yes/No can only claim to be measure of word recognition (Wesche & Paribakht 1993, personal communication). Yet this may not be such a limitation after all. It may be perfectly valid to test partial word knowledge (Takala 1985) or even simple word recognition, the *sine qua non* of vocabulary knowledge. Meara remarks that one cannot usually recognize a letter-string as an English word without knowing at least something about it (such as one of its meanings or its part

27

of speech). Conversely, a subject who fails to recognize a word as English probably cannot do anything with it (Meara 1990a).

In view of the Yes/No measure's exciting potential in several areas, it seemed worthwhile to continue exploring its validity. It was felt that the opportunity to study a fresh sample of learners using a new combination of Yes/No tests was very likely to shed new light on the feasibility of the Yes/No test instrument.

Yes/No Tests Used in Previous Research: Origin of Test Items

Real words. In earlier studies using Yes/No tests, real words have been selected from a variety of sources. Zimmerman et al. (1977) took words from an English dictionary (Webster's Seventh New Collegiate Dictionary,1963). Anderson & Freebody (1983) chose to take their real words (and their distractors) from the vocabulary subscale of the 1973 Stanford Achievement Test, reasoning that the items on the Stanford Test "are neither too easy nor too difficult and have good discriminating power" (p.235). Meara and his colleagues have used frequency-based word lists for real words (Meara & Buxton 1987, Meara & Jones 1988, Meara and Jones 1990). Meara (1991a) indicates that for the 1990 computerized Eurocentres test, Meara & Jones chose a series of samples from Kucera & Francis' 1967 frequency list. The first set of words tested came from Kucera & Francis' 1000 most frequent words; the second set from the 1000 next most frequent words, and so on, up to a ceiling of 10,000 words (Meara 1991a).

Meara, Lightbown and Halter (1994) compiled their pencil and paper yes/no tests

using the computerized Eurocentres Test (Meara & Jones 1990). In order to study the

effects on test results of varying degrees of cognacy, they altered the Yes/No tests'

original cognate content for both real words and distractors: "Of the two forms, one (...)

consisted entirely of items which were non-cognate with French; in the other form, (...)

half of the real word items were cognate with real French words. Half of the imaginary

words resembled French/English cognate words:"(Meara, Lightbown & Halter 1994, p.

303).

Distractors. Zimmerman et al.(1977) developed a list of distractors by starting

with a pool of 400 words four to ten letters in length, excluding compound words,

hyphenated words, and homographs. From these they made a random selection of

clusters of words having the same number of syllables. Nonwords were created by

interchanging "corresponding syllables of words within these clusters" (p. 7). One-

syllable words were arbitrarily split and the two parts interchanged. The authors operated

a second or third random interchange on items which were too difficult to pronounce or

in which the first transformation produced a real word. A further single-letter

modification was made to the five words which remained extremely difficult to

pronounce after all the previous modifications. "By this method, 100 nonwords were

created which had about the same average length, number of syllables, and letter

frequency as did the 100 real words from which they were derived"(p. 7).

As mentioned above, Anderson & Freebody (1983) took distractors as well as real words from the vocabulary subscale of the 1973 Stanford Achievement Test. Since their subjects were fifth graders, Anderson & Freebody retained all the test items from the fifth grade level of the Stanford vocabulary scale, augmenting it with items selected randomly from the two levels below and the two levels above the fifth grade level on that test.

Although Meara & Buxton (1987) do not elaborate on their means of choosing distractors, Meara & Jones (1988) discuss some aspects of distractor choice for the Eurocentres study: "The current version of the test uses imaginary words which are very carefully constructed so that they share the physical characteristics of the real words in the same set" (Meara & Jones 1988, p. 85). They note problems with these distractors: "(...) it is clear to us that some of the imaginary words are easier to handle than others: some can be rejected instantaneously, while others cause even native speakers of English to puzzle for a long time. We also think that some imaginary words cause difficulty to speakers from particular language backgrounds" (p. 86).

This chapter has reviewed previous studies on L2 Yes/No vocabulary size testing. Most have tested low to intermediate level European ESL learners and have made use of Yes/No test instruments based on frequency-based word lists. In each of these studies, certain language subgroups produced divergent results. The following chapter will describe the process leading to the present study, which seeks further validation for the Yes/No measure by testing higher level ESL subjects within a North American university setting.

CHAPTER 3: RESEARCH DESIGN

Rationale for This Study

Meara and Buxton (1987), Meara & Jones (1988,1990), Al-Hazemi (1993), and Meara, Lightbown & Halter (1994) have pioneered the testing of several Yes/No test instruments with L2 speakers. Of these, only the latter study used a North American setting, measuring a Yes/No test against an established diagnostic tool. The present study seeks to validate a Yes/No test instrument in a context as yet little explored.

Little work has been done using Yes/No tests based on higher level, more academically oriented word lists such as those produced by Nation (1986). In particular, no academically based Yes/No instrument has been administered to prospective and recently admitted university students in North America. In addition, it was possible in this study to make a direct comparison between a Yes/No test and the vocabulary subtest of a North American assessment measure. In the present study, the Yes/No was included in a regular English language test procedure, in which candidates complete part of the Michigan Test of English Proficiency (which includes a vocabulary subtest) and a writing task. Three language subgroups (French, Arabic, and Chinese/Cantonese) appeared to be large enough to permit separate analysis of their results.

Research Questions

The questions addressed by this project are:

1) How do the results of a Yes/No test based on a university entrance-level word list and a Yes/No test based on a general word list correlate with the Michigan Test's

multiple choice vocabulary test? Are the Yes/No and the multiple choice test types measuring the same thing?

2)   How do the results of the two Yes/No tests correlate with the multiple choice Michigan portion of the CELDT test (grammar, vocabulary and reading comprehension subtests), taken as a whole?

3)   Are there any patterns or tendencies revealed which may indicate the presence of more than one ESL population in the sample?

4)   If so, in what way(s) do these populations differ?

It was hoped that the answers to these questions would shed further light both on the Yes/No test's validity as a measure of vocabulary size, and perhaps on its ability to produce results similar to a more traditional instrument which is held to predict broader receptive skills (grammar, reading comprehension). The Michigan Test of English Language Proficiency (as part of Concordia University's CELDT battery) was selected as the major point of comparison against which the Yes/No tests in this study were measured. It has been in widespread use for several years in this particular setting. No presumption was made by the researcher that the Michigan should be seen as a criterion, or benchmark of unimpeachable validity, especially since the Michigan's validation process involved university students in the United States, not Canada. Interesting correlations between the Michigan and Yes/No test measures might reasonably be expected. Moderate to high correlations between the two measures could be an indicator of some confidence in the Yes/No test's construct validity.

It was expected that for Romance language groups the Yes/No tests would suffer from language effects involving Ll-L2 cognacy and visual frequency of the L2; for Arabic speakers problems were expected relating to Ll spelling patterns. Individual differences were expected to have an influence on all subjects, primarily the relative willingness to take risks (guess) in the face of uncertainty. It was expected that the resulting phenomenon of "negative" scores, to which the Yes/No is prone, would result in some unusable results. It was of special interest to see whether either or both of the two Yes/No tests measures selected would hold up as valid alternatives to multiple-choice vocabulary tests despite these expected limitations. Of special interest to this Montreal-based study was the further exploration of the divergent results obtained by French-speaking subjects.

Preliminary Testing

Preliminary testing was conducted during January and April 1993 with approximately 400 students at a community centre, in non-credit ESL courses staffed by volunteer teachers, mostly students from Concordia University. Data for just over 200 subjects were retained for study. Subjects were refugees or recent immigrants to Canada, who spoke a wide variety of languages. In both sessions Yes/No tests developed by Paul Meara (Meara 1992) were administered after preliminary placement (via oral interview) had been done by the community centre volunteers and coordinators from the Centre for Teaching English as a Second Language (TESL Centre) at Concordia University.

The ten individual Yes/No Tests administered at both sessions came from Meara's EFL Vocabulary Tests (1992). Real word items were based on two different word lists.

33

The real words on Yes/No Tests Levels 1 and 2 (four tests, two at each level) came from the first and second thousand word lists (described below) published in I.S.P. Nation's Vocabulary Lists: Words, Affixes & Stems (1986). The real words on Yes/No Tests Levels 3, 4, and 5 (again, two tests at each of these levels) came from the Cambridge English Lexicon (Hindmarsh 1980). For the first test session, an attempt was made to "match" the various Yes/No test levels to the proficiency levels which had already been established at the centre. This was not particulary fruitful, as it made the systematic examination between group results very difficult. For the second test session, all subjects were tested using one series of Yes/No tests spanning five frequency levels, in the hope that this arrangement would produce better profiles. The results seemed to indicate that Yes/No Test Levels 1 and 2, based on Nation's work, were able to discriminate between proficiency levels better than the Hindmarsh-based Levels 3,4, and 5.

For both sessions and all tests at all levels, nonwords or distractors came from one list, developed by Meara and his colleagues. The way in which this list was assembled is described later in this thesis. For the preliminary testing session, distractors and their influence were not studied. This is simply because it became apparent quite soon that the overall testing context held out little promise for valid results. It was decided to investigate the distractor phenomenon later, within the framework of a more rigorous testing situation.

These two test sessions at the community centre were valuable for discovering and rejecting problematical items and individual tests, for improving test instructions and the background information questions asked of each subject, and for mastering the Yes/No

34

scoring procedure. There were, however, drawbacks inherent in testing at this community centre which prevented valid conclusions from being drawn from the data. The most important drawback lay in the fact that the centre placed its students via a brief oral interview. The Yes/No test has always shown very weak correlations with measures of oral proficiency. In addition, little or no proficiency evaluation was made of the centre's students at end of term: once a student had completed a course at the centre, he or she was habitually placed one level higher for the following session, a kind of courtesy placement which is often found in this kind of community service organization. Thus, a student's placement in one of the higher level classes did not necessarily mean that the student's English proficiency was actually higher than that of a student in a low level class. Finally, it was not possible to observe the Yes/No testing directly, because tests were given in individual classes by each teacher. This made it impossible to know how and if the appropriate instructions had been given.

This experience prepared the way for measuring a Yes/No test against a known, established placement instrument. It also confirmed the inadvisability of comparing the Yes/No instrument with tests based only on oral proficiency, since the Yes/No has made no claim to correlations with that aspect.

Concordia University and the CELDT

The search for a reliable yardstick led to Concordia University's Concordia English Language Diagnostic Test (the CELDT). This test is administered on a routine, year-round basis to non-native speakers of English wishing to enter the university. Some

students write the examination as part of their admissions process to the university. Others, already admitted, write the exam for purposes of placement in university English courses only. Based on their CELDT results, candidates for admission are:

1) denied entrance to Concordia. They are encouraged to enroll in Concordia's Continuing Education Language Institute (CELI), a pre-university, non-credit ESL environment;

2) recommended for placement in university-level credit ESL courses (ESL 207, 208, or 209) which may be taken concurrently with other subject courses such as biology, commerce, communications;

3) exempted from ESL courses. These students may choose to enroll in mainstream university credit English courses.

The CELDT test is composed of non-current forms of the Michigan Test of English Language Proficiency (grammar, vocabulary, and reading comprehension multiple-choice subtests) and a pragmatic writing task (composition) developed by Concordia personnel. The multiple-choice Michigan subtests provided the primary point of comparison with this study's Yes/No tests.

The Michigan Test

Description

The Michigan Test of English Language Proficiency is part of a battery of tests designed to evaluate second language learners' ability to handle academic English

(University of Michigan, English Language Institute 1977). The complete battery includes a 30-minute written composition on an assigned topic and a test of aural comprehension, which may be replaced by an oral interview where the means to test aural comprehension are not available.

The Michigan Test of English Proficiency, hereafter referred to as the Michigan Test, consists of three multiple-choice subtests of grammatical usage, vocabulary and reading comprehension. The individual results of these three subtests, as well as their combined (and equated) results, are the yardstick against which the experimental Yes/No tests in this study were measured. Samples of each of these subtests may be found in Appendix A. For test security reasons, the samples shown are not from currently used tests and therefore are similar, but not identical, to Michigan Test Forms P and R, used in this study.

Reliability

The Michigan Test is considered to be a reliable means of discriminating among university applicants from various language backgrounds (University of Michigan, English Language Institute 1977). The University of Michigan's English Language Institute calculated the test reliability of Michigan Test Forms A, B, D, E, F, G, and H, using U.S. university subjects selected randomly. Reliability coefficients for these forms, computed by Kuder-Richardson Formula 21, was greater than .92, with a standard error of measurement for raw scores less than 4.70 (University of Michigan, English Language Institute 1977, p. 13). The reliability coefficient for this study's final sample ($N = 114$;

see Table 1 below), calculated using Kuder-Richardson Formula 21, was .936, with a standard error of measurement for raw scores of 4.56.

Validity

The Michigan Test's authors do not consider the grammar and vocabulary subtests as "a representative sample of the behavior required of a non-native speaker of English studying in an English language university" (University of Michigan, English Language Institute 1977, p. 14). They do affirm, however, that behavior required of subjects in the reading comprehension subtest is "closely analagous to the reading required of all university students" (p. 14). The vocabulary subtest items were taken from the Thorndike-Lorge general word count (Thorndike & Lorge 1944), and were taken to represent "fundamental lexical units basic to integrated language behavior in an academic situation" (p. 14). No such academic focus is claimed for the grammar subtest, which the authors describe as including items which have shown good discrimation among non-native speakers of English. However, the authors' belief in a link between vocabulary knowledge, grammatical ability and broader language skills is reflected in their description of lexical and grammatical ability as underlying "behavior necessary (although not sufficient) for integrated communication skills" (p.14).

The CELDT

Description

The Concordia English Language Diagnostic Test (CELDT) is composed of the multiple-choice Michigan Test (Part One) followed by a writing task (Part Two) which was developed by Concordia personnel.

CELDT Test procedure

Usual procedure. The CELDT is administered at regular intervals. Most test sessions are held on Friday evenings, from 6:00 p.m. until 9:00 p.m., in a large amphitheatre-style classroom. Before the test itself is distributed, the administrator guides the students through the completion of a background information questionnaire. The administrator then explains the instructions for the first part of the CELDT (standardized multiple-choice items on grammar, vocabulary, and reading comprehension: hereafter referred to as Part One), administers and collects this section, and then explains the instructions for the second part of the CELDT (the composition: hereafter referred to as Part Two), then administers and collects this section. The test administrator usually has two assistants. The first assistant helps the test administrator check the subjects' identification and collect examination fees. The second assistant distributes and collects forms, helps invigilate the test, and helps correct Part One while Part Two is underway.

Adding the Yes/No tests. It was arranged with the Director of the TESL Centre that at a certain number of test sessions, five individual Yes/No tests (described below) would be inserted into the CELDT battery between Parts One and Two. The subjects

were not given advance notice of this change to the regular testing pattern. Adding the Yes/No tests lengthened the overall test session by approximately one-half hour. The three occasions when the Yes/No tests were administered were Friday evenings, with the session beginning at 6:00 p.m. The three test sessions were administered by the same test administrator. A total of 135 candidates took the Yes/No tests along with the usual CELDT battery. Student background information such as primary language(s), current or proposed program of study, and test purpose (ESL placement or admissions/ESL placement) was gathered in the usual way via a CELDT information sheet. This information was then made available to the researcher. In return for using these CELDT testing sessions and Concordia personnel for research purposes, the researcher agreed to act as second assistant on evenings when the Yes/No test was administered. Invigilating and helping score the Michigan portion of the exam gave the researcher the opportunity to observe the test procedure first hand, and to tally each of the multiple-choice Michigan subtests separately.

On Part Two of the CELDT, candidates were required to write a composition 200 words or longer on a topic provided them in the form of conflicting arguments. Candidates were required to read the argument and counterargument provided, and to respond in writing. Although their personal opinions were not to be judged, their ability to integrate the arguments into a well-worked, original essay free of technical errors would be evaluated. This portion of the test took one hour.

<u>How the CELDT is scored</u>

Part One is scored at the testing session itself, while the students write Part Two. Scoring is done by two invigilators. Each test is scored twice, the second scorer verifying the first scorer's results. Part One's raw scores are then equated. In a process totally independent from the scoring of Part One, Part Two is evaluated by a minimum of two readers. A set of tests are scored by Reader 1, and then are sent to Reader 2, who scores them without knowing how the compositions fared under Reader 1. In cases where the first two readers disagree greatly, a third reader (head of ESL testing) makes the final decision. The overall recommendation to allow or refuse a student entry to Concordia and/or placement in university ESL classes is made by consulting a pre-established grid showing the various possible scores on Part One (the Michigan Test) and the various possible ratings on Part Two (the writing task). This grid, or algorithm, is shown in Appendix B.

Yes/No Tests used in this study

<u>Origin of Test Items</u>

All the Yes/No tests administered were chosen from Meara's <u>EFL Vocabulary Tests</u> (1992). The one-half hour time constraint allotted for Yes/No testing meant that only five individual tests could be administered. Of these the first two tests were from Level 5 (501 and 502), and the last three were from Level A (A01, A03 and A04). Examples of all five tests, and an explanation for the decision not to use Test A02 may be found in Appendices C and D. Each individual Yes/No Test sheet consisted of 40 real

words and 20 distractors. On Level 5 tests 501 and 502, the 40 real words were taken from Hindmarsh's Cambridge English Lexicon (1980), a word list using mainly frequency-based criteria. On Level A tests A01, A03, and A04, the 40 real words were taken from a university entrance level word list developed by Paul Nation and his colleagues in 1986 (Vocabulary lists: Words, affixes and stems).

Real Words: Level 5 Yes/No Tests

Hindmarsh's Cambridge English Lexicon (1980) was the list from which Level 5's real words were drawn. Using West's General Service List of English Words (1953) as a base, Hindmarsh added the contents of three commonly used English readers with a vocabulary range of 2000-2340 words), several basic English vocabulary lists with a range of 850-1490 words) and three frequency counts (each including the most common 2000 items). The result, after deletion of overlapping material, was the Intermediate English Word List (IEWL), composed of 4200 lexical items. Hindmarsh felt that the IEWL was an improved reflection of the vocabulary needs faced by candidates for the Cambridge First Certificate in English (FCE). These learners are at the intermediate level.

Hindmarsh augmented the IEWL with words from Thorndike & Lorge's frequency based list (1944) plus items from readers and additional word lists such as the New Oxford English Readers for East Africa (Books 1-5), all words from Longman Structural Readers Handbook list not already included, and words not already included from Kucera & Francis (67), a word list C.W. Wright (1965), and words not already

included from the IEWL. Hindmarsh described his lexicon as containing items which were frequent, based on teacher intuition and "reflecting the world today" (Hindmarsh 1980, p. ix). His finished list consists of 4500 lexical items, representing over 8000 semantic values, or gradations of meaning based on word frequency.

Appendix E lists all the Yes/No test items used in this study. After each item in Yes/No tests 501 and 502 is the word frequency level assigned it by Hindmarsh in his lexicon. All the words come from Hindmarsh's frequency level 5, an intermediate level.

Real Words: Level A Yes/No Tests

Three 1000-word vocabulary lists were developed and class-tested by Paul Nation and his colleagues at the English Language Institute (ELI) at the Victoria University of Wellington, New Zealand. Nation's third thousand-word vocabulary list (1986) provided the pool from which items on the Level A Yes/No tests were taken.

The "first thousand word list" was intended to give students starting out in English a means of functioning with a receptive vocabulary (defined as *all words which are understood when listening or reading*) and productive vocabulary (defined as *all words used for speaking and writing*). As can be seen below, the principles of selection for this basic word list included much more than frequency as criteria:

1) language needs: personal, social, thinking, labelling;

2) frequency: high frequency in language as a whole.

3) range: words occurring in many kinds of written English.

43

4)   economy: only essential words (excludes words which can be paraphrased; of two words with same meaning, only one is retained);

5)   regularity: focus is on words that have regular inflection and syntax (can be used in as many sentence patterns as possible);

6)   defining power: emphasis on words which are useful for defining or explaining other words;

7)   classroom and teaching needs: words needed frequently in class;

8)   loan words: inclusion of those requiring little learning effort. (Nation 1990)

Nation explains that "In a passage on a professional subject between 65% and 75% of the words will be in the first 1000-word list" (Nation 1986, p.4).

The second thousand word list assumes knowledge of the first. It was compiled using West's GSL (for frequency criteria) with the addition of most of a list developed at the Central Institute of English at Hyderabad, India. To be included in this list and the higher level list, words had to have fairly high frequency and occur in many kinds of written English (i.e. have a wide *range*). Further selection factors included *coverage* (words designating wide-range concepts, with no convenient substitutes); *meanings* (different meanings of a word were treated as separate words); *related meanings* (unless already present in the first 1000-word list, only more basic meanings were included so that they could be established first); *basic scientific concepts*; *contextual requirements* (words which are obligatory in certain contexts, such as *commit, perform*); *phrases* (such as *in case, carry out*, were included "when their inclusion is warranted on other grounds,

such as frequency, range, and coverage" (Nation 1986, p. 4). Four types of word were excluded from the second and third thousand word lists:

- words already occurring in the 1000 word list (for the third thousand-word list, words already in the second);

- technical or narrow-range words (limited to one or two specialist subject areas). Nation indicates that these words constitute about 5% of a professional passage.

- proper nouns and adjectives derived from proper nouns;

- literary words (colloquial expressions, slang, dialect words, archaic words, emotive words).

Of particular interest for the present study, in which the focus is on receptive vocabulary testing, is Nation's statement that "the 2nd and 3rd thousand-word lists are intended primarily as receptive vocabulary" (Nation 1986, p. 4).

Nation goes on to explain that, compared to the second thousand word list, the third thousand word list contains

"... more terms for difficult concepts and more terms used chiefly in scientific and professional material, including a few semi-technical terms. (...) It (...) excludes words which are frequent only in literary contexts and aims to give as wide a coverage as possible of words and concepts found in university textbooks and associated material. (...) On the other hand, [it] does not go beyond the vocabulary possessed by well-informed Englishmen (and Englishwomen) who lack professional training and specialist knowledge. *It covers not the concepts taught*

in scientific and professional courses, but the concepts whose knowledge is assumed by such courses" (emphasis mine) (Nation 1986 p. 81).

The focus on receptive vocabulary and on the lexical needs of pre-university ESL speakers make the Level A tests of particular interest as a source of Yes/No test items for use in a university entrance examination.

Distractors for Both Yes/No Test Types

The distractors (also called "nonwords" or "imaginary words") used in the Level 5 and Level A Yes/No tests come from one list of "orthographically possible words in the target language" (Meara, personal communication, 1991) which was compiled from two sources. Of the 20 distractors on each Yes/No test, 11 or 12 were taken from British phonebook listings. The remainder were formed by combining stems and affixes from real Romance words to form phonological and morphological combinations which are plausible in English. This proportion is meant to reflect the proportion of Romance words in frequent English vocabulary (Meara 1995, personal communication).

Distractors have an important impact on any Yes/No-type test. Since scoring is weighted to penalize a subject quite heavily for each wrong guess, distractors must not be overly "attractive". Unfortunately, appealing distractors seem to be a problem in Yes/No tests administered so far. Some distractors "fool" sizeable numbers of subjects into ticking them off as English words. Just three such "attractive" items on any one 60-item test may be enough to seriously lower its final score, especially if the subject's English proficiency is already poor.

46

Choice of Yes/No Tests

It was decided to administer one Yes/No test type (Level 5, described previously) which had already been administered during preliminary testing sessions at the community centre. The Level A Yes/No test type, new to this study, was chosen for its emphasis on the type of scientific and technical vocabulary required for university studies. The inclusion of the Level A Yes/No component meant running a pilot test before the core data gathering could begin. Subsequent analysis of variance tests of the pilot versus main data results allowed us to include the pilot test data in the main study data. This analysis will be described below.

Scoring the Yes/No tests

Signal detection and Yes/No Tests

Generally speaking, it can be said that signal detection theory

"provides a number of useful measures of performance in decision-making situations. (...) Essentially the measures allow us to separate two aspects of an observer's decision. The first of these is called <u>sensitivity</u>, that is, how well the observer is able to make correct judgements and avoid incorrect ones. The second of these is called <u>bias</u>, that is, the extent to which the observer favours one hypothesis over another independent of the evidence he has been given" (McNicol 1972, p. 11).

47

Meara and Jones (1988) give us a glimpse into the origins of signal detection methods. It appears that during the 1950's, the British Navy wished to test its radar operators' skill at detecting enemy submarines. The Navy

> ...was interested in three types of behaviour: times when an operator correctly identified a submarine that was actually there; times when an operator failed to identify a submarine that was actually there; and times when an operator identified a submarine that didn't actually exist" (Meara & Jones
>
> 1988, p. 82).

An analogous situation arises when a language learner is presented with an item on a Yes/No vocabulary test. The subject will enter into one of four possible situations, which can be described as follows:

1) The subject correctly decides that the item is a real word. This situation can be called a "hit" (HIT).

2) The subject mistakenly decides that the item is a real word. The item is actually a "nonword" word, or distractor. This situation can be called a "false alarm" (FA).

3) The subject mistakenly decides that the item is not a real word, when actually it is real. This situation can be called a "miss"(MISS).

4) The subject correctly decides that the item is not a real word (that it is a distractor). This situation can be called a "correct rejection" (CR).

Meara and Jones (1988) have produced a schema of these four possibilities:

	real words	distractors
response YES	HIT	FALSE ALARM
response NO	MISS	CORRECT REJECTION

Figure 3. Four possible response types, Yes/No tests (adapted from Meara & Jones 1988, p. 82).

They also point out the parallel between the radar operator's three situations and the Yes/No test taker's hits, false alarms, and misses: "all that is necessary is to replace 'submarines' by '[L2] words.' The mathematical model devised to handle the submarine situation (signal detection theory) should also apply to our vocabulary recognition task" (p. 82).

The grid used to score Yes/No tests allows test administrators to arrive at test results without using the rather complicated signal detection mathematical formulae (see Figure 4). The grid separates out a subject's sensitivity (d'), or actual ability to distinguish between the presence or absence of a stimulus, from the response criterion or bias (β'), or the behaviour the subject adopts under the influence of background "noise" before actually responding (McNicol 1972). The resulting measure, taking into account both sensitivity and bias, is called "Dm" (*delta-m*), or the Dm "score". In Yes/No testing terms, the Dm score reveals the subject's true hit rate: that is, the number of correct responses obtained once false alarms have been taken into account. The Dm

49

is in turn the basis on which a subject's overall vocabulary size may be estimated. "If

you score 75% on a set of sample tests based on the first 3000 words of Thorndike and

Lorge (1944) for instance, then we infer that you know approximately 75% of the 3000

words" (Meara 1994, p. 20). Figure 4 shows the signal detection formulae used to score

the Yes/No tests in this study, while Figure 5 presents an example of the scoring grid

based on these formulae.

$$Ag = .5 + \{[(h - f) \times (1 + h - f)] \div [4h \times (1-f)]\}$$

$$Dm = (4 \times Ag) - 3$$

h = proportion of hits
f = proportion of false alarms
Ag = a non-parametric index of discrimination necessary for calculating Dm

Figure 4. The signal detection formulae underlying the Yes/No scoring grid used in this study. Source: Pastorre, R.E., & Scheirer, C.J.(1974) in Meara (1991).

	mistakes	0	1	2	3	4	5	6	7	8	9	10	11	12	13
	40	100	95	90	85	80	75	70	65	60	55	50	45	40	35
	39	98	92	87	82	76	71	66	60	55	49	44	38	32	26
	38	95	90	84	78	73	67	61	56	50	44	37	30	24	17
	37	93	87	81	75	69	63	57	51	44	38	31	23	16	8
	36	90	84	78	72	65	59	52	46	37	32	24	16	8	
	35	88	81	74	68	62	55	48	41	34	26	18	9		
	34	85	78	72	66	59	52	43	37	28	20	11			
	33	83	76	68	61	54	46	39	31	22	14	4			
	32	80	73	65	58	50	42	34	26	17	7				
	31	78	70	62	54	46	38	29	11	1					
C	30	75	67	59	51	42	33	24	15	5					
O	29	73	64	56	47	38	29	19	9						
R	28	70	61	52	43	34	24	14	3						
R	27	68	59	49	40	30	20	9							
E	26	65	56	46	36	26	15	4							
C	25	63	53	42	32	21	10								
T	24	60	50	39	28	17	5								
	23	58	47	35	24	12									
Y	22	55	44	32	20	7									
E	21	53	41	28	16	3									
S	20	50	37	24	11										
	19	48	34	21	6										
A	18	45	31	17	2										
N	17	43	28	13											
S	16	40	24	8											
W	15	38	21	4											
E	14	35	17												
R	13	33	13												
S	12	30	9												
	11	28	6												
	10	25	1												
	09	23													
	08	20													
	07	18													
	06	15													
	05	13													
	04	10													
	03	8													
	02	5													
	01	3													

too many mistakes: do the test again but only say YES when you are really sure you know what a word means

at this level the tests aren't reliable: ask for an easier test

Figure 5. Grid used to score Yes/No Tests Level 5 and Level A (Meara 1992).

Scoring the Yes/No Test

These pencil-and-paper Yes/No tests were scored by the researcher in the following way:

1. The total number of words the subject checked off (ticks) were added up. Erasures and inked/pencilled-out marks were not counted as ticks. This was a change in procedure from the preliminary tests at the community centre, where erasures and inked/pencilled-out marks had been counted as ticks. Over time it became clear that counting the visible evidence of a subject's "second thoughts" showed only part of the picture: any boxes a subject had originally left blank and later decided to tick would remain undiscovered. This is why it was decided to add up only "final" decisions.

2. A mask made of stiff, opaque paper was created for each of the five tests, showing only the distractors and their item numbers. It was used to reveal how many distractors the subject had ticked on each test. The number of distractors ticked, or "false alarms", was entered under "f" (false alarm score).

3. The number of false alarms was subtracted from the total number of ticks to yield the "h" score, or total number of "hits".

4. Using the grid (see Figure 5) provided by Meara (updated from the grid in Meara 1992), the "Dm", or final score was calculated by aligning the total number of false alarms with the total number of hits. The resulting scores ranged from 0 to 100, unless the number of false alarms was so high and/or the number of hits so

low that the resulting "score" lay "off the grid". These negative scores were entered as N in the data.

5. Once all individual tests had been scored, the Dm scores, negative scores, and total number of hits and false alarms were recorded for each subject. Negative scorers were deleted from further statistical analysis (see discussion below). For each remaining subject, the two Yes/No Level 5 Dm results were averaged to yield his/her Level 5 score (referred to as Dm500). Each subject's three Yes/No Level A Dm results were averaged to yield his/her Level A score (referred to as DmA00). The subsequent correlations with CELDT scores used only the averaged Dm500 and averaged DmA00 scores, not the Dm scores from the five individual Yes/No tests.

Choice of Statistics

The Decision to Include the Pilot Test

Of a total of 135 participants in Yes/No testing, 55 attended the pilot test session. Only one change was made to the testing procedure after the pilot session: the 20 minutes allowed the subjects for the pilot Yes/No test had been reduced to 15 minutes for the remaining two sessions. (It was observed that even with the new 15-minute limit, most subjects had ample time to complete the Yes/No test). To investigate the possible effects of the 5-minute difference in time allotment, it was decided to run a one-way analysis of variance on the three test sessions. The results of this ANOVA showed no significant difference between the performance of the pilot and experimental groups, enabling us to

include the pilot group in the study data. This brought the total number of Yes/No subjects to 135. Negative scores would later reduce this number.

Analyzing the Data

Correlations

Multivariate, listwise Pearson product-moment correlations were calculated on both the total sample and on selected subgroups to compare the subjects' performance on various aspects of the CELDT with the two different Yes/No test types. The variables of most importance are indicated in bold in Figure 6.

	Yes/No Test Level 5 score (averaged)	Yes/No Test Level A score (averaged)
overall Michigan test (equated score)	--	--
Michigan grammar subtest (raw score)	--	--
Michigan vocabulary subtest (raw score)	--	--
Michigan reading comprehension subtest (raw score)	--	--

Figure 6. Correlation variables, Michigan versus Yes/No Tests.

Correlations were also calculated for the two Yes/No test types as compared to the CELDT writing task (PTS) and to the CELDT recommendation (REC). REC represents the final decision for admission and/or placement. It is based on total CELDT performance.

54

Groups and Subgroups

Although 135 subjects took the Yes/No test, 21 of these obtained a negative score on at least one of the five tests and had to be eliminated from further data analysis. Therefore, correlations were run for a total of 114 subjects as well as for selected language subgroups within that total. It should be noted that a subject was said to belong to a language subgroup only if he/she had listed only ONE language as his/her "primary language". Combinations of two or more languages were not included in the subgroups, since there was no way of confirming which of the languages listed was dominant. In Table 1 the Chinese subgroup (22 subjects) and the Cantonese subgroup (11 subjects) have been grouped for correlation purposes. This was done for two reasons. First, although the two languages are different, they share enough characteristics to be grouped for the purposes of this research. Secondly, although it might be supposed that the label "Chinese" meant "a Chinese language other than Cantonese", there remains the possibility that some of the subjects who listed their language as "Chinese" may have been speakers of Cantonese.

The Arabic subgroup had an extremely high negative score rate. In fact, negative scores so decimated the Arabic subgroup that further analysis of its data was not worthwhile. The similarly sized French subgroup lost only three subjects to negative scores, and its data were retained for analysis. The Chinese/Cantonese subgroup showed a much lower rate of negative scores than either the French and Arabic subgroups or the original sample. The factors which may give rise to these negative score rates will be treated in the Discussion.

Table 1

Percentage of Subjects (_Ss_) Obtaining Negative Scores per Subgroup (_n_) and per Original

Sample ($N = 135$)

Subgroup	_Ss_ (% of _N_)	Negative scorers (% of _n_, % of _N_)	_Ss_ retained
French	20 (15%)	3 (15% of _n_, 2% of _N_)	17
Arabic	17 (13%)	9 (53% of _n_, 7% of _N_)	8
Chinese/Cantonese	33 (24%)	2 (6% of _n_, 1% of _N_)	31
Mixed language[a]	98 (73%)	9 (9% of _n_, 7% of _N_)	89
Total _N_	135 (100%)	21 (-- , 16% of _N_)	114

[a]_N_ minus French & Arabic subgroups.

When subsequent analyses established that both the Arabic and French subgroups

produced novel results on the Yes/No tests, it was decided to take a look at an additional

subgroup: the total sample less the French & Arabic subgroups. This subgroup, which

included the Chinese/Cantonese subgroup as well as 36 other languages and language

combinations, showed a negative score rate which was well below that of the original

sample but slightly higher than that of the Chinese/Cantonese subgroup.

CHAPTER 4: RESULTS

The first question this study addresses regarding the Yes/No tests is the following: Are the Yes/No tests (Level 5, Level A) tapping lexical abilities similar to those measured by the Michigan vocabulary subtest? To explore this question, the correlation between the results of the two Yes/No tests and the Michigan Test's multiple choice vocabulary subtest were examined.

On all the correlations below, negative scorers have been deleted from the data. For the purpose of clarity, the original 135 subjects will be referred to as the "original sample", and the 114 subjects remaining after deletion of Yes/No scorers will be referred to as the "final sample", or N. Raw data and the complete Pearson correlations for the final sample and all subgroups can be found in Appendices F and G. Table 2 shows positive, significant, moderate correlations between the Yes/No results and the Michigan vocabulary subtest results for the final sample and the three subgroups. For the final sample, correlations are higher between Level A and the Michigan vocabulary subtest; a shared variance of 46% is obtained as opposed to 32% between Level 5 and the Michigan vocabulary subtest. The academic word list seems to capture more of what the Michigan vocabulary subtest is aiming for. Similar higher correlations are obtained for the "mixed language" subgroup which excludes the Arabic and French speakers and for the subgroup of Chinese/Cantonese speakers. For the French speakers alone, however, the opposite results are obtained, although the difference between the two is not as great. It will be suggested below that the extremely high level of English-French cognates in all the Level A tests plays a role in these results.

Table 2

Correlations between Michigan Vocabulary Subtest and Yes/No Tests Level 5 and

Level A.

	Michigan vocab vs Level 5 (c.d.)		Michigan vocab vs Level A (c.d.)	
N = 114 (final sample)	.568*	(32%)	.681*	(46%)
n = 89 (mixed language)	.577*	(33%)	.711*	(51%)
n = 31 (Chinese & Cantonese)	.604*	(36%)	.718*	(52%)
n = 17 (French)	.665*	(44%)	.613*	(38%)

Note. Level 5 = average of Yes/No test 501 and 502 scores
    Level A = average of Yes/No test A01, A03, and A04 scores
    c.d.   = coefficient of determination (shared variance)
    All figures are significant. Significance was determined using Table B, p.466, in
    R.L.D. Wright, Understanding statistics: An informal introduction for the
    behavioral sciences (1976), New York: Harcourt Brace Jovanovich.

*p < .01.

The second question this study addresses is: How do the results of these two

Yes/No tests correlate with the Michigan portion of the CELDT test (grammar,

vocabulary and reading comprehension subtests), taken as a whole?

A part-whole correlation between the Michigan vocabulary subtest and the overall

Michigan Test yielded the following results for this study's final sample: (.786, N=114,

p < .01, n = 114). Table 3 shows the correlations between the overall Michigan Test

scores and the Level 5 and Level A Yes/No test scores for not only the final sample but

also for the various subgroups which produced usable data.

58

Table 3

Correlations between Overall Michigan Test and Yes/No Tests Levels 5 and A.

	Overall Michigan vs Y/N Level 5 (c.d.)		Overall Michigan vs Y/N Level A (c.d.)	
N = 114 (final sample)	.622*	(39%)	.699*	(49%)
n = 89 (mixed language)	.619*	(38%)	.723*	(52%)
n = 31 (Chinese & Cantonese)	.607*	(37%)	.694*	(48%)
n = 17 (French)	.787*	(62%)	.661*	(44%)

Note. Level 5 = average of Yes/No test 501 and 502 scores
Level A = average of Yes/No test A01, A03, and A04 scores
c.d. = coefficient of determination (shared variance)
All figures are significant. Significance was determined using Table B, p.466, in R.L.D. Wright, Understanding statistics: An informal introduction for the behavioral sciences (1976), New York: Harcourt Brace Jovanovich.

$*p < .01.$

The correlations on this table are positive and significant. Some approach the strength of the relationship between the Michigan Test's other subtests and its vocabulary subtest. The Level A Yes/No test appears to cover more of the same Michigan Test ground than does the Level 5 test. Since the Level A is based on an academic word list, this may not be surprising. The French group again differs from the other groups: their Michigan results correlate much better with the Level 5 (.787) than with the Level A Yes/No test (.661). Tables 4 and 5 arrange the final sample and subgroup correlations differently to highlight each Yes/No test type's results against those of the Michigan vocabulary subtest and the overall Michigan Test.

Table 4

Correlations between Michigan Vocabulary Subtest and Yes/No Level 5 and between Overall Michigan and Yes/No Level 5.

	Michigan vocab vs Yes/No Level 5 ($c.d.$)		Overall Michigan vs Yes/No Level 5 ($c.d.$)	
N = 114 (final sample)	.568*	(32%)	.622*	(39%)
n = 89 (mixed language)	.577*	(33%)	.619*	(38%)
n = 31 (Chinese & Cantonese)	.604*	(36%)	.607*	(37%)
n = 17 (French)	.665*	(44%)	.787*	(62%)

Table 5

Correlations between Michigan Vocabulary Subtest and Yes/No Level A and between Overall Michigan and Yes/No Level A.

	Michigan vocab vs Yes/No Level A ($c.d.$)		Overall Michigan vs Yes/No Level A ($c.d.$)	
N = 114 (final sample)	.681*	(46%)	.699*	(49%)
n = 89 (mixed language)	.711*	(51%)	.723*	(52%)
n = 31 (Chinese & Cantonese)	.718*	(53%)	.694*	(48%)
n = 17 (French)	.613*	(38%)	.661*	(44%)

Note. Level 5 = average of Yes/No test 501 and 502 scores
Level A = average of Yes/No test A01, A03, and A04 scores
$c.d.$ = coefficient of determination (shared variance)
All figures are significant. Significance was determined using Table B, p.466, in R.L.D. Wright, Understanding statistics: An informal introduction for the behavioral sciences (1976), New York: Harcourt Brace Jovanovich.

*$p < .01$.

60

For the French subgroup, Yes/No Level 5 correlates better with the overall Michigan Test (.787) than it does with the Michigan vocabulary subtest (.665). There is a fairly strong shared variance of 62%.

The third research question, following from the first two, is the following: Are there any patterns or tendencies revealed which may indicate the presence of more than one ESL population in the sample, and if so, how do they differ from one another? The test results appear to indicate the existence of more than one ESL population in this sample of 135 subjects. Firstly, 21 of the 135 obtained negative scores which could not be analyzed, reducing the study sample to 114. Of the 21 negative scorers nearly half came from the small Arabic speaker subgroup: this calls into question the usefulness of these Yes/No tests for Arabic speakers.

The French subgroup, although not as plagued by negative scores as the Arabic subgroup, produced test results which ran against the current of all the other subjects' results. The higher correlations between the French subgroup's Level 5 (non-academic) Yes/No test results and both the Michigan Test's vocabulary subtest and overall scores may indicate a tendency for francophones to react to the Level A Yes/No tests' high French-English cognate content, either with overconfidence or with exaggerated caution. The mixed language subgroup, or the final sample minus the two "problem" subgroups, French and Arabic, shows results which are quite similar to that of the final sample. It should be noted that the mixed language subgroup is composed of subjects from 38 different primary languages and language "combinations" (some of the combinations include Arabic or French). Finally, the Chinese/Cantonese subgroup, whose languages

are very different from English, reacts to the Yes/No tests in the expected manner: like the final sample and the mixed language group (they are included in both these groups), the Chinese and Cantonese obtain Level A scores which correlate better with the overall Michigan and Michigan vocabulary subtest scores than do their Level 5 scores. Why this might be so is taken up in the Discussion.

For the great majority of subjects in this study, then, the A level Yes/No test seems to cover much the same ground as both the Michigan vocabulary subtest and the overall Michigan Test, although not enough to warrant replacing the Michigan with the Yes/No measure. Unfortunately, two important subgroups, the Arabic and French subgroups, behave erratically on the Yes/No tests administered. The problem with the French group is especially worrisome in the Quebec context, as the great majority of ESL speakers one might wish to test are francophones. The great need for a test which evaluates the ESL abilities of francophones may make it worthwhile to tailor a battery of Yes/No tests for the exclusive use of this population.

The fact that two language groups perform oddly on these Yes/No tests leads to another question: in a larger sample, where other language subgroups might be numerous enough to study, might further problems arise with these groups as well?

The issues of French-English cognacy, negative scores and the related problem of overly "attractive" distractors will be explored further in the Discussion.

CHAPTER 5: DISCUSSION, CONCLUSION, RECOMMENDATIONS

Discussion

Expected and Unexpected Results

As expected, most of the 114 subjects on the Yes/No tests obtained results on Yes/No test Level A which corresponded better with the Michigan Test measures than did their results on Level 5. The higher correlations may be influenced by the fact that three individual Level A tests were administered as opposed to two individual Level 5 tests. But they also may indicate that for many L2 subjects, a Yes/No test based on an academic word list may be a useful way to assess the size of their academic vocabulary. The better Level A-to-Michigan correlations are true for the final sample (less negative scores, or $N=114$, for a large su group ($n=31$) composed of Chinese and Cantonese speakers, and for another large subgroup ($n=89$) containing the Chinese and Cantonese and 36 other language groups who were not represented in large enough numbers to constitute subgroups on their own.

But most of the subjects does not mean all the subjects -- and the remainder pose some challenging problems indeed for the Yes/No test. The first roadblock came in the form of the 21 subjects, from a variety of language backgrounds, whose scores were "negative" and therefore useless as data. It would seem that both inordinate guessing (ticking off too many distractors) and exaggerated caution (leaving too many real words blank) can contribute to negative scores, although the more serious of the two causes is inordinate guessing. The second difficulty encountered was with two of the language subgroups. Test results and correlations for both the Arabic and the French subgroups

63

did not conform to expectations; their performances were quite different from one another, as well. One objective of this project has been to shed further light on these problems, so that future Yes/No test researchers may be able to avoid them. This study also addresses other, related riddles such as test takers' guessing or caution, the "attractiveness" of some of the Yes/No test's distractors, and L1 influences, especially with regard to cognacy.

Negative scores

Negative scores and this study

Some test takers ended up with "negative", unusable scores on their Yes/No tests. Negative scores represent an obvious hindrance to the Yes/No test's reliability. There is a positive side to the matter of negative scores, though: exploring the "why" behind them may break new ground.

Twenty-one of the 135 test takers obtained at least one negative score on the five Yes/No tests. Appendix H shows their performance. The negative scorers were generally of low proficiency. Sixteen out of the 21 obtained CELDT scores which precluded their admission to the university or their placement in university-level ESL courses, and none of the negative scorers who were admitted/placed by the CELDT were placed higher than intermediate level ESL.

Table 6 shows that about the same percentage of test-takers obtained negative scores on each of the test dates, while Table 7 shows that the frequency of negative scores increased along with the test sequence.

Table 6

Test Date and Frequency of Negative Scores.

Test Date	Mar. 18	Mar. 25	Apr. 8
Negative scores	8	5	8
n	55	29	51
% of n	15%	17%	16%

Table 7

Frequency of Negative Scores and Test Sequence.

Test:	501	502	A01	A03	A04
Negative scores:	2	4	6	9	13

While it appears that Level A was more difficult than Level 5 for about half the negative scorers (see Appendix H), the steady nature of the increase in negative scores throughout the testing sequence (which remained unvaried) rather than a sudden jump from level 5 to level A makes test fatigue a factor to consider. It will be necessary, in further tests of this type, to vary the order in which the tests are taken in each test session so that the fatigue factor may be better understood.

65

Why Do Negative Scores Occur?

As mentioned above, negative scores happen either when a subject ticks too many distractors (thinking they are real words) or leaves too many blanks where real words should have been ticked. Several factors may contribute to the negative-score problem.

The scoring formula. Is the signal-detection scoring formula too strict, pushing too many subjects "off the grid"? Does it take into account both risk-taking and cautious behavior? Meara (personal communication, 1994) and his colleagues are exploring the mathematical adequacy of the present formula. The present study contributes to the view that while a change in the formula might improve the situation, other factors also lead to negative scores.

L1 influence: the Arabic speakers. In their study on language transfer, Ard and Homburg (1983) studied the performance of two language groups (Arabic and Spanish speakers who were at about the same proficiency level) on the vocabulary portion of the Michigan Test (Form G, not used in this study). They noted that the Arabic speakers' performance on the test was considerably lower than that of the Spanish speakers. Ard and Homburg established that 60% of the key terms on this test resembled Spanish in form and meaning as opposed to only 1% for Arabic. They also found that the Spanish speakers did better even on items where no Spanish-English link was present. They wonder if the Spanish speakers' better results could be due not only to the more obvious

66

"overall closeness in lexical structuring" (p. 171) between Spanish and English, but also to a "finite effort" effect. This means that the Spanish speakers, finding many words comparatively easy to learn, would have more time than the Arabic speakers to concentrate on harder words.

Ard & Homburg's choice of a Romance language and Arabic offers some interesting parallels with the present study. On the CELDT's Michigan Test the Arabic speakers' results were much lower than those of the French speakers. This information is given in Table 8.

Table 8

Arabic and French Subgroups: Results on CELDT.

	not admitted/ not placed by CELDT	placed in low level ESL	placed in intermediate level ESL	placed in advanced level ESL
L1 Arabic (*n*=17)	10	5	2	--
L1 French (*n*=20)	2	7	5	6

Arabic speakers do have a problem with the CELDT. They form the largest language subgroup (36%) of candidates who have failed the CELDT three or more times since 1988 (personal communication, M. Gross, Test Administration Officer, TESL Centre, Concordia University, December 1994). Although the Michigan Tests used in the CELDT were not evaluated for cognate content in this study, the study's Yes/No tests do contain a high level of English-French cognates (see below). Did all these familiar-

looking lexical items enable the French speakers, like Ard & Homburg's Spanish-speaking subjects, to go through the Yes/No test more quickly, while the Arabic speakers had no such advantage? Might this also have happened on the Michigan Test?

Apart from the question of cognacy, the Arabs appear to have another difficulty on ESL tests, and especially on recognition-type tasks: the heavy influence of their L1. Previous research with Yes/No tests has shown that Arabic speakers (at least at low levels of English proficiency) may confuse English words with other English words having similar consonant patterns. In an attempt to confirm this, Ryan and Meara (1991) presented English words on a computer screen to a small group of Arabic speakers whose English was at the intermediate level. Each word appeared a first time for approximately one second. After a two-second blank interval, the word then reappeared in either identical form or altered by removal of one vowel from the second, third, fourth, sixth, or eighth position in the word. The subjects were asked to say which word pairs were identical. Figure 7 shows an example of this procedure.

1st presentation	2nd presentation	vowel deleted from
department	dpartment	second position
distribute	dstribute	second position
experiment	expriment	fourth position
automobile	autmobile	fourth position
management	managment	sixth position
revolution	revoltion	sixth position
sufficient	sufficint	eighth position
photograph	photogrph	eighth position

Figure 7. Examples of altered stimuli, Ryan & Meara (1991), p. 535.

68

The Arabic speakers were extremely slow at this task and found it much more difficult than did their non-Arabic speaking controls. Their problems did appear to relate directly to the missing vowels. These results combined with their previous pilot tests confirmed Ryan and Meara in their view that Arabic speakers have great difficulty in processing English words. They wonder if the root of the problem lies in the nature of the Arabic language, which relies to a great extent on its consonantal segments rather than on its vowels. Ryan and Meara conclude that the Arabic speakers' "word processing problems [in English]... may be related to the lexical structure and orthography of Arabic" (p. 539). A similar view is put forth by Al-Hazemi (1993). He suspects that test items may be misidentified by Arabic speakers, who are led astray by L1 spelling as well as pronounceability in the L2. He also hypothesizes that Arabic speakers, when confronted with uncertainty, may have a tendency to guess.

Distractors and Attractiveness

On Yes/No tests the false alarm rate is usually quite low: Meara (personal communication, 1991) has reported that in a test of 60 items, one or two false alarms can be expected. On the tests used in this study, the number of false alarms per test for each subject ranged from zero to 18. A high rate of false alarms does not necessarily result in a negative score. If a very high number of hits are ticked, false alarms may also be numerous without pushing the Dm score "off the grid" (although they will drastically lower the Dm score). One subject ticked 9, 12, 13, 13, and 12 false alarms respectively on the five tests without obtaining a negative score on any of them.

The ticking of large numbers of false alarms comes about when a test taker finds distractors inordinately "attractive", and is fooled into thinking that they are real English words or, as in the case above, when the test taker ticks virtually everything. The elimination of this problem is one of the major challenges for the Yes/No test's developers. There is a third factor, intrinsic to this study's test administration, which should not be overlooked. The Yes/No tests were inserted into the CELDT battery between Parts One and Two, because the researcher and the test administrator felt that beginning the test session or ending it with the experimental tests was not advisable. The Yes/No's location in the test series did present one unavoidable difficulty, however. Subjects had been actively encouraged to guess on Part One, as even a wild guess on a multiple choice test is preferable to leaving all the items blank. On the Yes/No tests followed immediately, subjects were urged not to guess. Although the Yes/No segment was preceded by clear, pointed, and repetitive instructions stressing the importance of *not* guessing (see Appendix C for the instruction sheet), it is obvious that the subjects may have found it difficult to adjust their "guessing" stance to the required "cautious" stance.

Of the original 135 subjects 40 were rated as "cautious", using the following criterion (arrived at intuitively): a "cautious" test taker ticked more than one false alarm no more than once during the five Yes/No tests. For example, a subject was considered cautious who ticked one false alarm on test 501, none on test 502, none on test A01, three on test A03, and one on test A04. The subject thus exceeded a single false alarm only once on the five tests.

Table 9 gives some information about these cautious subjects. When more than one false alarm was ticked, the test involved was listed, along with the "attractive" distractor(s) the subject ticked on that test.

Table 9

Performance of Cautious Subjects on the Yes/No Tests.

S #	CELDT REC	Language(s) listed	Test	Distractors ticked	Level 5 score	Level A Score
3	3	Spanish	A04	traduction harmonical multiplify	82	70
5	5	Greek/Eng[b]	none		90	84
6	1	French	A04	traduction gummer suddery	45	N[a]
17	1	Greek	none		42	29
18	4	Arab/Eng[b]	A03	diversal voluminary	93	89
20	1	Arabic	none		38	23
22	4	French	none		67	63
27	3	Polish	A04	traduction multiplify	80	67
28	2	Farsi	A01	faminisation break without	67	39
29	1	Chinese	502	aspection obsolation pungid	42	35
34	3	French	none		33	66
37	2	Japanese	none		79	57
39	2	Vietnamese	A04	traduction harmonical multiplify	87	46
41	1	Hungarian	none		69	62
42	4	French	A04	traduction harmonical	72	75
43	2	Cantonese	none		77	65
44	1	Chinese	none		41	54
48	2	Chinese	none		76	81
49	4	French	none		81	78

71

50	2	French	none		40	28
62	4	Bulgarian/English[b]	A04	harmonical multiplify	90	78
63	3	French	none		66	56
65	4	Bulgarian	A04	traduction harmonical multiplify	88	59
67	1	Cantonese	none		42	24
72	1	Cantonese	none		49	57
73	2	Chinese	A01	rudge break without	85	82
74	3	Arabic/French	A04	traduction harmonical	78	84
75	1	Vietnamese	A01	todd exemptation break without	73	63
92	2	French	A01	exemption spalding	47	31
94	4	French	none		69	50
101	2	Romanian	A01	rudge exemptation	62	65
105	3	Chinese	A01	misrequite exemptation faminisation	85	64
112	2	Korean	none		73	60
115	2	Arabic	none		56	34
125	5	English[b]	502	cicatration conceitful	90	91
126	3	Romanian/French	none		63	51
127	2	Cantonese	none		62	39
129	4	Chinese	none		77	90
132	2	Cantonese	none		69	62
133	2	Cantonese	A03	diversal participline	76	64

[a]This subject obtained a negative score on Level A and was therefore deleted from the final sample.

[b]Although these subjects listed English as their primary language or one of their primary languages on the CELDT information sheet, they had been required to take the CELDT based on additional other information on their language background available to the Test Administrator leading her to suspect that they were non-native speakers of English.

It can be seen that cautious test takers come from a range of language backgrounds and from all proficiency levels. Those subjects who departed from their usual cautious stance to tick more than one distractor on any given test were "trapped" by many of the same words. Chief among the attractive distractors are *traduction* (7), *harmonical* (6) and *multiplify* (5), all of which are on test A04.

If even cautious subjects are led to tick these items, then it may be reasonable to advocate removing these distractors from the Yes/No tests. Appendix I shows that these three distractors attracted 50%, 60% and 50% of the original 135 subjects.

An example of caution: the French speakers

This study's French speaking subjects appeared to react to the Yes/No tests with a considerable degree of caution. They showed a fairly low number of false alarms compared with the Arabic speakers.

It may be instructive to take a look at the French and Arabic subgroups' performances, especially since the two groups are approximately equal in size. Tables 10 and 11 compare the CELDT recommendations and false alarm rates for all Yes/No Tests for the L1 Arabic and L1 French subgroups. The subjects are separated into those whose scores were retained for correlation and those whose negative scores had to be discarded.

Table 10

L1 Arabic Subgroup (*n* = 17): CELDT Recommendations, False Alarms for Yes/No Tests Levels 5 + A; Usable and Negative Scores.

| | | USABLE SCORES | | | | | NEGATIVE SCORES | | | | |
| Subject # | CELDT REC | false alarms | | false alarms | | | false alarms | | false alarms | | |
		501	502	A01	A03	A04	501	502	A01	A03	A04
2	1						2	2	3	0	3
8	1	1	6	5	5	3					
19	1	4	4	1	2	1					
20	1	0	1	1	0	0					
31	1	1	0	3	2	3					
ª32	2						2	4	4	5	4
33	3	0	0	0	3	2					
40	1	1	3	2	1	1					
46	1						1	2	2	0	2
47	1						13	18	16	17	14
56	2	1	6	1	7	2					
ª57	3						14	13	9	11	11
86	1						0	2	2	3	5
ª89	2						4	6	2	2	5
90	1						2	2	4	3	5
ª100	2						5	8	5	5	3
115	2	1	1	0	1	1					

Note. CELDT recommendations: 1 = not recommended for admission to Concordia
2 = placed in low-level ESL classes
3 = placed in intermediate-level ESL classes
4 = placed in advanced-level ESL classes
Patterns: There was no particular pattern to incidence of false alarms.

ªThese subjects obtained negative scores on a Yes/No test but were recommended for admission to Concordia based on their CELDT results.

Table 11

L1 French Subgroup (n = 20): CELDT Recommendations, False Alarms for Yes/No Tests Levels 5 + A; Usable and Negative Scores.

Subject #	CELDT REC	USABLE SCORES					NEGATIVE SCORES				
		false alarms		false alarms			false alarms		false alarms		
		501	502	A01	A03	A04	501	502	A01	A03	A04
6	1						0	1	0	0	3
10	2	2	2	6	7	6					
*13	2						1	1	1	2	4
16	1						0	1	4	6	3
21	2	0	2	2	1	5					
22	4	0	1	1	0	1					
34	3	0	1	1	0	1					
38	3	1	3	5	3	5					
42	4	0	1	0	1	2					
45	4	0	2	2	2	2					
49	4	0	1	0	0	0					
50	2	0	1	0	0	1					
60	2	0	1	3	2	5					
63	3	0	0	1	0	1					
66	4	2	0	1	1	4					
83	2	1	3	3	1	4					
92	2	1	1	2	1	1					
94	4	0	0	0	0	1					
130	3	0	1	0	2	3					
131	3	1	3	2	2	2					

Note. CELDT recommendations: 1 = not recommended for admission to Concordia
    2 = placed in low-level ESL classes
    3 = placed in intermediate-level ESL classes
    4 = placed in advanced-level ESL classes
Patterns: - Subjects tend to guess more on level A than on level 5.
    - Within level A, the highest guessing rate occurred on test A04 (A01: 34 false alarms; A03: 31 false alarms; A04: 54 false alarms).
    - There was no false alarm rate over 7.

*This subject obtained a negative score on a Yes/No test but was recommended for admission to Concordia based on his/her CELDT results.

Tables 10 and 11 show which subjects were cautious and which were apt to take risks by guessing. The criterion described earlier was used to rate "cautious" test takers: subjects who ticked more than one false alarm no more than once during the five Yes/No tests. A typical cautious subject's series of false alarm rates might be 0, 1, 0, 1, and 3. A second criterion was used to rate subjects on the other end of the behavior scale as "risk takers". A "risk taker" ticks five or more false alarms on any two Yes/No tests. An example of a risk-taker's series of false alarm rates is 2, 2, 6, 7, and 6.

Applying these criteria to Tables 10 and 11 yields the following information for the Arabic and French subgroups:

Table 12

Arabic and French Subgroups: Risk-takers and Cautious Subjects.

	L1 Arabic ($n=17$)	L1 French ($n=20$)
cautious subjects	1 (6%)	9 (45%)
risk takers	5 (29%)	2 (10%)

The Arabic speakers appeared to be fairly willing to take risks: one-third of them accepted many distractors as English words. Of the remaining 12 Arabic speakers, only one was rated "cautious". For the French speakers the roles are reversed. They were cautious: when unsure of a word's status, nearly half the subjects decided against checking it off. Of the remaining 11 subjects, only two were rated "risk takers".

76

An anecdotal incident may lend further support to the "caution" theory regarding francophones. The same Yes/No tests administered at the CELDT were administered to J., a 13-year-old (the researcher's daughter) who is highly proficient in both English and French. Instructions were identical to those used at the CELDT. The test was scored immediately after J. took it. Her results showed very high scores on Level 5, dropping to a series of moderately high scores on Level A. This is consistent with the fact that J., at 13, was far from being a candidate for university admission! After the test J. engaged in a think-aloud session. She had ticked no false alarms at all, but missed 29 target words. Most of her reasons for not ticking real words had to do with their perceived connection to French. The nine words J. said she had never seen before were all Level A items. Of the remaining 20 misses, there were 12 for which she gave reasons that might best be described as showing "French-induced caution". These items are listed below along with her comments (Appendix J contains the complete text of the think-aloud):

Item	J. didn't check because:
arc	she had seen it in French; wasn't sure if it existed also in English.
criterion	it resembled *critère*; she wasn't sure sure if it existed also in English.
malleable	she "almost checked it off"; knew *malléable*; wasn't sure if it existed in English.
graph	it "sounded silly alone". "*graphic* would have been better".
morality	it "sounded too French."

accorder	*"accorder* is a French word."
equality	she thought only *equal* existed, and *égalité*.
sparse	she thought of *scarce*, *éparpiller*, *éparse*, but had never seen *sparse* before.
susceptible	she knew French *susceptible*, but didn't think it existed in English.
immune	she knew French *immuniser*, but was not sure *immune* existed in English.
urban	*urbain*.
basin	she "thought it should have another *s*: *bassin*.

In spite of the differences between J. and the CELDT subjects (her age, educational level, and comparatively high English proficiency), the results of her think-aloud are intriguing. It would be worthwhile, in further research of this kind, for think-alouds to be organized for test takers, to enable them, like J., to reflect on the choices they made on Yes/No tests.

The French Subgroup and Guessing

Unlike the Arabic speakers, whose false alarm rates fall into no particular pattern, the French speakers became slightly less cautious on the second, more academic Yes/No Level A (see Tables 10 and 11). Within the A level, they showed their highest tendency to guess on the last test of all, Test A04.

78

There are several ways to view this phenomenon. We have seen that one is the fatigue factor. But the Yes/No tests at Level A present another particularity which invites speculation.

It has been theorized that if a Yes/No test contains a level of French-English cognates which exceeds the level normally present in English, test results for native speakers of French may be unduly influenced (Meara, Lightbown & Halter 1994). In that same study, the level of English-French cognates occurring naturally in the first 3,000 words of English is estimated to be 40% or less. A survey was undertaken of the real words in all five Yes/No tests. The source consulted was <u>Webster's Third New International Dictionary</u> (1981). A complete listing of this study's Yes/No test real words and their English-French cognate status can be found in Appendix E. Table 13 shows to what extent the real words on this study's Yes/No tests are cognate with French. On the table, words called "primary cognates" are those the dictionary source links directly to French, Middle French, or Old French. Words called "secondary cognates" fall under one or more of the categories below (the examples come from the study's Yes/No tests):

a)   their entry into English is extremely old, making the English-French connection harder to distinguish, as *scorn*;

b)   although not listed by the dictionary as French cognates, these words entered French as well as English from Latin and are easily connected to French words today, as *technology, producer, criterion, subtract*;

c)   part of the word is cognate with French, as *cotton*wool, en*danger*;

d) the English word has been "borrowed" in modern French, as *snack-bar*, *wartime*, used in spoken French in the Hull, Quebec area to refer to houses built during the second world war: *des wartimes*).

Table 13

Percentage of Cognates on Yes/No Tests.

TEST	primary cognates	secondary ccgnates	total of 40	%
501	12	7	19	48%
502	14	6	20	50%
A01	18	9	27	68%
A03	21	13	34	85%
A04	26	8	34	85%

While the Level 5 tests contain a high cognate level (50%), the French-English cognate level for Level A is far higher (up to 85%). Perhaps this is enough to tempt some French speaking subjects to take more risks as they progress through the Level A tests.

The French speaking subjects in this study, then, are generally cautious in the face of unknown, "familiar"-looking words. This said, an 85% level of English-French cognates, perhaps in combination with test fatigue, may have prompted some guessing.

<u>A word on validity</u>

Pearson correlations for the total sample and the large Chinese/Cantonese subgroup show the expected higher correlation between the Michigan Test and Yes/No Level A than between the Michigan and Level 5. The French subgroup shows an opposite phenomenon: better correlations between the Michigan and Level 5 than between the Michigan and Level A. Why do the French speakers obtain higher correlations on Level A? It is tempting, based on previous research, to imagine that they guess wildly on Level A, thrown off by huge quantities of English-French cognates. Yet what this study's test results reveal is, instead, generalized caution. Although the guessing becomes more frequent towards the end of the Level A series, no French speaker ticks off more than 7 false alarms anywhere on the Yes/No tests, and for the vast majority, the false alarm rates are well under 7. Perhaps the question that may be asked at this point is whether it is the Yes/No or the Michigan Test which more accurately assesses these subjects' university-entrance-level vocabulary. If it turns out that the 85% cognate level of the Level A tests accurately reflects the English-French cognate level of academic English, then Yes/No Level A may be a good predictor of French speakers' lexical potential for university studies in English.

Conclusion

The moderate, positive correlations obtained for most of this study's subjects with the Michigan Test and its vocabulary subtest bode well for the Yes/No's future as a tool for measuring vocabulary size. For many subjects whose languages bear littl.

resemblance to ᴇᴎglish, such as those of the Asian countries, the Yes/No test measure based on an academic word list may be a quicker means of assessing pre-university level vocabulary size than a multiple-choice instrument like the Michigan Test's vocabulary subtest.

Before the Yes/No can replace existing tests, however, the problems certain language groups encounter with it must be solved. It remains disturbing that the Yes/No's target words and distractors are reacted to so differently by various language groups. In this study, only a handful of language subgroups were large enough to study. It will be necessary to successfully pilot an academically oriented Yes/No test with large single-language groups before the Yes/No can be used with confidence.

Because Arabic speakers appear to have major problems decoding English words, it is suspected that they will continue to have problems with both the Yes/No and other types of vocabulary tests. More progress can perhaps be expected with French speakers, if further research bears out this study's conclusion that they are cautious on Yes/No tests. If the French speakers at pre-university level are truly cautious, then they are coming to the Yes/No task with appropriate strategies, and should be able to perform well on the measure, although they will perform differently from other language groups. It will be up to the Yes/No tests' constructors to ensure that the cognate content of the real words matches the target English vocabulary, that overly attractive distractors, distractors too easy to eliminate, and recurrences of any type of item are dropped, and that their replacements are validated through testing. With such further validation, it does

not seem impossible to foresee a Yes/No test tailored especially for the Quebec French pre-university population.

Recommendations

A logical next step following this project might be to explore the fate of the 78 subjects from the study who, following the CELDT, went on to ESL classes at Concordia. It would be of interest to see how many of these students were placed accurately in their ESL classes. If level changes became necessary, it would also be interesting to see if the student had been moved to a level corresponding more closely to the Yes/No test results.

For future test administrations, it will be advisable to make the following changes:

- vary the order of the Yes/No tests at each test session to counterbalance the fatigue factor.

- eliminate any recurrence of real words or distractors.

- administer at least three tests at each level, to aid reliability.

- explore the possibility of relaxing the penalties for guessing inherent in the scoring formula.

- if French speakers are numerous among the subjects, examine the cognate content of all Yes/No real words, making sure it reflects the lexicon to be assessed.

- re-examine the distractors to eliminate those which are either too "attractive" or "unattractive". It may be advisable to rely more on "phone

book" type distractors, rather than on distractors constructed from word parts.

- have thinkaloud sessions after the tests.

- investigate placement changes made later in the semester.

References

Al-Hazemi, H. A. A. (1993). Low-level EFL vocabulary tests for Arabic speakers. Unpublished PhD Thesis, University of Wales, Swansea.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In Guthrie, J.T. (Ed.), Comprehension and teaching: Research reviews. Newark, Delaware: International Reading Association, 77-117.

Anderson, R.C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. Hutson (Ed.), Advances in reading/language research: A research annual. Greenwich, CT: JAI Press, 231-256.

Ard, J. & Homburg, T. (1983). Verification of language transfer. In S.M. Gass & L. Selinker (Eds.), Language transfer in language learning. Rowley, MA: Newbury.

Chen, Z. & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. Language Testing, 2,155-163.

Corson, D. (1983). Social dialect, the semantic barrier, and access to curricular knowledge. Language and Society, 12,213-222.

Cummins, J. (1981). Age on arrival and immigrant second language learning in Canada: a reassessment. Applied Linguistics, 2, 132-149.

Favreau, M. & Segalowitz, N. (1982). Second language reading in fluent bilinguals. Applied Psycholinguistics, 3, 329-341.

Goulden, R., Nation, P, & Read, J. (1990). How large can a receptive vocabulary be? Applied Linguistics, 11, 4, Dec., 341-363.

Gove, P.B. (Ed.)(1981). Webster's Third New International Dictionary of the English Language Unabridged. Springfield,MA: Merriam-Webster, Inc.

Harlech-Jones, B. (1983). ESL proficiency and a word frequency count. English Language Teaching Journal. 37, 62-70.

Hindmarsh, R. (1980). Cambridge English Lexicon. Cambridge: Cambridge University Press.

Levenston, E. A. (1979). Second language lexical acquisition: Issues and problems. Interlanguage Studies Bulletin, Utrecht, 147-160.

McNicol, D. (1972). A primer of signal detection theory. London: George Allen & Unwin Ltd.

Meara, P.M. (1990). A note on passive vocabulary. Second Language Research, 6, 2, 150-154.

Meara, P.M. (1990a). Some notes on the Eurocentres Vocabulary tests. In Tommola, J. (Ed.), Foreign language comprehension and production, Turku: AFinLA, 103-113.

Meara, P. M. (1991). Scoring a YES/NO vocabulary test.

Meara, P. M. (1991a). Vocabulary profiles. In S. Verlinde (Ed.), Proceedings of the Symposium on differentiation in LS Learning and Teaching. Leuven, 7-10 November 1990, Leuven, Instituut voor Levende Talen.

Meara, P. M. (1992). EFL vocabulary tests. Centre for Applied Language Studies, University College Swansea.

Meara, P. (1993). The bilingual lexicon and the teaching of vocabulary. In R. Schreuder, & B. Weltens, (Eds.), The bilingual lexicon. Amsterdam: John Benjamins Publishing Co., 279- 297.

Meara, P. (1994). The complexities of simple vocabulary tests. In F. G. Brinkman, J. A. van der Schee, & M. C. Schouten-van Parreren, (Eds.), Curriculum research: Different disciplines and skills in secondary education. Amsterdam: Instituut voor Didactiek en Onderwijspraktijk, Vrije Universiteit, 15-28.

Meara, P. M. (1994a). The dimensions of lexical competence. To appear in a volume by G. Brown.

Meara, P. M., & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. Language Testing, 4, 142-154.

Meara, P. M., & Jones, G. (1988). Vocabulary size as a placement indicator. In P. Grunwell, (Ed.), Applied Linguistics in Society. London: CILT.

Meara, P. M., & Jones, G. (1990). The Eurocentres Vocabulary Size Test: 10KA. Zurich: Eurocentres.

Meara, P. M., Lightbown, P.M., & Halter, R.H. (1994). The effect of cognates on the applicability of YES/NO vocabulary tests. Canadian Modern Language Review.50, 2, 296-311.

Melka Teichroew, F.J. (1982). Receptive versus productive vocabulary: a survey. Interlanguage Studies Bulletin, 6, 5-33.

Nation, I.S.P. (1986). Word lists: words, affixes and stems (Revised edition). Wellington: Victoria University English Language Institute.

Nation, I.S.P. (1990). Teaching and learning vocabulary. New York: Newbury House.

Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners. SSLA, 9, 2, 201-219.

Palmberg, R. (1988). On vocabulary-knowledge continua and foreign-language learners' mental lexicons. In Finlance, 1988, 7, 93-100. Abo Akademi, SF-20500, Finland.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. RELC-Journal, 19, 2, Dec., 12-25.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. Language Testing, 10, 3, 355-371.

Ringbom, H. (1987). The role of the first language in foreign language learning. Clevedon, UK: Multilingual Matters.

Ryan, A., & Meara, P. (1991). The case of the invisible vowels: Arabic speakers reading English words. Reading in a foreign language, 7, 2, 531-540.

Sharwood-Smith, M. (1984). Discussion of P. Meara: The study of lexis in interlanguage. in A. Davies, C. Criper & A.P.R. Howatt, (Eds.), Interlanguage. Edinburgh: Edinburgh University Press.

Takala, S. (1985). Estimating students' vocabulary sizes in foreign language teaching. In V. Kokonen, H. von Essen and Klein-Braley,(Eds.) Practice and problems in language testing. Tampere, Finland.

Thorndike, R. L. (1973). Reading comprehension education in fifteen countries. N.Y.: Wiley.

Thorndike, E. L. & Lorge, I. (1944). The teacher's word book of 30,000 words. New York: Teachers College, Columbia University.

University of Michigan, English Language Institute (1977). Manual: Michigan Test of English Language Proficiency.

West, M. (1953). The General Service List of English Words. London: Longman, Green & Co.

Wesche, M. & Paribakht, T.S. (1993). Assessing second language vocabulary knowledge: Depth versus breadth. In B. Harley (Ed.) (forthcoming).

Wright, R.L.D.(1976). Understanding statistics: An informal introduction for the behavioral sciences. New York: Harcourt Brace Jovanovich.

Zechmeister, E.B., D'Anna, C.A., Hall, J.W., Paus, C.H., & Smith, J. A. (1993). Metacognitive and other knowledge about the mental lexicon: Do we know how many words we know? Applied Linguistics, 14, 2, 188-206.

Zimmerman, J.P., Broder, K., Shaughnessy, J.J., & Underwood, B.J. (1977). A recognition test of vocabulary using signal detection measures and some correlates of word and non-word recognition. Intelligence, 1, 5-31.

Appendix A

Sample Questions and Answer Sheet, Michigan Test

PART I. GRAMMAR

EXAMPLE A.
"What is that thing?"
"That _____ a spider."
    a) to call
    b) for calling
    c) be called
    d) is called
The correct English sentence is: "That is
called a spider." To show that d, is called,
is the best answer to this example a cross
has been made next to d for Example A on
the answer sheet.

Problems
1. "When will Fred leave town?"
   "He _____ to leave tomorrow."
    a) has
    b) must
    c) will
    d) shall
2. "What did you do last night?"
   "I _____ at a concert."
    a) sing
    b) sung
    c) sang
    d) singing
3. "Is Jack a good student?"
   "No, he is _____ in the school."
    a) worse
    b) worst
    c) bad
    d) the worst
4. "How did you know Helen was here?"
   "She _____ by some of her friends."
    a) seen
    b) is seeing
    c) has seen
    d) was seen
5. "Has Chet started teaching?"
   "No, he is _____ in school."
    a) yet
    b) then
    c) as
    d) still

PART II. VOCABULARY
    There are two types of vocabulary items
in the test. In the first type you are given
a sentence followed by four words or phrases.
You are to find the word or phrase that is
closest in meaning to the underlined word (or
words) in the sentence and that could be used
in the sentence without changing its meaning
greatly.

EXAMPLE B.
It's too windy to go for a stroll.
    a) swim
    b) sail
    c) drive
    d) walk
The word "walk" means about the same thing
as "stroll" in this sentence. The sentence
"It's too windy to go for a walk" means about the
same thing as "It's too windy to go for a stroll."
To show that d, walk, is the correct answer, a
cross has been made in the space next to d for
Example B on the answer sheet.

    In the other type of item you are given
a sentence with one word omitted and list of
four words. You are to find the word that
would best complete the sentence.

EXAMPLE C. Because of the storm and
rough waves, it would be foolish to go out
sailing today in a small _____.
    a) automobile
    b) house
    c) boat
    d) beast
The word "boat" fits best in the sentence so
that it reads "Because of the storm and
rough waves, it would be foolish to go out
sailing today in a small boat." To show that
c, boat, is the correct answer, a cross has
been made in the space next to c for Example
C on the answer sheet.

Problems
6. When the boat left everyone was very joyful.
    a) happy
    b) sad
    c) tired
    d) angry
7. We found about a score in the box.
    a) one
    b) ten
    c) twelve
    d) twenty
8. Do not leave by this door; take the other _____.
    a) departure
    b) exit
    c) outline
    d) relay
9. Write carefully; don't _____.
    a) skulk
    b) scribble
    c) skimp
    d) skin
10. The floor wasn't clean so Joe _____ it.
    a) swept
    b) kindled
    c) erased
    d) smoothed

PART III. READING COMPREHENSION

EXAMPLE D. While getting ready to go to town one morning last week, my wife handed me a little piece of red cloth and asked if I would have time during the day to buy her two yards of cloth like that. I told her I would be glad to do it. And putting the piece of cloth into my pocket, I took a train for town.
The person telling the story is. . .
    a) a married lady.
    b) an unmarried lady.
    c) a married man.
    d) an unmarried man.
    You know that the person telling this story is a married man because he says ". . . my wife handed me . . ." Because c. a married man, is the correct answer, a cross has been made in the space next to c for Example D on the answer sheet.

EXAMPLE E. The author was given a red cloth. . .
    a) in the morning.
    b) at noon.
    c) in the afternoon.
    d) in the evening.
    The passage says, ". . . one morning last week, my wife handed me a little piece of red cloth . . ." To show that a. in the morning. is the correct answer, a cross has been made in the space next to a for Example E on the answer sheet.

The influenza virus is a single molecule built from many millions of individual atoms. You must have heard of the viruses, which are sometimes called "living molecules." While bacteria can be considered as a type of plant, secreting poisonous substances into the body of the organism they attack, viruses are living organisms themselves. We may consider them as regular chemical molecules, since they have a strictly defined atomic structure, but on the other hand we must also consider them as being alive, since they are able to multiply in unlimited quantities.

11. Bacteria are . . .
    a) poisons.
    b) larger than viruses.
    c) very small.
    d) plants.
12. The writer says that viruses are alive because they . . .
    a) have a complex atomic structure.
    b) move.
    c) multiply.
    d) need warmth and light.
13 The atomic structure of viruses . . .
    a) is variable.
    b) is strictly defined.
    c) cannot be analyzed chemically.
    d) is more complex than that of bacteria.

ANSWER SHEET

Part I. Grammar

Ex. A	a( )	3.	a( )
	b( )		b( )
	c( )		c( )
	d(x)		d( )
1.	a( )	4.	a( )
	b( )		b( )
	c( )		c( )
	d( )		d( )
2.	a( )	5.	a( )
	b( )		b( )
	c( )		c( )
	d( )		d( )

Part II. Vocabulary

Ex. B	a( )	8.	a( )
	b( )		b( )
	c( )		c( )
	d(x)		d( )
Ex. C	a( )	9.	a( )
	b( )		b( )
	c(x)		c( )
	d( )		d( )
6.	a( )	10.	a( )
	b( )		b( )
	c( )		c( )
	d( )		d( )
7.	a( )		
	b( )		
	c( )		
	d( )		

Part III. Reading Comprehension

Ex. D	a( )	12.	a( )
	b( )		b( )
	c(x)		c( )
	d( )		d( )
Ex. E	a(x)	13.	a( )
	b( )		b( )
	c( )		c( )
	d( )		d( )
11.	a( )		
	b( )		
	c( )		
	d( )		

KEY

1(a), 2(c), 3(d), 4(d), 5(d), 6(a), 7(d), 8(b), 9(b), 10(a), 11(d), 12(c), 13(b).

Algorithm for Arriving at CELDT Recommendations[a]		
Standardized Test Score	Pragmatic Task Score	Recommendation
295 or less	any	reject
300 -> 345	0	reject
	1	reject
	2	reject
	3	207
	4	retest
	5	retest
350 -> 375	0	reject
	1	reject
	2	207
	3	207
	4	208
	5	retest
380 -> 395	0	reject
	1	reject
	2	207
	3	208
	4	208
	5	retest
400 -> 425	0	retest
	1	207
	2	207
	3	208
	4	209
	5	209
430 -> 445	0	retest
	1	207
	2	207
	3	208
	4	209
	5	exempt
450 -> 500	0	retest
	1	retest
	2	retest
	3	208
	4	209
	5	exempt

Note. Formula for equating raw scores: $x/2*10$ = equated score, where x is raw score.

[a]Adapted from ESL testing procedure, Oct. 93

Yes/No Tests Used in This Study

Name _____ ___
     (family name)   (given name/s)

Concordia ID number (if known):_____

Date:_____

This test has five pages. You will find these instructions at the

top of each page:

    Read through the list of words carefully. For each word:

if you know what it means, place a check mark [✓] in the box

next to the word. If you don't know what it means, or if you

aren't sure, leave the box blank.

_____

Here is an example:

    man   [✓]   You know what this word means, and you

                        place a check mark in the box.

    atwater [   ]   Although you may have seen this word

                        before, you don't know what it means,

                        and you leave its box blank.

_____

It is important NOT to guess on this test, because some of the

words on the test do not have meaning in English. When you are

not sure of a word's meaning, leave the box next to it BLANK.

What you have to do:

Read through the list of words carefully. For each word:
   if you know what it means, place a check mark (✓) in the box.
   If you don't know what it means, or if you aren't sure, leave
   the box blank.

1 [ ] lessen	2 [ ] oak	3 [ ] mosquito
4 [ ] litholect	5 [ ] quorant	6 [ ] proceed
7 [ ] interfere	8 [ ] put up with	9 [ ] algebra
10 [ ] scurrilize	11 [ ] cottonwool	12 [ ] lobby
13 [ ] give away	14 [ ] trudgeon	15 [ ] bodelate
16 [ ] tighten	17 [ ] shady	18 [ ] bance
19 [ ] awkward	20 [ ] wartime	21 [ ] draconite
22 [ ] folksong	23 [ ] outskirts	24 [ ] technology
25 [ ] stand in for	26 [ ] victory	27 [ ] antique
28 [ ] chart	29 [ ] rot	30 [ ] manly
31 [ ] compose	32 [ ] risk	33 [ ] pea
34 [ ] tunnel	35 [ ] justal	36 [ ] call up
37 [ ] combustulate	38 [ ] democracy	39 [ ] opie
40 [ ] scudamore	41 [ ] homoglyph	42 [ ] abrogative
43 [ ] react	44 [ ] haque	45 [ ] nickling
46 [ ] bench	47 [ ] snack-bar	48 [ ] charlett
49 [ ] harden	50 [ ] scorn	51 [ ] equality
52 [ ] jewel	53 [ ] pass away	54 [ ] webbert
55 [ ] kiley	56 [ ] woolnough	57 [ ] hijack
58 [ ] baldock	59 [ ] farther	60 [ ] dose

h_____ f_____ Dm_____

What you have to do:

Read through the list of words carefully. For each word:
  if you know what it means, place a check mark (✔) in the box.
  If you don't know what it means, or if you aren't sure, leave
  the box blank.

1 [ ] tune	2 [ ] glory	3 [ ] porlock
4 [ ] cicatration	5 [ ] forecast	6 [ ] doll
7 [ ] powling	8 [ ] stove	9 [ ] dessert
10 [ ] rod	11 [ ] knock out	12 [ ] nursery
13 [ ] technology	14 [ ] limp	15 [ ] look into
16 [ ] permission	17 [ ] opponent	18 [ ] aspection
19 [ ] conceitful	20 [ ] obsolation	21 [ ] sneeze
22 [ ] cundy	23 [ ] restore	24 [ ] sip
25 [ ] pungid	26 [ ] catholic	27 [ ] haque
28 [ ] bee	29 [ ] producer	30 [ ] graph
31 [ ] investebrate	32 [ ] mechanic	33 [ ] inn
34 [ ] buttonhole	35 [ ] arkless	36 [ ] headlong
37 [ ] logam	38 [ ] mourant	39 [ ] weaken
40 [ ] violent	41 [ ] snell	42 [ ] whitelock
43 [ ] around	44 [ ] batstone	45 [ ] fade
46 [ ] awkward	47 [ ] cut out	48 [ ] incarminate
49 [ ] accord	50 [ ] saratogal	51 [ ] rake
52 [ ] provided	53 [ ] morality	54 [ ] overlook
55 [ ] peebles	56 [ ] enclose	57 [ ] confidential
58 [ ] screen	59 [ ] thicken	60 [ ] enigmanic

h_____ f_____ Dm_____

What you have to do:

Read through the list of words carefully. For each word:
    If you know what it means, place a check mark (✔) in the box.
    If you don't know what it means, or if you aren't sure, leave
    the box blank.

1 [ ] rudge	2 [ ] elphick	3 [ ] disgust
4 [ ] mass	5 [ ] hoard	6 [ ] bow
7 [ ] literate	8 [ ] misrequite	9 [ ] todd
10 [ ] hollow	11 [ ] podiast	12 [ ] privilege
13 [ ] correspond	14 [ ] intimant	15 [ ] rear
16 [ ] greer	17 [ ] subtract	18 [ ] symmetry
19 [ ] pulp	20 [ ] minimum	21 [ ] insurance
22 [ ] exemption	23 [ ] compass	24 [ ] spalding
25 [ ] carotic	26 [ ] wax	27 [ ] pegler
28 [ ] even	29 [ ] arc	30 [ ] degenerate
31 [ ] penalty	32 [ ] dispose	33 [ ] brief
34 [ ] score	35 [ ] squeeze	36 [ ] opie
37 [ ] tissue	38 [ ] endanger	39 [ ] keir
40 [ ] criterion	41 [ ] riot	42 [ ] watler
43 [ ] asbestial	44 [ ] altogether	45 [ ] faminisation
46 [ ] obscure	47 [ ] upkeep	48 [ ] arbus
49 [ ] concave	50 [ ] outlet	51 [ ] install
52 [ ] constagnate	53 [ ] break without	54 [ ] register
55 [ ] caste	56 [ ] execute	57 [ ] sparse
58 [ ] fragile	59 [ ] boobier	60 [ ] activate

swansea vocabulary tests: v1.1 1992
test no A01

h_____ f_____ Dm_____

What you have to do:

Read through the list of words carefully. For each word:
  If you know what it means, place a check mark (✓) in the box.
  If you don't know what it means, or if you aren't sure, leave
  the box blank.

---

1 [ ] cranicle	2 [ ] reclaim	3 [ ] federation
4 [ ] mode	5 [ ] gammonary	6 [ ] hand
7 [ ] license	8 [ ] activate	9 [ ] stain
10 [ ] coath	11 [ ] resign	12 [ ] intrinsic
13 [ ] speculate	14 [ ] numerical	15 [ ] manual
16 [ ] respect	17 [ ] pickard	18 [ ] precaution
19 [ ] ease	20 [ ] cardination	21 [ ] urban
22 [ ] susceptible	23 [ ] overhead	24 [ ] tiny
25 [ ] innoculism	26 [ ] ingredient	27 [ ] ashill
28 [ ] concrete	29 [ ] successive	30 [ ] gummer
31 [ ] beg	32 [ ] tearle	33 [ ] piccolotomy
34 [ ] generalise	35 [ ] centripetal	36 [ ] velocity
37 [ ] wookey	38 [ ] prospect	39 [ ] charlett
40 [ ] murray	41 [ ] diversal	42 [ ] arc
43 [ ] contempt	44 [ ] amplify	45 [ ] detergent
46 [ ] capacity	47 [ ] voluminary	48 [ ] equation
49 [ ] crash	50 [ ] maltass	51 [ ] cluster
52 [ ] descendant	53 [ ] sanitary	54 [ ] acklon
55 [ ] participline	56 [ ] distort	57 [ ] perceive
58 [ ] immune	59 [ ] pruden	60 [ ] varney

---

h_____ f_____ Dm_____

96

What you have to do:

Read through the list of words carefully. For each word:
    If you know what it means, place a check mark (✓) in the box.
    If you don't know what it means, or if you aren't sure, leave
    the box blank.

1 [ ] innocent	2 [ ] mystery	3 [ ] appertonal
4 [ ] diverge	5 [ ] suck	6 [ ] traduction
7 [ ] absurd	8 [ ] sacred	9 [ ] inertible
10 [ ] urge	11 [ ] asprey	12 [ ] peasant
13 [ ] cundy	14 [ ] economy	15 [ ] rhind
16 [ ] fuse	17 [ ] comprise	18 [ ] crole
19 [ ] strip	20 [ ] basin	21 [ ] jerram
22 [ ] camp	23 [ ] eventually	24 [ ] coil
25 [ ] reside	26 [ ] promote	27 [ ] moffat
28 [ ] harmonical	29 [ ] liable	30 [ ] multiplify
31 [ ] pestulant	32 [ ] grind	33 [ ] procession
34 [ ] restore	35 [ ] dip	36 [ ] delicate
37 [ ] malleable	38 [ ] vicinity	39 [ ] incorporate
40 [ ] gummer	41 [ ] baptistal	42 [ ] timber
43 [ ] suddery	44 [ ] obstacle	45 [ ] hammond
46 [ ] perish	47 [ ] effectory	48 [ ] caste
49 [ ] selfish	50 [ ] counterpart	51 [ ] aggressive
52 [ ] contest	53 [ ] tindle	54 [ ] architect
55 [ ] curify	56 [ ] jet	57 [ ] faith
58 [ ] yandle	59 [ ] surplus	60 [ ] radical

h_____ f_____ Dm_____

Appendix D

The Decision to Eliminate Test A02

In the six preselected tests, it was found that nine real words occurred more than once, a result of constructing the test via random sampling of items from word lists.

real word	test number	item number
awkward	501	19
	502	46
restore	502	23
	A04	34
technology	501	24
	502	13
arc	A01	29
	A02	1
	A03	42
caste	A01	55
	A04	48
activate	A01	60
	A03	8
strip	A02	26
	A04	19
successive	A02	46
	A03	29
urban	A02	55
	A03	21

Seven distractors occurred more than once in the same six tests.

distractor	test no.	item no.
opie	501	39
	A01	36
gummer	A03	30
	A04	40

charlett	501	48
	A03	39
cundy	502	22
	A04	13
haque	501	44
	502	27
peebles	502	55
	A02	31
greer	A01	16
	A02	53

The number of times real words and distractors recurred per test was compiled.

instances of recurrence

Test No.	real words	distractors
501	2	**3**
502	3	**3**
A01	3	**2**
A02	4	**2**
A03	4	**2**
A04	3	**2**
total:	19	**14**

Deleting Test A02 from the study eliminated the following recurring items:

AO2

arc	item 1	also occurred in	tests A01, A03
strip	item 26	"	test A04
successive	item 46	"	test A03
urban	item 55	"	test A03
peebles	item 31	"	test 502
greer	item 53	"	test A01

Deleting Test A02 balanced out the number of recurrences across tests:

test	recurrence(words)	recurrence(distractors)
501	2	**3**
502	3	**2**
A01	2	**1**
A03	1	**2**
A04	2	**2**
total:	10	10

Test A02 (not retained)

1	arc	**31**	**peebles**
2	**detailoring**	**32**	**nonagrate**
3	distant	33	refine
4	**snape**	34	municipal
5	constitution	35	intermediate
6	consent	**36**	**degate**
7	optimum	37	idle
8	**greenaway**	38	fine
9	diffuse	39	rival
10	vanish	40	flesh
11	chromosome	**41**	**surman**
12	inert	**42**	**convolition**
13	**liverick**	43	refuse
14	specimen	44	penetrate
15	preserve	45	aggregate
16	participate	46	successive
17	harness	47	enormous
18	climax	48	bleed
19	**bibby**	49	case
20	tip	50	decade
21	**sedgebeer**	**51**	**placorate**
22	march	52	scales
23	sustain	**53**	**greer**
24	**barmion**	54	abrupt
25	**flautism**	55	urban
26	strip	**56**	**candish**
27	province	57	dye
28	**allaway**	**58**	**ottery**
29	creed	**59**	**integrality**
30	limb	**60**	**give under with**

Appendix E

Items Used on the Yes/No Tests

Of six preselected yes/no tests, five were retained for this study: two Level 5 tests based on the Cambridge English Lexicon (Hindmarsh 1980), and three A-level tests based on Nation's third thousand-word list (Nation 1986). (See Appendix C for deleted Test A02 and the reasons for its deletion). The distractors for all tests came from a single list compiled by Paul Meara using one of two methods:

A)   names were selected from telephone directory listings;
B)   plausible-looking "words" were constructed by successive rearrangements of Romance word roots and affixes.

The real words and distractors on tests 501, 502, A01, A03, and A04 are listed below. Real words are indicated in regular type, distractors in boldface. Asterisks mark the first (or second) repetition of any item.

For tests 501 and 502, the table indicates:

1)   The frequency level(s) Hindmarsh (1980) assigned to each real word. A word's various meanings may have the same or different frequency levels.

2)   The existence (if any) of each real word oh Nation's 1986 first (1st), second (2nd), or third (3rd) thousand-word list. The symbol (--) indicates that the word cannot be found on any of the Nation lists; (--) followed by related words indicates that whereas the word is not on the Nation lists per se, words very closely related to it are; these items are considered to be included on Nation's lists.

Although only one real word (restore) is found on both Level 5 (502) and Level A (A04) tests, a percentage of words on each level 5 test can also be found on Nation's third thousand word list (17% of Test 501, and 10% of Test 502). This suggests that Nation considers very few of the words on tests 501 and 502 as key university entrance level vocabulary.

For tests A01, A03 and A04, the table indicates:

The existence (if any) of each real word on Hindmarsh 1980, and the frequency level(s) at which Hindmarsh has placed it. Again, the symbol (--) indicates a word not on Hindmarsh's list; (--) followed by related terms indicates that whereas the word is not on the Hindmarsh list per se, words very closely related to it are; these items are considered to be included on Hindmarsh's list.

101

The percentage of real words on the three A-level tests which can also be found on Hindmarsh's <u>Cambridge English Lexicon</u> is fairly high (55% of Test A01, 52% of Test A03, and 37% of Test A04). The wide range of frequency levels Hindmarsh has assigned to these words shows that high-frequency words may also be considered as university entrance level vocabulary. The usefulness of these words for university studies appears to depend on more than just frequency.

Cognates with French

The researcher has marked real words which are cognate with French with the symbols + + (primary cognates) and + (secondary cognates). An explanation of these symbols and further information on cognates may be found in the Discussion, on page 79.

Distractors

The table lists the distractors on all five tests. The letters A and B indicate the source from which each distractor was taken.

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
Test 501			
1 lessen	5,6	--	
2 oak	5	--	
3 mosquito	5	--	
4 litholect			B
5 quorant			B
6 proceed++	5	3rd	
7 interfere++	5,6,7	2nd	
8 put up with	5	--	
9 algebra	5	3rd	
10 scurrilize			B
11 cottonwool+	5	--	
12 lobby	5,7	--	
13 give away	5	--	
14 trudgeon			A
15 bodelate			B
16 tighten	5	2nd	
17 shady	5,6	--	
18 bance			A
19 awkward	5,6,6,7	3rd	
20 wartime+	5	--	
21 draconite			B
22 folksong	5	--	
23 outskirts	5	--	
24 technology+	5	2nd	
25 stand in for	5	--	

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
26 victory++	5	2nd	
27 antique++	5,7	--	
28 chart++	5,6,7	3rd	
29 rot	5,7,7	3rd	
30 manly	5	--	
31 compose++	5,6,7	2nd	
32 risk++	4,4,6	2nd	
33 pea	5	--	
34 tunnel++	5,6	3rd	
35 justal			B
36 call up	5	--	
37 combustulate			B
38 democracy++	5,6,7	2nd	
39 opie			A
40 scudamore			A
41 homoglyph			B
42 abrogative			B
43 react+	5,6,7	2nd	
44 haque			A
45 nickling			A
46 bench	5	--	
47 snack-bar+	5	--	
48 charlett			A
49 harden	5	2nd	
50 scorn+	5	--	
51 equality++	5	1st (equal)	

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
52 jewel++	5,6,7	3rd	
53 pass away+	5	--	
54 webbert			A
55 kiley			A
56 woolnough			A
57 hijack	5	--	
58 baldock			A
59 farther	5	1st	
60 dose++	5,6	--	

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
Test 502			
1 tune	5,5,6,6	2nd	
2 glory++	5,6	--	
3 porlock			A
4 cicatration			B
5 forecast	5	--	
6 doll	5,7	--	
7 powling			A
8 stove	5	--	
9 dessert++	5	--	
10 rod	5,6,7	2nd	
11 knock out	5,5,6	--	
12 nursery	5	--	
13 technology*+	5	2nd	
14 limp	5,5,7	--	
15 look into	5	--	
16 permission++	5	2nd	
17 opponent	5	2nd	
18 aspection			B
19 conceitful			B
20 obsolation			B
21 sneeze	5	--	
22 cundy			A
23 restore++	5,6,7	3rd	
24 sip	5	--	
25 pungid			B

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
26 catholic*++	5	--	
27 haque			A
28 bee	5	--	
29 producer+	5,6	2nd	
30 graph+	5	2nd	
31 investebrate			B
32 mechanic++	5	2nd	
33 inn	5,7	--	
34 buttonhole+	5,7	--	
35 arkless			A
36 headlong	5,7	--	
37 logam			A
38 mourant			A
39 weaken	5	2nd	
40 violent++	5	2nd	
41 snell			A
42 whitelock			A
43 around+	5,6,7	--	
44 batstone			A
45 fade++	5,5	2nd	
46 awkward*	5,6,6,7	3rd	
47 cut out	5,5	--	
48 incarminate			B
49 accord++	5	--	
50 saratogal			B
51 rake	5,5,7	--	

107

TESTS USED	In Hindmarsh 1980 (levels)	In Nation 1986 (I, II or III)	Distractors (A or B)
52 provided+	5	2nd	
53 morality++	5,6	2nd	
54 overlook	5,6,6,7	--	
55 peebles			A
56 enclose++	5,6	3rd	
57 confidential++	5	2nd (confidence, confidently)	
58 screen++	5	3rd	
59 thicken	5	1st (thick)	
60 enigmanic			B

Note: All words on tests A01, A03 and A04 are on Nation's third thousand word list.

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
Test A01		
1 rudge		A
2 elphick		A
3 disgust++	4	
4 mass++	3,4,6,6,7	
5 hoard	--	
6 bow	5,6,6,7	
7 literate+	-- (literature, literacy)	
8 misrequite		B
9 todd		A
10 hollow	4,5,5,5,6,6	
11 podiast		B
12 privilege++	--	
13 correspond++	4,6,6	
14 intimant		B
15 rear	5,6,6,7	
16 greer		A
17 subtract+	3	
18 symmetry+	--	
19 pulp++	--	
20 minimum+	4	
21 insurance	4,5,6	
22 exemptation		B
23 compass++	5,6	
24 spalding		A

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
25 carotic		B
26 wax	5	
27 pegler		A
28 even	2,5,7	
29 arc++	--	
30 degenerate+	--	
31 penalty+	--	
32 dispose++	--	
33 brief++	4,5,6	
34 score	2,3,6,7,7,7	
35 squeeze	4,4,7	
36 opie		A
37 tissue++	--	
38 endanger+	--	
39 keir		A
40 criterion+	--	
41 riot++	--	
42 watler		A
43 asbestial		B
44 altogether	--	
45 faminisation		B
46 obscure++	--	
47 upkeep	5	
48 arbus		A
49 concave++	--	
50 outlet	5,6	

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
51 install++	4,6,7 (spelled with one *l*)	
52 constagnate		B
53 break without		B
54 register++	4,4,5,7	
55 caste	--	
56 execute++	-- (executive)	
57 sparse+	--	
58 fragile++	--	
59 boobier		A
60 activate++	-- (act, action, active, actively, activity)	

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
Test A03		
1 cranicle		B
2 reclaim++	--	
3 federation++	--	
4 mode++	--	
5 gammonary		B
6 hand	1,4,5,5,6,6,6,6,7	
7 license++	5,7	
8 activate*++	-- (act, action, active, actively, activity)	
9 stain++	4,4,5	
10 coath		A
11 resign++	5,7	
12 intrinsic++	--	
13 speculate+	--	
14 numerical+	-- (number)	
15 manual++	-- (manufacture)	
16 respect+	3,4,6,7	
17 pickard		A
18 precaution++	--	
19 ease++	4,6,7	
20 cardination		B
21 urban+	--	
22 susceptible+	--	
23 overhead	--	
24 tiny	4	
25 innoculism		B

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
26 ingredient+	--	
27 ashill		A
28 concrete+	5,6,6,7	
29 successive+	-- (succeed, success, successful)	
30 gummer		A
31 beg	5,6	
32 tearle		A
33 piccolotomy		B
34 generalize++	-- (general)	
35 centripetal+	-- (centre, central, centigrade, centimetre)	
36 velocity++	--	
37 wookey		A
38 prospect+	--	
39 charlett		A
40 murray		A
41 diversal		B
42 arc*++	--	
43 contempt++	--	
44 amplify++	--	
45 detergent++	--	
46 capacity++	-- (capable)	
47 voluminary		B
48 equation+	-- (equal, equality, equator)	
49 crash	3,3,4,7	
50 maltass		A
51 cluster	--	

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
52 descendant + +	5	
53 sanitary + +	--	
54 acklon		A
55 participline		B
56 distort +	--	
57 perceive + +	--	
58 immune +	--	
59 pruden		A
60 varney		A

TESTS USED	In Hindmarsh 1980 (levels)	Distractors (A or B)
Test A04		
1 innocent++	4,5,5,6	
2 mystery+	4,7	
3 appertonal		B
4 diverge+	-- (divide, division, divorce)	
5 suck	--	
6 traduction		B
7 absurd++	--	
8 sacred++	5,5,6	
9 inertible		B
10 urge+	4,6	
11 asprey		A
12 peasant++	--	
13 cundy		A
14 economy++	4,7	
15 rhind		A
16 fuse+	5,5,7,7,7	
17 comprise++	--	
18 crole		A
19 strip	5,6	
20 basin++	3,4,5,7	
21 jerram		A
22 camp++	3,3,7	
23 eventually++	-- (event)	
24 coil++	5,5,6,7	
25 reside++	-- (resident)	

26 promote+	-- (promotion)	
TESTS USED	**In Hindmarsh 1980** **(levels)**	**Distractors** **(A or B)**
27 moffat		A
28 harmonical		B
29 liable++	--	
30 multiplify		B
31 pestulant		B
32 grind	5,6,7,7,7	
33 procession++	3	
34 restore*++	5,6,7	
35 dip	5,6,7,7	
36 delicate+	4,4,6,6,6,7	
37 malleable++	--	
38 vicinity++	--	
39 incorporate+	--	
40 gummer		A
41 baptistal		B
42 timber++	--	
43 suddery		A
44 obstacle++	--	
45 hammond		A
46 perish++	--	
47 effectory		B
48 caste	--	
49 selfish	4	
50 counterpart++	-- (counter, v.)	
51 aggressive++	5,6,7	

52 contest++	--	
TESTS USED	**In Hindmarsh 1980** **(levels)**	**Distractors** **(A or B)**
53 tindle		A
54 architect++	5	
55 curify		B
56 jet-+	3,5,6	
57 faith++	5,5	
58 yandle		A
59 surplus++	--	
60 radical+	--	

#	STS	PTS	Rec	Date	Prog	Lang	G	V	R	501 h	501 f	501 Dm	502 h	502 f	502 Dm	A01 h	A01 f	A01 Dm	A03 h	A03 f	A03 Dm	A04 h	A04 f	A04 Dm
1	215	2	1	1	4	6	11	11	5	34	6	43	35	9	26	32	9	7	28	8	20	31	7	20
2	205	3	1	1	1	2	14	7	4	17	2	13	20	2	24	9	3		8	0	88	6	3	
3	415	3	3	1	4	36	29	28	14	33	0	83	35	1	81	34	1	78	35	0	65	28	3	43
4	415	5	3	1	4	4	26	29	12	39	3	82	39	7	60	38	7	56	36	4	93	39	7	60
5	450	1.5	5	1	4	18	37	33	17	37	0	93	37	1	87	33	0	83	37	0	38	33	1	76
6	315	2	1	1	1	14	9	27	8	18	0	45	22	1	44	16	0	40	15	0	70	17		
7	305	1	1	1	1	9	20	15	7	34	1	78	33	0	83	30	2	59	31	1	46	29	3	38
8	360	1.5	1	1	5	2	29	19	8	32	1	73	36	6	52	31	5	38	33	5	40	28	4	43
9	265	3	1	1	2	10	17	12	6	23	3	24	33	5	46	27	4	30	27	7	60	24	3	5
10	325	2	2	1	2	14	17	20	9	31	2	62	33	2	68	32	6	34	39	3	47	35	5	48
11	325	1	1	1	5	10	24	12	10	29	1	56	34	4	59	28	1	52	29	3	66	27	6	49
12	405	2.5	2	1	4	14	27	26	13	30	1	67	30	3	51	29	1	64	34	2	62	27	2	49
13	350	2	2	1	2	1	18	24	11	25	1	53	25	1	53	22	1	44	31	1	81	19	2	
14	405	2	1	1	2	5	29	25	12	31	2	62	35	2	74	33	2	68	35	0	93	33	4	61
15	445	1.5	1	1	2	14	32	38	15	35	1	81	36	3	72	35	1	81	37	6	43	35	3	68
16	205	2	1	1	1	17	8	12	5	18	0	45	21	1	41	22	4	7	19	0	84	21	3	16
17	225	1	4	1	2	4	17	9	3	16	0	40	17	0	43	17	1	28	17	2	17	14	1	17
18	435	4	1	1	2	2	36	30	15	37	4	93	39	1	92	37	1	87	38	2	33	38	0	95
19	320	1.5	1	1	1	2	21	17	7	30	0	42	28	4	34	18	1	31	18	0	81	17	1	28
20	280	1	1	1	2	2	16	15	6	22	0	55	15	1	21	17	1	28	13	0	75	12	5	9
21	410	2	2	1	4	14	26	26	16	31	0	78	33	2	68	28	2	52	35	2	74	26	7	15
22	420	4	4	1	5	14	29	31	14	24	0	60	32	1	73	30	1	67	30	5	5	23	1	47
23	355	2.5	2	1	2	9	23	19	12	33	3	83	36	4	65	34	1	78	35	1	53	38	2	56
24	290	1.5	1	1	4	23	16	16	7	27	1	68	34	0	85	23	2	47	24	1	84	26	3	56
25	315	2	1	1	2	40	20	17	7	29	0	47	30	0	75	21	3	28	25	0	75	20	2	24
26	405	3	3	1	2	39	33	26	8	37	3	87	40	2	90	37	1	75	36	3	53	37	1	75
27	410	3	2	1	6	28	29	22	17	31	1	78	35	1	81	31	2	70	30	1	45	29	1	56
28	345	1.5	2	1	2	13	24	15	11	25	0	63	28	0	70	22	0	32	25	1	46	18	6	31
29	220	2	1	1	2	10	17	7	4	22	0	55	24	3	28	22	2	55	18	0	62	11	3	6
30	335	3	1	1	1	38	21	23	4	35	3	68	35	5	55	31	4	46	31	4	36	32	4	34
31	310	4	1	1	2	2	18	21	4	30	1	67	32	0	80	33	3	61	31	2	80	27	2	40
32	340	3	2	1	5	2	25	18	6	19	2	21	27	4	30	25	4	21	20	5	32	15		
33	390	2.5	3	1	2	14	32	17	13	27	0	68	26	0	65	25	0	63	26	3	66	18	5	17
34	460	2	3	1	4	10	34	37	18	34	0	85	35	1	81	25	1	53	32	3	55	29	7	64
35	350	2	2	1	1	10	30	13	9	31	2	78	36	2	74	28	3	43	25	3	75	27	1	20
36	355	3	2	1	2	21	23	17	14	34	0	72	32	4	65	34	1	78	34	0	50	32	5	26
37	370	2	2	1	4	14	30	16	13	31	1	78	35	0	80	24	0	60	22	3	37	26	3	56
38	420	2	3	1	1	40	33	27	12	30	0	67	32	3	68	35	5	55	37	1	61	30	2	33
39	350	2.5	1	1	2	2	34	18	10	34	1	85	35	0	88	26	1	56	24	1	81	25	1	32
40	280	4	2	1	2	20	24	11	6	28	0	61	35	3	54	21	2	28	20	1	85	13	0	13
41	360	2	1	1	3	14	20	24	11	27	1	59	31	1	78	28	1	61	28	1	40	25	2	63
42	420	2.5	4	1	4	9	22	29	17	28	1	70	34	1	73	31	0	78	35	1	68	32	1	65
43	350	2	2	1	2	10	21	25	7	27	0	68	32	0	85	27	1	59	34	0	15	24	0	50
44	265	1.5	2	1	2	14	13	15	7	11	0	28	21	0	53	25	0	63	16	0		23	2	58
45	435	4	4	1	4	14	34	34	13	29	0	73	35	2	74	28	2	52	33	2		31	1	62
46	255	1	1	1	6	2	16	12	6	14	1	17	12	2		11	2		6	0		4	2	

#	STS	PTS	Rec	Date	Prog	Lang	G	V	R	501 h	501 l	501 Dm	502 h	502 l	502 Dm	A01 h	A01 l	A01 Dm	A03 h	A03 l	A03 Dm	A04 h	A04 l	A04 Dm
47	230	1	1	1	2	2	10	13	7	38	13	17	39	18	76	38	16	84	38	17	73	34	14	85
48	355	2	2	1	2	10	24	16	14	30	0	75	33	1	84	36	1	78	32	0	80	34	0	75
49	470	4	4	1	4	14	39	36	18	31	0	78	36	1	37	31	0	33	15	0	38	30	0	13
50	315	3	2	1	4	14	15	23	10	17	0	43	20	1	75	13	0	43	32	2	65	13	1	24
51	225	1.5	1	1	2	10	14	15	7	29	2	56	37	3	75	34	6	65	38	3	78	28	5	67
52	435	2	2	1	2	32	32	30	18	34	1	78	37	3	51	36	4	68	23	3	24	38	5	41
53	400	3	3	1	1	10	30	25	9	28	2	52	30	3	87	27	0	65	36	3	72	21	1	78
54	425	4	4	1	4	10	35	24	16	39	4	76	37	1	87	32	2	81	40	1	95	31	0	52
55	465	3	3	1	3	36	35	37	19	38	0	95	39	6	9	35	1	64	32	7	26	34	5	46
56	320	3	2	2	2	2	22	16	7	29	14	64	27	13	82	29	1	49	40	11	45	26	2	23
57	360	4	3	2	2	2	29	19	8	36	3		39	3	64	37	9	75	39	3	82	37	11	87
58	445	5	5	2	4	18	39	32	14	39	0	82	39	1	50	21	3	28	32	0	80	37	1	28
59	415	4	4	2	2	3	27	29	14	25	0	63	29	1	73	23	2	24	30	2	59	21	2	33
60	350	3	2	2	3	14	19	23	10	31	1	78	24	4	84	37	3	75	38	1	90	30	5	90
61	415	3	3	2	2	9	32	31	12	38	0	90	38	1	68	31	3	75	37	1	87	38	1	68
62	405	4	4	2	3	8	32	28	3	38	0	95	36	0	81	24	0	78	29	0	73	33	2	44
63	400	3	3	2	4	14	21	31	13	25	1	63	27	1	85	29	1	50	35	4	62	22	1	78
64	335	4	2	2	3	33	25	15	8	34	0	78	35	0	75	32	4	38	29	0	73	38	3	32
65	435	4	4	2	3	7	32	30	18	36	2	90	34	0	50	25	1	73	27	1	59	25	3	26
66	400	4	4	2	4	14	25	26	13	33	0	68	30	12	24	15	1	53	13	0	13	26	4	21
67	295	2	1	2	2	9	20	17	3	13	9	33	20	3	61	37	0	38	39	13	26	15	12	24
68	305	2	3	2	2	40	20	16	6	40	0	55	38	2	81	32	13	8	30	5	33	38	4	59
69	350	2.5	1	2	1	23	24	21	7	33	1	83	33	3	61	33	3	58	36	2	78	34	4	65
70	385	2	2	2	2	23	29	22	10	33	1	76	37	1	50	29	1	76	29	3	47	32	2	59
71	350	2	1	2	2	10	24	18	11	37	0	87	33	1	87	29	4	38	21	0	53	30	2	44
72	325	2	1	2	3	10	28	11	7	19	0	48	24	1	81	29	0	73	37	0	93	22	1	85
73	370	3	3	2	1	3	27	22	10	33	0	83	37	18	78	33	2	68	37	0	93	34	0	81
74	410	2	1	2	4	40	25	13	15	30	0	75	35	2	28	31	0	78	32	0	80	37	2	56
75	295	2	2	2	2	13	21	22	6	27	6	68	34	8	30	31	3	54	32	5	42	26	1	11
76	395	3	1	2	2	3	30	33	11	36	4	52	34	4	53	35	6	48	16	4		31	8	24
77	225	1	1	2	4	10	11	19	6	27	1	30	27	0	87	23	6		19	2	21	28	5	54
78	320	2.5	1	2	2	16	16	20	9	22	1	44	21	2		18	1	31	32	1	73	9	2	19
79	445	3	3	2	4	10	37	30	18	38	6	90	39	8	52	33	3	61	30	5	33	31	3	44
80	290	2	1	2	2	10	16	15	8	29	0	19	28	2	90	34	5	52	28	2	52	29	6	65
81	315	3	2	2	4	2	22	13	9	23	6	58	28	2	51	27	0	68	37	4	69	22	1	34
82	450	5	5	2	4	11	36	33	18	40	1	70	40	3	40	38	5	67	34	1	78	36	4	
83	300	3	2	2	3	14	14	22	5	29	4	64	30	3		29	3	47	14	2		28	4	9
84	255	1.5	1	3	3	13	17	12	5	24	14	17	27	18	62	23	0	58	37	18	28	12	2	
85	185	0	1	3	1	26	8	8	5	37	0		37	2	66	35	18		24	3	48	25	12	
86	340	2	2	3	1	2	20	17	12	32	6	80	31	6	34	27	2	49	35	6		18	5	26
87	385	2.5	1	3	1	19	29	20	12	34	2	43	39	1	48	33	8	22	16	5	32	35	9	
88	230	0	1	3	1	3	12	14	10	20	4	24	19	6	52	21	4	3	22	2	16	15	5	
89	355	3	1	3	1	2	28	16	8	32	2	50	35	2	52	29	2	56	21	3	34	23	5	42
90	285	1	1	3	1	27	17	13	8	29	0	56	28	5	56	25	4	12	28	4	56	21	5	34
91	295	1.5	1	3	1	14	10	22		38		95	34			23	5		28			32	5	42
92	360	2.5	2	3	2		23	21	12	20	1	37	26	1	56	15	2	4	26	1	56	19	1	34

#	STS	PTS	Rec	Date	Prog	Lang	G	V	R	501 h	501 l	501 Dm	502 h	502 l	502 Dm	A01 h	A01 l	A01 Dm	A03 h	A03 l	A03 Dm	A04 h	A04 l	A04 Dm
93	360	2.5	2	3	2	29	25	25	7	34	3	66	38	8	50	40	5	75	40	3	85	38	3	3
94	440	5	4	3	3	14	36	32	15	27	0	68	28	0	70	22	0	55	22	0	55	21	1	41
95	415	4	4	3	4	34	34	24	14	38	3	78	40	7	65	38	7	56	38	7	56	38	8	50
96	495	5	5	3	1	10	40	39	20	40	2	90	40	4	80	40	2	90	40	1	95	39	0	98
97	325	0.5	1	3	4	24	20	17	9	40	9	55	40	12	40	38	12	24	37	13	8	38	11	30
98	470	5	5	3	3	25	39	39	15	40	6	70	40	5	75	40	4	80	39	5	71	40	7	65
99	460	4	4	3	5	15	36	35	18	32	1	73	36	3	72	29	2	56	37	1	87	29	2	56
100	355	2	2	3	5	2	28	18	9	34	5	52	37	8	44	32	5	42	32	5	42	14	3	
101	390	2	2	3	3	30	19	29	14	27	1	59	29	1	64	30	2	59	30	0	75	24	0	60
102	410	3	3	3	1	36	26	28	14	31	2	70	30	3	51	27	3	40	29	3	73	22	2	32
103	275	0	1	3	2	13	16	17	3	32	1	65	26	1	56	20	1	37	23	0	24	21	3	16
104	490	5	5	3	4	36	40	39	19	40	0	95	37	4	69	38	1	90	37	0	93	38	2	84
105	405	3	3	3	1	10	29	23	14	34	1	85	34	0	85	33	3	61	26	0	65	26	0	65
106	430	3	3	3	1	21	37	26	16	39	4	92	37	3	75	38	1	90	39	4	76	37	1	87
107	325	3	2	3	5	40	17	18	11	38	4	73	37	7	51	37	5	63	40	7	65	34	8	28
108	305	1.5	1	3	1	21	23	13	6	36	0	65	39	5	71	30	4	42	28	7	3	28	6	14
109	445	4	4	3	5	34	33	33	19	36	3	90	38	3	78	32	1	73	35	2	74	36	3	72
110	320	1.5	1	3	2	40	22	13	10	37	2	75	35	4	62	26	1	56	28	4	34	32	8	17
111	305	1.5	1	3	1	10	18	14	10	21	1	28	25	2	42	19	0	48	17	0	43	18	0	45
112	345	3	2	3	2	23	27	13	11	32	8	73	29	8	73	29	0	73	22	1	44	25	7	63
113	325	1.5	1	3	2	9	23	12	11	31	0	11	37	8	44	33	8	22	34	10	11	33	2	31
114	255	1.5	1	3	2	10	21	7	5	25	3	63	30	4	42	26	3	36	24	0	60	30	1	59
115	355	2	2	3	2	2	26	20	9	22	0	44	30	1	67	22	0	55	19	1	34	13	6	13
116	465	5	5	3	4	37	38	37	16	39	3	90	40	6	70	37	4	69	38	2	84	40	3	70
117	410	1.5	3	3	3	36	33	28	8	29	0	73	32	3	58	31	1	62	30	1	67	30	3	51
118	275	1.5	1	3	2	22	14	17	5	34	3	66	35	7	41	32	4	50	34	7	37	28	2	43
119	325	1.5	1	3	3	28	23	15	8	32	1	73	35	3	68	34	3	66	31	1	70	30	5	59
120	240	0	1	3	2	35	28	13	8	31	1	70	33	0	83	29	3	47	29	5	29	28	7	33
121	235	1.5	1	3	2	23	11	10	6	34	5	52	31	7	11	31	9	1	26	10	3	34	2	3
122	410	2	2	3	3	36	25	26	17	32	1	73	35	3	55	37	3	75	37	2	81	36	2	72
123	460	5	5	3	1	12	39	33	18	40	1	100	40	0	90	37	0	87	37	1	87	32	2	78
124	450	5	5	3	4	16	34	33	20	38	0	95	38	3	84	32	1	80	33	0	76	38	1	65
125	500	5	3	3	4	11	40	40	20	39	1	92	39	0	87	38	1	90	37	0	93	13	0	90
126	405	3	3	3	5	31	23	27	16	24	1	50	30	0	75	26	1	56	25	0	63	17	1	33
127	350	2	2	3	6	9	22	19	10	23	0	58	26	0	65	20	1	37	21	0	53	28	5	28
128	335	2	1	3	3	28	18	23	7	29	3	47	33	3	61	29	0	73	33	3	61	36	1	24
129	430	4	4	3	2	10	28	32	18	31	0	78	33	1	76	36	0	90	38	0	95	24	3	84
130	425	3	4	3	3	14	32	26	18	27	0	68	30	3	67	24	0	60	29	2	56	35	2	28
131	405	3	3	3	4	14	34	22	10	34	1	78	34	0	66	32	2	65	34	2	72	24	0	74
132	355	3	2	3	1	9	22	19	13	25	0	63	30	0	75	23	0	58	27	0	68	31	1	60
133	350	2	2	3	1	9	18	23	9	29	1	64	35	0	88	29	1	64	30	2	59	26	2	70
134	360	3	2	3	2	9	24	24	9	27	0	68	34	0	85	31	2	62	35	0	88	16	2	46
135	255	2.5	1	3	2	10	15	10	9	23	0	58	30	1	67	28	2	52	20	2	24	16	2	8

Native Languages: Subjects

Code	Description	No. of speakers (total 135)
01	Amharic	1
02	Arabic	17
03	Arabic/French	4
04	Arabic/English	2
05	Armenian/Arabic	1
06	Bengali	1
07	Bulgarian	1
08	Bulgarian/English	1
09	Cantonese	11
10	Chinese	22
11	English	2
12	English/French	1
13	Farsi	4
14	French	20
15	French/Arabic	1
16	German	2
17	Greek	1
18	Greek/English	2
19	Hebrew	1
20	Hungarian	1
21	Japanese	3
22	Khmer/Chinese	1
23	Korean	5
24	Korean/French	1
25	Krio	1
26	Kurdish/Russian/Armenian	1
27	Nepali	1
28	Polish	3
29	Punjabi	1
30	Romanian	1
31	Romanian/French	1
32	Russian	1
33	Serbo-Croatian	1
34	Slovak	2
35	Somali	1
36	Spanish	6
37	Spanish/French	1
38	Tamil	1
39	Urdu/English	1
40	Vietnamese	6

Variables for Fields of Study, CELDT Recommendations, Test Dates

CODE   Description

1   **Commerce and Administration**: includes accountancy, administration, business administration, business, commerce, economics, international business, management information systems, marketing

2   **Engineering and computer science**: includes building engineering, civil engineering, computer science, computer engineering, electrical engineering, engineering, mechanical engineering, industrial engineering

3   **Fine Arts**: includes art education, art history, cinema, design art, film production, fine arts - theatre, fine arts, music, painting & drawing, photography, studio art

4   **Humanities**: communications, communications studies, early childhood education, geography, German, history, library studies, modern languages, political science, psychology, TESL certificate, TESL, urban studies

5   **Science**: includes actuarial math, biochemistry, biology, chemistry, statistics

6   **unknown**

CELDT RECOMMENDATIONS

CODE   Description

1   rejected (not admitted to Concordia/placed in ESL)
2   admitted to Concordia and/or placed in ESL 207
3   admitted to Concordia and/or placed in ESL 208
4   admitted to Concordia and/or placed in ESL 209
5   exempted from ESL courses

TEST DATE

CODE   Description

1   March 18, 1994
2   March 25, 1994
3   April 8, 1994

Pearson correlation matrix: final sample ($N = 114$)

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	1.000							
PTS	0.770	1.000						
REC	0.867	0.930	1.000					
G	0.897	0.740	0.816	1.000				
V	0.905	0.731	0.834	0.731	1.000			
R	0.851	0.667	0.760	0.720	0.735	1.000		
DM500	0.622	0.519	0.602	0.602	0.568	0.550	1.000	
DMA00	0.699	0.588	0.655	0.638	0.681	0.601	0.729	1.000

BARTLETT CHI-SQUARE STATISTIC: 1118.126 DF= 28 PROB= 0.000

MATRIX OF PROBABILITIES

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	0.000							
PTS	0.000	0.000						
REC	0.000	0.000	0.000					
G	0.000	0.000	0.000	0.000				
V	0.000	0.000	0.000	0.000	0.000			
R	0.000	0.000	0.000	0.000	0.000	0.000		
DM500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
DMA00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note. STS = overall Michigan Test result (equated)
PTS = results of pragmatic writing task, CELDT
REC = recommendation for or against admission to
Concordia & placement in ESL classes based on CELDT results (Michigan Test & writing task)
G = Michigan grammar subtest result (raw score)
V = Michigan vocabulary subtest result (raw score)
R = Michigan reading comprehension subtest result (raw score)
DM500 = Yes/No Level 5 test results (501 and 502 averaged)
DMA00 = Yes/No Level A test results (A01, A03 and A04 averaged)

Pearson correlation matrix: mixed language subgroup (final sample minus French & Arabic:
$n = 89$)

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	1.000							
PTS	0.798	1.000						
REC	0.872	0.938	1.000					
G	0.906	0.807	0.852	1.000				
V	0.915	0.749	0.843	0.768	1.000			
R	0.840	0.687	0.759	0.721	0.730	1.000		
DM500	0.619	0.562	0.626	0.603	0.577	0.549	1.000	
DMA00	0.723	0.636	0.67ʋ	0.662	0.711	0.626	0.709	1.000

BARTLETT CHI-SQUARE STATISTIC: 890.506 DF= 28 PROB= 0.000

MATRIX OF PROBABILITIES

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	0.000							
PTS	0.000	0.000						
REC	0.000	0.000	0.000					
G	0.000	0.000	0.000	0.000				
V	0.000	0.000	0.000	0.000	0.000			
R	0.000	0.000	0.000	0.000	0.000	0.000		
DM500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
DMA00	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note. STS = overall Michigan Test result (equated)
PTS = results of pragmatic writing task, CELDT
REC = recommendation for or against admission to Concordia & placement in ESL classes based on CELDT results (Michigan Test & writing task)
G = Michigan grammar subtest result (raw score)
V = Michigan vocabulary subtest result (raw score)
R = Michigan reading comprehension subtest result (raw score)
DM500 = Yes/No Level 5 test results (501 and 502 averaged)
DMA00 = Yes/No Level A test results (A01, A03 and A04 averaged)

Pearson correlation matrix: Cantonese/Chinese subgroup (*n* = 31)

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	1.000							
PTS	0.780	1.000						
REC	0.890	0.904	1.000					
G	0.879	0.760	0.816	1.000				
V	0.871	0.744	0.864	0.656	1.000			
R	0.823	0.731	0.811	0.678	0.690	1.000		
DM500	0.607	0.572	0.628	0.510	0.604	0.544	1.000	
DMA00	0.694	0.603	0.700	0.561	0.718	0.669	0.723	1.000

BARTLETT CHI-SQUARE STATISTIC: 267.260 DF= 28 PROB= 0.000

MATRIX OF PROBABILITIES

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	0.000							
PTS	0.000	0.000						
REC	0.000	0.000	0.000					
G	0.000	0.000	0.000	0.000				
V	0.000	0.000	0.000	0.000	0.000			
R	0.000	0.000	0.000	0.000	0.000	0.000		
DM500	0.000	0.001	0.000	0.003	0.000	0.002	0.000	
DMA00	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

Note. STS = overall Michigan Test result (equated)
PTS = results of pragmatic writing task, CELDT
REC = recommendation for or against admission to Concordia & placement in ESL classes based on CELDT results (Michigan Test & writing task)
G = Michigan grammar subtest result (raw score)
V = Michigan vocabulary subtest result (raw score)
R = Michigan reading comprehension subtest result (raw score)
DM500 = Yes/No Level 5 test results (501 and 502 averaged)
DMA00 = Yes/No Level A test results (A01, A03 and A04 averaged)

Appendix G-4

Pearson correlation matrix: French subgroup ($n = 17$)

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	1.000							
PTS	0.428	1.000						
REC	0.757	0.843	1.000					
G	0.930	0.445	0.708	1.000				
V	0.830	0.525	0.715	0.684	1.000			
R	0.854	0.255	0.553	0.699	0.707	1.000		
DM500	0.787	0.309	0.583	0.707	0.665	0.590	1.000	
DMA00	0.661	0.381	0.644	0.639	0.613	0.428	0.785	1.000

BARTLETT CHI-SQUARE STATISTIC: 135.532 DF= 28 PROB= 0.000

MATRIX OF PROBABILITIES

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	0.000							
PTS	0.086	0.000						
REC	0.000	0.000	0.000					
G	0.000	0.073	0.001	0.000				
V	0.000	0.030	0.001	0.002	0.000			
R	0.000	0.323	0.021	0.002	0.002	0.000		
DM500	0.000	0.228	0.014	0.001	0.004	0.013	0.000	
DMA00	0.004	0.131	0.005	0.006	0.009	0.087	0.000	0.000

Note. STS = overall Michigan Test result (equated)
PTS = results of pragmatic writing task, CELDT
REC = recommendation for or against admission to Concordia & placement in ESL classes based on CELDT results (Michigan Test & writing task)
G = Michigan grammar subtest result (raw score)
V = Michigan vocabulary subtest result (raw score)
R = Michigan reading comprehension subtest result (raw score)
DM500 = Yes/No Level 5 test results (501 and 502 averaged)
DMA00 = Yes/No Level A test results (A01, A03 and A04 averaged)

Appendix G-5

Pearson correlation matrix: Arabic subgroup: $n = 8$

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	1.000							
PTS	0.493	1.000						
REC	0.715	0.884	1.000					
G	0.944	0.483	0.684	1.000				
V	0.539	-0.044	0.120	0.278	1.000			
R	0.839	0.618	0.849	0.880	0.068	1.000		
DM500	0.409	0.228	0.145	0.399	0.403	0.158	1.000	
DMA00	0.361	0.335	0.206	0.223	0.640	-0.071	0.559	1.000

MATRIX OF PROBABILITIES

	STS	PTS	REC	G	V	R	DM500	DMA00
STS	0.000							
PTS	0.215	0.000						
REC	0.046	0.004	0.000					
G	0.000	0.225	0.062	0.000				
V	0.168	0.917	0.778	0.505	0.000			
R	0.009	0.102	0.008	0.004	0.874	0.000		
DM500	0.315	0.587	0.732	0.327	0.322	0.708	0.000	
DMA00	0.380	0.417	0.625	0.596	0.087	0.867	0.150	0.000

Note. STS = overall Michigan Test result (equated)
PTS = results of pragmatic writing task, CELDT
REC = recommendation for or against admission to Concordia & placement in ESL classes based on results of the CELDT (Michigan Test & writing task)
G = Michigan grammar subtest result (raw score)
V = Michigan vocabulary subtest result (raw score)
R = Michigan reading comprehension subtest result (raw score)
DM500 = Yes/No Level 5 test results (501 and 502 averaged)
DMA00 = Yes/No Level A test results (A01, A03 and A04 averaged)

Negative Scorers

#	STS	REC	LANG	hits					false alarms				
				501	502	A01	A03	A04	501	502	A01	A03	A04
1	215	1	6	34	35	32	28	31	6	9	9	8	7
2	205	1	2	17	20	9	8	6	2	2	3	0	3
6	315	1	14	18	22	16	15	17	0	1	0	0	3
13	350	2	14	25	25	22	31	19	1	1	1	2	4
16	205	1	14	18	21	22	19	21	0	1	4	6	3
32	340	2	2	19	27	25	20	15	2	4	4	5	4
46	255	1	2	14	12	11	6	4	1	2	2	0	2
47	230	1	2	38	39	38	38	34	13	18	16	17	14
57	360	3	2	36	39	39	40	37	14	13	9	11	11
77	225	1	3	27	27	23	16	28	4	4	6	4	5
78	320	1	10	22	21	18	19	9	1	0	1	2	2
80	290	1	10	29	28	34	30	29	6	8	5	5	6
84	255	1	13	24	27	23	14	12	4	3	0	2	2
85	185	1	26	37	37	35	37	25	14	18	18	18	12
86	340	1	2	32	31	27	24	18	0	2	2	3	5
88	230	1	3	20	19	21	16	15	2	1	4	5	5
89	355	2	2	32	35	29	22	23	4	6	2	2	5
90	285	1	2	29	28	23	21	21	2	2	4	3	5
91	295	1	27	38	34	23	28	32	0	5	5	4	5
100	355	2	2	34	37	32	32	14	5	8	5	5	3
121	235	1	23	34	31	31	26	28	5	7	9	10	7

Explanation of symbols:

STS: overall (equated) Michigan Test score (grammar, vocabulary, reading comprehension subtests); out of a possible 500.

V: Vocabulary subtest score, Michigan Test: out of a possible 40.

REC: Recommendation code for admission/placement to Concordia University based on results of Michigan Test and writing task.

> 1 = failed the CELDT
> 2 = admitted to lowest level ESL
> 3 = admitted to intermediate level ESL.

LANG: Primary language(s) as listed by the candidate on the CELDT information sheet.

> 2 = Arabic   14 = French
> 3 = Arabic/French   23 = Korean
> 6 = Bengali   26 = Kurdish/Russian/Armenian
> 10 = Chinese   27 = Nepali
> 13 = Farsi

Dm scores: Dm scores for each Yes/No test. N = negative score.

false alarms: False alarm rate for each Yes/No test. False alarms belonging to tests which resulted in negative scores are printed in boldface.

Appendix I

Distractors

The number of ticks per distractor for each test session are listed below. Subjects' final decisions were tallied, not erasures or crossouts. When a distractor "attracted" 30% or more of the 135 subjects, this percentage has been noted to the right of the list, and these distractors have been printed in bold.

	Mar. 18 $n=55$	Mar. 25 $n=29$	Apr. 8 $n=51$	Total $N=135$	%
501					
4. litholect	3	--	3	6	
5. quorant	3	6	5	14	
10. scurrilize	2	2	3	7	
14. trudgeon	--	5	5	10	
15. bodelate	--	1	2	3	
18. bance	8	3	6	17	
21. draconite	3	4	7	14	
35. justal	3	2	4	9	
37. combustulate	9	4	5	18	
39. opie	--	--	3	3	
40. scudamore	--	--	--	--	
41. homoglyph	4	5	11	20	
42. abrogative	11	2	10	23	
44. haque	1	4	5	10	
45. nickling	10	7	13	30	
48. charlett	1	4	8	13	
54. webbert	3	3	8	14	
55. kiley	1	2	2	5	
56. woolnough	2	3	4	9	
58. baldock	3	4	4	11	

	Mar. 18 $n=55$	Mar. 25 $n=29$	Apr. 8 $n=51$	Total $N=135$	%
502					
3. porlock	2	2	2	6	
4. cicatration	3	4	9	16	
7. powling	11	6	14	31	
18. **aspection**	15	11	15	41	30%
19. conceitful	10	6	13	29	
20. obsolation	11	5	19	35	
22. cundy	5	4	4	13	
25. pungid	2	1	5	8	
27. haque	2	2	5	9	
31. investebrate	5	4	8	17	
35. arkless	10	5	4	19	
37. logam	--	1	5	6	
38. mourant	5	3	6	14	
41. snell	9	5	16	30	
42. whitelock	6	3	8	17	
44. batstone	4	3	9	16	
48. incarminate	5	3	8	16	
50. saratogal	3	--	4	7	
55. peebles	9	9	10	28	
60. enigmanic	11	8	16	35	

	Mar. 18 $n=55$	Mar. 25 $n=29$	Apr. 8 $n=51$	Total $N=135$	%
A01					
1. rudge	9	6	7	22	
2. elphick	4	1	2	7	
8. misrequite	5	3	4	12	
9. todd	8	4	8	20	
11.podiast	3	1	2	6	
14.intimant	12	6	9	27	
16.greer	2	1	4	7	
22.exemptation	13	16	25	54	40%
24.spalding	6	4	11	21	
25.carotic	7	2	7	16	
27.pegler	2	2	1	5	
36.opie	1	2	6	9	
39.keir	2	--	1	3	
42.watler	4	2	6	12	
43.asbestial	2	3	7	12	
45.faminisation	10	6	12	28	
48.arbus	3	4	9	16	
52.constagnate	3	2	3	8	
53.break without	28	13	28	69	51%
59.boobier	1	3	2	6	

132

	Mar. 18 $n=55$	Mar. 25 $n=29$	Apr. 8 $n=51$	Total $N=135$	%
A03					
1. cranicle	2	3	2	7	
5. gammonary	--	--	5	5	
10.coath	10	5	10	25	
17.pickard	4	3	5	12	
20.**cardination**	14	7	19	40	30%
25.innoculism	2	2	4	8	
27.ashill	2	1	5	8	
30.gummer	4	3	9	16	
32.tearle	3	3	1	7	
33.piccolotomy	1	--	1	2	
37.wookey	3	2	5	10	
39.charlett	1	7	13	21	
40.murray	7	1	9	17	
41.**diversal**	24	16	24	64	47%
47.**voluminary**	12	9	19	40	30%
50.maltass	2	1	3	6	
54.acklon	1	1	2	4	
55.participline	10	7	10	27	
59.pruden	11	6	8	25	
60.varney	4	2	3	9	

	Mar. 18 $n=55$	Mar. 25 $n=29$	Apr. 8 $n=51$	Total $N=135$	%
A04					
3. appertonal	1	3	2	6	
6. **traduction**	23	15	29	67	50%
9. inertible	9	3	9	21	
11.asprey	5	4	5	14	
13.cundy	2	2	4	8	
15.rhind	1	1	--	2	
18.crole	4	2	3	9	
21.jerram	--	--	1	1	
27.moffat	4	--	4	8	
28.**harmonical**	30	18	33	81	60%
30.**multiplify**	30	15	23	68	50%
31.pestulant	2	3	4	9	
40.gummer	5	2	7	14	
41.baptistal	10	4	11	25	
43.suddery	2	4	3	9	
45.hammond	8	3	4	15	
47.effectory	12	7	12	31	
53.tindle	3	2	4	9	
55.curify	7	6	14	27	
58.yandle	1	2	2	5	

Total false alarms ticked, $N=135$ (all sessions):

Test 501	236
Test 502	393
Test A01	360
Test A03	353
Test A04	429

134

Comments on distractors

1. **traduction** attracts far more false alarms (64) than **mourant** (6). Both are French words, but whereas **mourant**'s shape is an uncommon one for an English word, many English words (*induction, reduction, seduction, deduction, production,* etc...) look extremely similar to **traduction**.

2. All the distractors that attracted 30% or more of subjects were constructed from the Romance word list.

Appendix J

J. (a bilingual francophone 13-year-old highly proficient in French and English) took the Yes/No test administered at CELDT at home, in March 1994.

We had been discussing the distractor *break without*, so she was aware from the start that this was not a real word.

Before beginning the test, I read the cover page to J., including the bit about the example.

During the test, J. commented on a word which occurred twice; I told her to make her decision and keep going. She also mentioned that she "knew a word in French." Without asking what the word was, I just said, "Remember, it's an English test".

J. was finished in 11 minutes.

Her scores:

Test	hits	false alarms	Dm
501:	37	0	93
502:	37	0	93
A01:	32	0	80
A03:	30	0	75
A04:	32	0	80

J.'s performance on the frequency-based Level 5 tests was much higher than her scores on the Level A tests, which had been constructed from university entrance level vocabulary lists. She checked off no false alarms.

J. engaged in a "think-aloud" session after the test was over. She described what she was thinking, if anything, regarding certain words (usually words she had decided not to tick).

Target words missed and reasons why (L1-related misses are typed in bold):

arc   -   had seen in French; wasn't sure if existed also in English.

concave   -   knew in French; wasn't sure if existed also in English. This is the "French" word she had asked me about during the test.

criterion	-	resembled *critère*; wasn't sure if it existed also in English.
malleable	-	almost checked off; knew *malléable*; wasn't sure if existed in English.
peasant	-	thought it needed an *h* (*pheasant*).
graph	-	it "sounded silly" alone. "*graphic* would have been better"
morality	-	sounded too French.
accord	-	didn't check because *accorder* is a French word.
equality	-	thought only *equal* existed, and *égalité*.
hijack	-	had seen *hijacked* in a text, but thought it needed a *gh* = *highjack*.
stand in for	-	had never heard this before, had only heard *stand for*.
caste	-	thought of the word that means "a bunch of stars for a movie", knew it couldn't be that because of final *e*.
compass	-	was unsure of its pronunciation.
symmetry	-	thought she saw an *n* in it, and knew that couldn't be right.
sparse	-	thought of *scarce*, *éparpiller*, *éparse*, but had never seen *sparse* before.
susceptible	-	knew French *susceptible*, but didn't think it existed in English.
immune	-	knew *immuniser*, but not sure *immune* existed in English.
urban	-	*urbain*.
detergent	-	in her mind, pronounced it with a hard *g*.
basin	-	thought it should have another *s*: *bassin*.

There were nine words she had never encountered before. All were on the Level A tests:

reclaim	comprise
counterpart	diverge
surplus	intrinsic
degenerate	centripetal
numerical	

J. had some comments also on some of the distractors.

crole:	it's missing an *e*: *creole*.
peebles:	it's *pebbles*, with a *b* in place of an *e*.
greer:	"looks like *gréer*."
traduction:	didn't have trouble with this because her father is a translator and she knows that the word is *translation* in English.

J.'s general approach to the test therefore seems to been one of caution: when confronted with a word she either did not know or recognized as looking French, she did not check it off.