

MX 0519368 0



Development, Implementation, and Evaluation of an Online English Placement Test at College Level: A Case Study

A project submitted to Middlesex University in partial fulfilment of the requirements for the degree of Doctor of Professional Studies (Applied Linguistics – English Language Online Testing)

S. Ioni Papadima-Sophocleous

Master of Letters, The University of New England

Master of Education, La Trobe University

Postgraduate Diploma in Computer Assisted Language Learning (CALL),
The University of Melbourne

Graduate Certificate in Education, La Trobe University

Diploma in Education, La Trobe University

Bachelor of Arts, National and Kapodistrian University of Athens



National Centre for Work Based Learning Partnerships
Middlesex University

March 2005

p1278430

Site TP	MIDDLESEX UNIVERSITY LIBRARY
Accession No.	0519368
Class No.	428.2407 PAP
Special Coll. <input checked="" type="checkbox"/>	

The Sheppard Library
Middlesex University
The Burroughs
London NW4 4BT
020 8411 5852
<http://lr.mdx.ac.uk>



Middlesex
University

This book is due for return or renewal by the date stamped
below. Fines will be charged for overdue items. R. C.

SHORT LOAN COLLECTION

Salisbury University

Salisbury University
Salisbury, MD

Salisbury University
Salisbury, MD

Salisbury University
Salisbury, MD

“There is only one thing more painful than learning from experience, and that is not learning from experience.”

Archibald MacLeish
(1892-1982)

For the things we have to learn before we can do them, we learn by doing them.

Aristotle (384 BC – 322 BC)

I grow old learning something new every day.

(Solon 636? ~ 558? B.C.)

To the memory of my grandmother Salomi Zachariadou who has been a source of inspiration for lifelong learning; to my high school English teacher Nitsa Zachariou who has been the perfect model to follow in language education.

Declaration

I hereby declare that this research project has been conducted by myself. This work is a record written by myself except where assistance and help have been acknowledged. No other person's work has been used without due acknowledgement in the main text of the research project.

Except where reference is made in the text of the research project, this research project contains no material published elsewhere or extracted in whole or in part from a research project by which I have qualified for or been awarded another degree or diploma. This research project has not been submitted for the award of any degree or diploma in any other tertiary institution.

Some of the results obtained in this research project have been presented as follows:

Meeting, Seminar or Conference presentation

28/01/2005 *Development and Implementation of a work based Online English Placement Test at College level: A case study.* Hungarian Society for the Study of English (HUSSE) 7 conference, the University of Veszprém, Hungary.

12/01/05 *NEPTON report to Intercollege Administration,* Intercollege Languages Departmental Conference, Nicosia.

14/01/05 *NEPTON report to Intercollege Faculty,* Intercollege Languages Departmental Conference, Nicosia.

03/06/04 *NEPTON report to Intercollege Administration,* Intercollege Languages Departmental Conference, Nicosia.

29/04/04 *Departmental Meeting Report: New English Placement Test Online: NEPTON*

19/11/2003 DPS 4521 Programme Planning, 40 credits at level 4, Programme Approval Panel (PAP), Development and Implementation of an Online English Placement test: A Case Study.

11/06/03 *NEPTON's First Steps*, Intercollege Languages Departmental Conference, Nicosia.

Electronic Publication

Papadima-Sophocleous S. (2004). *NEPTON (New English Placement Test Online)*, Intercollege, Nicosia, Cyprus. Website.

Papadima-Sophocleous S. (2005). *NEPTON (New English Placement Test Online) Sampler*, Intercollege, Nicosia, Cyprus. CD Rom.

Printed

Papadima-Sophocleous, S. (2005) 'Intercollege English Placement Test goes ONLINE' in *InBrief. The Intercollege Staff and Faculty Newsletter*. Issue No 10. Intercollege, Nicosia.

Papadima-Sophocleous, S. (Under consideration) 'Development and Implementation of a work based Online English Placement Test at College level: A case study, in *Proceedings of Hungarian Society for the Study of English (HUSSE) 7 conference, the University of Veszprém, Hungary*.

Salomi Papadima-Sophocleous

TABLE OF CONTENTS	Page
Declaration:	i
Contents:	iii
Acknowledgements:	x
Summary:	xv
Table of Abbreviations:	xviii
List of Figures:	xx
List of Tables:	xxii
Chapter 1. General Introduction	1
1 General introduction	2
1.1 Research project overview	2
1.2 Current needs	3
1.3 NEPTON's conception	3
1.4 What I bring into the project	6
1.5 Project components	7
1.6 Project report structure	7
1.7 My role in the various parts of the project	8
1.8 Research project aspirations	10
Chapter 2. Aims and Objectives	11
2 Aims and Objectives	12
2.1 Rationale	12
2.2 Research aim	12
2.3 Research objectives	13

Chapter 3. Literature Review	10
3 Literature Review	15
3.1 Introduction	15
3.2 L2 Teaching Approaches: A brief history of L2 Learning and Teaching	15
3.3 Second Language Assessment	17
3.4 Second Language Testing	18
3.5 Language areas	19
3.5.1 Reading Comprehension	19
3.5.2 Vocabulary	19
3.5.3 Grammar	19
3.5.4 Listening Comprehension	20
3.5.5 Writing	20
3.5.6 Speaking	20
3.6 Language Test Design	21
3.7 (L2) testing and computers	22
3.8 Electronic Test	23
3.9 CBT and CAT	24
3.10 Current EPT Practices: Pen-and-paper or Electronic Testing	25
3.11 Electronic testing device profiles	25
3.12 Review of current English Placement Tests	26
3.13 Computer Based Test (CBT) Advantages	27
3.14 Computer Adaptive Test (CAT) advantages	29
3.15 CBT and CAT Disadvantages	30
3.16 Conclusion	31

Chapter 4. Methodology and measurement techniques	33
4 Introduction	34
4.1 Factors that have shaped my project	34
4.2 Discussion of approaches and methods relevant to the project	35
4.2.1 Postmodern Approaches to Research	35
4.2.2 Work based research	36
4.2.3 Applied linguistics research	37
4.2.4 Work based and applied linguistics Research Methods	37
4.2.5 Case Study	38
4.2.6 Quantitative and Qualitative Methods and Data Collection Techniques	40
4.2.7 The Identity of the research - Rationale for choice of research approach	41
4.2.8 The Identity of the research - Rationale for choice of data collection techniques: Qualitative and Quantitative data analysis: measurement of test validity and reliability	41
4.3 Required Resources	42
4.3.1 Human Resources	42
4.3.2 Other Resources	43
4.4 Ethical Considerations	44
4.5 My Role as worker / researcher	44
4.6 Ways of overcoming subjectivity involved in approaching research as insider researcher	45
4.7 Major influences (positive and negative) on me as a worker researcher.	46
4.8 Political issues related to my research project.	47
4.9 Conclusion	48

Chapter 5. Project Account	50
5 Introduction	51
5.1 Needs Analysis	52
5.1.1 Context	52
5.1.2 Setting	53
5.1.3 Students	53
5.1.4 Lecturers	54
5.1.5 Intercollege English language programme and existing EPT	54
5.1.6 International English Language Equivalences	56
5.1.7 Test choice	57
5.2 NEPTON Test Specifications	59
5.2.1 NEPTON purpose	59
5.2.2 NEPTON description	59
5.2.3 NEPTON test sections	60
5.2.4 NEPTON item number per section	61
5.2.5 NEPTON test types	62
5.2.6 NEPTON language skills, test methods and instructions	63
5.2.6.1 Writing: Structure and Vocabulary	63
5.2.6.2 Reading Comprehension	64
5.2.6.3 Writing	64
5.2.7 NEPTON item criteria selection	65
5.2.8 Language Elements	66
5.2.8.1 Grammatical points	66
5.2.8.2 Lexical points	66
5.2.8.3 Topics	67
5.2.8.4 Settings	67
5.2.9 Timing	67
5.2.9.1 NEPTON	67
5.2.9.2 NEPTON (P)	68

5.2.10 Computer Based Tests (CBT)	68
5.3 NEPTON Development	68
5.3.1 NEPTON project committee	68
5.3.2 Test item writing and moderation	68
5.3.3 The item bank	69
5.3.4 Moderation	69
5.4 NEPTON field-testing	70
5.4.1 The computerized testing software	70
5.4.2 NEPTON Pen-and-Paper field-testing	71
5.4.3 NEPTON field-test participants' evaluation	72
5.4.4 NEPTON Pen-and-Paper field-testing: improvement needs	74
5.4.5 Revised NEPTON slide paradigm	77
5.4.6 Criterial levels of performance and placement procedures	78
5.4.7 NEPTON objective component cutoff point paradigm	78
5.5 NEPTON Implementation	88
5.5.1 NEPTON sitting	89
5.5.2 NEPTON and the test-takers	89
5.5.2.1 NEPTON test-taker demographic information an computer familiarity	90
5.5.2.2 Test taker gender, age, linguistic background and country of origin.	90
5.5.2.3 Test familiarity and test-taker equity	92
5.5.3 Taking NEPTON	93
5.5.3.1 New English Placement Test ONLINE	93
5.5.3.2 Subjective Component: hand written assessment task	95
5.5.4 Final NEPTON score and result reporting	95
5.5.5 Conclusion	96

6.	Project Findings: Results & Discussion	97
	6 Introduction	98
	6.1 Test reliability	98
	6.1.1 The NEPTON overall performance	98
	6.1.2 General factors influencing reliability	99
	6.1.3 General and individual factors influencing reliability	111
	6.1.4 Situational factors influencing reliability	114
	6.2 Test validity	114
	6.2.1 Internal, face validity:	114
	6.2.1.1 Test-taker choice of the test mode of delivery	115
	6.2.1.2 Pre and post NEPTON TEST test-taker questionnaires	115
	6.2.1.3 Post NEPTON Test test-taker questionnaire	120
	6.2.2 Internal, content validity	123
	6.2.2.1 Experts biographical data	123
	6.2.2.2 Experts Post NEPTON / NEPTON (P) and Test Specification	124
	6.2.3. External, construct validity	133
	6.2.4. Conclusion	133
7.	Conclusions and Recommendations	134
	7 Introduction	135
	7.1 Work based and applied linguistics research theories combined	135
	7.2 Research project evidence	138
	7.3 Conclusions	139
	7.4 Reflection	141
	4.5 Directions for future research	142

References	145	
Appendices	162	
Appendix 1	Review of previous learning	163
Appendix 2	Collated data from British universities	165
Appendix 3	Profile of University or College English Placement Tests	167
Appendix 4	Profile of other electronic assessment devices used in L2 testing	171
Appendix 5	Profile of Assessment Electronic Authoring Tools as part of learning online environments	174
Appendix 6	Text type inventory	176
Appendix 7	Grammar inventory	178
Appendix 8	Vocabulary inventory	181
Appendix 9	Topic inventory	184
Appendix 10	Settings inventory	187
Appendix 11	Intercollege NEPTON writing task sample	189
Appendix 12	Criteria for marking the writing task	192
Appendix 13	Calculation of final placement	194
Appendix 14	Letter to Faculty (a)	196
Appendix 15	Letter to Faculty (b)	198
Appendix 16	Student post field-test questionnaire	200
Appendix 17	Student pre-test NEPTON questionnaire	202
Appendix 18	Student post-test NEPTON questionnaire	204
Appendix 19	Experts NEPTON Test Specification questionnaire	206
Appendix 20	Printed samples of the NEPTON (electronic test)	212
Appendix 21	Samples of the NEPTON (electronic test) editor environment	217

Acknowledgements

This research project has been in the making for a long time, even many years. It is the culmination of knowledge and experience with people and students from all corners of the earth (including Europe, Asia, America, The Middle East, Africa, The Pacific) and contemplation, conversation and reading on the topic.

Thus a fascination with second language learning, and later on with teaching, testing, computer assisted language learning, and curriculum development (both printed and electronic) has been a preoccupation that has touched my life since I was a very young girl serving English-speaking customers in my mother's novelty shop in Cyprus and reading *Ladybird* books, which were given to me as presents by a primary-school teacher who used to come to our shop as a customer.

Indeed, many people have made the completion of this research project possible, and some of those who have had an obvious and direct influence are acknowledged here in appreciation.

From the depth of my heart, I first wish to thank my husband Abraham, for always being there for me throughout this challenging period. Without his continuous support, understanding, love and endless faith in me, this task could have never been completed. I am forever grateful!

My son Orestis has been a source of strength and passion for perseverance. Although he was facing change of personal, social and educational environment, having to move with the family from Australia to Cyprus at the age of sixteen, complete his final high school years, take exams and face the challenge of military service at the age of eighteen, Orestis has always been there for me, with a smile... and a cup of hot chocolate during those long hours of night study.

My parents, Andreas and Maroula, have always encouraged me in education and have continually shown that they value highly my intellectual lifelong pursuits. During difficult after-war times in Cyprus in 1974, they even migrated to Australia to make tertiary studies possible for me. A smile full of affection and pride from them makes it always all worthwhile.

My mother-in-law has also been a source of inspiration: although she had the opportunity to only complete two years of primary school education, she struggled all her life to educate her ten children and continues to the present day to be the source of encouragement for education to the whole family, her grandchildren and their children. She has also been there for me, proud and happy to see me progress.

The positive encouragement of Dr Kathy Doncaster, post-graduate Curriculum Leader for Work Based Learning Studies, and Joint Programme Leader Masters/Doctorate in Professional studies, has made all the difference during the initial stage of my DProf studies. I wish her all the best in her new job.

I thank Dr Carol Costley, Part2 Dprof Programme Leader, and Head of Work Based Learning Research Centre, for her advice and support throughout the programme. Her guidance has been very valuable.

Veronica Bailey, NCWBLP Programme and Student Administrator in Cyprus, has been continuously supportive with anything related to administrative matters of a distance-learning programme. Her warm and friendly rapport with people made it always easier to deal with administrative processes.

Savvas Kontos helped me greatly with the application procedures, and Steven Watt was always there for me with the various administrative matters that had to be dealt with from the Middlesex campus. I also extend my gratitude to any other Middlesex faculty and staff who helped me through different stages of my studies.

My Intercollege colleagues Dr Paul Gibbs, and Dr Gregory Makrides contributed to updating my knowledge in current research methods by offering such a course at Intercollege, for Middlesex university students in Cyprus. I found these sessions and their interdisciplinary aspects very enlightening and rewarding

Intercollege gave me the opportunity to pursue a work based research project that was of importance to both the institution and myself. It also supplied me with doctorate subsidy and time release during the final year of my study. For these opportunities, I will be forever grateful.

I thank Mr. Emiliou Solomou, the Intercollege Nicosia campus director, and Dr James Leigh, head of the Intercollege Language Department and Intercollege for entrusting me with the task of researching the area of online English placement testing and develop an end product to be used by Intercollege for incoming students.

I am also very thankful to Mr Stathis Mavros, the Intercollege Limassol campus director for his continuous co-operation, encouragement and support.

Dr James Leigh, who also acted as my advisor and consultant provided me with guidance and efficient scientific support throughout my studies. He has been constantly there to advise, listen, and discuss issues during all stages of the project, during easy and not so easy times. I have been very fortunate to have him as my consultant. He has not been a great advisor and consultant but a caring person as well.

I would like to express my great appreciation to Dr Dmitry Apraksin, the IT guru of Intercollege who has accepted enthusiastically to take part in the project and develop the in-house tailored electronic environment, which now hosts Intercollege's New English Placement Test Online (NEPTON).

I am deeply grateful to my colleague Chris Alexander who has been my shadow in everything that the NEPTON test needed. He showed great interest in the project and helped me immensely in all its stages, but particularly in the iteration process, the

formation of the slide paradigm, the trials, just name it, he was there. I will never forget those moments when we were both dead-tired but still found the courage to support each other and carry the project through, particularly the field-testing and the trials. Thank you, Chris... you have been a true colleague and friend.

Dr Andreas Alexandrou came as a lifesaver to the project's rescue just when the project needed his valuable knowledge of iteration processes. I'll never forget his kindness, wisdom and support.

My colleague Dr Mihalis Koutsoulis taught me how to use the Statistical Package for Social Sciences (SPSS). I entered this project with a little knowledge of statistics and I come out of it with enough knowledge to understand and use statistics for my research.

Dr Marios Hadjicharalambous entered the NEPTON scene in the final stages of the data statistical analysis and gave me a hand during those difficult times when an academic semester is about to finish and final exams need to be prepared and marked, and results to be handed in and a thesis to be dealt with at its final stages. Marios, you have been a lifesaver. Thanks to you I have become an expert in table and chart presentations.

My very good colleague Despina Telemachou from the Larnaca campus took on board the huge task of editing the whole content of the test item bank, twice. Together with Sophia Michael, who I also thank greatly, both contributed to many other stages of the NEPTON project (for example, field testing administration, language level descriptor writing, and evaluating) from the Larnaca campus, any time it was needed.

Dr Chris Pavlides from the Limassol campus also undertook the huge task of editing the whole test and checking different aspects such as cultural bias. His contribution is also greatly appreciated.

I am also thankful to Skevi Ioannou, who coordinated the administrative part of the test trial. I thank her for the times she was supposed to go home and yet she stayed at work to make sure everything was running smoothly. I also appreciate the cooperation of

Andreas Tzionis, who organized very efficiently the administration of the NEPTON test at the Limassol campus.

Without the support of all the language faculty (from all three campuses) for this programme in numerous ways (for example, editing, field-testing, trialling, giving feedback, and marking the written component) and their continuous help, a high stake project such as NEPTON would have never been able to get off the ground and this research project would have not been successful.

The Information and Technology (IT) and the administrative staff have tirelessly supported the NEPTON programme during its trial for two whole months. The administrative staff took pride in being the first ones to know about how NEPTON works. They happily came to the three in-service sessions organized for them before the test sessions, and they themselves enjoyed taking the test and having a firsthand taste of it! They organized diligently the times and venues for the test and invigilated the test sessions with great success. The IT staff were always there to set up the labs, deal with unexpected technological matters or just be supportive to students and invigilators with the use of technology. Special thanks go to Elena Papastylianou from the IT department, for entering all the data in the SPSS programme for analysis and for contributing to the NEPTON Sampler creation, together with Andreas Kekkou.

Without the students at Intercollege in all three campuses, the NEPTON test and this research project would not have been realized. Their commitment, enthusiastic and participation made the NEPTON test and this research project possible.

I cannot forget the contribution of my sister Ero Soleas and my brother-in-law Tonis Soleas. It was their laser printer, which printed the precious research project copies which were sent to the examiners for assessment.

Finally I would like to express my sincere appreciation to Middlesex University for making it possible for me to pursue doctorate studies from Cyprus.

Summary

The primary purpose of the present project was to research the case study of current English placement practices at Intercollege in view of incorporating change, improvement and efficiency, within the framework of current work based learning and applied linguistics (and more particularly English online language testing) research discipline.

The review of work based learning and current theories and practices in applied linguistics research discipline helped establish the characteristics of an insider researcher and the research approach and research techniques that would best serve such a project.

The review of current theories and practices in second language (L2) teaching and learning in general, and in L2 testing in particular revealed that there is an extensive range of practices: these range from testing discrete points to integrative tasks. Tests are also delivered both in pen-and-paper as well as in electronic form, the latter being either computer based testing (CBT) or computer adaptive testing (CAT).

The review of current English placement practices at Intercollege indicated the need for a new English placement test, developed in a scientific way, informed by current theories and practices, based on current test design models and taking advantage of more efficient methods of delivery, and placement. This review also revealed the need for more efficiency in the mode of delivery, administration, marking, reporting and test duration. Finally, this study of the current English placement practices at Intercollege established the need for a placement test that would incorporate a mechanism of continuous testing of reliability and validity as well as improvement.

The detailed study of the specific context, setting, particular language programme, resources, test-takers, instructors, etc. informed by current theories and practices in second language (L2) testing online, helped in the development of the New English Placement Test Online (NEPTON) test specifications, and as a consequence, the

development of the proposed test itself. The study of test delivery modes and the consideration of the specific work based conditions and requirements, for example administration, delivery, time and money efficiency, urgent need of an improved and more efficient English placement test (EPT) resulted in the selection of computer based testing delivery, with many features of the computer adaptive testing delivery mode incorporated in it such as randomized selection of test items and fewer items.

The test item writing and item moderation process resulted in the formation of a substantial pool of varied items in different skills, text types, topics, settings, and covering a variety of lexical and grammatical points and communicative, authentic-like situations in all six levels.

The field test which was took place in May 2004 in pen-and-paper form by almost 1200 students in all three Intercollege campuses helped check the content and the test trial which took place in the period of August-September in its electronic form helped come up with the test cutoff points, and the fine-tuning of the test. The item analysis ensured the appropriateness of all items. Pre-test questionnaires established test-takers' biographical data and information about test-taker computer familiarity. The test face validity (stakeholders' attitudes and feelings about the NEPTON) was established through the use of pre and post-test questionnaires. Experts in the area coming, from the three campuses, also studied the test specifications and the test itself (both in its electronic and pen-and-paper format) and completed a questionnaire, thus contributing to the establishment of the test content and construct validity. The test reliability was established through a split half reliability index process and a series of other aspects or processes such as the size of the item bank, the instructions, the moderation process, and the item analysis, which are explained in chapter 5 in more details.

The research project consists of two components:

- (a) The report, which describes the way work based and applied linguistics research approaches were used to investigate the case study of English placement test at

college level at Intercollege in Cyprus and to what extent this has broad change, improvement and efficiency to current practices; and

- (b) The evidence of such a research project, which is the New English Placement Test Online (NEPTON), in other words, the test itself, developed, implemented and evaluated in order to materialize this change, improvement and efficiency aimed at by this project.

Table of Abbreviations

ACT:	Australian Capital Territory
ALL Guidelines:	Australian Language Levels Guidelines
ALTE:	Association of Language Testers in Europe
BENG:	Beginner English
CALL:	Computer Assisted Language Learning
CALT:	Computer Assisted / Adaptive Language Testing
CAT:	Computer Adaptive test
CBT:	Computer based test
CEF:	Common European Framework
COBLE:	Committee of BENG level English
CSF:	Curriculum and Standards Frameworks
DEET:	Department of Employment Education and Training
EFL:	English as a Foreign Language
ENGL:	English
EPT:	English Placement Test
ESL:	English as a second language
FL:	Foreign Language
ICT:	Information and Communication Technology
IT:	Information Technology
IELTS:	International English Language Testing Service
IGCSE:	International General Certificate of
IRT:	Item Response Theory
GCE:	General Certificate of Education
GCSE:	General Certificate of Secondary Education
KR20:	Kuder Richardson 20
KR21:	Kuder Richardson 21
L1:	First Language
L2:	Second Language
LOTE:	Language Other Than English

Table of Abbreviations (cont.)

LOTE P-10:	Language Other Than English from Preparatory to Year 10
NCWBLP:	National Centre for Work Based Learning Partnerships
NEPTON:	New English Placement Online
QPT:	Quick Placement Test
P & P:	Pen-and-paper
SPSS:	Statistical Package for the Social Sciences
TAFE:	Technical And Further Education
VBOS:	Victorian Board of Studies
VCAB:	Victorian Curriculum and Assessment Board
VCE:	Victorian Certificate of Education
TOEFL:	Testing of English as a Foreign Language

List of Figures

Figure	Title	Page
5.1.	NEPTON raw data	75
5.2.	Bounds on cutoffs	76
5.3.	Lower bound cutoffs – iteration 1	77
5.4.	Cutoffs found from iteration 1 and iteration 2	78
5.5.	Cutoffs found from iteration 2 – iteration 3	79
5.6.	Lower bound, actual, and upper bound comparisons	80
5.7.	Class numbers based on lower bound cutoffs	81
5.8.	Class numbers based on derived cutoffs	81
5.9.	Class numbers based on upper bound cutoffs	82
5.10.	Lower bound cutoffs	82
5.11.	Derived cutoffs	83
5.12.	Upper bound cutoffs	83
5.13.	Test-taker biographical data	86
5.14.	NEPTON test interface	89
6.1.	NEPTON raw data	94
6.2.	Class Number based on derived cut-offs	95
6.3.	Item Analysis: Facility Value (F.V.)	102
6.4.	Comparison of electronic and written components of NEPTON results	107
6.5.	The NEPTON tutorial and computer familiarity	109
6.6.	Test-taker familiarity with technology and basic computer skills: mouse and clicking	112
6.7.	Test-taker familiarity with technology and basic computer skills: scrolling and the use of technology	113
6.8.	Test-taker familiarity with computer and the Internet use	114
6.9.	Test-taker familiarity with the use of email and chat	115
6.10.	Experts Post NEPTON / NEPTON (P) and Test Specification (a)	120
6.11.	Experts Post NEPTON / NEPTON (P) and Test Specification (b)	122

6.12.	Experts Post NEPTON / NEPTON (P) and Test Specification (c)	124
6.13.	Experts Content validity evaluation	126

List of Tables

Table	Title	Page
5.1:	Language level descriptors	52
5.2:	Exit levels for Intercollege courses as measured by IELTS and TOEFL	53
5.3:	Test questions and test items	57
5.4:	Revised, final NEPTON slide algorithm of randomization (After the test-field / trial)	58
5.5:	First programme slide algorithm	67
5.6:	Student NEPTON post field-test questionnaire: Instructions & length	69
5.7:	Student NEPTON post field-test questionnaire: activities	70
5.8:	Student NEPTON post field-test questionnaire: topics and visuals	70
5.9:	Test-taker NEPTON post-field test questionnaire: feelings before and during the test	71
5.10:	Revised, final NEPTON slide algorithm of randomization (After the test-field / trial)	73
5.11:	NEPTON cutoff point paradigm	75
6.1:	NEPTON test-taker post test questionnaire – Clear instructions	96
6.2:	NEPTON test-taker post-test questionnaire – computer clarity, activity layout and visuals	97
6.3:	Item Analysis: Facility Value – sample	101
6.4:	Item Analysis: Facility Value	102
6.5:	Item Analysis: Discrimination Index – sample	104
6.6:	Item Analysis – Facility Value and Discrimination Index Report – sample page	105
6.7:	NEPTON Correlations	106
6.8:	NEPTON test-taker post-test questionnaire – Feelings before and during the NEPTON test.	108
6.9:	NEPTON post-test questionnaire – Skill sections	116
6.10:	NEPTON post-test questionnaire – Topics and variety of activities	117

6.11:	NEPTON post-test questionnaire – Topics of hand-written component	117
6.12:	NEPTON post-test questionnaire – test length: electronic and hand written component	117
6.13:	NEPTON post-test questionnaire – time spent on the test	118

Chapter one

General Introduction

1. General Introduction

Testing has always been an important aspect of education. Educators use test to determine a student's knowledge and skill with respect to the curriculum being taught. Many important decisions are made based on the results of particular tests including placement in specific educational programmes, provision of remedial instruction, the awarding of diplomas, and eligibility for higher education and related financial assistance (Carson 2003 pp. 4-5)

In recent years, the use of computer-generated testing for placement purposes has become an increasingly appealing and efficient method of assessment. Many colleges and universities all over the world have been switching from pen-and-paper testing to computer based testing (CBT) and computer adaptive testing (CAT) for the sake of efficiency and effectiveness (Dunkel 1997). The Test of English as a Foreign Language (TOEFL), which is a proficiency test is now being offered in both computer and pen-and-paper format. Since technology generated tests are increasingly used as a method of assessment in our century in more and more educational institutions, it becomes vital to explore the capability and the potential of the use of new technologies in testing. A closer look indicates that some educators use generic, commercial software such as *Question Mark* in order to construct valid and reliable electronic tests. This may be because needs are satisfied with such tools or it is due to constraints of time and / or resources or lack of expertise in language testing or new technologies. Others choose to develop their own testing electronic environments to serve their particular setting and testing context.

1.1 Research project overview

This research project is primarily concerned with the investigation of a case study, the existing English placement test at Intercollege, with the view of offering change, improvement and efficiency. This test places students at six different English language levels, according to students' language ability prior to commencing college courses at Intercollege in Cyprus. Students study English at the level they are placed parallel to their course of studies, for example Business Administration. The investigation results in the need for the development, implementation, delivery and evaluation of a computer-

based test (CBT), an Online English Placement Test, which would place students better at their appropriate English language level and would be more efficient in administration and delivery.

1.2 Current needs

English is the language of instruction at Intercollege; therefore it is important for students to have a satisfactory knowledge of the language to be able to cope with academic English for their studies. For many years, Intercollege has been using its current English Placement Test (EPT) to place students in the appropriate levels of Intercollege English courses, which help the students support their academic college studies.

The increasing number of students studying at Intercollege, the need for more efficient EPT delivery, administration and student placement, based on current theories and practices and the need to accommodate prospective Intercollege students' specific needs in a more efficient way indicated a necessity for investigation, change and improvement. As a result, in 2003, the Intercollege Languages Department embarked on the New English Placement Test ONLINE (NEPTON) project.

1.3 NEPTON's conception

I was given the responsibility to develop the NEPTON test based on the abovementioned needs. After an initial study of current theories and practices in L2 electronic testing, I decided to devise a totally new test, especially made for the case of Intercollege and not use a test package off the shelf or adapt an already existing test for use in my organization. I decided to develop a tailored-made test that would reflect the learning context, the students' profile and the English programme of Intercollege. In other words, it would reflect the students' background, the type of English they use, their studies at Intercollege and the social and educational context in Cyprus where Intercollege is hosted. In addition, I decided to embark on the design and the

development of an in-house made electronic environment, which would host the content of this electronic test. This would be possible with the collaboration of Intercollege Information and Technology experts. I compared computer adaptive tests (CAT) with computer based tests (CBT). Although there were some advantages in CAT, I decided to adopt the CBT model for the NEPTON project for the following reasons:

- (a) CBT is already an improved form of testing over pen-and-paper. This means that it offers the change required as one of the expected outcomes of this project, together with improvement and efficiency.
- (b) NEPTON is required to cater for immediate needs. CAT needed more time, and costs to be explored and implemented. However, many of its advantages have been incorporated in the CBT NEPTON format, and often as a better solution to the CAT.
- (c) While it is definitely shorter than the pen-and-paper test but longer than a CAT could be, I believe it is fairer because it gives the opportunity to the test-taker to be tested at all levels and in various styles and question types in more than one item per level compared with what is happening in a CAT situation.
- (d) Compared with traditional pen-and-paper tests, it still provides test-takers more time per question.
- (e) The NEPTON uses the randomization process for selecting test items in three ways, item-based, test-based and level. It is based on a slide (subtest) algorithm and the cut-off points are set through a process of iteration. This gives the opportunity to test-takers to be tested in more than one type of activity and skill per level in order to arrive at the decision of their achievement at that level, instead of the smaller number of items a CAT has.
- (f) Moreover, it has a big pool of items as CAT does. These elements ensure test security, similar to that of a CAT.
- (g) NEPTON allows test-takers to revisit items during the whole time of the test, something that a CAT does not.
- (h) The NEPTON computer based test allows easy revisions of the test as a computer adaptive test would.

- (i) We have capitalized on the in-house expertise and resources.
- (j) We avoided the IRT constraints, mentioned in the literature (Canale 1986, Tung 1986). Rudner (1998) mentions that item banking and item response theory are not cure-all for measurement problems. Persistence and good judgment must remain vital aspects in any test construction and test usage effort.
- (k) Although there are CBT development costs, it was foreseen that the future costs for running and maintaining the test would be reduced compared to the pen-and-paper test, and perhaps less than the CAT, which requires more expenses for the development compared with the CBT.
- (l) An in-house made test offers the advantages of continuous improvement, both in terms of content and technology, without relying on external providers.

Most of the disadvantages of CBT and CAT, physical and performance considerations (Brown, 1997) do not seem to be issues for the NEPTON project for the following reasons: there are plenty of computers (with good size screen and graphics capabilities) and computer labs, and sufficient number of lab assistants to keep the computer equipment in working order. The testing software is developed in-house. Most test-takers are familiar with computers. There are at least two mechanisms to familiarize test-takers with the test format and mode of delivery prior to taking the test (information booklet and electronic tutorial). Moreover, specific concerns have been raised relating to Item Response Theory (IRT) constraints which support CAT (Canale 1986, Tung 1986) and its various models used (Goldstein 1979, Lord 1983, Henning 1987 and 1989, Divgi 1986). Finally, a characteristic of both types of tests (CBT and CAT), which is the potential to use objectively scored language tests (Alderson 1987, Smith 1983), is used for the NEPTON test.

The NEPTON test is therefore an original test devised for the specific needs of Intercollege students. Its content reflects and serves the particular English language placement needs of Intercollege students and the electronic environment that hosts it is

developed in-house, tailored specifically to accommodate those needs with the use of New Technologies.

1.4 What I bring into the project

I have been a senior lecturer in Second Language (L2) Education and Computer Assisted Language Learning (CALL) at Intercollege, Cyprus since the beginning of 2002. Moreover, I have been involved in the existing Intercollege English Placement Test programme. My 24-year involvement in teaching and learning languages (at both secondary and tertiary level), course delivery, both on campus and online¹, in L2 testing², programme evaluation³, curriculum development, both printed and electronic⁴, L2 research in general⁵ and postgraduate studies (see appendix 1) provided me with a solid background for such a research project. All these helped me shape my present knowledge and skills and contributed to their evolution. At the same time, I saw the changes in L2 testing approaches and modes of delivery and the need for constant change and improvement. During my curriculum writing, I came across varied ways of using the technology for teaching and learning purposes, which impressed me to the point I decided to extend my expertise in L2 printed and electronic curriculum development by further exploring the area of online testing. This opportunity was given

¹ Melton, Lalor and Eltham High Schools: French, Greek, English and Italian as L2/FL; La Trobe and RMIT Universities and Intercollege: L2 teaching, L2 Methodology and CALL, on campus and online: 1979-present.

² Leaving Victorian Certificate of Education (VCE) chief assessor, examiner, vetter, and curriculum writer, Victorian Board of Studies (VBOS), Australia (AU): 1982-2000; *Ça Alors*, CIS & Heinemann Publications, Au, 1998-2000; *Voilà 1 & 2*, John Wiley and Sons Publications Au, 2002; Intercollege New English Placement Test Online, 2002-present.

³ Greek L2 material evaluation, (MEd preliminary dissertation, 1991; EDIAMME programme: Evaluation of teaching material for the teaching of Greek as a foreign language in Australian schools, (1996-98); Accreditation: Victorian Course Design and Curriculum Standards Frameworks (CSF), P-10, (1998-2000); LOTE/ESL curriculum comparison (1998); French teaching material review, Australia (2002)

⁴ Papadima-Sophocleous: *Voilà 1* (2002); *Voilà 2* (2002): printed, CD and Web; Gault and Papadima-Sophocleous *Voilà 1* and *Voilà 2* Teacher Resource Books. 2004; *Ελληνικά! Γιατί όχι; (Greek! Why not?)* (2001): printed; *Ça Alors 2* (2000), *Ça Alors 1* (1998): printed and audio; Sophocleous S. *Γεια σου (Hi)* (1992): printed.

⁵ Papadima-Sophocleous, Development of an online graduate Diploma in Educational Studies: LOTE: Greek, RMIT University, (1999-2000); Collaborative school language project: *French-Australian Links* (1998-2000); Multicultural Education, 1998; *Sister school program Templestowe Primary School and 1st Experimental School of Ioannina*, (1996); *Fabien, un Jeune Stéphanois* (1994). European programme – Socrates, Comenius 2.1: development of 3 modules, Teacher training in teaching Greek as an L2 and the use of ICT in L2, 2003-2006.

to me at Intercollege where I started working three years ago. The development of a New English Placement Test Online (NEPTON) for the student English Placement Testing needs and the cooperation of Intercollege with Middlesex work based programmes offered me the new challenge of embarking on such a research project within the framework of both work based and applied linguistics research traditions.

The Head of Languages Department assigned me to lead the review Committee of Beginner level English (COBLE) (Papadima-Sophocleous et al. 2004) of the existing Intercollege EPT programme, and also develop NEPTON, thus contribute to change and improvement (see more details on pages 8-9: 1.7).

On a practical level, this meant improving efficiency in running, and delivering EPT practices, in reporting results and in placing students more efficiently, based on current theories and practices in L2 online testing at Intercollege. On a research level, it gave me the opportunity to conduct an applied linguistics research as an insider researcher and gain insights into developing, implementing and evaluating a model of an effective placement programme for a specific tertiary (college) English learning context.

1.5 Project components

The project consists of two components: the first one is a case study. It comprises a research in the area of work based learning and applied linguistics research, in the specific area of L2 online testing: the case study of the specific language testing context at Intercollege, a development, implementation and evaluation of an English Placement Test Online. The second component is the NEPTON, the actual test, designed and developed for Intercollege in my capacity as a senior lecturer in L2 education, CALL and Testing.

1.6 Project report structure

The project report is preceded by a synopsis and consists of seven chapters.

Chapter 1 sets the parameters of the research project.

Chapter 2 contains the aims and the objectives and with the statement of the problem, the topic is placed in its historical framework.

Chapter 3 cites the pertinent literature on the theme from various sources worldwide.

Chapter 4 refers to the investigation methodology and measurement techniques.

Chapter 5 presents the project activity.

Chapter 6 presents the project analysis, interpretation and discussion of the results.

Chapter 7 summarises the conclusions and cites recommendations regarding the different stages of NEPTON as a work based / applied linguistics research project.

1.7 My role in the various parts of the project

This is a work based project, in other words, as a researcher I had to act as an inside researcher. This involves carrying out my duties at my workplace and carrying out research at the same time. I was given by the head of my department the task to review the English placement practices of our institution and improve them, taking advantage of New Technologies. To do this, I had to review the whole of the English language programme of Intercollege, because the English placement test constitutes an integral part of it and should reflect it. For this reason, I was also given the responsibility to lead the Committee of Beginner Level English (COBLE). Its aim was to review the English language programme at Intercollege at all three campuses. This involved organising meetings, delegating tasks, coordinating work, running meetings at all three campuses, informing colleagues about the aim of the review committee and the latest theories in L2 teaching, learning and testing (including L2 language level descriptors), organising committee members into subgroups (based on different levels), and compiling the subgroup suggestions together into a report. This task also involved managing differences of opinion and leading the group into some kind of consensus. It also involved management of different needs and tendencies within and between campuses, with the aim to come up with a report that would satisfy all colleagues, and all needs and would offer improvement to the English language programme. In addition, I had to

create a new test and lead all the people who would be involved in the creation and the implementation of it, people such as colleagues, administrators, experts, and students. I studied the theories and practices in L2 testing online in order to gain the necessary knowledge needed for the task. I developed the test specifications, and the test design tailored to the unique needs of Intercollege. I created an item bank, in other words I wrote the test items. I then came up with the algorithm, which was used to choose different test items for each unique test (originally two 9-item slides for each level). I coordinated the test item moderation process and made the appropriate changes suggested by the moderators (the moderators were colleagues from all three campuses). I organized and coordinated the field-testing with the help of a colleague from my campus who acted as my assistant during the whole duration of the project. I delegated the coordination of the field-testing at each other campus to a colleague from the respective campus. The field-testing involved testing the items in pen-and-paper form with more than 1400 students at all three campuses. My assistant and I organized the test papers according to the groups and the colleagues acting as invigilators, distributed them, then we collected them after the test, and sorted them out. Information technology assistants scanned the answer sheets and entered the results in the computer. Then, I analysed the data in terms of content, and after some discussion with my assistant, I made some changes. These included the change of the slide algorithm to one 9-item algorithm per level, which included a more representative choice of items (reflecting for example different types of activities, and skills). Based on the findings, my assistant, together with a statistician contributed greatly in the conception of the cut-off point algorithm and ran the iterations to come up with alternatives for the first cut-off points. These cut-off points were later on changed after the trial of the electronic version of the test and the inclusion of the written component. The decision for the final cut-off points was taken as a group, by the head of the department, my assistant, the IT expert and myself. Another aspect of my role was to continuously work very closely with the IT expert, in order to come up with an in-house, tailor-made electronic environment that would best host the designed test and run it. With the help of my assistant, I organized and ran in-service training sessions for people such as the administrators, the invigilators, and the markers. I compiled in-service training booklets. I wrote questions for the written

component of the test, and with my assistant, we compiled a set of criteria for the written task. Moreover, I had to carry out item analysis. With the help of IT assistant and a statistician, I entered the data in SPSS and Excel. Then I studied the data, came up with an item analysis method and ran an item analysis to find the item facility value and discrimination index. I also researched the various ways to test reliability and validity. I chose the ones that seemed to be the most appropriate for the particular project and used them to analyse the data in order to see the effectiveness of the testing programme, its validity and reliability. I compiled an information booklet about the New English Placement Test Online (NEPTON) for incoming students. To be able to implement such a project I had to lead, coordinate and organize in different ways more than 2200 participants (such as test-takers, faculty, experts, information technology personnel, and administrators). I had to continually face the unexpected and come up with the quickest and the best possible solutions. I had to make decisions, take responsibilities, make difficult and delicate negotiations, present and discuss the project progress and results, and make novelty accepted by involving as many stakeholders as possible.

1.8 Research project aspirations

With the findings, the recommendations and the conclusions of this project, alongside the New English Placement Test Online (NEPTON) (see evidence online / on CD ROM), which I have developed as part of my research, I hope to have offered new knowledge in the area of work based learning and the school of applied linguistics research, change, improvement and more efficiency in placing prospective Intercollegic students at the appropriate English language level courses in a more scientific, and efficient way.

Chapter two

Aims and Objectives

2. Aims and Objectives

2.1 Rationale

Intercollege, a private tertiary institution, needs a new placement test to replace the existing one. Its purpose would be to assign new students to classes at six English levels: English Beginner (BENG): BENG-50, BENG-80, BENG-90, BENG-100, and English (ENGL): ENGL-100 and ENGL-101. Existing course objectives at all levels are expressed in rather general 'communicative' terms. There is a need for test improvement and sufficient and accurate information on overall ability in the language for placement purposes. More than eight hundred new students enroll within a matter of days at the beginning of the each academic year. The test needs to be efficient, quick and easy to administer, score and interpret.

2.2 Research aim

The aim of the NEPTON research project is to bring change, improvement and more efficiency to the existing Intercollege English Placement practices, by developing a new English Placement Test, based on current theories and practices in L2 testing, including the use of new technologies. The research project investigates a case study of an English Placement Test programme developed for specific college needs, and the possible understanding and new knowledge it can contribute to the already existing general knowledge about language testing using new technologies. Its originality rests in the fact that: (a) it deals with the specific English placement needs of a particular context in contrast to the generic English Placement Tests (EPTs) that exist and are broadly used (in other words, it takes into consideration aspects such as the setting, the students, the syllabus, and the student ethnic background) (b) it is also a work based (in other words the researcher is an inside researcher) and applied linguistics research project. (c) although it is only a computer-based and not a computer adaptive test, it is based on a 9-item 6-slide paradigm, which permits randomization of item selection from the whole test item battery as well as a randomization within a choice of two tests for

both the electronic and the pen-and-paper version. Finally, the tangible product of this work based learning and applied linguistics research project is significant to Intercollege as it aims to provide changes in English placement practices, in placing students more efficiently in English courses and administering the test more efficiently, in other words facilitating better practices at a specific setting more efficiently.

2.3 Research objectives

The following are the objectives of this research:

- (a) To explore possibilities of combining work based learning and applied linguistics (L2 online testing) research method in order to investigate change and improvement in English placement testing, with the use of new technologies.
- (b) To produce evidence of such a work based and applied linguistics research project in the form of an online English placement test, catering for the specific case of a college.

Chapter three

Literature Review

3. Literature Review

3.1 Introduction

The aim of this chapter is to review the research literature in the area of applied linguistics research (available range of research approaches and techniques) over the last two decades in order to use them as a researcher within my own practice-based context (work based learning) to:

- (a) Acquire knowledge and an understanding of L2 teaching approaches, L2 assessment and testing, L2 learning areas and skills, language test design, L2 testing and computers, Electronic testing, Computer based testing and computer adaptive testing, English placement practices, both pen-and-paper and electronic, their advantages and disadvantages and investigate the possibility of using such knowledge for a research project, which aims to bring change, improvement and efficiency in placement practices through the development, implementation and evaluation of an Online English Placement Testing programme.
- (b) Develop an understanding of the main criticisms that have been made of work in these areas.
- (c) Develop skills and abilities expected from an insider researcher in applied linguistics.
- (d) Read analytically and synthesize with the aim of using this new knowledge to research the present project.

3.2 L2 Teaching Approaches: A brief history of L2 Learning and Teaching

Effective (L2) Testing in the 21st century entails the in-depth knowledge of current theories and practices in L2 language teaching, learning and testing. It also involves in-depth examination of current theories and practices in the use of new technologies in L2 testing.

I went through an extensive study of L2 learning and teaching theories⁶, references⁷, and curriculum design⁸. Furthermore, I studied the Common European Framework of Reference for Languages (CEF), (Council of Europe 2001)⁹ the language level descriptors of the Association of Language Testers in Europe (ALTE)¹⁰(UCLES 2003) and other sources such as Common Standards Frameworks (CSF) (1988), VCE Course Design (1990). In addition, I examined computer assisted language learning and information and communication technology (ICT)¹¹ in the area of L2 testing, which followed similar route as the L2 teaching approaches¹². Lastly, I examined testing model designs suggested by various researchers¹³. As a result, I established a basic

⁶ *Grammar-Translation Method*, (1840-1940); *Direct Method*, (end of the nineteenth century and the beginning of the twentieth); *Audio-lingual* approach, end of the 1950; (Larsen-Freeman 1986; Thanasoulas 2002); *Silent way*, 1970 (Cattegno 1972); *Suggestopedia*, *Community Language Learning* (Curran 1976). *Communicative Language Methods: Functional-Notional Approach*, 1980, (Finocchiaro and Brumfit 1983); Total Physical Response (Asher 1977); the Natural approach; Cooperative Language Learning; the content-Based Teaching; the Task-Based Teaching, 1990; the task-oriented language learning; the content-oriented language learning; the cognitive language learning; the process-oriented language learning; the learner autonomy; the situated learning; the collaborative learning. Learner-Centred (Nunan 1991), knowledge construction and collaboration, problem solving, holistic learning, multi-disciplinary work, etc. followed into the 2000s with interactivity as one of its main characteristics, particularly towards meaningful interaction in multimedia programmes for language teaching (Felix 1997) and the post-Methods Era (Richards, Rodgers (2001).

⁷ Brumfit and Johnson (1979); Canale and Swain (1980); Krashen (1982); Stern (1983); Johnson and Porter (ed.) (1983); Brumfit (1984); Larsen-Freeman (1986); Richards and Rodgers (1986); Nunan (1987, 1988, 1989, 1991, 1995); Johnson (1989); Lo Bianco (1989); Harmer (1991); Ministry of Education (1987, 1990); Lightbown and Spada (1993); Yule (1994); Richards and Rodgers (2001);

⁸ Munby (1978); Brumfit (1979, 1984); Johnson (1982); Yalden (1983, 1987); Breen (1984); Brumfit (1984); Richards (1984, 1985); Nunan (1985a); Dubin and Olshtain (1987); Scarino, Vale, and McKay (1988); Richards (2001); Kitao and Kitao (accessed 2003).

⁹ Council of Europe (2001); this document provides a practical tool for setting clear standards to be attained at successive stages of learning and for evaluating outcomes in an internationally comparable manner. The Framework provides a basis for the mutual recognition of language qualifications, thus facilitating educational and occupational mobility. It is increasingly used in the reform of national curricula and by international consortia for the comparison of language certificates. The European Union Council Resolution of November 2001 recommended the use of this Council of Europe instrument in setting up systems of validation of language competences.

¹⁰ ALTE Framework (2004) ALTE is an association of providers of European foreign language examinations, including some of the major international providers in the field of language testing. One of ALTE's main aims is to make clear how qualifications achieved in different languages correspond to each other and what they actually mean in practice. This has the effect of making qualifications more usable and of increasing people's potential mobility.

¹¹ O'Shea and Self (1983); Kenning and Kenning (1984); Wyatt (1983, 1984); Higgins (1983); Kenning, and Kenning (1990); Warschauer (1996); Godwin-Jones (1996); Tella (1996); Levy (1997); Patrikis (1997).

¹² Behaviouristic CALL (1960s & 1970s), Communicative CALL (late 1970s and 1980s), Integrative CALL (1990s) and Interactive CALL (current).

¹³ Spolsky (1968); Savignon (1972); Kelly (1978); Rea (1978); Morrow (1979); Oller (1979); Moller (1982b); Skehan (1988); Hughes, (1989); Weir (1990, 1995); Heaton (1998); Godwin-Jones (2001); Fulcher (2002).

background and frame of reference for the theoretical basis needed for a New English Placement Test Online development. In general, current theories in L2 test development suggest the following:

3.3 Second Language Assessment

In current assessment practices, a broad range of student performance is covered in order to maximise the discovery of personal strengths. Each student is asked to demonstrate what he or she can do in the form of specific and clear tasks rather than competing with each other. Opportunities are provided to participate in communicative and interactive use of the target language in a wide range of activities. Assessment in languages involves the assessment of language performance based on the communication activities, which are central to classroom learning (Scarino et al. ALL Guidelines, book 3, p. 46). The teaching content and the learning processes form the basis for assessment, with the proportion of time spent on the teaching and learning of a particular topic, deciding the amount of emphasis that is given to it (LOTE Frameworks, year 7-10, Course Development Guide).

Consideration is also given to all components of language learning (for example speaking, listening, reading, writing, pronunciation, vocabulary, grammar, and cultural awareness) in different realistic, contextualised and interactive situations. Students are exposed to communicative data, which is comprehensible and relevant to their own needs and interests. They are exposed to sociocultural data and direct experience of the culture(s) embedded within the target language. Students are given the opportunity to become aware of the role and nature of language and of culture. Great attention is given to help students become aware of their present stage of development, build on their strengths and acknowledge the limitations they should try to overcome. Provision is given so that assessment would assist teacher and learner(s) to monitor progress in both the 'process' as well as the 'product' of learning (Scarino et al. 1988). The aim of assessment is to also motivate learners and teachers, to inform the teaching and learning process, and to guide future curriculum planning, to inform other relevant people, to encourage cooperation styles of work, to encourage responsibility and involvement

(Clark 1987). There are different ways in which selection of the most appropriate methods of collecting information can occur and students' performance can be assessed. These should be similar to good classroom exercises and activities, which can be used as excellent means of assessment. (Scarino et al. 1988). Assessment needs to take into account the principles and practices in learning in general, and those in L2 learning in particular, in order to be in accordance with the general aims of learning, and, as an integral part of learning, with the rest of the components of the particular L2 learning programme. New principles and practices in L2 assessment, therefore, have become an integral part of L2 teaching and learning in order for the latter to be more successful and meaningful.

3.4 Second Language Testing

The same applies for testing. It is an integral part of teaching and should reflect the process and expected outcomes of learning. As Heaton mentions in *Writing English Language Tests* (1998, p. 8) "... ways of assessing performance in the four major skills may take the form of tests of:

- Listening (auditory) comprehension, in which short utterances, dialogues, talks and lectures are given to the test takers;
- Speaking ability, usually in the form of an interview, a picture description, role play, and a problem-solving task involving pair work or group work;
- Reading comprehension, in which questions are set to test the students' ability to understand the gist of a text and to extract key information on specific points in the text; and
- Writing ability, usually in the form of letters, reports, memos, messages, instructions, and accounts of past events, etc..."

Many current tests still isolate the language areas learned. Test sections generally include: structure and usage, vocabulary, reading comprehension, writing, and sometimes speaking (Heaton 1998). The following are some examples of test items in those areas suggested in the theory of L2 testing:

3.5 Language areas

3.5.1 Reading Comprehension

Test items testing reading comprehension can take the form of: multiple-choice questions, multiple-choice short or longer texts, unique answer, short answer questions, guided short answers, cloze, cloze elide, information transfer summary cloze, selective deletion, gap filling, C-Tests, identification of order of events, topics, or arguments, referents, and guessing the meaning of unfamiliar words from context word matching, sentence matching, picture and sentence matching, true/false questions, completion items, rearrangement items, and open-ended items (Weir 1990, 1995; Hughes 1989; Heaton 1998).

3.5.2 Vocabulary

Test items testing vocabulary can take the form of: multiple-choice questions, short answer questions, cloze, cloze elide, selective deletion gap filling, C-Tests, information transfer multiple-choice for recognition of synonyms, definitions; gap filling, picture item identification multiple-choice items, sets (associated words), matching items, completion items, and more objective items (word formation test items, items involving synonyms, rearrangement items, definitions) (Weir 1990, 1995; Hughes 1989; Heaton 1998).

3.5.3 Grammar

Grammar test items can take the form of: paraphrase, completion, modified cloze (Hughes 1989); multiple-choice grammar items, rearrangement items, completion items, transformation items, items involving the changing of words, 'broken sentence' items, pairing and matching items, combination and addition items (Heaton 1998).

3.5.4 Listening Comprehension

Test items testing listening comprehension can take the form of: multiple-choice questions, short answer questions, information transfer, dictation, partial dictation, listening recall note taking, use of recordings or live presentations phoneme discrimination tests, tests of stress and intonation, statements and dialogues, testing comprehension through visual material, and understanding talks and lectures (Weir 1990, 1995; Hughes 1989; Heaton 1998).

3.5.5 Writing

Test items testing writing can take the form of, for example, editing task, essay test, controlled writing tasks, and summary (Weir 1990, 1995); Hughes (1989) also suggests letters, postcards, notes, forms, announcements, descriptions, narration, comment, expressing thanks, requirements, opinions, comment, attitude, confirmation, apology, want/need, and information. Heaton (1998) separates the written tasks he suggests into levels: At basic level he suggests postcard, diary entry, and forms. At intermediate level, he proposes similar written task as in the basic level as well as guide and a set of instructions. For the advanced level he recommends similar test items as in the intermediate level, plus newspaper report and notes.

3.5.6 Speaking

Test items testing speaking can take the form of: verbal essay, oral presentation, free / controlled interview, information transfer: description of a picture sequence or a single picture, interaction tasks, information gap student – examiner, role play oral interview, interaction with peers, response to tape recordings, questions and requests for information, interpreting, tape-recorded stimuli, imitation, prepared monologue, and reading aloud, conversational exchanges, using pictures for assessing oral production, short talk, (group) discussion and role-playing (Weir 1990, 1995; Hughes 1989; Heaton 1998).

3.6 Language Test Design

In order to decide on the most appropriate test design, it is useful to be aware of the various language testing approaches and their limitations in terms of the criteria of validity, reliability and efficiency. A review of language testing approaches reveals three main approaches: the psychometric (structuralist approach), the psycholinguistic (sociolinguistic approach) and the communicative paradigm. Although their distinction is neither real nor absolute, it describes testing in terms of the particular focus each approach represents (Weir 1990).

The first one mainly tests 'discrete' linguistic items. Many researchers (Savignon 1972; Kelly 1978; Spolksy 1968; Rea 1978; Morrow 1979) argued that although testing a candidate's linguistic competence is a necessary component of a test battery, it is not sufficient to test language proficiency, as it has no relationship to the use of language in real life contextualised situations.

The second one claims that, contrary to the 'discrete point' tests, 'global integrative tests' measure the ability to integrate disparate language skills in ways, which more closely approximate the actual process of language use (Oller 1979). A number of aspects of this approach were criticised, aspects such as the validity of the tests, (Alderson 1978a), and their indirectness (Kelly 1978; Morrow 1979; Moller 1982b; Rea 1978).

Finally, the third approach, the communicative paradigm claims to test language ability. It claims to evaluate samples of performance, in certain specific contexts of use, created under particular test constraints. Such tests replicate to some degree conditions of actual performance (Skehan 1988). They claim to assess communicative language ability in a meaningful setting and context: linguistic, discoursal and sociocultural (Weir 1990), through authentic tasks and genuine texts. It is commonly accepted that there is a lot of research to be carried out in various aspects of the communicative paradigm in language testing: for example, the effect of the techniques of measurement of language

performance, the holistic and qualitative assessment of productive skills, the tests' validity, and reliability.

3.7 (L2) testing and computers

Computers have played a key role in language testing since 1935 (Fulcher 2002). However, the use of computers in language testing did not become widespread and generally available until the expansion of the personal computer in the late seventies and early eighties (Godwin-Jones 2001). Since then, there have been many computer programmes developed for L2 electronic testing¹⁴. Moreover, testing mechanisms built into Web course management systems such as *WebCT* or *Blackboard* also facilitate testing over the Internet (Godwin-Jones 2001). It is therefore evident that from early scoring devices to the latest Computer Adaptive Tests (CATs), computers have come to play a major role in test construction, item banking, test administration, scoring, data analysis, report generating, research, and the dissemination of research.

In recent years, testing theories and practices in L2 have been changing in a number of ways. In line with political and economic pressure to demonstrate 'efficiency and effectiveness', educational institutions, and instructors develop tools and procedures for testing learners' linguistic ability in a more efficient way. At this stage, however, relatively little research has been done about the various aspects of current language testing practices, particularly with the implications added by the use of technology.

Today there are various different practices enabling L2 testing: testing software practices, testing mechanisms built into web course management systems, and testing programmes developed by institutions or instructors to tailor particular needs. English Placement or English Language Performance tests such as IELTS, TOEFL, QPT, and GCSE exist in both pen-and-paper and / or electronic, adaptive form. Although these

¹⁴ Calis, Dume University; Dasher, University of Iowa; HyperCard, Apple; Toolbook, Asymetrix; WinCalis; MaxAuthor, University of Arizona, (Godwin-Jones B, 2001) as well as Internet applications which enable test delivery over the Internet. Examples of such Internet applications are WebPractest, Hot Potatoes, Interactive Quiz Makers, Matchmaker2, Placemaker3, etc.

tests are recognized worldwide for their efficiency and quality and used widely, on the one hand they are mainly generic and on the other, research conducted around their implementation is not exhausted.

Many L2 researchers, practitioners, and experts in testing organizations are actively engaged in designing, developing, and using computerized placement and achievement computer tests or investigating various aspects of L2 computer tests¹⁵. Many aspects of electronic L2 testing in general are yet to be researched for example CALT Design Issues, CALT scoring issues, CALT logistical issues (Brown 1997), validation procedures for different types of media use, oral testing over the web, the possibilities of virtual reality for near-perfect task authenticity and performance-based testing (Roever 2001), testing writing online, and work based online college language placement testing.

3.8 Electronic Test

The two main types of electronic tests are:

- (a) Computer-Based Tests (CBT): this refers to a test, which is delivered electronically using a personal computer. A CBT can be administered independently of time and location. It provides new types of questions, for example, point-and-click, scroll, drag and drop, usually multiple-choice questions presented in a linear form. The maximum marks give the score. CBTs usually have many questions and are rather long. However, they reduce answering errors compared with filling in the small circles or squares on answer forms for pen-and-paper tests. They are less expensive, for example there is no paper waste, and no markers are used. They allow for better design and continuous item improvement. Lastly, they enhance test-taker motivation.

¹⁵ Eignor, Way, Stocking, and Steffen (1997); Larson and Madsen (1985); Kaya-Carton, Carton and Dandoli (1991); Stevenson and Gross (1991); Carlson (1994); Henning, Johnson, Boutin, & Rice (1994); West (1995); Yao (1995); Yoes (1996); Johnston (1996); Linacre (1996); Luecht (1996); Larson (1996); Weiss (1996); Brindley (1997); Brown (1997); Taylor, Jamieson, Eignor, and Kirsch (1998); Negata (1998); Chalhoub-Deville (1996, 2001); Godwin-Jones (2001); de Jong (1996); Dunkel (1996, 1997, 1999, 2000); Fulcher (1999, 2000); Roever (2001).

(b) Computer Adaptive Tests (CAT) Computer Adaptive Testing (CAT) and in languages Computer Adaptive Language Testing (CALT) refer to the use of tests, which are tailored to the needs of individual test takers. CALTs make use of technology in such a way that each item is selected from a large item bank depending on the previous response of the test taker. This means that all test-takers are given an item of "average" difficulty when they start the test. After the answer is given, the question is scored immediately. If a test-taker gets the answer wrong, a less difficult item is selected by the programme for the second item. If the test-taker gets the first item right, an item of similar difficulty will be selected. If this is also answered correctly, a more difficult item is selected. This process continues until it is clear that a test taker has performed consistently at a particular level and it is established that that is the level, which can confidently be used to report the test taker's attainment. Low levels of measurement error are required for high-stakes certification tests and indicate that the test would likely produce a similar score if re-administered right away, so reliability is high. Because it is unclear exactly when the test will end, a CAT usually presents a variable number of questions, and a minimum and maximum number of questions are typically set. The score is not based on the number of correct answers but is derived from the level of difficulty of the questions answered correctly. The score is computed statistically and is based on the principles of Item Response Theory (Lord 1980).

3.9 CBT and CAT

The main advantage of a CAT over a traditional computerized test design is efficiency. Because it avoids presenting questions that provide no help in determining the person's score (that is, questions that are too easy or too hard), the CAT can determine a person's score with fewer questions, sometimes reducing the length of the test by more than 50 percent. This efficiency is the main reason why certification candidates overwhelmingly prefer adaptive tests and why many institutions have adopted this measurement technology (Grist 1989).

3.10 Current EPT Practices: Pen-and-paper or Electronic Testing

A placement test, and in this case an English placement test, is a test that pairs a student with an appropriate English course. At two- and four-year institutions in many countries, English placement tests are used to measure readiness for college-level course work and help faculty decide in which English course students should enroll. Institutions often employ scores obtained from a variety of tests such as Test of English as a Foreign Language (TOEFL), Quick Placement Test, International English Language Testing System (IELTS), General Certificate of Education (GCE), or locally designed English Placement tests, in order to make admission and placement decisions for students, including students for whom English is a second or foreign language.

English placement tests existed in pen-and-paper form long before the electronic English placement tests. They have mainly been in the form of multiple-choice questions and essay writing. Little by little many of them are gradually being transformed into electronic form. Many such English Placement Tests (in pen-and-paper or electronic form) are currently being used by various universities or colleges for incoming students, particularly those of non-English speaking background. The review of a number of them established how current theories and practices in L2 testing in general and in L2 electronic testing in particular have been put into practice.

3.11 Electronic testing device profiles

The English Placement Tests of a number of tertiary institutions were looked at¹⁶. The testing tools of a number of electronic assessment devices were also examined¹⁷. These devices offer a substantial variety of techniques which can be used for electronic

¹⁶ Norwalk Community College, San Francisco State University, Columbia University, The California State University, University of Cincinnati, Pennsylvania College of Technology, Oxford and Cambridge Universities; QPT: Quick Placement Test, TOEFL: Test of English as a Foreign Language; IELTS: International English Language Testing System (IELTS): Jointly managed by the University of Cambridge Local Examinations Syndicate, the British Council and IDP Education Australia; The Michigan; ACT Educational Services Division; PEAK English, the online Interactive English School.

¹⁷ *Question Mark*: Courseware Developer with the School of Biological Sciences, University of Sydney; *Hot Potatoes*; *Quia*; *WebAssign*, *WebCT*; *Blackboard*.

language testing, such as true/false and yes/no, matching, multiple choice, random selection question, question groups and selection by group, fill in the blank and short answer, essay, multimedia support such as audio, video, MIDI, Shockwave, Flash, and streaming media.

3.12 Review of current English Placement Tests

A review of the above representative types of English Placement Tests revealed the following:

Many universities and colleges have developed their own English Placement tests, primarily for non-English speakers, for their own English Placement testing needs. They are mainly pen-and-paper tests, testing reading comprehension, and grammar in mainly multiple-choice question form and writing in essay form. They are taken and assessed on campus. Their purpose is to place students according to their level of English (see appendix 2).

English Placement tests such as the *Quick Placement Test* of Oxford and Cambridge Universities, *TOEFL* (Test of English as a Foreign Language), or *IELTS* (International English Language Testing System), are used to evaluate the English proficiency of people whose native language is not English for admission to undergraduate and postgraduate studies. These English tests exist in both pen-and-paper and computer-based form. They mainly assess listening, structure, reading, and writing. They primarily assess in multiple-choice question form with some using other types of activities such as completing a sentence, selecting a synonym, and fill in the blanks, and they are of a generic type. Although there's criticism of levels such as CEF and standardised tests such as IELTS and TOEFL (Alderson 1987, Fulcher 2003), it is widely recognized that they are still one practical way of comparing with generic widely used levels and standardised tests (see appendix 3).

There are also independent agencies, which offer English Placement Testing services, agencies such as ACT Educational Services Division and PEAK English the online Interactive English School.

Moreover, there are free or commercial electronic assessment devices, which are used in L2 testing, and more particularly for placement. Examples of such devices are *Question Mark*, *Hot Potatoes*, and *Quia*. As we have seen, these devices offer a substantial variety of techniques, which can be used for electronic language testing. However, they are generic, restrictive in their functions because they only provide templates of specific test activity types and, more importantly, they depend on central, external control and can prove quite expensive because there is a user fee (see appendix 4).

Finally, there are Assessment Electronic Authoring Tools, which form part of comprehensive learning online environments such as *WebCT* and *Blackboard*. They also offer a wide range of testing techniques mentioned above but have the same restrictive characteristics as the commercial electronic assessment devices, in that they only provide templates of specific test activity types and they also depend on central, external control for which user fee needs to be paid (see appendix 5).

3.13 Computer Based Test (CBT) Advantages

From what we can establish so far, many of the testing techniques used in electronic form derive from the techniques suggested in the L2 testing theory literature by, for example, Hughes (1989), Weir (1990), and Heaton (1998), for pen-and-paper tests. Computerized testing, however, whether adaptive or not, provides some significant advantages (Brown 1997, Dunkel 1999, Roever 2001) over pen-and-paper ones for the test taker, the test publisher, and anyone else who uses the test results. These benefits are:

- (b) Computer tests usually take less time and are therefore more efficient. Research has shown that a 50-question computerized test takes much less time to administer than a 50-question pen-and-paper test.

- (c) They allow test-takers to work at their own pace, without time limit.
- (d) Many test-takers like computers are more motivated and even enjoy the testing process (Stevenson and Gross 1991). It seems that taking a test on the computer is more interesting and less intimidating than taking the test on paper.
- (e) Test-takers may find that computer-based tests are less overwhelming (as compared with equivalent pen-and-paper tests) because the questions are presented one at a time on the screen rather than in an intimidating test booklet with hundreds of test items.
- (f) Electronic tests can also provide more variety in kinds of testing techniques, which can prove more interactive and authentic like.
- (g) New types of questions (point-and-click, drag and drop, and simulations) improve the test's ability to measure the important skills.
- (h) The computer allows for better test designs with more flexible item selection routines. CATs use better item selection algorithms.
- (i) CB Ts can provide improved test security and as a result test results are more meaningful. Random ordering of test questions, adaptive testing, and other innovations expose test questions less often and make it impossible for one examinee to successfully cheat by copying another's correct answers. New performance-based test questions are difficult (and sometimes unnecessary) to memorize for future test coaching.
- (j) Apart from including text, graphics, photographs, computerized tests can even include interactivity, sound, and full-motion video clips.
- (k) Test-takers make fewer extraneous errors, answering computerized test questions, than they do when filling in the small circles on answer forms for pen-and-paper tests.
- (l) Computerized tests are much more accurate at scoring selected-response test results than are pen-and-paper tests, or oral exams, and at reporting scores, than human beings are.
- (m) Computers score everyone the same way and do not consider such factors irrelevant to the score as examinee gender or culture.

- (n) The test can be scored immediately, providing instantaneous feedback for the examiners and the examinees.
- (o) Computers can also provide statistical analysis of results.
- (p) Computer based tests are superior in terms of reliability and validity (Dunkel 1991).
- (q) Computer based tests can be administered quicker.
- (r) Computerized testing makes it easy for a test publisher to insert experimental and unscored questions into a test.
- (s) The increased use of computers in test development and delivery may reduce the testing costs in the future for the test developer, test publisher, test user, and examinee.
- (t) Computerized tests can be administered at times and locations more convenient to the test taker, in other words, they can be administered online at any place and at any time.

3.14 Computer Adaptive Test (CAT) advantages

According to the literature, CATs have some additional advantages (Dunkel 1991):

- (a) Although computerized tests have proved to be more efficient than pen-and-paper ones, CATs seem to be even more efficient. They are shorter and quicker because they use even less test items to arrive at the test-taker's level.
- (b) Compared with traditional exams and typical time limits set for them, experience has shown that CATs, even with their shorter time limits, provide more time per test question, compared to both pen-and-paper and computer based tests.
- (c) CATs improve security in several ways. First, CATs expose items at a much-reduced rate, allowing the items to be effective for a longer period of time. Second, test-coaching efforts that focus on individual items are less effective because it is not clear which items a person will be presented with.
- (d) An adaptive test is as accurate as any other test at determining a pass/fail decision; however, for providing a score for high- and low-ability individual, it is actually more accurate. Traditional tests are less accurate at these extremes.

- (e) Being able to keep the testing times down reduces the test delivery costs. For some test publishers this means being able to reduce testing costs to test-takers.
- (f) CATs allow easier revisions of the test. If a test question is not functioning well, it can be removed without a complete republishing of the test. Development and publishing costs are saved without affecting the integrity of the test or resulting scores.
- (g) CATs produce a similar psychological test-taking experience for everyone. Each test-taker answers questions that are personally challenging without being too hard or too easy. Boredom from answering many easy questions and frustration from answering too many hard questions can be avoided.
- (h) Test-takers prefer adaptive tests for many of the advantages described above.

3.15 CBT and CAT Disadvantages

Research also refers to some disadvantages using computers in L2 testing. Brown (1997), for example, divides these disadvantages into two categories: physical considerations (computer equipment not always available or in working order; reliable sources of electricity; screen size limitations; graphics capabilities) and performance considerations (student familiarity with using computers or keyboards). Dunkel (1991) also talks about high cost of computer hardware and software, the time required to acquaint test-takers with the computer and the potential of evaluating not only language but also computer skills. Cohen (1984) and Madsen and Larson (1986) are concerned with the potential bias in computerized exams against those unfamiliar with the new technology.

Despite all these limitations, CBTs and CATs offer many, useful advantages that could be further explored or exploded.

3.16 Conclusion

Current L2 teaching and testing approaches, L2 test designs, and the use of technology in these areas, were investigated. Computer based tests advantages over pen-and-paper tests were discussed. These include: measurement of reliability, validity, repeatability, efficiency, and standardization. Such aspects are which are difficult to achieve with traditional pen-and-paper tests. CBT benefits range from ensuring individualized assessment to making possible the efficient measurement of one or more broad dimensions of an individual's aptitude or achievement (e.g. structure, vocabulary, and reading) for example, as in the NEPTON test's case, placement purposes. Computers also offer a relatively inexpensive, secure, and standardised alternative to the traditional L2 proficiency test, particularly for the kinds of tests that can be mass administered (e.g. structure, vocabulary, reading comprehension proficiency test). The design of a computer-based test also offers the possibility of creating a more efficient and positive testing experience for the test takers. CBTs also save test time, are more efficient in test management and can provide immediate test reporting.

It was decided to develop a tailor, in-house made computer-based English Placement Test for the following reasons: As mentioned before, although widely used, existing such tests do face criticism and need to be researched more. On the other hand, they are generic, and expensive. An in-house made test would be tailored to the specific needs of its setting, and its test-takers. It would draw from and reflect the institution's curriculum and as a result would cater for the specific needs of the test-takers. It would have initial development costs but would run efficiently after that. It would have the advantage of being changed according to changing needs, at any time. Furthermore, it would provide hands on research knowledge.

In Cyprus, there is a number of public and private tertiary institutions. Some of them have English as the language of instruction. For this reason, they have English placement test to place their students in their English language level courses. All of them, however, use pen-and-paper English placement tests. The NEPTON test at

Intercollege is the first English placement test used in Cyprus tertiary level in an electronic form and based on the latest theories and practices in L2 testing. Moreover, within the wider international perspective, it offers an example of an English placement test offered online, which has been designed and developed for specific, local needs. This means that the NEPTON Test is based on current theories and practices in L2 electronic testing and at the same time it reflects the students' needs and the English language programme of the specific organization and that its mode of delivery has been especially tailored, and in-house made to cater for those needs. Moreover, the NEPTON test offers different suggestions of test item presentations, through the locally devised algorithm and alternative ways of reaching cut off points through a second algorithm devised for this reason. The NEPTON test is an example of the design, development, and implementation of an English placement test in an institution which is relatively small in student numbers and human and other resources; however, the NEPTON test proves that it is possible for such institutions to make the most of current theories and practices at a local level, without having to involve large numbers of students and experts and spend large amounts of money. Instead, it creates the environment where local expertise is involved in real life research problem solving, and is given the opportunity to contribute to research in general.

Chapter four

Methodology and Measurement Techniques

4. Introduction

In this chapter I will describe the research approach, the research methodologies and the measurement techniques I chose for the project. More particularly, I will refer to the following: factors that have shaped the project; a discussion of approaches and methods relevant to the project, the identity of the research (the rationale for choice of the research approach and the data collection), validity and reliability; required resources; ethical issues; and my position in the research.

Research in L2 usually arises directly from the disciplinary frameworks in which knowledge has been traditionally ordered within the university. This project moves beyond this framework and attempts to research English language Online Testing within the applied linguistics framework combined with work based learning, in other words, within my own practice-based context.

The project consisted mainly of two components:

The first component was a case study of the development and the implementation of an Online English Placement Test at college level.

The second component, which was the evidence, was the *New English Placement Test Online* (NEPTON) developed in my capacity as a senior lecturer and project NEPTON writer (see evidence in electronic form and samples of the electronic test in printed form in appendices 20 and 21)

4.1 Factors that have shaped the project

The factors that have shaped this research project have been the needs of the college I worked at for a more efficient, updated and improved English Placement Test, the work based learning nature of my programme of study and my 25-year accumulated knowledge and experience in L2 teaching (English, French, Greek and Italian), teacher-

training, curriculum development and evaluation and testing, computer assisted language Learning (CALL), and information and communication technology (ICT), both on campus and online.

4.2 Discussion of approaches and methods relevant to the project

4.2.1 Postmodern Approaches to Research

All approaches to research are a reflection of cultural beliefs about the world we live in and want to live in. Postmodern approaches to research reflect the contemporary decline of absolutes and a questioning of the belief that following the correct methods guarantees true results. They reflect an awareness of the complexity and socio-historical contingency of the practices through which knowledge is constructed about the world and us. They emphasize the need for science to be self-reflective about its limitations. They acknowledge the socio-cultural contexts, the intimate inseparability of knower and know, the known and the means of knowing, the impossibility of separating the subjects and the objects of research challenges, the assumption of an 'objective' world and the existence of complexity, uncertainty, heterogeneity and difference. This kind of approach was mostly appropriate for this project since the latter reflected a specific socio-cultural, work based context and was based on a process of experiential learning (Kolb 1996, Houle 1980, Jarvis 1987) through uncertainty, reflection, and a quest for a specific world, in this case, that of placing students at the appropriate English language levels at Intercollege in Cyprus. Reflective learning (Brockbank and McGill 1998), as opposed to propositional learning used in traditional, academic environments where knowledge is acquired in a technically rational value framework (Schon 1983, 1987) is a constructionist view of reflective practices which enhances a learner's critical and reflective abilities, where there is action on which to apply the knowledge and where realities of professional life and practice are taken into account.

The literature in both work based and applied linguistics was primarily written in English. This reflects the long history of research in both areas.

4.2.2 Work based learning

Work based learning is research carried out by experienced and valued professionals while working at their workplace, without taking a career break (Boud 2001). The work based researcher is an insider researcher who draws from the work place in order to conduct research. Work based learning has introduced new ways of researching. The research degrees constitute an integral part of the researcher's professional work. As a result researchers obtain high quality postgraduate research degrees (Masters, and Doctorate). In support of such practices, in his critique of the neglected epistemological assumptions of educational research, Usher (1996) claims that to be effective practitioners we ought to try and be researchers of our practice.

In work based learning, working and learning are coincidental, however not the same. Learning and work influence, reinforce and complement each other. The learner / worker learns to be able to manage both roles. Work based learning aims are twofold: the production of an end product directed by work and the acquisition of knowledge directed by learning. As Boud (2001) claims, "... most work-related learning involves the development of knowledge of use in improving present practices or processes or in developing practices or processes for the future..." Knowledge is generated through work. The work place is a potential site of knowledge production, as only universities used to have been thought of traditionally. Work based learning is not based on standard curriculum. Instead, it equips learners to be continuing learners and productive workers through engagement with research tasks that extend and challenge them, taking them beyond their existing knowledge and expertise, in a way that enables them to deal with unknown situations. Unlike traditional university learning which expects students to focus on a particular issue or problem and practise solving it until they become experts, work based learning provides students with the opportunity to work out what the problem is. Moreover, work based learning is developing an appreciation of what was actually learned, in other words, stand back and reflect on the learning. It is developing the ability to reflect on knowledge and learning and be able to transfer it to other situations. For, as Bowden and Marton claim (1998, p. 25), one "... can never re-

enter the very situation which gave birth to learning. Transfer is involved in every instance of learning; questions of transfer are simply questions of learning...” Further Davis and Chisholm (2003) say that “...if learning is the true goal of higher education, and the awarding of the qualification the evidence of this learning... then that learning does not have to be done ‘in the academy’. Greater recognition should be given to that learning achieved in the home, at work and in the community...” All in all, it is the combination of academic and professional knowledge.

4.2.3 Applied linguistics research

Research in applied linguistics, and more particularly in L2 usually arises directly from the disciplinary frameworks in which knowledge has been traditionally ordered within the university, and acquired mainly in a technically rational value framework. However, as Hudson (1999) points out, “Applied linguistics provides, the theoretical and descriptive foundations for the investigation and solution of language-related problems, especially those of language education (first-language, second-language and foreign-language teaching and learning)... it is concerned with professional activities whose aim is to solve ‘real-world’ language-based problems...” The section of applied linguistics dealing with the research in English language teaching played a pioneer role in the search for effective and efficient second language (L2) online testing. This project attempts to research English language Online Testing by combining this aspect with the framework of work based learning. The approach of the research methodology of this project has been shaped by three factors: the needs of my institution for a New Online English Placement Test, the work based learning nature of my study programme and my background knowledge and involvement in L2 (English, French and Greek) teaching, curriculum development, testing and evaluation.

4.2.4 Work based learning and applied linguistics Research approach

I have studied the various research approaches used in work based learning (Armsby 2000) such as Action Research, Soft Systems Methodology, Case Study, Experiments,

Surveys (Fink 1995, Funk 1995), and Ethnography¹⁸, and those used in applied linguistics research (experimental method, ethnography and case study) in order to choose the most appropriate one for the specific research project. The Case Study used in both work based learning and applied linguistic research approach seemed to be the most appropriate since this research project was a case study of the development and the implementation of an English Placement Test Online at a specific tertiary institution, in other words, at the place where I worked.

4.2.5 Case Study

Case Study is a strategy for doing research. It involves carrying out postgraduate level empirical investigation on a particular, current phenomenon within the context of a specific workplace (Usher 1996), using multiple sources of evidence. The 'contemporary phenomenon', in other words, the 'case', can be virtually anything (individual, multiple cases, group, institutions, an innovation, or a programme). The sampling strategy involves questions like WHO (which persons are observed, interviewed, etc.), WHERE (in or about) which setting are data collected?), WHEN (at what times?) and WHAT (which events, activities or processes are to be observed, etc.?) The Case Study is an opportunity for one aspect of a problem to be studied in some depth within a limited time scale (Davey 1991). It allows the researcher to concentrate on a specific instance or situation and to identify various interactive processes at work. The majority of case studies are carried out as freestanding exercises. The researcher identifies an 'instance' (Bell 2003), which could be, for example, as in the case of this work based and applied linguistics research project, the introduction of a New English Placement Test Online (NEPTON), and observes, questions, and studies. Gillham (2000) defines Case Study as a unit of human activity embedded in the real world, which can only be studied or understood in context, which exists in the here and now that merges in with its context so that precise boundaries are difficult to draw. It can be a single case or an educational programme, such as an English Placement Testing

¹⁸ Bell, J. (1999); Cohen, L. & Manion, L (2000); Gilham, B (2000); Yin, R. M. (1989, 1993); van Maanen, J. (1995); Gillham, B. (2000); Funk A. (1995); Eignor, D. & al. (1997); Salkind, N.J. (2000).

Programme. The *Case Study* approach is ideally suited to the resources and environment of a work based researcher. Any research technique may be used within case studies. It is vitally important to provide a full context for cases studied. A case study investigates the above to answer specific research questions and seeks a range of different kinds of evidence which is there in the case setting, and which has to be abstracted and collated to get the best possible answers. It takes place at particular times in particular places with particular people.

I could fully capitalize on my existing skills and knowledge with the case study research method from the applied linguistics tradition, however this research project gave me the opportunity to (a) use it within the tradition of work based learning and applied linguistics research discipline and (b) add to these skills and knowledge in using more quantitative strategies and become familiar with statistics and electronic data analysis devices.

The chief limitation on the value of case study is the question of how far understanding of a specific case, with its rich context, can be transferred to other situations (Denscombe 1998), to the extent to which findings from the case study can be generalized to other examples, that is, the findings may not have reliability. Work based research projects are often very context specific, therefore this approach can be revealing of specific work based research situations, particularly when these are compared with others. A large number of subjects used in a case study is one way with which this limitation can be softened. In the present project, it is envisaged to use between 800 to 1200 subjects. Another way is to follow Bassey's (1981) preference to use the term 'relateability' rather than 'generalizability'. In this option, an important criterion for judging the merit of a case study is the extent to which the details are sufficient and appropriate in a similar situation to relate decision making to that described in the case study. According to this theory, the relateability of a case study is more important than its generalisability.

4.2.6 Quantitative and Qualitative Methods and Data Collection Techniques

Research methods are the ways used to solve a problem, to reach an objective, and to accomplish an end result. According to Leedy (1989), “all research methodology rests upon a bedrock axiom: *the nature of the data and the problem for research dictate the research methodology...*[and in general terms,] if the data is verbal, the methodology is *qualitative*, if it is numerical, the methodology is *quantitative*”.

Quantitative methods

According to the literature, Quantitative methods involve statistics, in other words, counting and measuring. There are two types of statistics: descriptive statistics are like averages (usually called ‘means’), which ‘describe’ data in a summary fashion. Inferential statistics are those, which enable to draw potentially meaningful and significant inferences from quantitative data. In case study research, Quantitative methods take a wider view. Questionnaires are the most common data collection technique used. For this project, inferential statistics, questionnaires and the data from the New English Placement Test online (NEPTON) will be used.

Qualitative Methods

Qualitative Methods focus primarily on the kind of evidence (what people tell the researcher, what they do) that will enable the researcher to understand the meaning of what is going on. This can only be seen from the perspective of those involved through interviews, diaries, and observation, notes, open-ended questions, which are the most common data collection techniques used. For this project the insider researcher diary, open-ended questions and participants’ notes are used as data collection techniques.

Case study and Quantitative and Qualitative Methods

Case study research is not exclusively concerned with qualitative methods: all evidence needs to be pulled into the case study research data collection. Quantitative and qualitative methods are often used together. Statistical results, for example, need to be

described and interpreted. This present research project will employ the case study approach and will try and reach its objectives through the use of both quantitative and qualitative methods and their respective data collection techniques.

4.2.7 The Identity of the research - Rationale for choice of research approach

David Nunan (1995) talks about qualitative and quantitative research, discusses various research methods used in applied linguistics (action research, experimental method, ethnography, and case study) and describes various data collection techniques such as observation, diary keeping, surveys, and questionnaires. These are research methods and techniques also used by work based research learning, and therefore I will pursue the research project using the methods and techniques utilized in both work based learning and applied linguistics research. I will adopt the case study because the research questions deal with a particular case in a particular context, that of English Placement Test at a specific college in a specific country. Although the research project involves participation from a representative number of stakeholders (students, administrators, and lecturers), and it aims at bringing programme change, the expected change does not deal with them as members of a particular social group. From the six types of case studies that exist (Illustrative, Exploratory, Critical Instance, Programme Effects, Cumulative and Implementation) I will choose to use Programme Effects case study. It is related to the improvement of a programme that would work for all stakeholders more efficiently.

4.2.8 The Identity of the research - Rationale for choice of data collection techniques: Qualitative and Quantitative data analysis: measurement of test validity and reliability

Institution documents and academic literature in both printed and electronic form such as policies, syllabi, course outlines, EPT guidelines, tests, journal, and research articles were studied and analysed in a qualitative way. Discussion of the comparison with the old English Placement test was another source of qualitative analysis, which provided information about possible change, improvement and efficiency (in placement, time, and

administration) of test practices. Case study research, however, is not exclusively concerned with qualitative methods. Quantitative analysis and the use of various analysis techniques such as classical measurement theory for item analysis and iteration for working out the test cut-off points have been used. Quantitative analysis of the test results established reliability and some forms of validity. Inferential quantitative analysis of students and colleagues' pre- and post- test data also established internal (face) validity, and external (construct) validity. With the use of inferential quantitative methods, it was possible to view the case from the perspective of those involved. These were the methods and techniques, which have been used for data collection and analysis.

So far in my research endeavours, I have used mainly qualitative research methodology. The present research project gave me the opportunity to enrich my knowledge and skills in the following ways: I researched L2 online testing as a work based and applied linguistics project. Case study was my major research approach, however, although it was one case that did not mean one individual or one small group, but the whole population of a tertiary institution sitting for the English Placement Test developed and evaluated. Although a case study, the results would be, in some ways able to be generalised due to the large number of participants (more than 800), relateability, however, seems to be more important than generalisability. Moreover, I enriched my knowledge in quantitative research, learned more about statistics, statistical analysis, and the use of computerised programmes such as Excel, SPSS, and VISIO, to analyse and present data.

4.3 Required Resources

4.3.1 Human Resources

Various people of different areas of expertise were involved in the project:

- (a) I, the work based, insider researcher, worked as the L2 / English as a foreign language (EFL) language specialist, who researched, planned, and designed the

test, wrote the test items, gave the technological requirements of the test to the Information Technology expert to develop the online test electronic environment, developed the test, coordinated the various test stages and ensured test evaluation.

- (b) The research assistant helped with some of the test item editing, the test iteration process, the conception of the cut-off point system, and the trial.
- (c) A collaborator helped with the iteration process and the cut-off point system.
- (d) Authorities in the field provided guidance on test design and data analysis, and critiques of the individual items: Language faculty members with authority in the field of L2 and statisticians agreed to take up those roles.
- (e) The Intercollege Information and Technology expert, computer programmer and instructional-technology designer created the computer software to implement the test design in its computerized form.
- (f) Lab assistants helped with the NEPTON trial.
- (g) A research assistant helped with the uploading and the tagging of the test items.
- (h) Intercollege English language lecturers from all three Intercollege campuses provided feedback on the strengths and weaknesses of particular items and invigilated the test pen-and-paper field-testing.
- (i) All Intercollege English students from all three Intercollege campuses field-tested the prototype items in the item pool in pen-and-paper form.
- (j) All Intercollege English students trialled the NEPTON in its electronic form.
- (k) Administrators helped with the NEPTON trial

4.3.2 Other Resources

- (a) Some time allowance was allocated to the test writer, (three hours per week for three academic semesters and one summer session) to research and write the test.
- (b) Some budget was allocated to a computer programmer to create the computer software to implement the test design in its computerized form.
- (c) Intercollege computer department provided hardware and software necessary for the project.

- (d) Intercollege language lab and all other computer laboratories were used to accommodate trialling and running of NEPTON.

4.4 Ethical Considerations

Test-takers and other participants were provided with the necessary information about the nature of the research and their consent to participate was asked. Most of the test material was unique and original, developed by the test writer. References were made to any other material used for the project. Acknowledgements have been made to various people who have contributed to this project in one way or another and without their input this project would have never been possible to get off the ground (see acknowledgments). Thought was also put into the specific skills, qualities, and demands required of my role as a worker / researcher. Ways of overcoming subjectivity involved in approaching research as an insider and as a researcher were sought.

4.5 My Role as worker / researcher

I have always drawn research questions from my various working environments, from everyday work and research needs and interests. I have always been a firm believer of marrying theory and practice, doing research for a purpose, for improving practices (what some may call today *quality assurance*), and deal particularly with research questions that I could draw from my daily practices and work experience. From all my accumulated experience, I have found that it is somewhat a delicate situation to be a worker researcher. On the one hand, one needs to take into consideration the constraints posed by the working environment, the work administration as well as the relation to other colleagues and as a consequence the special skills and knowledge you need to develop to deal with all of that. On the other hand, however, a worker / researcher is in the fortunate position to have immediate access to research and research implementation context. Delicate handling is needed to meet the requirements of the work and research aspects of the research project, in a way that would serve both work and research in the best possible manner. Keeping the balance between the two is an added challenge to both roles of worker and researcher. A research-based outcome, however, which would

serve real purposes is something that definitely makes sense to me. Co-operational collaboration with other colleagues and other staff in general can be both beneficial and essential. Moreover, if the particular research and research end-product (in this case the New English Placement Test Online) is implemented in the workplace, collaboration offers the opportunity to make other colleagues a part of it and feel ownership, thus giving them the opportunity to feel comfortable and eager to accept change and innovation the innovation would bring to the already existing practices. I believe that research derives and serves everyday activities. Meeting both work and research requirements, in the best possible way, is a challenge by itself. I couldn't see myself researching something that is completely foreign to my professional daily needs. Part of my work as an academic is research therefore I consider answering research questions that relate to my professional endeavours a natural element of my professional practice.

4.6 Ways of overcoming subjectivity involved in approaching research as insider researcher

In this research project I had a dual role, that of worker (English educator and tester), and that of insider researcher. This had the potential to impact the research. The following were some of the procedures I used in an attempt to overcome the subjectivity involved in approaching research as an insider:

- (a) I took into consideration as many resources and references as possible which deal with work based research and the area of Second Language Online Testing.
- (b) I took into consideration the English language syllabus of the institution and tailored the new test to its needs.
- (c) I collaborated with as many colleagues as possible and considered their input as part of qualitative analysis; for example, I had colleagues validating the content of the test.
- (d) I had colleagues and other interested groups (e.g. administrators, colleagues from other disciplines interested in the English Placement Processes of our institution) trial the test.

- (e) I consulted and involved experts in statistical analysis and used quantitative methods to support findings.
- (f) I had as many students as possible trial the test and used that in quantitative and qualitative analysis (e.g. had students giving their opinion about the test).
- (g) I presented the project findings in relevant meetings, seminars and conferences and published the projects findings to receive feedback.

4.7 Major influences (positive and negative) on me as an insider researcher.

I identified the following major influences (positive and negative) on me as a worker researcher.

Positive:

- (a) I am challenged by projects which combine academic research and offer change and improvement to academic practices
- (b) Although I fear the unknown, and novelty, they challenge me and I always find myself involved in some kind of project that brings substantial change in specific areas.
- (c) I always like to have someone ready and positive to discuss issues that concern the research work I'm involved in, someone like a mentor, a colleague who likes sharing academic quest. So far in my life I have been lucky in having someone like that wherever I worked. In fact, I have found two very valuable colleagues at Intercollege as well, and that made this research work much more enjoyable and effective.
- (d) I always feel the need to be academically challenged by my work. I always like to do research related to my work and contribute towards understanding and improving current practices.
- (e) My workplace offered me the opportunity to get involved in a research project that would both challenge me academically and research wise and be beneficial for the institution at the same time.

- (f) For this research project, I had most resources needed in hand, in other words, subjects, other experts, necessary information, and key people.

Negative:

- (a) My dual role as a worker / researcher made it difficult at times to work subjectively on the project.
- (b) I felt challenged having to combine specific political and other needs of the organization I was developing the NEPTON, with principles in research. This helped me develop a skill of balancing, judging and deciding on serious issues entailing a considerable degree of responsibility.
- (c) Project and work-related time constraints were always a worry, and made me anxious but eventually they turned into challenges.
- (d) Ignorance of other stakeholders of what the particular area and research involved helped me arrive at challenging ways of informing and explaining without offending.
- (e) Reluctant colleagues (feeling being threatened by change or even by me as the person in charge of the project) and administrators (feeling uncomfortable towards change and the unknown it entails). I found it adventurously challenging to gradually arrive at implementing change and proving it is beneficial.

4.8 Political issues related to my research project.

Students (accurate) placement in the various English courses at Intercollege has serious and considerable ramifications in the following ways:

- (a) The time students need to spend studying English during their studies at Intercollege, as a result of their NEPTON placement is of considerable concern to both students and administrators.
- (b) The amount of money students have to pay for each English language level they are placed in, particularly for those students who start at the lower levels.

- (c) The compatibility of the NEPTON results with other well known EPTs (English Placement Tests) used by other Universities.

The stakeholders at the private tertiary institution I work are the administration, faculty and students who will benefit from this research project. The intention was to involve as many of them as possible in the validation and reliability process so that NEPTON's quality was ensured and a sense of belonging to these participants was developed.

4.9 Conclusion

This research project was based on the needs of the college I work at for a more efficient, and improved English Placement test, the work based learning nature of my programme of study and my 25-year accumulated knowledge and experience in L2 teaching, teacher-training, curriculum development, evaluation and testing, and computer assisted language learning.

The area of Epistemologies and Methodologies was investigated, with particular reference to the professional area of (L2) language teaching and learning, and more specifically to the working area of English Language Testing Online. The research was conducted within the area of postmodern, naturalistic epistemologies. This approach was mostly appropriate since it reflected a specific socio-cultural, work based context and was based on a process of experiential learning, through uncertainty, reflection, and a quest for a specific world, in this case, that of the English placement testing programme of Intercollege. The project combines work based learning (where the researcher is an insider researcher) and the applied linguistics research discipline: work based learning draws from a work situation (Intercollege) where the researcher is not given a problem to solve, but is required to identify the problem (inefficiencies of the existing English placement testing programme), conduct research as in insider worker and work towards producing an output, a tangible product or a service (a New English Placement Test Online). In work based research projects such as the present one, work and learning go hand in hand and is not confined only in the setting of a university. This

research project work was carried out at Intercollege. Although Intercollege is a tertiary institution, in the case of this project it was the work place where the research project was carried out. On the other hand, this research also drew from applied linguistics in order to learn how to bring change, improvement and efficiency in the current Intercollege English placement testing practices. Case study was chosen as a strategy to carry out this research. It is a strategy used in both work based learning and the applied linguistics research discipline and investigates a specific case, in this case, that of Intercollege English placement testing practices. A combination of qualitative and quantitative techniques was used in order to measure the test validity and reliability. I developed skills in managing human and other resources, and my dual role as a worker / researcher; I developed knowledge in managing positive and negative influences and political issues involved; I developed ways to overcome subjectivity and I considered ethical issues related to human and other resources.

Chapter five

Project Account

5. Introduction

In the previous chapter we established the research methodology and data collection techniques needed for this project. This chapter provides a detailed commentary on the project process: the needs analysis, the development and implementation of the New English Placement Test Online (NEPTON), and the trial.

The first phase evolved around the choice made as a result of the study of current theories and practices in L2 testing, and more particularly in the study of the particular setting of Intercollege existing English placement testing practices in order to establish the necessary background needed for the development, implementation and evaluation of NEPTON.

The second phase is the development of the NEPTON Test Specifications.

The third phase dealt with item writing, item bank creation and the moderation of the test items.

The fourth phase evolved around the field-testing of the first format of the test, which was pen-and-paper, and the analysis of the results of this first trial.

The fifth phase provided a detailed description of the implementation of the NEPTON, its format, and the computer-based trial.

The project account is presented under the following sub-headings:

- (a) Needs Analysis
- (b) NEPTON Test Specifications
- (c) Test Development
- (d) Pen-and-Paper Field-Testing and Reviewed Test Specifications
- (e) NEPTON Implementation

5.1 Needs Analysis

In this section the particular academic setting in which this project was carried out was examined. The context, in which this English placement test was carried out, and the characteristics of the particular setting was looked at, in other words, Intercollege, the characteristics of the students, the lecturers, the English programme and its relation with international language equivalences. This information determined the choice of the type of test needed to cater for the particular needs of this case.

5.1.1 Context

As we have seen in chapter 1 (1.2), at Intercollege, incoming students sit for an English placement test at each of the three campuses at the beginning of every semester. According to their result, they are placed in English for Beginners (BENG-50), Academic Language Skills I (BENG-80), Academic Language Skills II (BENG-90), College English (BENG-100), Basic Writing (ENGL-100) or English Composition (ENGL-101) course. For many years Intercollege used an English placement test that had a bank of 4 different exam papers. That EPT covered the areas of English structure, vocabulary through sentence-based multiple-choice questions, reading comprehension through text-based multiple-choice questions and writing through essay writing. It was based on traditional pen-and-paper testing methods, administration, correction and result reporting. Various standardized exams were also used for placement in Intercollege English classes¹⁹. The need for change and improvement, due to the increase in student numbers and the changes in current theories and practices in L2 testing, resulted in embarking on the New English Placement Test ONLINE (NEPTON) project. On a practical level, it meant improving efficiency in running, and delivering an EPT, in reporting results and in placing students more efficiently, based on current theories and practices in L2 online testing. On a research level, it meant giving the opportunity to

¹⁹ Students with grades in these exams are placed as follows: GCE "O" level English Language/IGCSE "C" ENGL-100; GCE "O" level English Language/IGCSE "A" or "B" ENGL-101; Cambridge First Certificate "A" or "B" ENGL-100; Cambridge Proficiency "A" or "B" or "C" ENGL-100; TOEFL (paper) 500-549 ENGL-100; TOEFL (paper) 550+ ENGL-101; TOEFL (computer) 173-212 ENGL-100; TOEFL (computer) 213+ ENGL-101.

conduct a work based and applied linguistics research project and gain insights into developing, implementing and evaluating a model of an effective placement programme for a specific college English learning context.

5.1.2 Setting

Intercollege is a private tertiary institution in Cyprus. It offers undergraduate programmes in the following broad disciplines: Business Administration, Computer Science & Engineering, Early Childhood Education, Humanities, English Language and Literature, Liberal Arts, Accounting and Banking, Law, Office Administration/Secretarial, Pre-Medicine, Culinary Arts and Aesthetics and Beauty Therapy. It also offers postgraduate studies (Masters) in Management, Business Administration, Education, Special Education, Psychology, International Relations, Sociology, Hotel and Tourism Management, and Computer Science. The language of instruction is English.

5.1.3 Students

The majority of Intercollege students are of Greek-Cypriot background²⁰. Other Cypriot students include Turkish and Armenian (from, for example, mixed marriages or naturalization). The non-Cypriots who reside in Cyprus include nationalities such as Russians, Ukrainians, and Tunisians. These people learn English at both primary and secondary level in schools in their respective countries. International students at Intercollege currently make up one-third of the student body and represent fifty nations²¹. All these students are faced with a very significant learning challenge. They

²⁰ The Cypriot Greek-speaking students are at very different levels in their English language development; they are not a homogenous group. They include many different kinds of students, for example: students who have learned English at both primary and secondary level in government schools; students who have learned English at both primary and secondary level in private English language schools; students who have learned English at both primary and secondary level in both government schools and after school tuition schools called "Frontistiria"; students who have been born in English speaking countries such as England, Australia or USA and came to live in Cyprus.

²¹ Armenia, Australia, Bangladesh, Bahrain, Bulgaria, Canada, China, Denmark, Equador, Egypt, Estonia, Finland, Georgia, Greece, Iran, Iraq, Israel, Japan, Jordan, Kazakstan, Kuwait, Latvia, Lebanon,

continue to learn a second language, English, but they are also learning in it and through it as well in their courses. In other words, they must continue to learn English, and at the same time they must learn to learn in English, they must learn to study all their course subjects, for example, secretarial studies, hotel and tourism management, through the medium of English as their second language. Consequently, at Intercollege, continuing to learn English as a second language goes hand in hand with studying college courses. English language teaching is therefore parallel with other academic course teaching.

5.1.4 Lecturers

Lecturers hold Doctorates or Master degrees, mainly from English-speaking universities from all over the world. They are a multicultural mix of lecturers from Cyprus, the United Kingdom, the United States of America, Bulgaria, Rumania, Poland and Australia. There are about twenty-four full-time and twenty-one part-time instructors who work at the three different Intercollege campuses (Nicosia, Limassol and Larnaca). Lecturers of English are specialists in teaching English at the various levels of English language competence. They run their courses in English. They are experts in fields such as English literature, teaching English as a Second Language, Computer Assisted Language Learning, L2 teacher training, intercultural, multicultural and peace education in second language learning phonetics and phonology. They are involved in scholarly work and present papers and attend conferences in their fields.

5.1.5 Intercollege English language programme and existing EPT

Current theories in L2 teaching and learning, testing and language level descriptors were researched. The various Intercollege documents related to the existing English programme and English placement test were studied (English syllabi, course outlines, levels, existing placement test pen-and-paper format and test guidelines and criteria for assessing the written component, student and faculty population) in order to establish the

Libya, Malawi, Mauritius, Mexico, Nigeria, Pakistan, Palestine, Philippines, Poland, Romania, Russia, Saudi Arabia, Serbia, South Africa, Somalia, Sudan, Switzerland, Syria, Tanzania, United Arab Emirates, Ukraine, United Kingdom, United States, Venezuela, Yemen, Zambia, Zimbabwe.

context and the relevant current needs. The textbooks and other material used were also studied to establish the context of the existing English programme in relation to the placement test. It was established that the English programme (syllabi, language levels, existing placement test²², assessment criteria, teaching and learning material) needed to be reviewed in light of current theories and practices and the changing needs of students. Although the previous Intercollege English Placement pen-and-paper test consisted of four parts testing structure, vocabulary, reading comprehension and writing, the descriptors in the Intercollege *Placement Test: Marking Guidelines and Sample Placement Test* concentrated only on the written component of the Intercollege English Syllabus²³. This indicated the need for the development of a comprehensive set of descriptors of language levels for the English programme at Intercollege, one that would be informed and compatible with CEF and ALTE ones. For this reason, a committee was formed to review the current needs of the English programme, develop language descriptors, and combine change and improvement in the programme with the development of the New English Placement Test Online. As the NEPTON project leader, designer, developer, and writer, I was asked to also lead this Committee of BENG Level English (COBLE). The syllabi were updated with current theories in L2

²² As stated in the *Placement Test: Marking Guidelines and Sample Placement Test* Intercollege document, "...the purpose of the existing Intercollege EPT is to place students in to the appropriate level of English in order to support their academic studies at the College..." It consists of four parts. The first three sections are multiple-choice questions, consisting of questions on English Structure (30 questions), Vocabulary (15 questions) and a Reading Comprehension section (19 questions). The fourth part consists of writing short essays (approximately 250 words) on a topic assigned on the day of the examination (e.g. 'Smoking is bad for your health. Discuss.'). The time spent on the test is 90 minutes (60 minutes for the multiple choice and 30 minutes for the essay). There are four variations of the same test items in the format of four test papers. The results of the EPT help place students in the six different English courses mentioned earlier. The test places students in six language level courses ranging from beginner to advanced.

²³ *BENG-50* is for the true beginner of English. *BENG-80* students demonstrate some fluency but they have limited communicative ability, and limited vocabulary. They frequently misuse elementary vocabulary, and are often unable to use coordinating conjunctions correctly to form compound sentences. They demonstrate inconsistency in the use of elementary grammar structures. They often omit or misuse articles and prepositions. *BENG-90* students demonstrate more sophisticated writing, however it is often characterized by weaknesses in intermediate-advanced grammar. *BENG-100* students have more grammar control than those placed in the lower levels, but their sentence structure remains unsophisticated and their vocabulary elementary. This course begins with a grammar review to access student weaknesses, however, it focuses on the writing of sentences and paragraphs as well as academic vocabulary development. At *ENGL-100* level, students' writing has a few serious problems, but their writing often reflects a weakness in the development of ideas. *ENGL-101* students demonstrate near-native English ability, fluency and clarity. The students master basic skills of writing in English and moves into writing for academic purposes.

teaching and learning and were complemented by the newly developed level descriptors. Table 5.1 shows how Intercollege language levels compare with the ALTE (Association of Language Testers in Europe) and the Council of Europe Levels.

Table 5.1: Language level descriptors

ALTE Level	ALTE Level Description	Council of Europe CEF	CEF Description	Intercollege Levels
0	Beginner (Breakthrough)	A1	Basic User: Breakthrough	BENG 50
1	Elementary (Waystage)	A2	Basic User: Waystage	BENG 80
2	Lower Intermediate (Threshold)	B1	Independent User: Threshold	BENG 90
3	Upper Intermediate (Independent user)	B2	Independent User: Vantage	BENG 100
4	Advanced (Competent user)	C1	Proficient User: Effective Operational Proficiency	ENGL 100
5	Very Advanced (Good user)	C2	Proficient User: Mastery	ENGL 101

5.1.6 International English Language Equivalences

Table 5.2 indicates the equivalence of Intercollege BENG levels with IELTS and TOEFL:

Table 5.2: Exit levels for Intercollege courses as measured by IELTS and TOEFL

Intercollege	International Terms (Course content level)	IELTS	TOEFL
BENG50	Beginner	< 4.0	NA
BENG80	Pre-intermediate	4.0	400
BENG90	Intermediate	4.5	450
BENG100	Upper intermediate	5.0	500 (Also Cambridge First Certificate English)

Inventories such as grammar, vocabulary and topic inventories were compiled for each level. The review of the textbooks and other material resulted in the replacement of some old ones, which appeared on the course list, with some currently used and some new ones. This whole process aimed at improving current practices in the English

language programme and the English placement test and informing them with the latest developments in the area. This also ensured the content validity of the test in question.

5.1.7 Test choice

After reviewing the current theories and practices in L2 teaching and testing in general, and that of the particular teaching and testing context especially of Intercollege, it was decided to adopt a form of the 'loosely' used Communicative Paradigm as the approach for the design of the New English Placement Test Online, needed for the current English placement needs of students at Intercollege. This was based on the following facts: NEPTON needed to be context specific, in other words, it had to serve the specific context of Intercollege and the particular needs of its students. A communicative language test is based on a description of the language that the test-takers need to use. Such a test should reflect communicative situations in which the test-takers are likely to find themselves. In the case of Intercollege students, the communicative situations they study and are likely to find themselves in are:

- (a) Everyday situations: Although Intercollege is in Greek-speaking Cyprus, English is mainly the common language of communication amongst people who don't speak Greek and live or study in Cyprus. In other words, English is very often the language of communication in students' everyday life.
- (b) Academic setting: Although Cyprus is a Greek-speaking country, Intercollege uses English as the language of instruction therefore students need an adequate knowledge of English for the needs of their academic studies.

According to relevant literature, Communicative language tests are those, which make an effort to test language in a way that reflects how language is used in real communication (Kitao and Kitao 1966). Since it is not always possible to make language tests fully communicative, it may often be possible to incorporate communicative elements in them. NEPTON included text types which were authentic or authentic like, similar to the ones students would come across in their academic, personal and social settings in Cyprus and overseas. These text types covered relevant topics, and incorporated vocabulary, structure and sociolinguistic elements which

derived from the literature review and the level descriptors, which reflected the various levels students, would be placed in. The test assessed writing using two different approaches: First, writing was divided into discrete levels, vocabulary and grammar. These elements were tested separately by the use of objective electronic testing activities. These tested knowledge of vocabulary and grammar in the form of sentence-level multiple-choice items. The test also assessed vocabulary and grammar knowledge in a more contextualised mode, in the form of text with four or five multiple-choice questions of dropdown menu selections. Reading comprehension was tested in two forms: in the contextualised and situational form of signs (accompanied by visual.) and texts with multiple-choice questions. Finally, the test assessed writing through a more direct, communicative and extended global integrative writing task, in a non-electronic mode, that is, hand written. All the electronic components were automatically assessed and the results were automatically recorded in the test *Results* section of the administrator's testing electronic environment, with the use of an algorithm. The hand-written component was marked by experienced and trained markers with the use of a set of criteria. Then, the two results were put together to place each student.

The use of a 'loosely' communicative approach was also used for the following reasons:

- (a) The intention was that the new test would not be too different to the existing pen-and-paper current practices, so that it would be more easily accepted by all the participants concerned (students, lecturers, administrators). It was therefore decided to introduce a New Placement Test Online, which would offer some improvements in content, delivery (for example, inclusion of interactivity), and administration. Such improvements seemed adequate to be accepted and handled by stakeholders as part of a first phase of innovation and change.
- (b) Time and urgent pressure to meet current needs as quickly as possible permitted us to work during very tight time limits. A more communicative approach to language testing needed more time to be investigated.
- (c) This was considered phase 1 of the implementation of the New English Placement Test Online. Further improvement and development is planned.

A draft of the NEPTON Test Specifications document was written. It included mainly the following sections: purpose, test description, format and content, item criteria selection, test methods, test items, kinds of tasks, topics and text types, timing, criteria levels of performance, scoring, results, and description of the computerized test function.

This Test Specifications document was then improved after the item writing and moderation process and pen-and-paper field-testing of the NEPTON. The following is the revised document:

5.2 The NEPTON Test Specifications

The NEPTON test specifications were a document that described the test in details; it stated what the test tested and how it tested it. It was followed by the test developers (test and item writers) as a blueprint in order to write the test. It was then used by experts for the test evaluation in order to establish its content and construct validity (Kitao and Kitao 2002; Alderson 1995, Clapham and Wall 1995).

5.2.1 NEPTON purpose

The New English Placement Test Online (NEPTON) and New English Placement Test On Paper (NEPTON P) were developed for Intercollege to assess English language proficiency of incoming students, and place them in appropriate college level English courses appropriate to their level of English competence, namely: BENG-50, BENG-80, BENG-90, BENG-100, ENGL-100 and ENGL-101.

5.2.2 NEPTON description

Test-takers could take the NEPTON in two forms:

- (a) The NEPTON form, which was a Computer-Based Test (CBT), accompanied by a hand written component and
- (b) NEPTON (P), which was its “equivalent” Pen-and-Paper Test.

5.2.3 NEPTON test sections

Each randomly generated NEPTON and NEPTON (P) consisted of 33 discrete questions, totalling 54 test items and a writing non-electronic task of 120 words in length. The discrete items were randomly chosen from a level-based pool of the following activity types:

- (a) Sentence-based Structure (SB-S)
- (b) Text-based Structure (TB-S),
- (c) Sentence-based Vocabulary (SB-V),
- (d) Text-based Vocabulary (TB-V),
- (e) Sign-based Reading Comprehension (SB-RC), and
- (f) Text-based Reading Comprehension (TB-RC).

There was a substantial pool of each one of these types of items:

Table 5.3: Test questions and test items

QUESTIONS							TOTAL
Level	SB-S	TB-S	SB-V	TB-V	SB-RC	TB-RC	
BENG-50	50	10	22	7	21	10	120
BENG-80	59	8	29	14	20	10	140
BENG-90	40	11	29	12	0	10	102
BENG-100	52	5	13	5	0	5	80
ENGL-100	46	5	55	6	0	4	116
ENGL-101	33	5	27	5	0	4	74
TOTAL	280	44	175	49	41	43	632
ITEMS							
Level	SB-S	TB-S	SB-V	TB-V	SB-RC	TB-RC	
BENG-50	50	50	22	28	21	40	211
BENG-80	59	40	29	56	20	40	244
BENG-90	40	55	29	48	0	40	212
BENG-100	52	25	13	20	0	20	130
ENGL-100	46	25	55	24	0	16	166
ENGL-101	33	25	27	20	0	16	121
TOTAL	280	220	175	196	41	172	1084

5.2.4 NEPTON item number per section

The test was generated through electronic and random choice, following a nine-item slide design per level. There were two combinations of random selection: one for test 1 and one for test 2. This was how it worked: three one-item questions were chosen from the SB-S BENG-50 pool, then one text (including 5 test items) from the TB-S BENG-50 pool and one question (one item) from the SB-RD BENG-50 pool (total of 9 test items). The random selection continued from the BENG-80 item pools to ENG-101: (nine items per level):

Table 5.4: Revised, final NEPTON slide algorithm (After the field-testing)

	TEST 1	TEST 2
BENG-50		
9 items per slide	SB-S: 3 TB-S: 5 SB-RC: 1	VB-S: 4 VT or RC: 4 SB-RC: 1
BENG-80		
9 items per slide	VB-S: 4 VB-T / TB-RC: 4 SB-RC: 1	SB-S: 3 TB-S: 5 SB-RC: 1
BENG-90		
9 items per slide	SB-S: 4 SB-T: 5	SB-V: 5 TB-V or TB-RC: 4
BENG-100		
9 items per slide	SB-V: 5 TB-V or TB-RC: 4	SB-S: 4 TB-S: 5
ENG-100		
9 items per slide	SB-S: 4 TB-S: 5	SB-V: 5 TV-V or TB-RC: 4
ENG-101		
9 items per slide	SB-V: 5 TB-V or TB-RC: 4	SB-S: 4 TB-S: 5
9 items x 6 slides	54 items (33 questions)	54 items (33 questions)

The total length of each generated test consisted of 9 level-based slides of 6 items each, equaling 54 items per student. In the NEPTON, each item was presented one at a time. Students had the opportunity to move from one question and type to another freely, reviewing, skipping, going backwards and forwards, until they decided they were finished with all of them. At that point, they could click on the *Finish* Button to end the test. Once they did that, they could not go back to the test. In the NEPTON (P) all items were presented together in a form of a test booklet. Students had the opportunity to freely go from one question or item to the others as in the NEPTON test until they completed and submitted their test paper. In the NEPTON the answers were electronically recorded and the results were calculated automatically. The NEPTON (P)

was accompanied by an answer sheet, which was then scanned and electronically scored.

All this ensured that:

- (a) Each test was unique.
- (b) Students were assessed in more than one way at each level before they moved on to the next level.
- (c) The different activity types catered for different learning styles.
- (d) Test security such as item disclosure, and item theft were minimized.
- (e) The test-taker had test control and could review the test items.
- (f) Pen-and-paper test are usually very long. They can sometimes be 112 test items long and can last for more than two and a half hours. On the other hand, computer adaptive test are very short. They often consist of twenty items and are ten to fifteen minutes long. The NEPTON test was long enough to assess test-takers language competence in a satisfactory manner in many different ways and at many different levels. It consisted of 54 items and it lasted about an hour, including the electronic tutorial.

5.2.5 NEPTON text types

As the writer of the test content, I had to choose and write the different text types to be included in the test. The passages I chose consisted of complete and meaningful texts drawn from authentic sources. These included newspapers, magazines, advertisements, literary books, letters, and short stories (see appendix 6 for text-type inventories). I also developed authentic-like material, reflecting the language levels tested, similar to ones found in teaching and learning material of the respective levels, which were based on current theory and practice in L2 teaching, learning and testing. These represented different kinds of writing such as informative, persuasive, and summative. After the moderation process, I made some minor editing for length or to simplify particularly difficult vocabulary or structures. This portion of the test was designed to stress variety by including a sampling of different topics, functions, situations, styles, level of

difficulty, lengths of passages, and types of questions. Contextual clues were strongly evident, so that items were functionally and semantically explicit.

5.2.6 NEPTON language skills, test methods and instructions

The test helped determine test-takers preparedness for academic course work. The first, electronic part of the test measured student skills in certain areas of English: Writing in a discrete form (structure, vocabulary) and reading comprehension in a discrete form. The second, hand-written part of the test measured students' writing skills in an integrative way. In order to meet these criteria, a variety of test item formats were selected to measure a number of different types of language competence (see appendix 11). Effort has been made for the instructions to be unambiguous, short and clear.

5.2.6.1 Writing: Structure and Vocabulary

In the test, this section was designed to measure the candidate's ability to recognize structure and vocabulary that is appropriate for Standard English.

There were two types of questions in these two sections:

- (a) The first type consisted of a sentence with one or more words missing. For each sentence, students selected the word or phrase that best completes the sentence. There were five multiple choices for each sentence to choose from. This activity assessed students' knowledge of grammatical structure and vocabulary in a sentence context, in a discrete way.
- (b) The second type consisted of a text given with five blanks of one or more words missing. For each blank, test-takers chose from a dropdown menu selection of five multiple choices the word or phrase that best completed it. This activity assessed students' knowledge of grammatical structure and vocabulary in a context of a passage, in a discrete way.

5.2.6.2 Reading Comprehension

There were two types of questions in this section:

- (a) The first type consisted of a number of written signs, accompanied by a relevant visual. For each sign at BENG-50 level, test-takers selected the word or phrase that best described the setting indicated by the sign. For each sign at BENG-80 level, test-takers selected the word or phrase that best reflected the meaning of the sign. There were three multiple choices for each sign to choose from. This type of question was used for the first two lower levels. This activity assessed students' ability to understand signs aided by visual. The main focus was the understanding of the main idea.
- (b) The second type consisted of a presentation of a text type with four questions, and four multiple choices for each question to choose from. This activity assessed, for example students' ability to find the main idea, and skim for details. In the test, this section was designed to measure the candidate's ability to read and understand (for example skim, and scan.) English from various English-speaking settings.

Once this discrete electronic part of the test was completed, students moved on to the hand-written component, which consisted of a written task. This part determined students written fluency.

5.2.6.3 Writing

All written tasks were based on tasks drawn from and reflecting meaningful written text communication test-takers would come across during their college life as well as similar to ones found in teaching and learning material of the respective levels. Students were given a selection of the following types of written tasks (descriptive, narrative, informative, and persuasive): story telling, account, letter, essay on personal, college, experiences and social issues (see appendix 11). The written task required was of 120 words in length. The written task of the test measured students' writing skills in an integrative way.

5.2.7 NEPTON item criteria selection

Each item was selected according to the following criteria:

- (a) The test purpose: the NEPTON test was a placement test, which placed incoming students in the college at six English language levels.
- (b) Topic: the topics in the NEPTON test were chosen according to each level. They were similar to topics covered in textbooks of the same levels, suggested by theoretical books for those levels²⁴, drawn from the suggested topics of L2 level frameworks such as CEF, ALTE, LOTE CSF and VCE LOTE²⁵, or from the personal, social and educational settings of the students.
- (c) Skills: The NEPTON test covered the writing competence (structure, vocabulary, and fluency), and reading comprehension (skimming and scanning), appropriate at each level. They were similar to topics covered in textbooks of the same levels, suggested by theoretical books for those levels, drawn from the suggested topics of L2 level frameworks such as CEF, ALTE, LOTE CSF and VCE LOTE, or from the personal, social and educational settings of the students.
- (d) Format: a variety of format such as sentence, signs and text, catering for different learning styles, and reflecting the respective levels, skills and content was used in the NEPTON test.
- (e) Sociocultural context: Appropriate sociocultural aspects, reflecting the English language all over the world was reflected in the NEPTON test.
- (f) Test-taker background and study setting were also reflected in the text types, situations, settings and contexts depicted in the NEPTON test.

All tasks in writing (structure, vocabulary) and reading comprehension sections were objective. Hand-writing tasks were mostly of an authentic simulated nature and were subjective (see examples of the types of questions covering the above skill areas in NEPTON and NEPTON (P) tests in Appendices 19 and 10 respectively).

²⁴ Nunan (1985a, 1987, 1988, 1989, 1991); Richards (2001); Ur (1997); Harmer (1991); Larsen-Freeman (1986); Brumfit (1984); Littlewood (1986); Munby (1978); Johnson (1982); Wallace (1991); Richards and Renandya (2002).

²⁵ Language Other Than English Curriculum and Standards Framework II, 2000, Victoria, Board of Studies; LOTE VCE Study Design and sample units.

The differences between NEPTON and NEPTON (P) were:

- (a) The mode of delivery of the discrete items. In the case of NEPTON, each discrete point was presented electronically. Each question was chosen randomly at a time and presented singly. In the case of NEPTON (P), each question was chosen randomly, and all 33 questions (representing 54 test items) were presented all together in a printed form.
- (b) Both the NEPTON and the NEPTON (P) tests were unique.
- (c) The NEPTON test was automatically scored. The NEPTON (P) was accompanied by an answer sheet, which was then scanned and electronically scored.

5.2.8 Language Elements

The following is a description of the grammatical, and lexical points covered in the NEPTON test, as well as the topics and settings reflected in it.

5.2.8.1 Grammatical points

The grammatical structure section tested a variety of morphological and syntactic forms deriving from the items represented in the course description competence levels, which were compatible with the course syllabus, learning material used and topics covered, and compared with other descriptors such as the ALTE (Association of Language Testers in Europe) and the European Council Levels (see appendix 7 for more detailed table of grammatical points that have been used in the test at the various levels)

5.2.8.2 Lexical points

The lexical section tested a variety of vocabulary deriving from the items represented in the course description competence levels, which were compatible with the course syllabus, learning material used and topics covered, and compared with other descriptors such as the ALTE (Association of Language Testers in Europe) and the European

Council Levels (see appendix 8, for more detailed table of lexical points that have been used in the test at the various levels)

5.2.8.3 Topics

The topics included in the test covered those suggested by current theories and practiced in English courses of the respective levels through the use of textbooks, and authentic sources, such as newspapers, magazines, advertisements, literary books, letters, short stories, Internet sites, and email messages (see appendix 9, for more detailed table of topics that have been used in the test at the various levels).

5.2.8.4 Settings

The settings included in the test covered those suggested by current theories and practiced in English courses of the respective levels through the use of textbooks, and authentic sources, such as newspapers, magazines, advertisements, literary books, letters, short stories, Internet sites, and email messages (see appendix 10, for more detailed table of settings that have been reflected in the test at the various levels).

5.2.9 Timing

5.2.9.1 NEPTON

The electronic component of the test included items, which tested the writing competence (structure and vocabulary) and reading comprehension. These test items were designed as a test of skill and not speed. Ample time was allowed to answer questions. The electronic component was given 80 minutes to complete but the average time it took students to complete this section was 41 minutes. The written component of the test was given approximately 30 minutes to finish, but the average time it took students to complete this section was 20 minutes. The whole test took between 80 to 100 minutes to complete, including the electronic tutorial, at the beginning of the test.

5.2.9.2 NEPTON (P)

Students were given 90 minutes for the multiple-choice section and 30 for the written task. The whole test took between 90 to 120 minutes to complete.

5.2.10 Computer Based Tests (CBT)

The NEPTON test was designed using electronic techniques. This meant that the computer automatically and randomly determined how questions were presented to students based on the six-level-based slide and two test paradigm. This technique allowed test-takers to answer questions in three different skill areas through six different styles of activities from lower to higher levels. Students completed all sections. Examples were given in the NEPTON sample booklet as well as in the NEPTON Tutorial.

5.3 NEPTON Development

5.3.1 NEPTON project committee

As the person responsible for the NEPTON project, I formed a project committee for the New English Placement Test ONline (NEPTON) consisting of: the language faculty head; an Information Technology member; a research assistant; a number of other English language lecturers who acted as editors, consultants, trial participant; statisticians; and myself who acted as the NEPTON project leader and also played various major roles during the whole of the development of the project (see introduction, page 8, point 1.7)

5.3.2 Test item writing and moderation

The item writing was based on the test specifications. I was the main writer. The research assistant wrote some single sentence items for some of the higher levels. The texts I chose were meant to provide a representative sample of the kinds of spoken and written English found at the various levels.

5.3.3 The item bank

A pool of about 750 questions, totalling more than 1500 items was originally developed for the six English course levels and served as an item bank. The items measured writing (structure and vocabulary), and reading comprehension skills (scanning and skimming) in English for students applying to study at the college. The items were widely chosen from the whole area of content (for all six levels) and were presented in various forms.

5.3.4 Moderation

An editing - moderation process proceeded in order to test the validity of the test. About 42% of the English programme-teaching faculty from all three campuses was involved in this process (see appendix 14). It included the following: editing, commenting on the clarity of each task, appropriateness of each item in terms of the level, appropriateness of what each item tested (vocabulary, and structure; reading comprehension: main idea, scanning for main ideas, comprehension with the aid of visual, make inferences, scanning for specific details.), the skill, the suitability of texts in general and for each level in particular, formats and items, the uniqueness of each correct answer, the content, the context and the topic, as well as the appropriateness of the sociocultural aspects the content imbedded. Based on the moderators' suggestions, I edited the items in the test bank, improved the choice or length of the texts, and made some deletions. There was only one acceptable response for all items. The pilot test was then presented to a group of native speaker professional in the field (language experts and instructors, with experience in testing), consisting of four experienced lecturers. Again, based on their suggestions I revises the test items. Through the moderation process the original number of questions and items went down to 632 questions consisting of 1084 items. All these chosen items were moderated to qualify according to the criteria above.

5.4 NEPTON field-testing

5.4.1 The computerized testing software

The college Information Technology unit was asked by the NEPTON writer to develop an Intercollege in-house made computer testing environment, tailored to the specific needs of the college English placement programme, already described (see appendices 20 and 21). This computer testing shell ran on an IBM personal computer and allowed the test developer to create, administer, and analyze tests. Items were stored together in the computer memory along with descriptive information through the use of common classifiers (six levels; structure sentence and vocabulary sentence with a blank and five multiple choices; structure passage and vocabulary passage with four or five blanks and dropdown menus of five choices; reading comprehension - signs, and reading comprehension – passage with four questions and four multiple choices each, text divided into four sections with four titles to match). These classifiers permitted ready retrieval and rapid presentation for testing purposes, according to the programmed slide algorithm. Originally these were as follows:

Table 5.5: First programme slide algorithm

BENG-50 and BENG-80: Set a	
Slide 1	Slide 2
<ul style="list-style-type: none"> • Structure sentence 5 • 1 Vocabulary text cloze 4 	<ul style="list-style-type: none"> • Vocabulary sentence 3 • 1 Structure text cloze 5 • Signs 1
BENG-90 and ENG-101: Set a	
Slide 1	Slide 2
<ul style="list-style-type: none"> • Structure sentence 4 • 1 Vocabulary text cloze 5 	<ul style="list-style-type: none"> • Vocabulary sentence 4 • 1 Structure text cloze 5
BENG-50 and BENG-80: Set b	
Slide 1	Slide 2
<ul style="list-style-type: none"> • Structure sentence 5 • 1 Reading comprehension text 4 	<ul style="list-style-type: none"> • Vocabulary sentence 3 • 1 Structure text cloze 5 • Signs 1
BENG-90 and ENG-101: Set b	
Slide 1	Slide 2
<ul style="list-style-type: none"> • Structure sentence 5 • 1 reading comprehension text 4 	<ul style="list-style-type: none"> • Vocabulary sentence 5 • 1 skimming 4
2 slides per level (9 items per slide)=18 items per level x 6 levels= 108 each P & P field test paper Set a OR set B	

From the item banks presented earlier, items were randomly selected to firstly produce the pen-and-paper field-testing version, of the college New English Placement Test. This included items containing text and graphics. This process created a large number of unique tests of the slide design presented above. The test-generation programme randomly chose set (a) or (b) slide 1 and 2 in order, from BENG-50 to ENGL-100, choosing a different item from each of the nine items per slide each time. This process randomized the choice of items in two ways, item based and set based. The generated test consisted of 108 items (two slides per level equaling eighteen items by six slides equaling 108 items). Test-takers received the test in a booklet form, accompanied by a 3-page answer sheet.

5.4.2 NEPTON Pen-and-Paper field-testing

The NEPTON test was initially field-tested in pen-and-paper form in May 2004. A sample of about one thousand and six hundred students took part in the initial field-testing at three testing sites: Nicosia, Larnaca and Limassol Intercollege campuses (see appendix 15). The test-takers were incoming and continuing students who had enrolled at Intercollege during the Academic year 2003 - 2004, and were placed in one of the 6 levels of English instruction offered at Intercollege on the basis of their performance on the existing Intercollege English Placement pen and paper entry test. Most of them, continued English at a higher level than the one placed in the previous semester. A small percentage was repeating that level. Each test-taker was given a NEPTON pen-and-paper booklet, a 3-page answer sheet and a post-test questionnaire. Each test consisted of approximately 65 randomly chosen questions representing approximately 108 test items (text-based questions included between four to five test items each). The mode of item presentation followed the slide design mentioned earlier, from lower to higher levels, and each item was randomly chosen from the different test-item type pools (sentence, cloze-text, signs, text, matching text and titles). Students recorded their answers on the 3-page answer sheet. Students were given two hours to complete the test. The answer sheets were then scanned, and the results analysed. This process helped improve the content of the test and formulate the cut off points.

5.4.3 NEPTON field-test participants' evaluation

Post field-test questionnaires were given to test-takers to reveal their attitudes and reactions, and their feelings about the NEPTON test. The questionnaire items covered the following topics: instructions and length, activities, topics, visual, feelings before and during the test. I summarized statistically the scanned results, and the acceptability of items and test components was determined:

Post field Testing Student questionnaire (see appendix 16)

Table 5.6: Student NEPTON post field-testing questionnaire: Instructions and length

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 The instructions were clear.	3.3%	5.4%	19.6%	43.7%	27.6%
2 The length of the test was appropriate.	10.4%	15.8%	24.6%	33.0%	13.2%

Based on the data above, most test-takers found the instructions clear (43.7% agree and 27.6% strongly agree, a total of 71.3%). 26.2% strongly disagree or disagree that the length of the test was appropriate, 24.6% were neutral and 46.2% agree or strongly disagree. English lecturers, who acted as invigilators, found the test long. The test writer was aware that the test was long at that stage, but it was the only way to trial each item by as many test-takers as possible in order to have a satisfactory item analysis. For these reasons, we had the test long for the field pen-and-paper test, with the intention of investigating this aspect of the test and find ways to improve it for the final version.

Test-takers were also asked to comment on the test activities, topics and visual: Tables 5.7 and 5.8 sum up their answers:

Table 5.7: Student NEPTON post field-testing questionnaire: activities

		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I found the activities, which required filling the blank in a sentence manageable.	3.5%	8.1%	26.6%	42.6%	15.8%
2	I found the activities, which required choosing the best possible answer from a menu in a text manageable.	4.0%	9.6%	25.7%	43.4%	15.5%
3	I found the reading comprehension multiple-choice texts manageable.	6.6%	10.6%	27.2%	40.4%	13.9%
4	I found the activities, which required matching each part of a text with an appropriate heading manageable	4.8%	10.1%	31.4%	39.4%	11.1%
5	There was enough variety of activities.	4.5%	9.1%	21.5%	38.6%	23.4%
6	The activity layout was clear.	5.6%	9.9%	21.0%	41.9%	20.1%

From table 5.7, it is suggested that most test-takers felt comfortable with the sentence-based activities. We took into consideration some of their suggestions and made improvements. We also improved the layout of the text-based activities to facilitate test-takers' understanding. As suggested, we shortened the reading comprehension multiple-choice texts, which seemed long. We took the activities requiring matching each part of a text with an appropriate heading completely out of the test because the layout in both the electronic and the pen-and-paper versions of the test seemed confusing. Test-takers also seemed to have felt that there was enough variety of activities.

Table 5.8: Student NEPTON post field-testing questionnaire: topics and visuals.

		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I found the topics interesting.	7.6%	11.1%	28.9%	34.0%	16.5%
2	The visuals are appropriate.	5.1%	8.7%	29.0%	39.6%	14.0%

From the results in table 5.8, it seems that on the whole, test-takers found the topics interesting and the visual appropriate. The signs indicated by some test-takers as unclear will be redesigned.

Table 5.9: Test-taker NEPTON post-field testing questionnaire: feelings before and during the test

		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	Before the test I felt tensed.	22.6%	18.3%	26.2%	19.8%	10.4%
2	I felt comfortable during the test.	7.9%	7.6%	24.9%	31.5%	26.2%

Test-takers were also asked to express their anxiety level before and during the test. From what we can see in Table 5.9, the test-takers didn't feel particularly tensed before the test, and during the test they seemed to have felt quite comfortable. The comparison of the two indicates that the nature of the test made test-takers feel more and more comfortable during the test.

5.4.4 NEPTON Pen-and-Paper field-testing: improvement needs

The following difficulties were encountered during the test development and field-testing:

- (a) There was pressure from the administration to speed test preparation processes so that the test would be ready for the new academic year. This meant having to trial even at the end of semester 2, just before the exams and during exam preparation time, so that we could have some data analysis to the test. That was not the best of times to ask both faculty and students to field-test the pen-and-paper format of the proposed test.
- (b) There were some IT related problems: the uploading items took longer than expected; there were some mistakes in the item tagging during the uploading process; the agreed slide design was not consistently followed, and that needed to

be looked at and fixed; sometimes some editing done electronically was not refreshed in the electronic test environment and as a consequence was not included in the final version of the test when it was trialled electronically.

- (c) There were delays in the uploading of the test items and the preparation of the electronic test at the beginning of the semester. As a result, the trial took place almost at the end of the semester, just before the final examinations.
- (d) As a result, some students did not take it seriously and some instructors did not trial it with their classes.
- (e) Scanning had to wait because of other Intercollege administrative scanning priorities. This delayed the data analysis and the preparation for the electronic test trial.

The data analysis had two aims:

- (a) To improve the test and the test items
- (b) To come up with a method to work out the cut-off points for NEPTON.

NEPTON qualitative data:

The outcomes of the input from the instructors-invigilators, the test-takers' post-test open-ended questions and the test team observations indicated the following:

- (a) Some mistakes in the item tagging had to be corrected (for example, some questions were tagged as structure when in actual fact they were reading comprehension. These mistakes occurred during the data entering and uploading of the items)
- (a) Some checking needed to be done on possible cultural bias of some items
- (b) More / general editing needed to take place
- (c) The layout needed to improve
- (d) There was a need for more improvement of items, particularly in reading comprehension activities. For example, some texts needed to be shortened and the skills assessed needed to be defined more clearly.
- (e) Instructors felt that there had to be a written component to assure placements

- (f) The agreed slide design wasn't strictly followed at all times. The algorithm had to be checked.
- (g) Both test-takers and instructors felt the test was too long. It had to be, however, to ensure that all items on the item bank were substantially used.

All these points were taken into consideration and amendments were made accordingly: The mistakes in the tagging were corrected. Some possible cultural bias of some items was discussed and decisions were made. For example, where the word lift was used (a word used in England and Australia), there was a dash and the word 'elevator' (used in the United States of America) added next to it to cater for possible cultural bias. Two added editors were asked to look at the text again. They made suggestions, which were taken into consideration and amendments were made. The layout of the text-based tasks, which appear as dropdown menus in the electronic version, was improved in its pen-and-paper form. Some texts were shortened. A written component was included in the test to assure placements. The algorithm for the 9-item slide design was checked and fixed to ensure the agreed design. Improvements were made to some items, based on the suggestions made by the instructors-invigilators and the test-takers' post-test open-ended questions and the team observations.

Moreover, an improved format was used to generate new tests:

- (a) After discussion with my main assistant, I worked out a new slide design, which resulted in a new slide algorithm and the shortening of the test from 65 questions (108 items) to 33 questions (54 items).
- (b) I also reduced the types of activities: the matching skimming activity (the activity which asked the test-takers to match the four different paragraphs of the text with one of the four titles given to them) was taken completely out of the test.

5.4.5 Revised NEPTON slide paradigm

The following table presents the new slide paradigm, which was used to generate the final version of the NEPTON test, in both its electronic and pen-and-paper form.

Table 5.10: Revised, final NEPTON slide paradigm (After the test-field / trial)

	TEST 1	TEST 2
BENG 50		
9 items per slide	SB-S: 3 TB-S: 5 SB-RC: 1	SB-V: 4 TB-V or TB-RC: 4 SB-RC: 1
BENG 80		
9 items per slide	SB-V: 4 TB-V / TB-RC: 4 SB-RC: 1	SB-S: 3 TB-S: 5 SB-RC: 1
BENG 90		
9 items per slide	SB-S: 4 TB-S: 5	SB-V: 5 TB-V or TB-RC: 4
BENG 100		
9 items per slide	SB-V: 5 TB-V or TB-RC: 4	SB-S: 4 TB-S: 5
ENG 100		
9 items per slide	SB-S: 4 TB-S: 5	SB-V: 5 TB-V or TB-RC: 4
ENG 101		
9 items per slide	SB-V: 5 TB-V or TB-RC: 4	SB-S: 4 TB-S: 5
9 items x 6 slides	54 items (33 questions)	54 items (33 questions)

Total test length = 9 x 6 = 54 items²⁶ one-side answer sheet

A test version was randomly chosen, either Test 1 or Test 2. For each test, each test-taker was then presented with a series of six slides, one per level (BENG-50, BENG-80, BENG-90, BENG-100, ENG-100 and ENG-101). Each slide included nine items. These items were chosen from different pools of items, representing different types of activities and different skills. For example, in Test 1, the BENG-50 slide randomly chose three sentence-based structure items from the sentence-based structure item pool, five text-based structure items from the text-based structure item pool, and one sign-based reading comprehension item from the sign-based reading comprehension item pool. The same process continued for the random selection of items for all slides in the rest of the five levels. Test-takers ended up answering thirty-three randomly chosen questions, including fifty-four items.

²⁶ Activity types: Sentence-based Structure (SB-S), Text-based Structure (TB-S), Sentence-based Vocabulary (SB-V), Text-based Vocabulary (TB-V), Sign-based Reading Comprehension (SB-RC), and Text-based Reading Comprehension (TB-RC).

NEPTON also improved in all those areas including its layout. This allowed the electronic, randomized generation of a set of unique NEPTON (P) test, with each item randomly selected, compatible to the NEPTON version. A new one-side page answer sheet was prepared for NEPTON (P), for scanning. The improved item bank and as a consequence the reviewed test specifications made it possible to create a NEPTON version of the test, electronically delivered.

The results of the data analysis of the pen-and-paper field-testing also provided the method, which helped generate the NEPTON test cut-off points that were necessary for the automatic placement of test-takers. Statistical analysis of the responses to the pen-and-paper field-testing version of the test yielded the parameters that were used by the algorithm to drive the electronic version of the test.

5.4.6 Criterial levels of performance and placement procedures

NEPTON scores were interpreted meaningfully against the Intercollege English Language Programme Levels, which reflected the syllabus framework, the teaching material and the assessment procedures.

5.4.7 NEPTON objective component cut-off point paradigm

Both electronic and pen-and-paper objective test components were based on the following cut-off points:

Table 5.11: NEPTON cut-off point paradigm

Level	Points (out of 9)
BENG-50	5
BENG-80	4
BENG-90	4
BENG-100	5
ENG-100	4
ENG-101	5

These cut-off points were calculated for each slide and each level, and were the result of a long process of iteration of students' NEPTON results. Although mainframe computer programs were available, it was decided to use local expertise to find a system, tailored to the project needs to arrive at the cut-off points. This was developed by Professor Andreas Alexandrou and Chris Alexander. The data was entered on the spreadsheet grid and was analysed. A series of iterations were made in a statistical effort to refine the cut-off points.

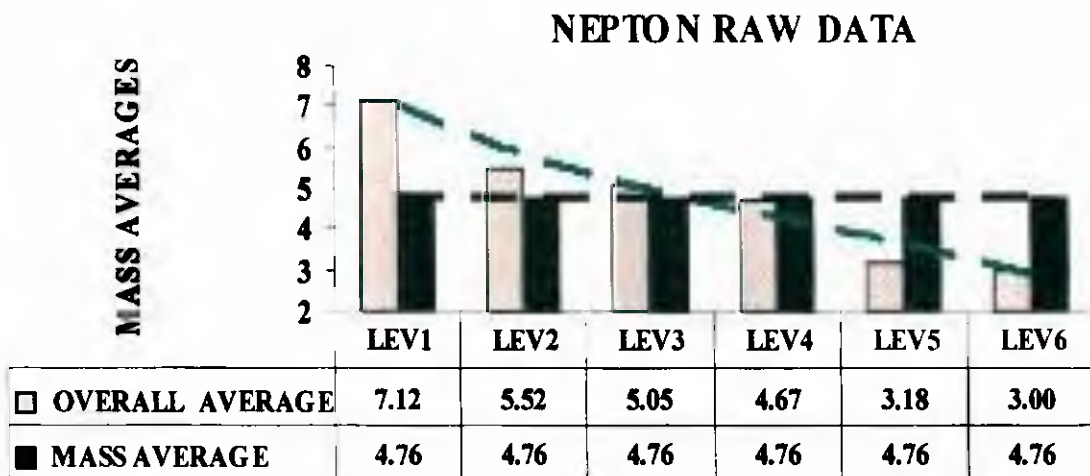


Figure 5.1. NEPTON raw data

Figure 5.1 explains this process: firstly we analysed the NEPTON raw data (this indicated the average of all test-takers taking all questions): we added the overall averages of all students taking all questions at each of the six levels and divided them by six to come up with the mass average (the total of the overall average). We made 4.76 into an integer, as we had to have whole numbers as cut-off points. The average cut-off point was five. Once we worked out the mass average we set our bounds, that is, what was to be acceptable.

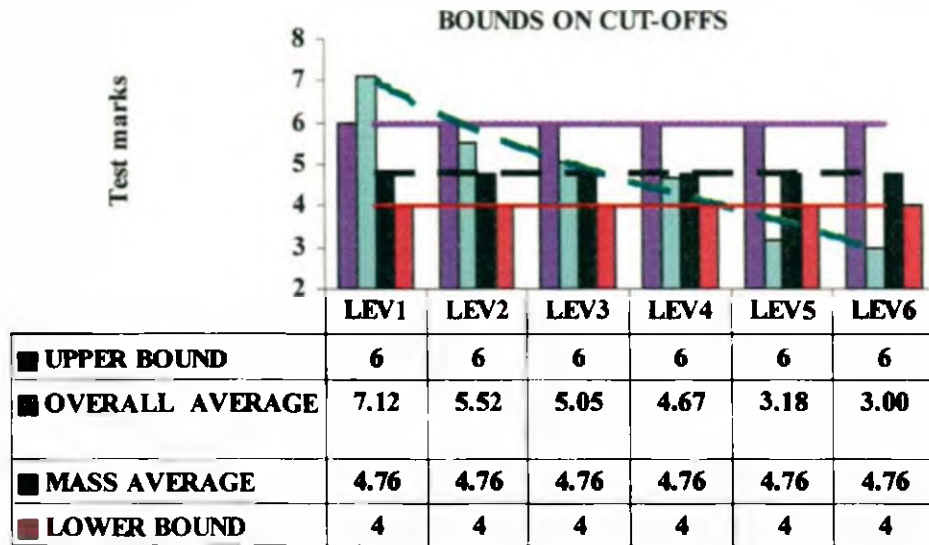


Figure 5.2. Bounds on cut-offs

The next step was to find the upper and lower bound. As figure 5.2 indicates, the mass average was five. Therefore, it was decided that the upper and lower bounds would be one above and one below, that is the upper bound would be six and the lower bound would be four. Once we found the upper and lower bounds we started the iteration process to find acceptable cut-off points, which would indicate the group average.

LOWER BOUND CUT-OFFS- ITER 1

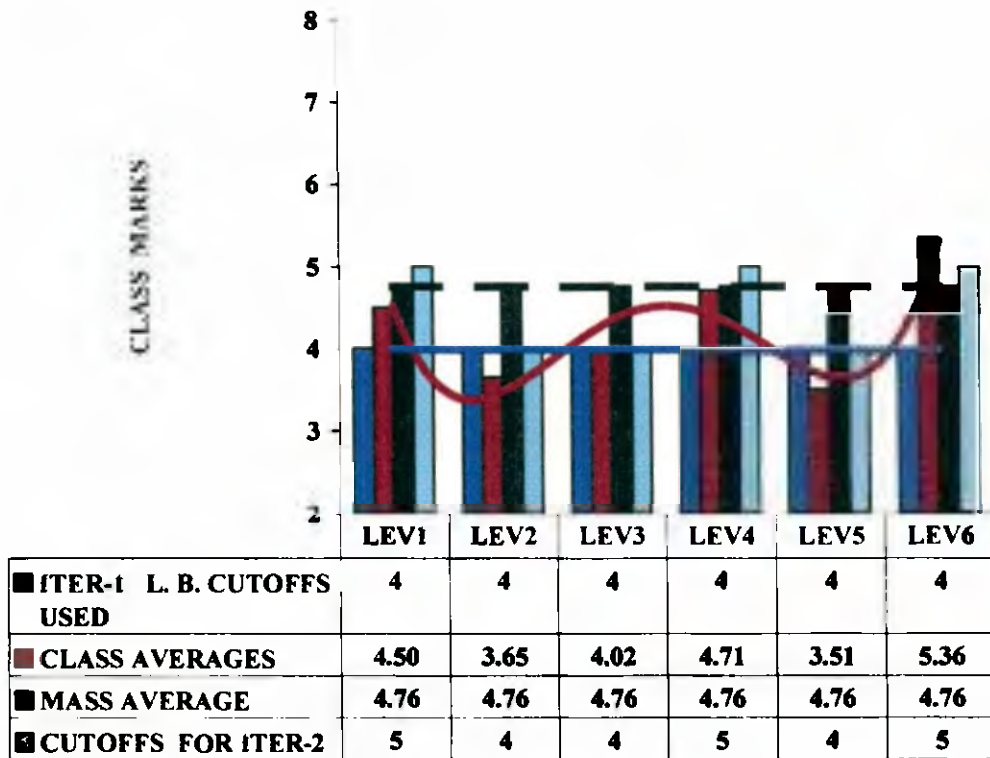


Figure 5.3. Lower bound cut-offs – iteration 1

Iteration 1 came up with the lower bound cut-off points. These were four at each of the six levels (figure 5.3).

A second iteration was carried out to try and improve this cut-off point. The result of Iteration 2 was: five, four, four, five, four, five (figure 5.4).

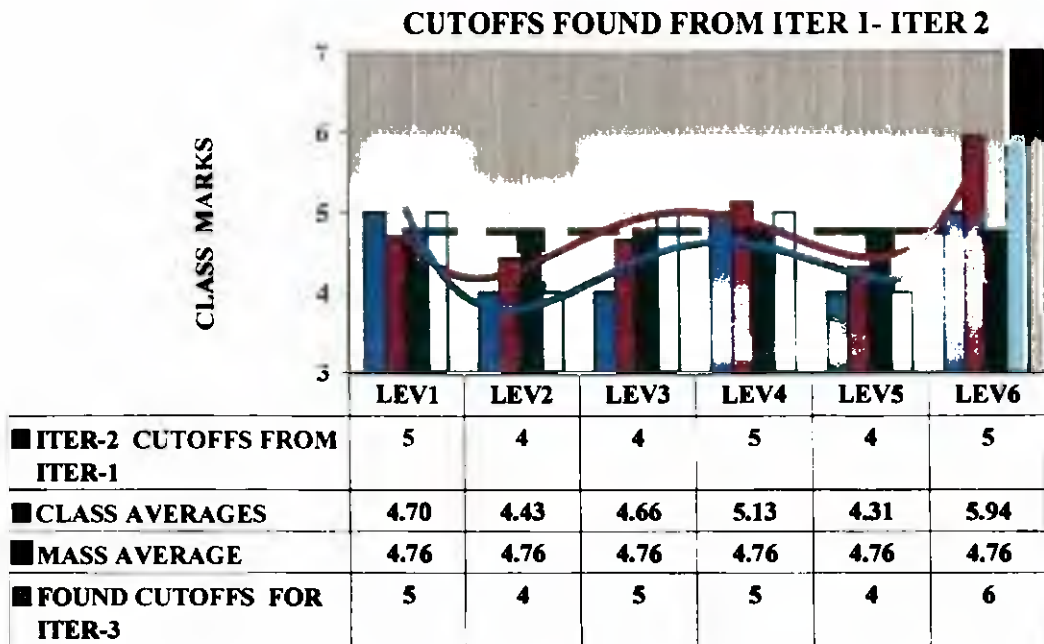


Figure 5.4. Cut-offs found from iteration 1 and iteration 2

The result of Iteration 2 (five, four, four, five, four, five) was again used in a third iteration to see if we could come up with better cut-off points.

CUTOFFS FOUND FROM ITER 2 - ITER 3

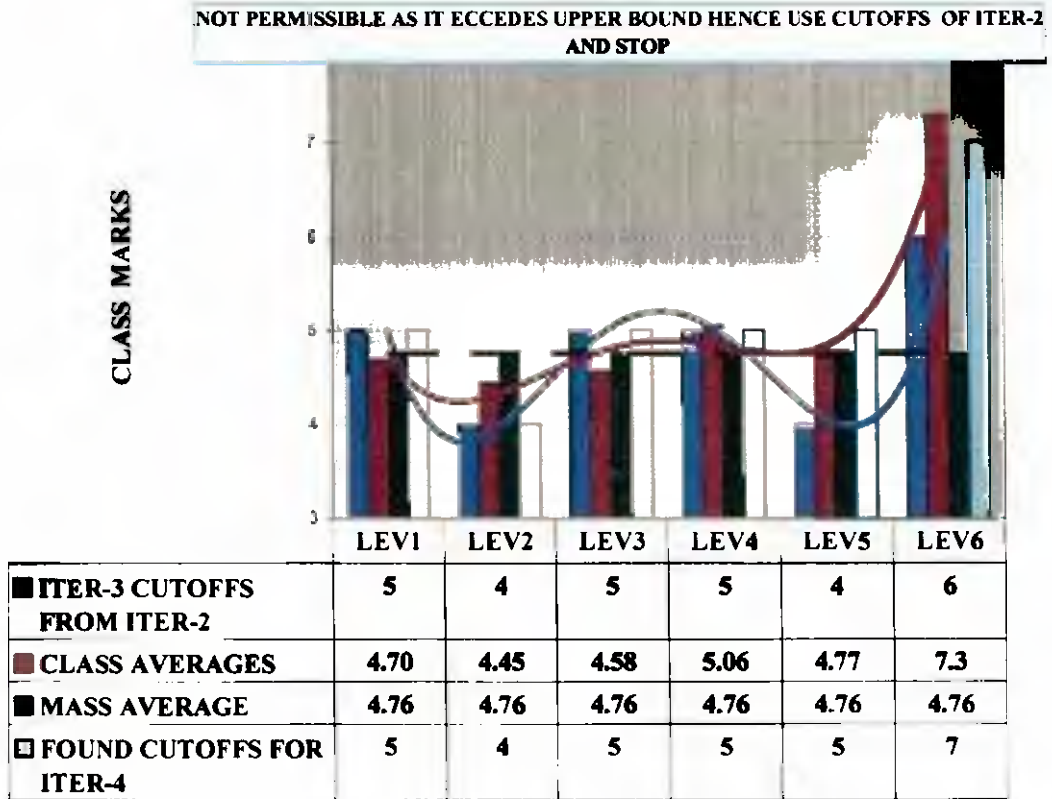


Figure 5.5. Cut-offs found from iteration 2 – iteration 3

The result of Iteration 3 was: five, four, five, five, four, six. These cut-off points were unacceptable because they went above the average. As a result, this would lead to an unacceptable distribution at the highest levels.

L. B., ACTUAL, U.B. COMPARISONS

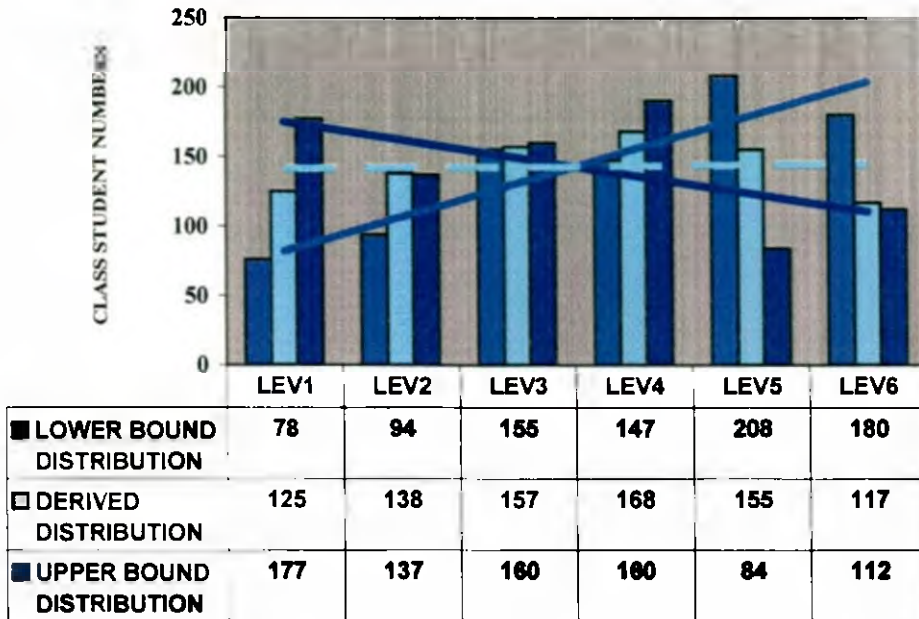


Figure 5.6. Lower bound, actual, and upper bound comparisons

The result of Iteration 3 (five, four, five, five, four, six.) was iterated again to come up with new cut-off points. The result of Iteration 4 was: five, four, five, five, five, seven. These cut-off points were again not permissible as they exceeded the upper bound.

Therefore, the best cut-off points were the ones found in iteration 2 that is five, four, four, five, four, and five: they fell within the boundaries that were set which were four for the lower bounds and six for the upper bounds.

As we have seen so far, the lower bound distribution at all level is four and the upper is five. Figure 5.6. compares the lower, upper and derived distribution of those test-takers at each level. The figures that follow illustrate the class numbers based on the various cut-off points we derived:

CLASS NUMBERS BASED ON L.B CUTOFFS

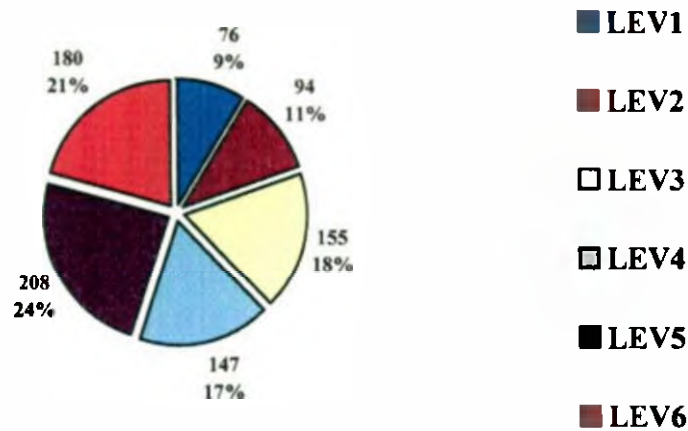


Figure 5.7. Class numbers based on lower bound cut-offs

Figure 5.7. indicates the class distribution based on the four, four, four, four, four, four cut-off points.

CLASS NUMBERS BASED ON DERIVED CUTOFFS

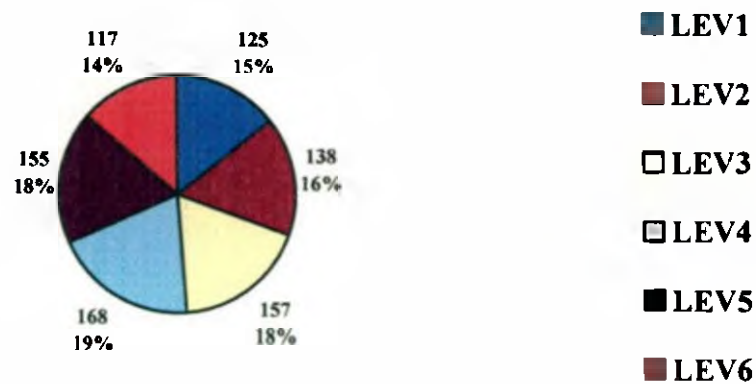


Figure 5.8. Class numbers based on derived cut-offs

Figure 5.8. indicates the class distribution based on the five, four, four, five, four, five cut-off points.

CLASS NUMBERS BASED ON U.B CUTOFFS

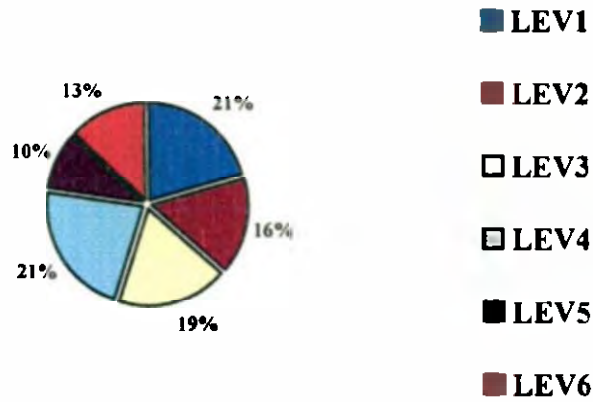


Figure 5.9. Class numbers based on upper bound cut-offs

Figure 5.9. indicates the class distribution based on the six, six, six, six, six, six cut-off points.

LOWER BOUND CUT-OFFS (4,4,4,4,4,4)

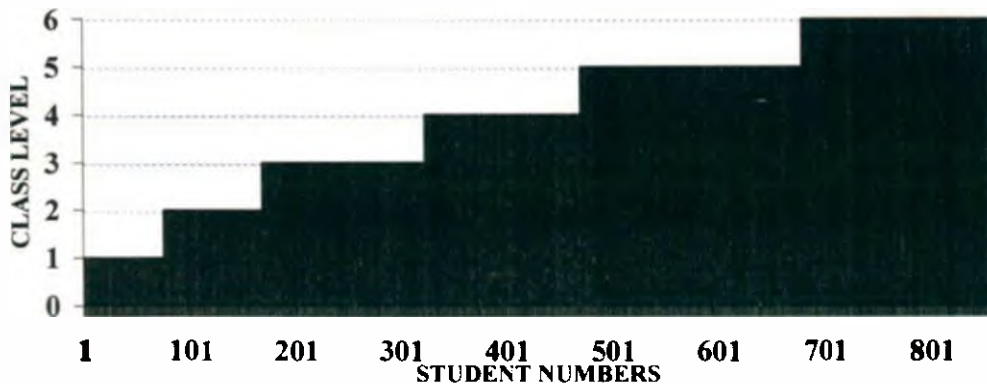


Figure 5.10. Lower bound cut-offs

Figure 5.10. indicates the number of student distribution at each level based on the lower bound cut-off points (four, four, four, four, four, four).

DERIVED CUT-OFFS (5,4,4,5,4,5)

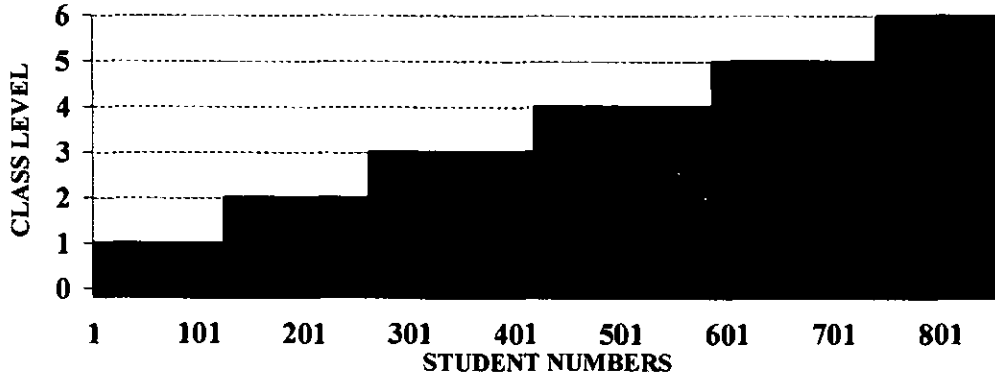


Figure 5.11. Derived cut-offs

Figure 5.11. indicates the number of student distribution at each level based on the lower bound cut-off points (five, four, four, five, four, five).

UPPER BOUND CUT-OFFS (6,6,6,6,6,6)

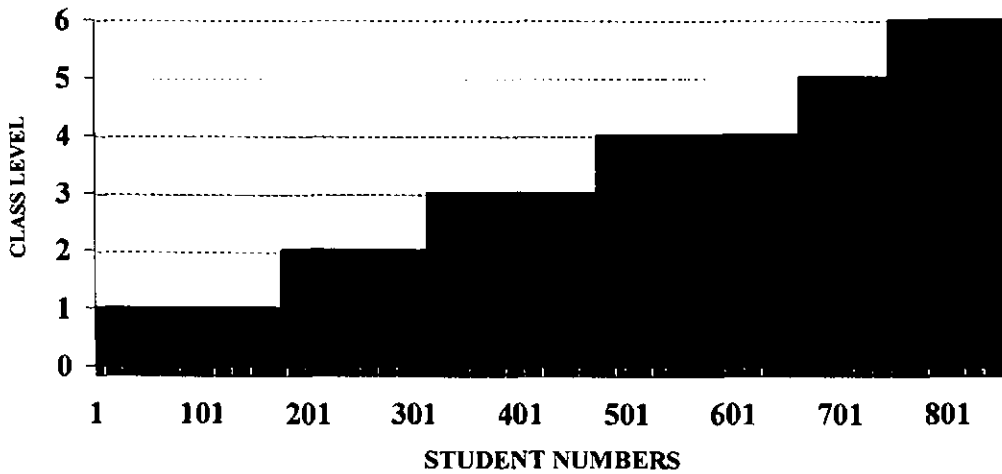


Figure 5.12. Upper bound cut-offs

Figure 5.12. indicates the number of student distribution at each level based on the lower bound cut-off points (five, five, five, five, five, five).

As a result, The NEPTON placement was computed statistically and based on the principles of the nine-item six-level slide design and the iteration process described above. There were certain cut-off points for each slide. As we have seen, the best cut-off points were the ones found in iteration 2, that is five, four, four, five, four, five, because they fell within the boundaries set (four for the lower bounds and six for the upper bounds). Test-takers moved from one level slide to another if they reached the designated cut-off point of each slide. In other words, if a test-taker reached the cut-off point of five for BENG-50, he or she moved on to the next level. If he or she reached the cut-off point of four for BENG-80, he or she moved on to the next level. If he or she reached the cut-off point of four for BENG-90, he or she moved on to the next level. If he or she reached the cut-off point of five for BENG-100, he or she moved on to the next level. If he or she reached the cut-off point of four for ENG-100, he or she moved on to the next level. If he or she reached the cut-off point of five for ENG-101, he or she moved on to the next level. This formula calculated the final placement and converted it to an Intercollege Placement English competence level.

5.5 NEPTON Implementation

As we have mentioned before, an online testing environment was created in-house at Intercollege for the implementation of NEPTON (see 5.5.3.I and appendices 20 and 21). The item bank was uploaded in the editor section of the online testing environment, and onto the web server at Intercollege. The item bank was firstly used to randomly generate the pen-and-paper format of the test, which was field-tested in May 2004. As discussed earlier, as a result the items were improved, the test was shortened and improved, a new slide algorithm and a system to arrive at cut-off points for the electronic version were worked out. The changes were made based on the test-takers' post-test questionnaires, the instructor-invigilators' comments and the NEPTON test writer observations. Then, the electronic format of the test was trialled. The test was randomly generated following the new slide algorithm. The following is a description of the process of the electronic test trial:

5.5.1 NEPTON sitting

Eight computer labs were used, a total of 127 computers at a time. There were continuous test-taker sittings between late August and the end of October. Varied numbers of test-takers took the test on a daily basis. The total of test-takers was about eight hundred (only three opted to take the test in its pen-and-paper form). The test was hosted on the college server and delivered via the Internet. Test-takers were given access permissions, which were switched on just before the beginning of the test, and switched off as soon as it ended. The test-takers were invigilated in the computer laboratories at the college, by college staff who helped test-takers through the tutorial session, log on to the test, and ensured that each test-taker submitted the answers for scoring on completion of the test. These invigilators were familiar with the test format because, prior to the test-takers taking the test, they had three hands-on in-service training sessions which helped them familiarise themselves with the test. I, as the project leader, a project team member and two lab assistants were on stand-by for all laboratories at all times. The test-takers accessed the test by clicking the NEPTON test shortcut on the computer screen and using the Internet Explorer browser. Test-takers were given thirty minutes for the tutorial, eighty minutes for the electronic test and forty minutes for the hand-written essay. We gave test-takers the maximum of time, knowing that this total time was not necessary. We wanted to make sure we could collect all the necessary data. We also wanted to be flexible in terms of the duration of the test, during this phase of the project. Test-takers were allowed to leave the room once they had finished with both parts of the test, electronic and written. We electronically monitored the time it took each test-taker to take the electronic test to establish the average time needed for the test in the future.

5.5.2 NEPTON and the test-takers

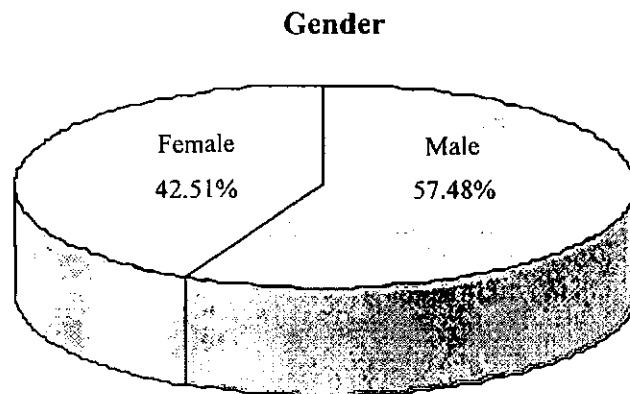
There were about eight hundred test-takers involved in the autumn NEPTON administration. All test-takers took the NEPTON test (as well as the non-electronic written task). Test-takers were asked to complete pre and post test questionnaires: Two

hundred and sixty three test-takers volunteered to complete the pre-test questionnaire and one hundred and thirteen test-takers chose to complete the post-test questionnaires.

5.5.2.1 NEPTON test-taker demographic information and computer familiarity

Testing began with the pre-test questionnaire (see appendix 17). From that, we extracted the demographic information about the test-takers (age, gender, English language background, and the subject students intended to study at college) and their basic computer skills and familiarity with computers and the new technologies.

5.5.2.2 Test taker gender, age, linguistic background and country of origin.



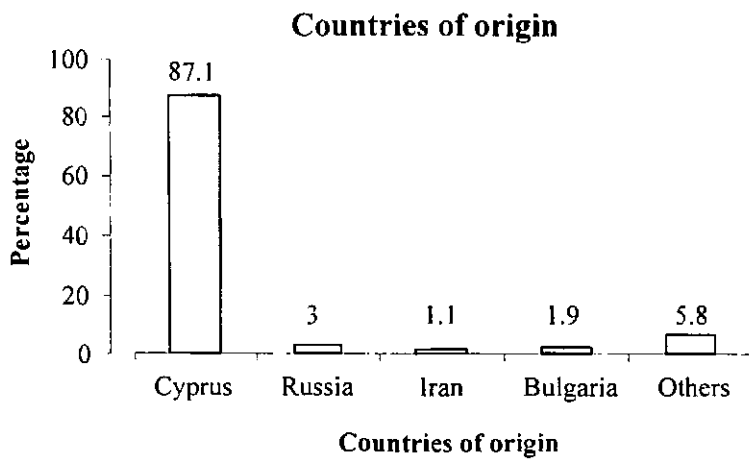
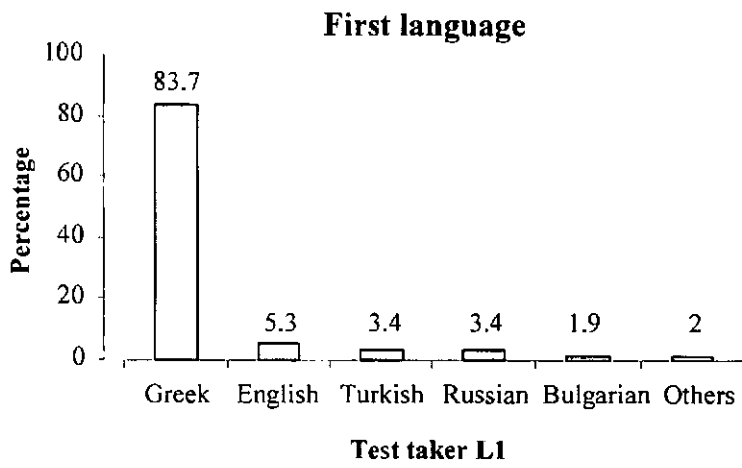
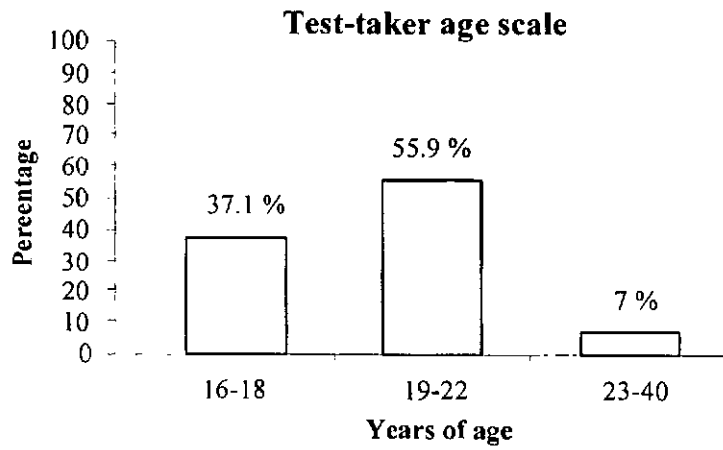


Figure 5.13. Test-taker biographical data

The sample consisted of more male test-takers. They ranged from seventeen to forty years of age. Most of them were of Greek language background with smaller percentages of English, Turkish, Russian, Bulgarian and other linguistic background. Most of them were from Cyprus with a smaller percentage from Russia, Bulgaria, Iran and other countries.

5.5.2.3 Test familiarity and test-taker equity

The aim of NEPTON and NEPTON (P) was to ascertain the students' level of English at the time of testing; therefore Intercollege did not expect that the test-takers would need to spend time preparing for it. However there was provision for test-takers to familiarize themselves with the test format, the skills required to take the test and prepare before taking the test by doing one of the following:

- (a) For both NEPTON and NEPTON (P), test-takers had previous access to the NEPTON booklet, which informed them about the purpose of the test, and its format, and provided them with examples of the test activity types.
- (b) For NEPTON, students could go through the electronic Tutorial. This provided a tutorial of the computer basic skills they needed to take the test (mouse clicking, and scrolling). It also provided test-takers with a warm-up and test familiarization session. This included a description of the electronic environment, question and activity type samples, how the test and the interface layout looked like and the number of questions and items test-takers needed to answer to complete the test. Finally, through the tutorial, students also had direct access to the test and first hand experience of how it looked and worked. At the end of the tutorial, invigilators supplied each test-taker with a unique username and password. Access permission was then switched on.

5.5.3 Taking NEPTON

Test-takers took the test using a personal computer. They used their credential. (user name and password) to enter the NEPTON test area. The first page they came across was the registration area. They were asked to give their surname, first name, and date of birth. Once they finished with the registration, they entered the test area where they took the test. They communicated with the computer by using the keyboard and the mouse to enter their answers and to supply other information. The computer screen allowed the test-takers to see the instructions, questions, multiple choice response options, their responses and other information presented to them.

5.5.3.1 New English Placement Test ONLINE

This is how the test interface looked:

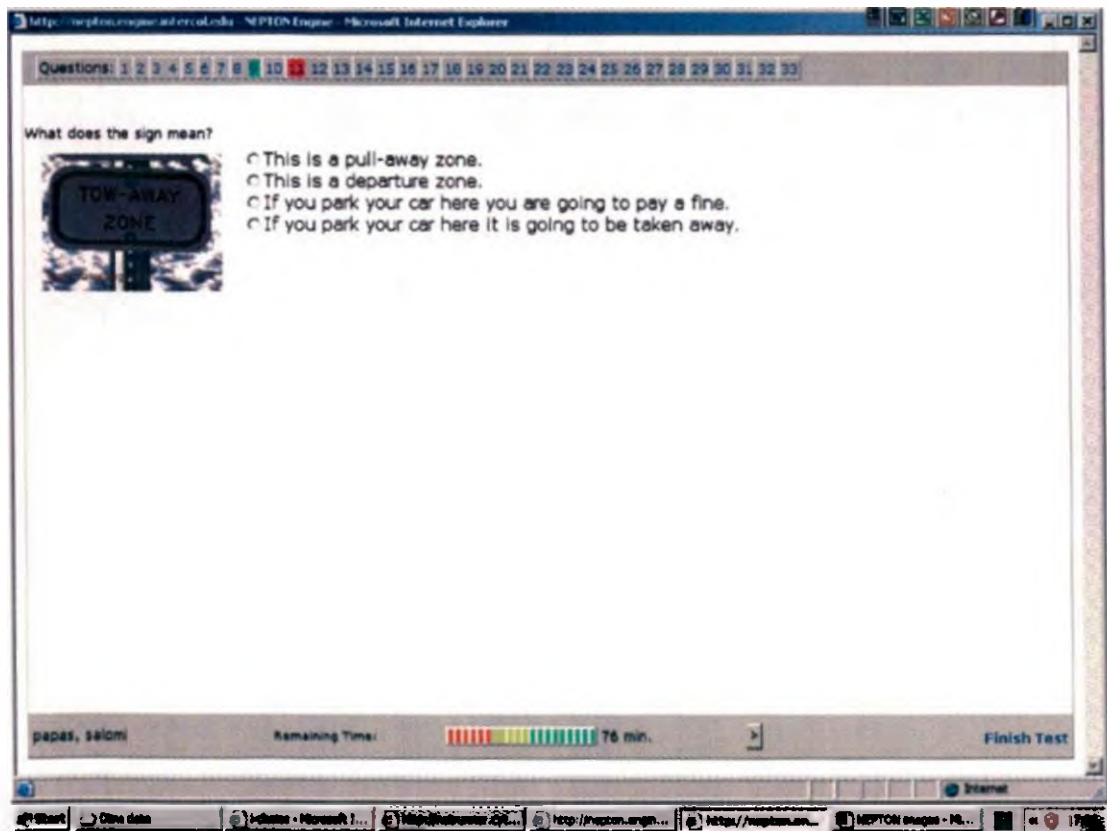


Figure 5.14. NEPTON test interface

At the top of the screen, test-takers could see the number of questions they would go through, from number 1 to number 33. The red colour indicated which question test-takers were at. The green colour indicated the questions test-takers had answered. When test-takers completed a question and verified their answer, they clicked on a new question and a new screen appeared with their next question. Test-takers could move from one question to another by clicking on the different question numbers or by clicking on the > button, at the bottom of the screen. Test-takers could go backwards and forwards, and revisit questions by using the number question buttons at the top of the screen. The thirty-three questions represented an algorithm of six level-based slides, incorporating nine test items each (equalling 54 items in total), each in the form of, a sentence, a text with dropdown menus, and a sign, or a text with multiple-choice questions. At the end of the test, when they completed all 33 questions, test-takers clicked on the *Finish Test* button. If they clicked on *Next (>)* *Last Page* appeared next to the *Finish Test* button. This meant it was the last page of the test and test-takers had to click on the *Finish Test* button to exit the test. Another window popped up and asked test-takers if they were sure they wanted to exit the test area. If they wanted to go back to the test and continue working on it, they had to press *No* and that would take them back to the test. If test-takers were sure they had completed the test and wanted to exit it, they pressed *Yes*. In that way, the electronic test was completed. Test-takers could not re-enter the test area, once they clicked on the *Finish Test* button. If during the test the power was cut off, the test-takers could still enter the test area using their username and password and continue the test where they left it, once the power was restored. There are two more test features worth noting. At the bottom left side of the screen, test-takers could see their name, as they entered in when they registered. On the right of their name, in the middle, bottom part of the screen, test-takers could monitor the time they had left to complete the electronic part of the test (see appendix 20 for more printed samples of the NEPTON electronic test).

When the test-takers finished the test, access permission was automatically switched off. The results of the electronic part of the test were recorded and stored automatically but were not revealed to the test-takers. Once the test-takers completed the electronic

component of the test, they were given the written component. They were given five essay topics to choose from. At the end of the hand-written component of the test, test-takers were given a post-test questionnaire to establish their feeling about the NEPTON.

5.5.3.2 Subjective Component: hand written assessment task

Like its pen-and-paper predecessor, the NEPTON was for placement purposes. What was needed, therefore, was an efficient and accurate placement test. To assure proper placement while the NEPTON test was in its formative stages, supplementary evaluation in the form of essays was carried out. With this assurance, it seemed satisfactory enough to proceed with the NEPTON experiment, using the electronic activity formats presented earlier: sentence-based structure or vocabulary, text-based structure or vocabulary, reading comprehension (sign with text and text with multiple-choice questions).

Experienced markers assessed the writing task. In general, the written test was assessed on how well test-takers met the set criteria, covering each of the six language levels: natural use of language, range of vocabulary and structure, content organization, clear communication and completion of task (see appendix 12 for more detailed marking criteria). Each written task was marked by at least two markers and if there was disagreement, a third marker was used. They placed the test-takers at the appropriate level.

5.5.4 Final NEPTON score and result reporting

Test-takers were only placed in one of the six language levels by the NEPTON test if they reached a minimum level at each level, according to the cut-off points calculated. The reported score indicated the English language level test-takers were placed. According to experts (Dunkel 1991), electronic tests should be used in conjunction with a guided oral interview and an essay. The NEPTON followed part of Dunkel's suggestions and was therefore considered in conjunction with a written task. The oral interview was found practically impossible to implement at this stage. Both electronic

and written results were considered and the level of English of each test-taker was determined, to which each test-taker was placed (see appendices 13 and 21).

5.5.5 Conclusion

The needs analysis provided the necessary information about the context, setting, students, lecturers, English programme, and International English language equivalences, which helped in choosing the appropriate test type and write the test specifications for this particular case study. A process of test item writing and moderation, online test environment development, field-testing and trialling gave new insights in the area of English online testing through the use of combined work based / applied linguistics research theories.

Chapter six

Project Findings: results and discussion

6. Introduction

This chapter provides an analysis and discussion of the findings. The questions were whether the NEPTON test was reliable, valid, and brought change and improvement. The assessment of successes, failures and lessons learned at each stage are discussed.

6.1. Test reliability

6.1.1 The NEPTON overall performance

The following are some general findings related to the NEPTON overall performance:

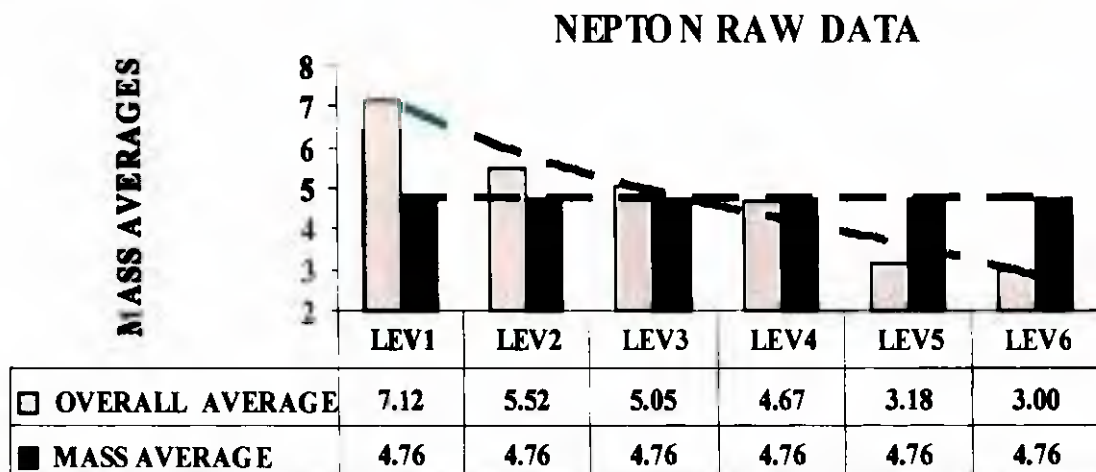


Figure 6.1. NEPTON raw data

The NEPTON raw data indicated the average of all test-takers taking all questions. Each overall average indicated all students taking all questions at each level. The mass average indicated the total of the overall average divided by six. The raw data suggested an acceptable reliability of the items at each level, starting from higher at level 1 (BENG-50 and going to lower at level 6 (ENG-101), however the results for level 5 and 6 indicate that there was not much discrimination between these two levels. This meant that the items in those two levels needed to be investigated further and in more detail during the item analysis process.

CLASS NUMBERS BASED ON DERIVED CUTOFFS

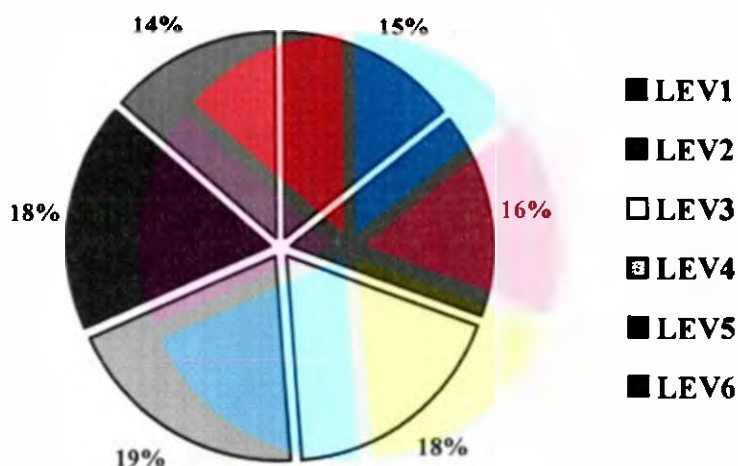


Figure 6.2. Class Number based on derived cut-off points

Figure 6.2. indicates the class distribution based on the five, four, four, five, four, five cut-off points which was found to be the most acceptable cut-off point system (five was the average cut-off point, six was the upper bound and four was the lower bound; five was the cut-off point for BENG-50 / level 1, four for BENG-80 / level 2, four for BENG-90 / level 3, five for BENG-100 / level 4, four for ENG-100 / level 5 and five for ENG-101 / level 6). It is an even distribution, both educationally (test-takers are placed in six language levels, according to their results) and administratively (there is an even distribution of students in the classes at each language level).

6.1.2. General factors influencing reliability

As Bachman (2003) mentioned, a fundamental concern in the development and the use of a language test is to identify potential sources of error in a given measure of communicative potential ability and to minimize the effect of these factors on that measure. In other words, we must be concerned about errors of measurement, or unreliability. We must minimize the effects of sources of such unreliability. The following are some ways that were used to make the NEPTON more reliable and to

check to what extent its objective, electronic component was sufficiently reliable for its purpose. Most of them are based on Hughes' suggestions (1989) and Dunkel's (1999) factor categorization:

Large item bank: The NEPTON test bank included 635 questions, a total of 1084 items. From that pool, the fifty-four items were randomly chosen to form a test for each test-taker. These items reflect the test objectives stated in the test specifications, for example, in content, context, text types, and skills. The large item pool strengthens the test's reliability (Hughes 1989).

Instructions: Every effort was made to develop short and clear instructions for the NEPTON test. This was how test-takers expressed themselves on that issue.

Table 6.1: NEPTON test-taker post-test questionnaire – Clear instructions (see appendix 17)

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The instructions were clear.	1.8%	4.4%	10.6%	38.1%	40.7%

From the test-takers answers it seems clear that most test-takers (78.8%) felt quite comfortable with the test instructions. It is evident that the type of instructions used help in the reliability of the test.

No question choice: In the NEPTON, test-takers do not have a choice of questions. Moreover, in the test structure types of questions, five multiple-choice items were used (five choices of distractors to choose from). These are considered more reliable.

Moderation: As we have seen, a moderation process took place in the initial stages of the item bank development. This resulted in the elimination of ambiguous items, the agreement of acceptable responses, and the editing of questions and texts and helped as a factor ensuring the test reliability.

Discrete items, objective scoring: The use of only items that permitted objective scoring (multiple-choice questions in different forms) in the NEPTON test also contributed to the test reliability.

Test specification and scoring key: In the Test Specification document, which described the NEPTON test, there was provision of a detailed cut-off point system.

Unique tests: All items were randomly chosen to form a unique test for each test-taker.

Clear computer screen layout: The NEPTON computer screen layout was clear and easy to understand. It did not include many features, which would make it look very cluttered and as a result confusing to the test-taker.

Table 6.2: NEPTON test-taker post-test questionnaire – computer clarity, activity layout and visual

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 The computer screen was clear.	1.8%	2.7%	8.8%	39.8%	44.22%
2 The activity layout was clear.	1.8%	3.5%	18.6%	36.3%	35.4%
3 The visuals were appropriate.	1.8%	6.2%	18.6%	40.7%	28.3%

Most NEPTON test-takers were happy with the clarity of both the computer screen (39.8% agree and 44.22% strongly agree) and the activity layout (36.3% agree and 35.4% strongly agree). Visuals were generally found clear and of sufficient quality. Some of the signs, which were indicated as unclear are to be redesigned.

Item analysis

Item analysis is used to construct reliable measurement scales and evaluate their reliability. The item analysis “tests the test”. The item analysis report analyzes each test item and the overall test reliability. Test reliability measures the accuracy, stability, and

consistency of the test scores (The Scoring Statistics Guide, 2005). The item analysis helps eliminate ambiguous and inappropriate items. There are three major measurement theories, which can estimate reliability: the classical true score theory, the generalizability theory and the item response theory (IRT) (Bachman 2003). Although modern measurement theory has developed beyond the classical measurement theory (CMT), CMT is still useful in many situations where neither generalizability theory nor item response theory (IRT) can be applied due to practical or other limitations. For example to use the IRT measurement theory, the test-takers must exceed the number of 1000. We used the classical measurement theory (CMT) for item analysis to measure the reliability of the NEPTON test. The number of test-takers was less than 1000. There was an urgent need to implement the NEPTON test for more efficient delivery, administration and placement. Although the NEPTON test was not a computer adaptive test, its items were randomly chosen, based on the algorithm developed for the specific needs of the NEPTON test. For this reason CMT offered flexibility in the analysis of the items.

Due to the constraints and contingencies of the institution, which had an urgent need for the introduction of the NEPTON, the item analysis had to be delayed. When item analysis was carried out, there were 1084 items in the item bank. Each item was analysed to establish its facility value. We examined relevant data to arrive at the system to be used to analyse item Facility Value:

Item analysis statistics

The basic purpose of a placement test is to spread test-takers out along a general continuum of language abilities. The test items, however, which test the test-takers in that continuum need to be analysed. Two item statistics are typically used in the item analysis of placement tests: facility value (FV) and discrimination index (DI) or item facility (IF) and item discrimination (ID) as referred to by Brown (2003).

Facility Value (F.V.)

Facility Value (FV) is defined as the proportion of students who answered a particular item correctly (Brown 2003, *Special Connections*, *Item Analysis* 2005). If 90 out of 100 students answered a particular item correctly, for example, the proportion would be $90/100 = .90$. An FV of .90 means that 90% of the students answered the item correctly, and by extension, that the item is very easy. Ideal items in a placement test should have an average FV of .50. Such items would thus be well centered, in other words, 50 percent of the test-takers would have answered correctly, and by extension, 50 percent would have answered incorrectly. In reality however, items rarely have an FV of exactly .50. According to Brown (2003, pp. 16-19) those that fall in a range between .30 and .70 are usually considered acceptable for placement purposes. In the *Scoring Office, Michigan State University* 'Item Analysis' website (2005), the desired indices of difficulty by test constructors are no lower than 20 and not higher than 80, with an average index of difficulty from 30 or 40 to a maximum of 60. In the *Test Scoring Statistics Guide* (2005) the overall measure ranges between 60 to 80. According to Alderson et al. 1955, p 81), "... if examiners want a wide spread of scores from an exam, that is if they want students' scores to range from very high to very low, then they will select items which are as near to an F.V. of 50% as possible because such items provide the widest scope for variation among the individual students..."

Discrimination Index (D.I.)

The discrimination index measures the ability of an item to distinguish between lower and upper scoring students taking the test. One way the index is computed is by taking the number of students in the upper 27% group who answered correctly minus the number in the lower 27% group who answered correctly and dividing by 27% of the total number of students (*Test Scoring Statistics Guide, Interpreting the Reports* 2005).

Another way is to sort out the tests by total score and create two groupings of tests, the high scores, made up of the top half of tests, and the low scores, made up of the bottom

half of tests. For each group the difficulty index of each item is calculated. Then the difficulty index for the low scores group is subtracted from the difficulty index for the high scores group (Spccial Connections 2005).

In the *Test Scoring Statistics Guide, Interpreting the Reports* (2005), the discrimination index ranges in the following way:

The discrimination index ranges between 1.00 and -1.00. 1.00 is achieved if all upper students got the item right and all lower students missed it. 0.00 indicates that the same number of students from each group marked the item correctly. A negative number indicates that the lower students did better than the upper students on the item. Ideally, test items will have a positive discrimination index above 0.30. Very easy or very difficult items cannot discriminate well between students. Overly difficult items are missed by both upper and lower scoring students while overly easy items are correctly answered by both groups. As a result, overly easy and overly difficult items will have a low discrimination index. You may still want to include these items (to cover the test content, for example) even though they do not assist in assigning grades (*Test Scoring Statistics Guide, Interpreting the Reports* 2005, electronic reference).

Once those items that fall within, for example the .25 to .80 range of F.V. are identified, the items among them that have the highest D.I. (for example >0.30) should be further selected for inclusion in the revised test. This process would help the test designer to keep only those items that are well centered (for example between .25 and .80) and discriminate well (for example, above .30) between the high and the low scoring students.

The NEPTON item analysis

The item analysis was done using the tests of 584 test-takers and the 1084 test items included in the test battery. Although the intention was to have every item used by a considerable number of test-takers so that all items would be sufficiently analysed, the item analysis process revealed that, because of the randomised item selection, the number of times each item was selected varied, with an average of thirty-five. We proceeded with the item analysis, bearing in mind that this would be a continuous process for item improvement.

The NEPTON Facility Value (F.V.)

Based on the literature on item analysis and the research context, it was decided that an acceptable facility value for each item would be between 25% to 80%. To establish the facility value of the items, each item was ranked from high to low level. The total of students who took each item was recorded next to each item, together with the number of correct answers per test item. The total number of correct answers was then divided by the total number of students who took the item to establish the item's facility value. Table 6.3 derives from the data used for item analysis. The table indicates how the test item data have been recorded and the way the facility value of each item (1084 test items in total) was found.

Table 6.3: Item Analysis: Facility Value - sample

Item Level	Item Type	Item Identity	Total number of students who took the item (S)	Total number of Correct Answers (C)	Facility Value (C/S)	Items with good facility value: 25%-80% (✓)
ENGL-101	Reading Comprehension	767 /i4	14	12	0.85	
ENGL-101	Reading Comprehension	767 /i2	13	11	0.84	
ENGL-101	Vocabulary Text	640 /i2	34	27	0.79	✓
ENGL-101	Reading Comprehension	767 /i3	14	11	0.78	✓
ENGL-101	Structure Text	556 /i4	23	18	0.78	✓
ENGL-101	Structure Sentence	548 /i1	22	15	0.68	✓
TOTAL of test items with good Facility Value						665 / 1084

All the NEPTON test items were analysed and the following Facility Value Index was revealed:

Item Analysis: Facility Value

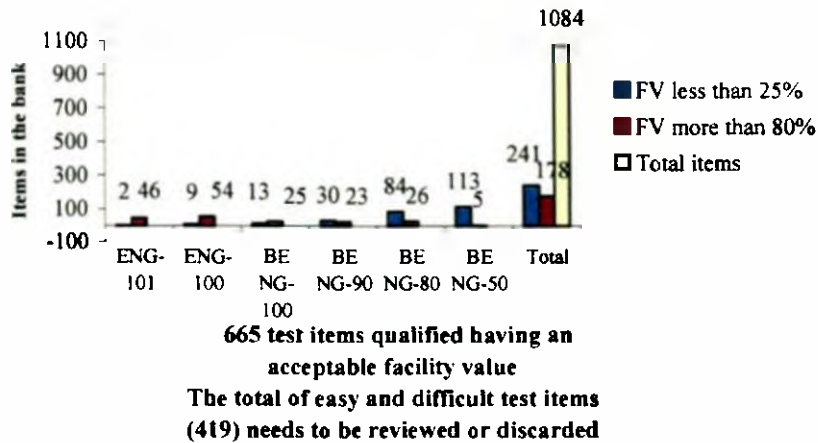


Figure 6.3 Item Analysis: Facility Value (F.V.)

As seen above in figure 6.3, two hundred and forty-one items were easy (2 ENGL-101 items, 9 ENGL-100, 13 BENG –100, 30 BENG-90, 84 BENG-80 and 113 BENG-50), in other words they ranged between 80% to 100%). One hundred and seventy-eight items were difficult (46 ENGL-101 items, 54 ENGL-100, 24 BENG-100, 23 BENG-90, 26 BENG-80 and 5 BENG-50), in other words they ranged between 0% and 25%. The total of 241 easy items and 178 difficult items was 419 items. These were outside the range of 25% to 80% and therefore were to be reconsidered for revision, improvement, fixing or deletion, and as a consequence establish new cut-off points through a new iteration process. Out of the total of 1084 items in the test item bank, 665 qualified having an acceptable facility value between 25% and 80% that is 2.5 to 8.0.

Table 6.4 Item Analysis: Facility Value (F.V.)

Item Analysis: Facility Value				
Percentages of items				
Easy items	%	Difficult items	%	TOTAL
ENG-101	4.24%	ENG-101	0.18%	4.42%
ENG-100	4.98%	ENG-100	0.83%	5.81%
BENG-100	2.21%	BENG-100	1.19%	3.40%
BENG-90	2.12%	BENG-90	2.76%	4.88%
BENG-80	2.39%	BENG-80	7.74%	10.13%
BEMG-50	0.46%	BEMG-50	10.42%	10.88%
Total	16.40%	Total	23.12%	TOTAL 39.52%

As a result, table 6.4 indicates that 23.12% of all items at all levels had FV more than 80% and 16.40% had FV lower than 25%. This brought the total of items that needed to be retested and / or revised to 39.52%.

The NEPTON Discrimination Index (D.I.)

According to Alderson et al. (1995), "... the highest discrimination possible is +1.00, which is achieved if all the students in the top group get an answer right and none of the students in the bottom group does. Such items are very unusual. Often item writers are content with D.I.s of +.4 or above, but there are no rules as to what D.I.s are acceptable, since the possibility of getting high D.I.s varies according to the test type and range of ability of the examiners..." *Test Scoring Statistics Guide, Interpreting the Reports* (2006) argues that "... Ideally, test items will have a positive discrimination index above 0.30..." The latter was used to calculate the discrimination index of each test item of the NEPTON test.

To establish the discrimination index of the NEPTON test items, all 1084 items were ranked from high to low level. The total of students who took each item was recorded next to each item, in two categories (high and low score groups). The number of correct answers (CH) in the high score group and the total number of students who took each item at high score group (SH) were recorded next to each item. The difficulty index was then calculated for the high score group (DH). In addition, the number of correct answers (CL) in the low score group and the total number of students who took each item at low score group (SL) were recorded. The difficulty index was calculated for the low score group as well (DL). The low and high difficulty indices were then calculated to arrive at the discrimination index of each item. Four hundred and thirteen out of 1084 test items had an acceptable discrimination index above .30. Table 6.5 is a sample page out of the 12-page report that demonstrates this process of establishing the discrimination index of each test item.

Table 6.5: Item Analysis: Discrimination Index – sample

Item Id	Total number of students	Facility Value	Total number of correct answers at high level (CH)	Total number of students at high level (SH)	Difficulty Index for high score group (DH-CH/SH)	Total number of correct answers at low level (CL)	Total number of students at low level (SL)	Difficulty Index for low score group (DL-CL/SL)	Discrimination Index (DH-DL)	Acceptable Discrimination Index: above .30 (✓)
787 /A4	14	0.93	7	8	0.28	3	6	0.33	0.04	
767 /A2	13	0.24	6	8	0.73	3	9	1	-0.29	
648 /A2	34	0.79	13	13	1.00	34	21	0.67	0.33	✓
767 /A3	14	0.78	7	8	0.22	4	6	0.67	0.21	
536 /A4	33	0.78	11	17	0.92	7	33	0.64	0.28	
548 /A1	22	0.68	10	11	0.91	9	11	0.45	0.45	✓
956 /A1	24	0.66	10	12	0.23	6	12	0.50	0.38	✓
818 /A1	47	0.65	18	23	0.78	13	24	0.54	0.26	
942 /A1	31	0.63	7	12	0.58	7	18	0.70	-0.12	
649 /A4	23	0.60	7	8	0.22	7	15	0.47	0.41	✓
955 /A1	29	0.58	12	13	0.92	8	16	0.31	0.61	✓
947 /A1	19	0.57	8	18	0.28	3	9	0.33	0.47	✓
958 /A2	21	0.57	6	14	0.37	4	7	0.57	0.08	
763 /A1	58	0.56	11	15	0.73	6	15	0.40	0.33	✓
958 /A5	38	0.56	12	14	0.26	5	16	0.31	0.54	✓
955 /A5	38	0.55	13	17	0.76	6	21	0.38	0.38	✓
Total number of test items with acceptable Discrimination Index										413
Total number of test items										1884

The test items that fell within the .25 to .80 range of facility value and the items among them that had the highest discrimination index (>0.30) were further selected for inclusion in the revised test. This process helped keep in the test only those items that were well centered and discriminated well between the high and the low scoring students. Table 6.6 is a sample page of the 12-page report of this process:

Table 6.6 Item Analysis – Facility Value and Discrimination Index Report – sample page

Item Level	Item ID	Total number of students	Facility Value	Discrimination Index	Items well centred and discriminating well: Items with FV between 25% and 80% and discrimination index above .30	Items not well centred and / or not discriminating well: Items not with FV between 25% and 80%, and not with discrimination
BENG-90	133 /A1	39	0.84	0.26		✓
BENG-90	118 /A1	38	0.84	0.35	✓	
BENG-90	720 /A3	6	0.83	0.20		✓
BENG-90	143 /A4	6	0.83	0.50	✓	
BENG-90	144 /A1	6	0.83	0.20		✓
BENG-90	144 /A4	6	0.83	-1.00		✓
BENG-90	648 /A3	6	0.83	0.50	✓	
BENG-90	290 /A1	17	0.81	0.43	✓	
BENG-90	718 /A3	16	0.81	0.31	✓	
BENG-90	200 /A1	31	0.80	0.22		✓
BENG-90	319 /A5	23	0.80	0.11		✓
BENG-90	145 /A1	15	0.80	0.16		✓
BENG-90	140 /A1	5	0.80	-0.50		✓
BENG-90	280 /A1	27	0.77	0.17		✓
BENG-90	312 /A4	22	0.77	-0.18		✓
BENG-90	272 /A1	17	0.76	0.26		✓
Total of items with good FV and DI					390	
Total of items with bad FV and DI						694
Total number of test items						1084

Three hundred and ninety items (35.98%) out of 1084 were found well centred and discriminating well and 694 items (64.02%) were found not well centred and not discriminating well. In total, three hundred and ninety items were found to have good facility value as well as good discrimination index. The rest of the test items will be reviewed in order to be improved and included in the test at a later stage.

The NEPTON inter-item consistency: split-half reliability index: There are various methods of measuring reliability: test-retest reliability, parallel-form reliability, split half reliability index, KR 20 formula and KR 21 formula. The first two are considered impractical, time consuming and unsatisfactory. Although the split half method is itself not totally satisfactory since the strength of the correlation will depend on which items are chosen to go into each of the two halves, it is reported as the easiest of the last three, although KR20 and KR21 are considered better). Moreover, the split half method

generally produces similar results like them (Alderson et al. 1995). The NEPTON test reliability was estimated measuring the *inter-item consistency*. We simulated the parallel forms method by calculating the *split-half reliability* index. This involved dividing a test into two, using the odd-even method for splitting the items, treating these two halves as being parallel versions, and correlating these two halves. A perfectly reliable test would have a reliability index of +1.0. From what the table below indicates, the two halves of the NEPTON test correlated strongly, thus suggesting a high NEPTON test reliability.

Table 6.7: NEPTON Correlations

		Odd number	Even number
Odd number	Pearson Correlation	1	.822(**)
	N	866	866
Even number	Pearson Correlation	.822(**)	1
	N	866	866

** Correlation is significant at the 0.01 level (2-tailed).

Other factors taken into consideration:

- (a) The test was to place test-takers at various language levels and that the construction process of the item bank was based on the needs analysis, on the COBLE review and on the Test Specification. We therefore had a general expectation that the cut-off points would be around 4-5 and indeed they were.
- (b) We also had the pen-and-paper cut-off points as a starting point.
- (c) We studied the gross proportion of statistics which indicated from previous years the proportional percentage of students allocated to each level as there was no indication there were differences in this year's intake
- (d) We took into consideration the marking of the written component by trained and certified markers. We compared the electronic placement with the written component placement results to find out how they correlate. The figure below indicates how they compare:

Comparison of Electronic and Written component placement results

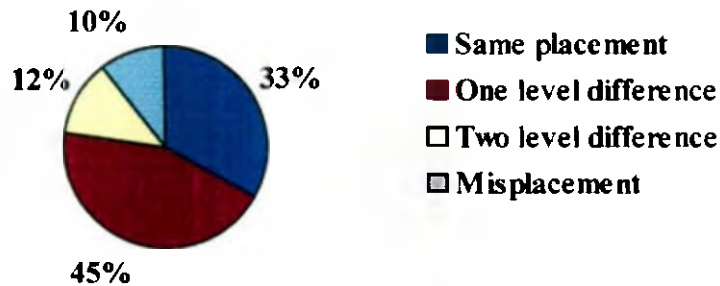


Figure 6.4. Comparison of electronic and written components of NEPTON results

Thirty-three percent of both results were exactly the same. Forty-five percent had one level difference, but the test being pro-student, the test-taker was placed in the higher of the two levels. This total of 57% indicates a good correlation between the two components of the test, although, higher correlation of the same placement would have been more satisfactory. Twelve percent had two levels difference. In this case test-takers were placed in the middle level by a personalized decision. Ten percent of the test-takers were found to be in some way misplaced by usually one level. This is of a concern and we planned to investigate the possible reasons why we think it happened, for example improvement of items through the item analysis and better marking of written task, based on more clearly defined criteria and better trained markers. As part of all this process, we had some indication that the markers needed more training in marking, to be more consistent in the use of the criteria given to them, and perhaps give them clearer criteria, which had to be developed.

6.1.3 General and individual factors influencing Reliability

The test-takers need to be familiar with the test format before taking it: being familiar with the test format makes test-takers feel more comfortable in taking it. Individual factors such as transient factors (e.g., the physical and psychological health of the test-

taker) and stable factors (e.g., their experience with similar tests) may influence test-takers' performance. For these reasons, test-takers have been given a number of opportunities to become sufficiently familiar with the test format before taking the test. They were given the opportunity to familiarize themselves with the test format, and types of activities in the form of a booklet given to them upon registration for the NEPTON test and in the form of an electronic tutorial by the invigilators, just before taking the test. This gave hands on experience to the test-takers of the electronic form of the test before taking the test. This was a type of orientation to the test in an effort to sufficiently prepare the test-takers to take the test and reduce the tension of taking it as much as possible. Test-takers were asked how they felt before and during the test (see appendices 17 and 18). This is what they answered:

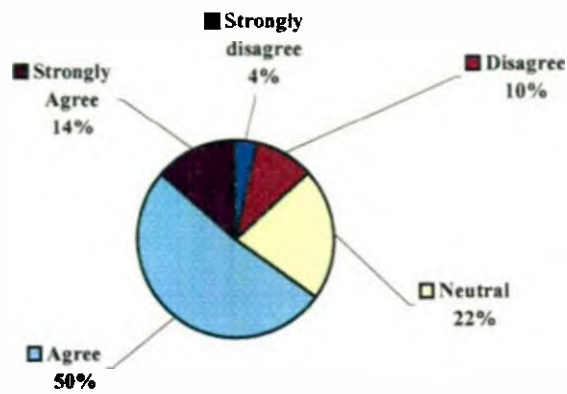
Table 6.8: NEPTON test-taker post- electronic test questionnaire – Feelings before and during the NEPTON test. (see appendix 18)

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 Before the electronic test I felt tensed.	13.3%	13.3%	24.8%	24.8%	12.4%
2 I felt comfortable during the electronic test.	7.1%	8.8%	23.9%	27.4%	29.24%

Based on the data above, the test-takers tension rate before the test averages between 24.8% neutral and 28.8% agreeing and 12.4% strongly agreeing. On the other hand, they expressed they felt quite comfortably during the test, ranging from 27.4 agreeing and 29.2% strongly agreeing.

Test-takers were also asked whether the tutorial helped them.

The NEPTON tutorial helped me



Although I had some previous computer experience, the Tutorial helped me

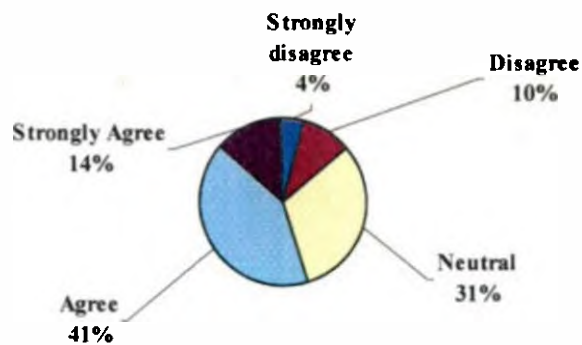


Figure 6.5. The NEPTON tutorial and computer familiarity

The student NEPTON post-test data indicates that 14% strongly agree and 50% agree that the tutorial was helpful to them. Even the majority of those who had some previous computer experience strongly agreed (14%) or agreed (41%). We should point out here that, although there was a considerable number of test-takers who were neutral, disagreed or strongly disagreed about this, only three out of more than eight hundred opted to take the test in its pen-and-paper format. This suggests that the majority of test-takers, preferred to take the placement test electronically.

6.1.4. Situational factors influencing reliability

The conditions of the test administration were uniform and non-distractive: test-takers were sent to specific computer laboratories; invigilators helped students through the tutorial. English language test experts and lab assistants were available at all times. There were no major distractions for most test-takers, apart from one occasion where we had system failure due to excessive demand of the system. This was quickly dealt with.

6.2. Test validity

Another aspect we needed to look at was the test validity, in other words the extent to which the objective component of the NEPTON test was sufficiently valid for its purpose and the reactions and feelings of test-takers and faculty towards it.

The sample used in the test trial was of adequate size for the test validation. There were more than 800 test-takers. About 260 answered the pre-test questionnaire (see appendix 16) and about 131 answered the post-test questionnaire (see appendix 18). They are representative of the population for which the test is intended in age, experience and background. The language levels of the test provide an adequate basis for validating the instrument. The large size of the item pool (1084 items) also secured higher test validity.

6.2.1 Internal, face validity:

Firstly, we established the extent the test-takers were familiar with computers and found out about their attitudes and feelings towards the computer delivery of the NEPTON. We wanted to find out whether the eventual English language placement would be significantly affected by the test-takers' familiarity with computers.

6.2.1.1. Test-taker choice of the test mode of delivery

Test-takers had a choice of taking the test electronically or in pen-and-paper format. Out of more than 800 students, only 3 opted the NEPTON (P) option. Only five students said they wanted to do the NEPTON (P) but when they did the NEPTON tutorial they changed their mind and did NEPTON instead.

6.2.1.2. Pre and post NEPTON TEST test-taker questionnaires (see appendices 17 and 18)

Before taking the NEPTON, data was collected from each test-taker on computer familiarity. The data were analysed to establish test-takers' prior computer-familiarity, and attitudes toward taking the test electronically in comparison with the pen-and-paper test. This involved asking test-takers whether they felt comfortable with basic computer skills such as the use of mouse, clicking, scrolling and the use of technology in general. It also involved asking them about further use of technology such as computer use in general, use of the Internet, the email and chat.

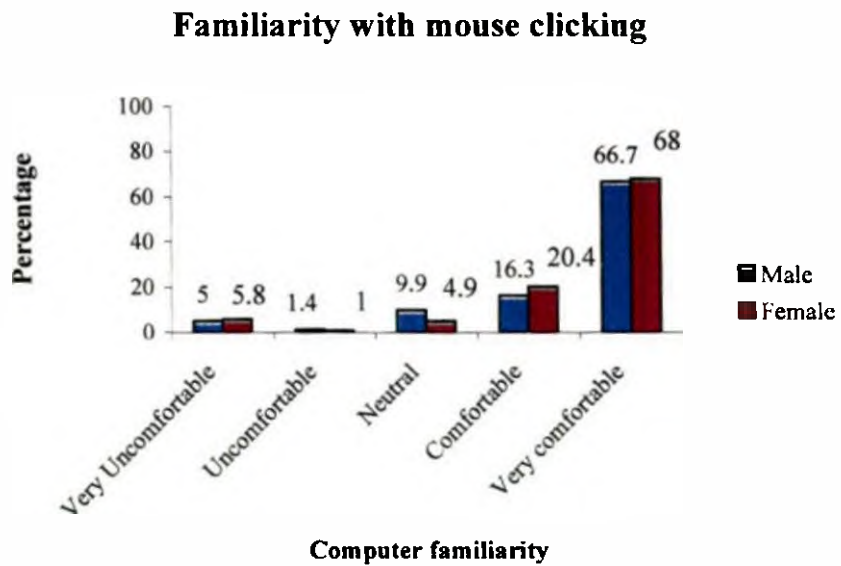
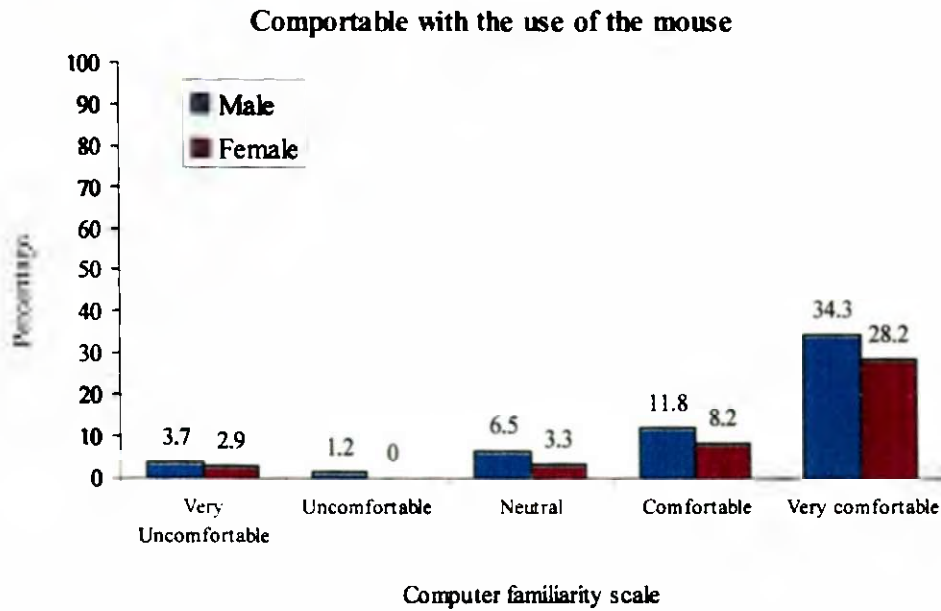


Figure 6.6. Test-taker familiarity with technology and basic computer skills: mouse and clicking

A considerable percentage of male (34.3%) and female (28.2%) test-takers felt very comfortable and comfortable (11.8% male and 8.2% female) using the mouse. It is

interesting to note, however, that there was a percentage of 3.7% male and 2.9% female who felt very uncomfortable and 1.2% male who felt uncomfortable using the mouse. Sixty-six point seven percent of male and 68% of female test-takers felt very comfortable with mouse-clicking. Similarly, however, 5% males and 5.8% females felt very uncomfortable and 1.4% males and 1% females uncomfortable with mouse-clicking.

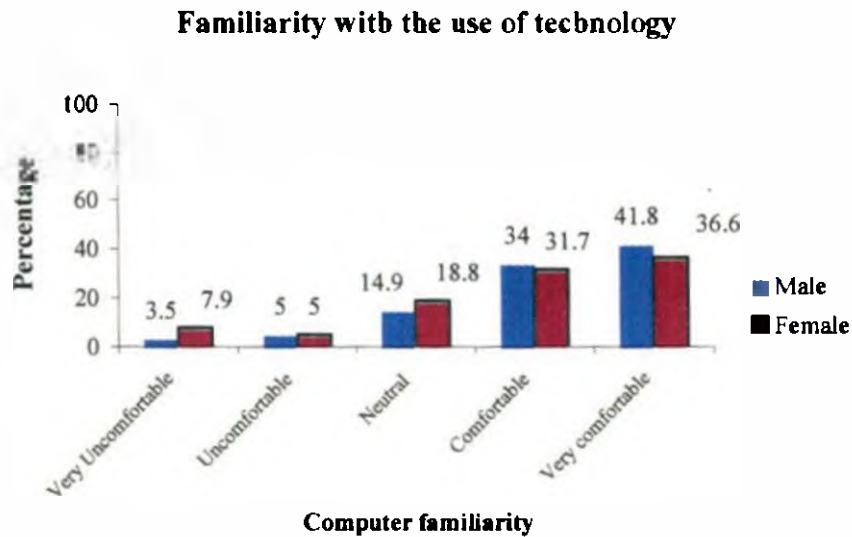
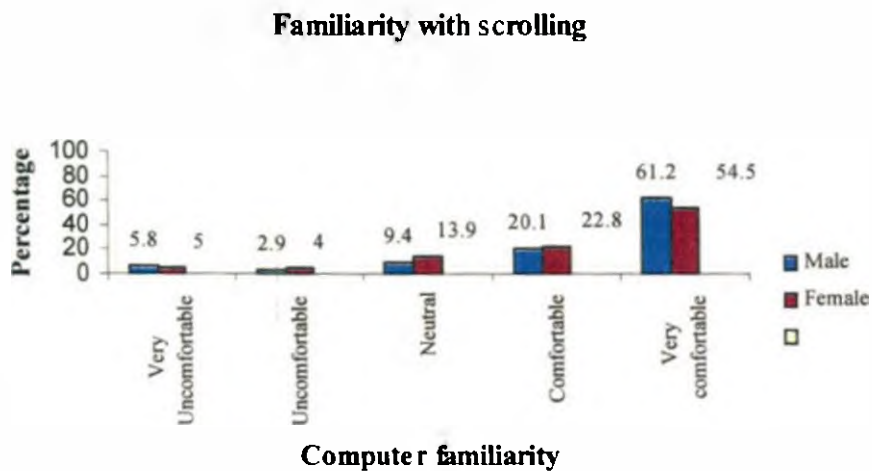


Figure 6.7. Test-taker familiarity with technology and basic computer skills: scrolling and use of technology

Sixty-one point two percent male and 54.5% female test-takers felt very comfortable with scrolling. On the other hand, there is a 5.8% male and 5% female very

uncomfortable and 2.9% male and 4% female uncomfortable with scrolling. Forty-one point eight percent male and 36.6% female test-takers felt very comfortable and 34% male and 31.7% female comfortable with the use of technology, while 3.5% male and 7.9% female felt very uncomfortable and 5% male and 5% female felt uncomfortable.

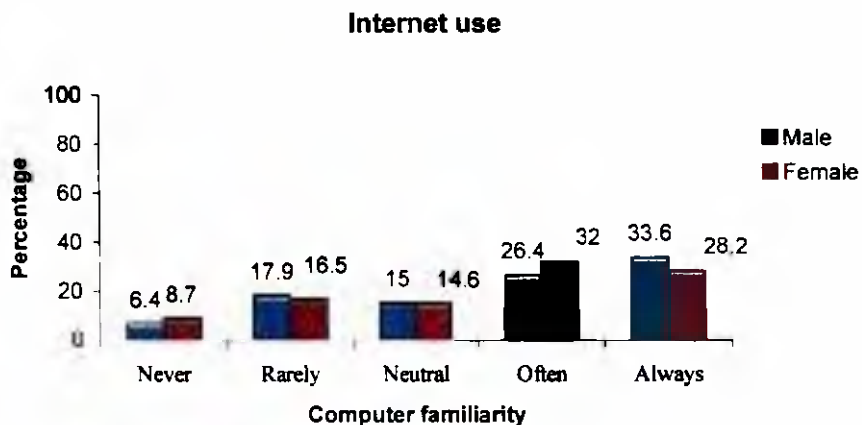
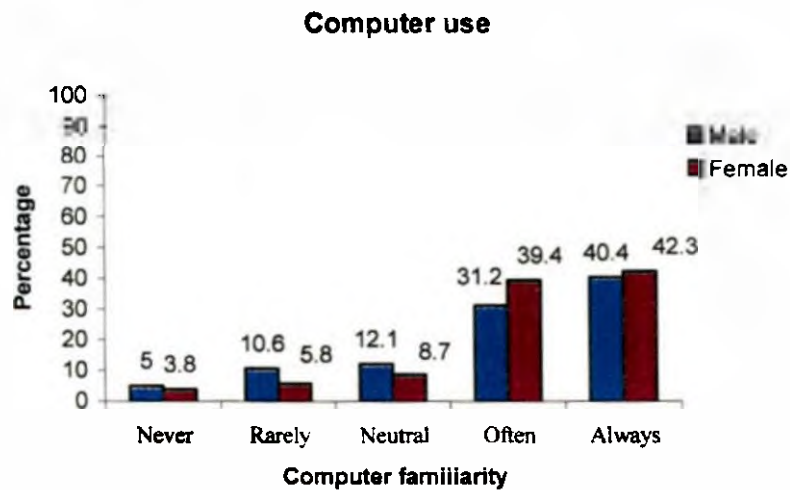


Figure 6.8. Test-taker familiarity with computer and the Internet use

The majority of test-takers always (40.4% male and 42.3% female) or often (31.2% male and 39.4% female) used the computer while there was a considerable percentage of them who never (5% male and 3.8% female) or rarely (10.6% male and 5.8% female) used the computer. Many test-takers always (33.6% male and 28.2% female) or often (26.4%

male and 32% female) used the Internet. However, there was a considerable percentage of test-takers who never (6.4% male and 8.7% female) or rarely (17.9% male and 16.5% female) used it.

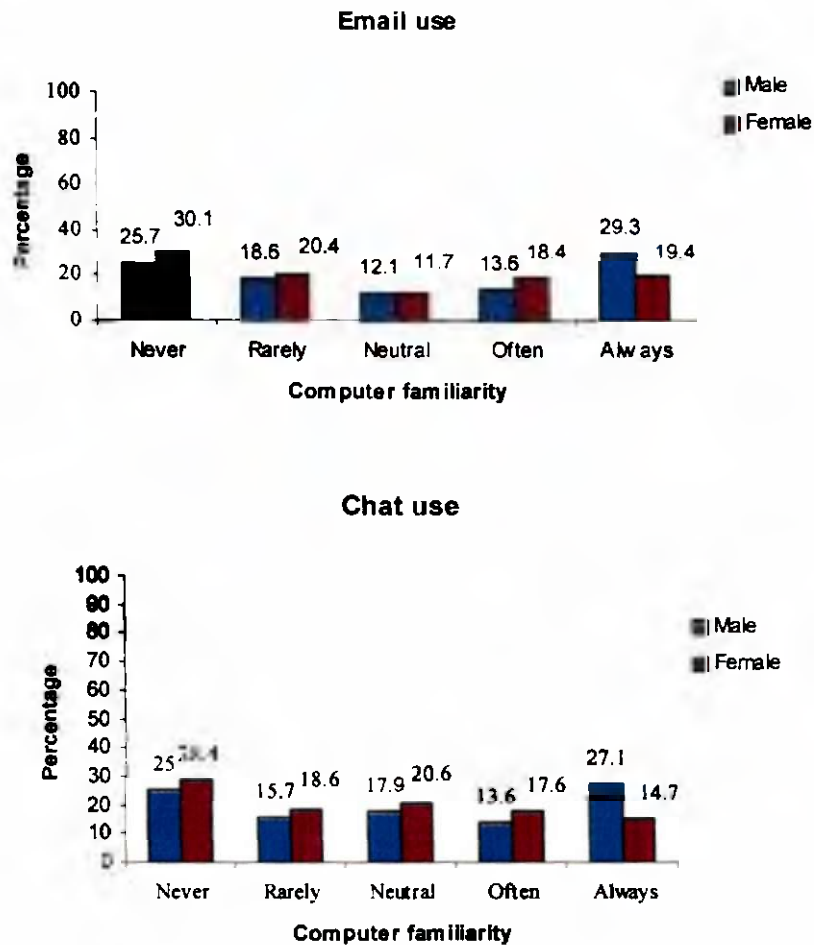


Figure 6.9. Test-taker familiarity with the use of email and chat

Quite a few test-takers always (29.3% male and 19.4% female) or often (13.6% male and 18.4% female) used the email, whereas there was even more percentage of male (25.7%) and female (30.1) test-takers who never or rarely (18.6% male and 20.4% female) did so. Fewer test-takers always used chatting (27.1% male and 14.7% female) a percentage used it often (13.6% male and 17.6% female often) and most did not (25% male and 28.4% female never and 15.7% male and 18.6% female rarely). From the above data, it was evident that many test-takers felt adequately comfortable in the use of the

basic computer skills (mouse use, clicking and scrolling) needed to take the NEPTON test and used various tools on a regular basis, however there was a considerable number who didn't seem to be.

On the whole, the test-takers did not seem to feel disadvantaged with the use of technology in the delivery of the NEPTON test. On the contrary, as we have seen, the majority preferred to take the electronic test version. Another indication was their comments when the electronic test had finished and they were given the hand-written task to complete. They expressed their preference towards the electronic component of the test rather than the handwritten task. However, attention needs to be given to those test-takers who feel uncomfortable with the basic computer use they need to take the NEPTON test. They need to be given the opportunity to practice before the test. This is already done as part of the tutorial. At a later stage, it is also planned to be part of the non-secure component of the NEPTON test, where prospective test-takers can practice taking the test online in their own time even weeks or months before the test.

6.2.1.3. Post NEPTON Test test-taker questionnaire (see appendix 18)

After exiting from the objective NEPTON, the test-takers completed a questionnaire. The items covered the following topics:

Table 6.9: NEPTON post-test questionnaire – Skill sections

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I found the structure/vocabulary (sentence-based) section manageable.	1.8%	10.6%	21.2%	44.2%	21.2%
I found the structure/vocabulary (text and dropdown menu selection) section manageable.	1.8%	3.5%	27.4%	44.2%	20.4%
I found the reading comprehension sign section manageable.	1.8%	7.1%	23.0%	43.4%	20.4%
I found the reading comprehension (text and multiple-choice) section manageable	0.9%	8.0%	19.5%	49.6%	17.7%

Based on the data above, most test-takers found the different types of questions covering the different skills manageable. The majority also found the instructions clear (43% agree and 46% strongly agree). If a comparison was also made between similar data taken in the preliminary pen-and-paper format of the field-testing of the test (see table 5.7) it is suggested that the test-takers felt there was improvement in the final electronic form of the test.

Table 6.10: NEPTON post-test questionnaire – Topics and variety of activities

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 I found the topics of the electronic part of the test interesting.	2.7%	8.0%	22.1%	38.9%	23.0%
2 There was enough variety of activities.	1.8%	9.7%	23.9%	46.0%	15.9%

The test-takers also found the NEPTON topics of the electronic part of the test interesting (38% agree and 23% strongly agree), while 22.1% were neutral, 8.0% disagreed and 2.7% strongly disagreed. The test takers also found that there was enough variety of activities (46% agree and 15.9% strongly agree), however, 23.9% were neutral, 9.7% disagreed and 1.8% strongly disagreed. Again, if similar comparison was made between table 5.8 (field-testing in preliminary pen-and-paper format) and table 6.10 (trial of electronic form of test), it is clear that there is improvement in test-takers' perceptions of the test.

Table 6.11: NEPTON post-test questionnaire – Topics of hand-written component

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 I found the topics of the written part interesting.	6.2%	7.1%	22.1%	36.3%	25.7%

A great number of test-takers found the topics of the written part very interesting (36.3% agreeing and 25.7% strongly agreeing), whereas as a 22.1% remained indecisive. More investigation is needed as to what the test-takers would prefer.

Table 6.12: NEPTON post-test questionnaire – test length: electronic and hand written

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1 The length of the electronic test was appropriate.	1.8%	7.1%	26.5%	38.1%	24.8%
2 The length of the written test was appropriate.	1.8%	10.6%	27.4%	35.4%	21.2%

Test-takers seem to be happy with the length of both the electronic (26.5% neutral 38.1% agreeing and 24.8% strongly agreeing) and the written component (27.4% neutral, 35.4% agreeing and 21.2 strongly agreeing) of the test. Again, the proportion of neutral responses is substantial and may need consideration.

Table 6.13: NEPTON post-test questionnaire – time spent on the test

	Very long	Long	Neutral	Little	Very little
1 How much time did you spend on this test as a whole (electronic and written)?	2.7%	29.2%	47.8%	13.3%	4.4%
2 The amount of time for this test was	13.3%	21.2%	46.9%	8.8%	1.8%
3 The test items were	13.3%	12.4%	53.1%	6.2%	1.8%

The majority of the test-takers found neither the test (46.9% neutral) nor the test items (53.1% neutral) long. Forty-seven point eight percent of the test-takers felt that the time they spent on the test as a whole was appropriate while 29.2% felt it was long and only 2.7% felt it was too long.

The results above as a whole indicate a general acceptance and a feeling of comfort in regards to the NEPTON test. The test-takers attitudes and reactions indicate a general acceptability of items and test components.

6.2.2. Internal, content validity

At Intercollege there are 24 full-time and 23 part-time practicing English lecturers across all three campuses. Six of them (two from each campus) with expertise in the area were asked to examine NEPTON's content validity by doing the following:

- (a) Study the NEPTON Test Specifications
- (b) Study two sample Pen-and-Paper tests
- (c) Study NEPTON
- (d) Compare all three tests (The NEPTON and two sample pen-and-paper tests) with the Test Specification by rating the test on a questionnaire (see appendix 19) according to the degree to which it met certain criteria

6.2.2.1. Experts' biographical data (see appendix 19)

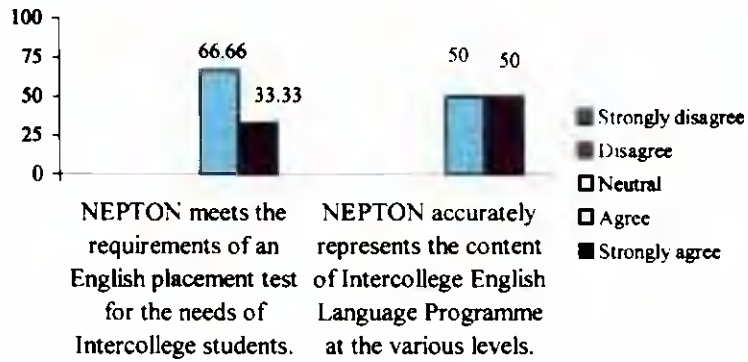
The following is a summary of the experts' biographical data: Four experts were female and two male. Most of them were Greek-speaking with English as their second language. Four came from Cyprus; one came from England and another one from Greece. Most of them had years of experience in teaching the various levels, except in two occasions where one had not taught BENG-50 and another who had not taught ENGL-101.

Before evaluating the NEPTON, data was also collected to establish experts' prior computer-familiarity, and attitudes toward taking the test electronically. This recorded the following: All experts demonstrated high familiarity with computers. All of them said they were very comfortable with basic computer skill (the use of the mouse, with clicking, scrolling and the use of computer in general.) All experts routinely used the Internet and the email and only one never used chat.

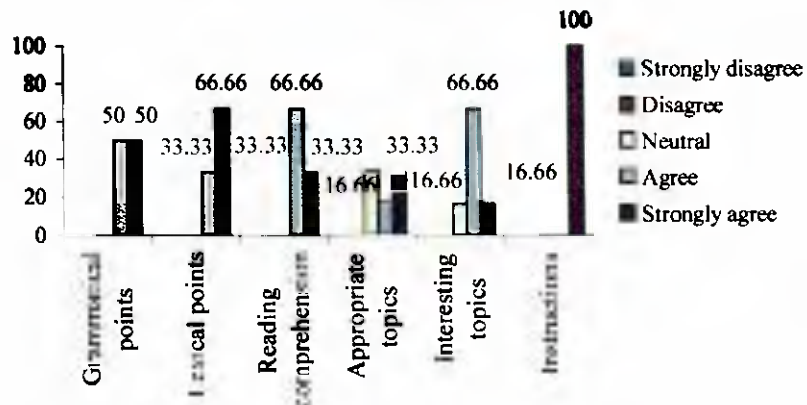
6.2.2.2. Experts' Post NEPTON / NEPTON (P) and Test Specification (see appendix 19)

NEPTON in general

Experts NEPTON content validity evaluation



Experts NEPTON content validity evaluation



Experts NEPTON content validity evaluation

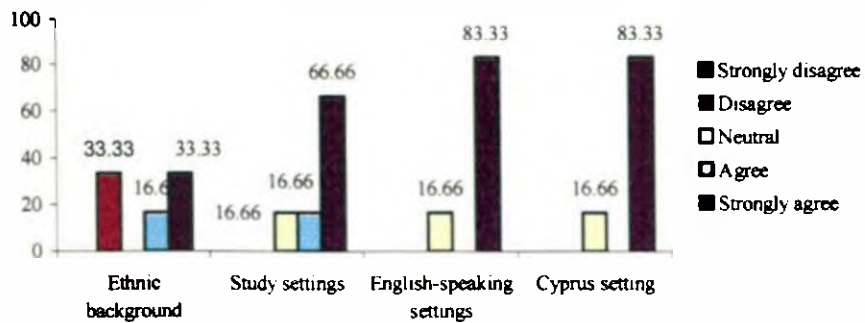


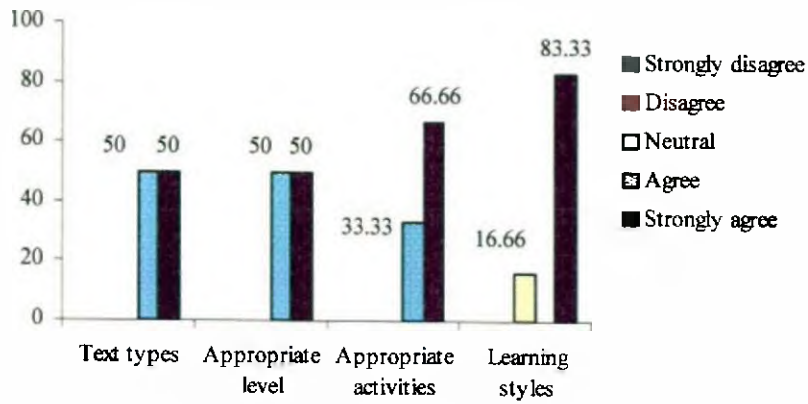
Figure 6.10. Experts Post NEPTON / NEPTON (P) and Test Specification (a)

The data in figure 6.10 (illustration 1) show that 33.33% of the experts strongly agreed and 66.66% agreed that the NEPTON test met the requirements of an English placement test for the needs of Intercollege students. Fifty percent strongly agreed and the other 50% agreed that the NEPTON accurately represented the content of Intercollege English language programme at the various levels.

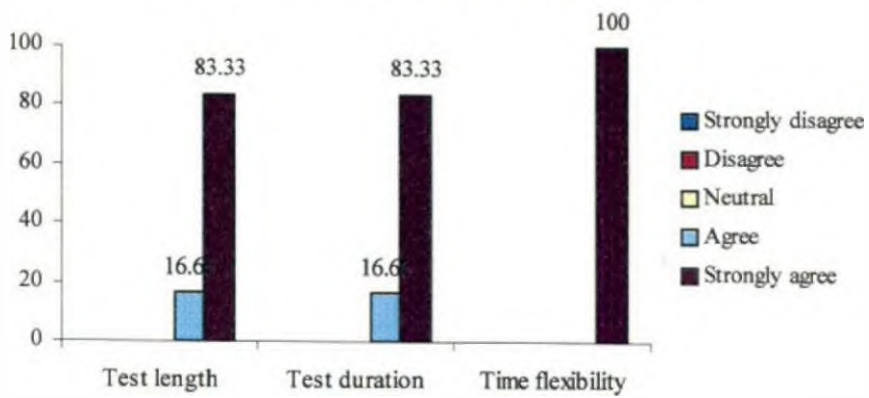
According to the data in figure 6.10 (illustration 2), fifty percent strongly agreed and 50% agreed that the grammatical points represented the structure content of the respective levels. Sixty six point sixty six percent strongly agreed and 33.33% agreed that the lexical points represented the structure content of the respective levels. Sixty six point sixty six percent strongly agreed and 33.33% agreed that the reading comprehension sections represented the structure content of the respective levels. Sixteen point sixty-six percent strongly agreed and 33.33% agreed that the topics were appropriate for the respective levels, while another 16.66% were neutral. The majority of experts (66.66%) agreed that the topics were interesting, whereas 16.66% respectively strongly agreed or were neutral. All experts strongly agreed that the instructions were clear.

In figure 6.10 (illustration 3), the experts' opinions varied in the following areas: 33.33% strongly agreed that the test reflected the test-takers ethnic background, 16.66% agreed and 33.33% disagreed. Sixty six point sixty six percent strongly agreed that the test reflected settings test-takers would come across in their studies, whereas 16.66% agreed and another 16.66% were neutral. Eighty-three point thirty-three percent strongly agreed that the test reflected settings and contexts in English –speaking settings. Another 16.66% felt neutral. Finally, 83.33% strongly agreed that the test reflected the setting where test-takers would be studying whereas the rest of them (16.66%) felt neutral about this issue.

Experts NEPTON content validity evaluation



Experts NEPTON content validity evaluation



Experts NEPTON content validity evaluation

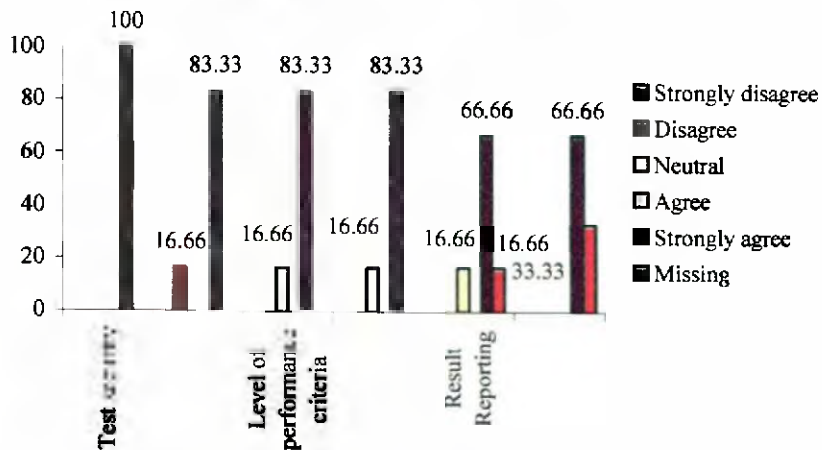


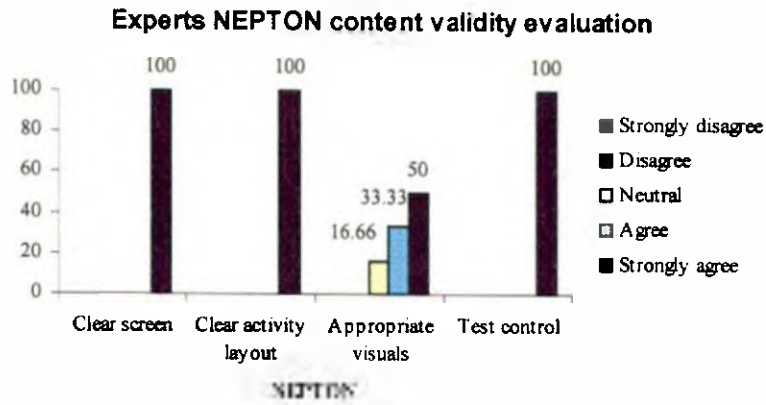
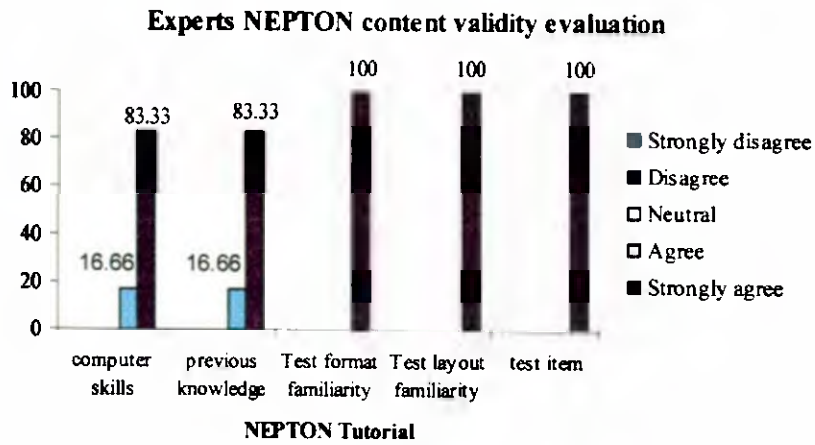
Figure 6.11. Experts Post NEPTON / NEPTON (P) and Test Specification (b)

According to the data in figure 6.11 (illustration 1) Fifty percent of experts strongly agreed and another 50% agreed that there was sufficient variety in text types, and that the NEPTON test items generally targeted the appropriate level. Sixty six point sixty six percent strongly agreed and 33.33% agreed that the kinds of activities were appropriate. Eighty-three point thirty-three percent strongly agreed that the activity types covered variety of learning styles. On the other hand, 16.66% felt neutral about.

From the data in figure 6.11 (illustration 2), it is evident that eighty-three point thirty-three percent of the experts strongly agreed and 16.66% agreed that the time given to the test-takers to take the test was appropriate and all of them strongly agreed that there was time flexibility to cater for test-takers' difference in time needs.

In figure 6.11 (illustration 3), all the experts strongly agreed that the NEPTON secured test and item security and item disclosure. Eighty-three point thirty-three percent strongly agreed that item selection criteria and the levels of performance criteria were appropriate, and that the scoring system was satisfactory. Sixteen point sixty-six percent disagreed on the item selection criteria and 16.66% felt neutral about the rest. Finally, 66.66% strongly believed that the test results were appropriately reported to the test-takers and that the NEPTON test booklet helped test-takers familiarize themselves with the test, the format, the test layout and the test items. Thirty-three point thirty-three percent did not give any answer about the test result report and 66.66% did not give any answer about the familiarity booklet.

NEPTON



NEPTON (P)

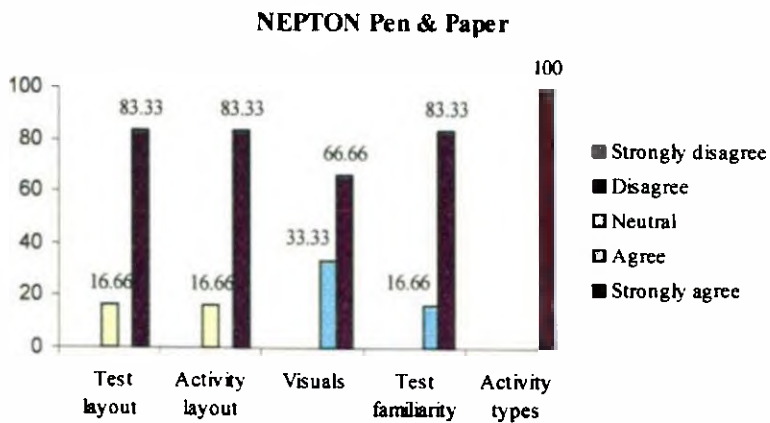


Figure 6.12. Experts Post NEPTON / NEPTON (P) and Test Specification (c)

NEPTON

The experts were then asked to evaluate specific aspects relevant to the NEPTON (electronic component) only. According to the figure 6.12 (graphic 1), eighty-three point thirty-three percent strongly agreed and 16.66% agreed that the tutorial helped test-takers familiarize themselves with the computer skills they needed to take the NEPTON test, even those with some previous experience with computers. All experts strongly agreed that the tutorial helped the test-takers familiarize themselves with the NEPTON test format, layout and test items.

According to the figure 6.12 (graphic 2), Fifty percent of the experts strongly agreed that the visuals were appropriate, 33.33% of them agreed and 16.66% were neutral. All of them strongly agreed that the computer screen and the activity layout were clear and that test-takers were given opportunities to control the test, for example, review items, go backwards and forwards.)

NEPTON (P)

The experts were then asked to evaluate specific aspects of the NEPTON (P). Figure 6.12 (graphic 3) indicates that sixty-six point sixty-six percent strongly agreed and 33.33% agreed that the visuals were appropriate. Eighty-three point thirty-three percent strongly agreed that the pen-and-paper test and activity layout was clear, while 16.66% felt neutral. They also responded the same about satisfactory provision of test familiarity (test format and types of activities). All of them strongly agreed that the test-takers were given opportunities to control the test, that is review items, go backwards and forwards)

NEPTON and NEPTON (P)

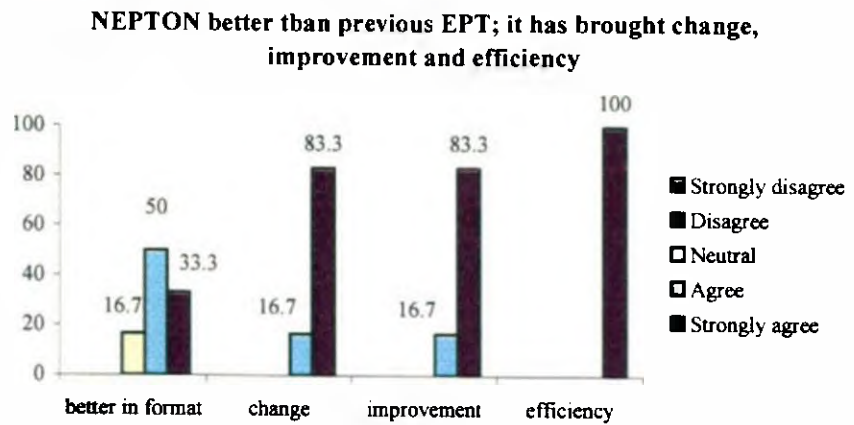
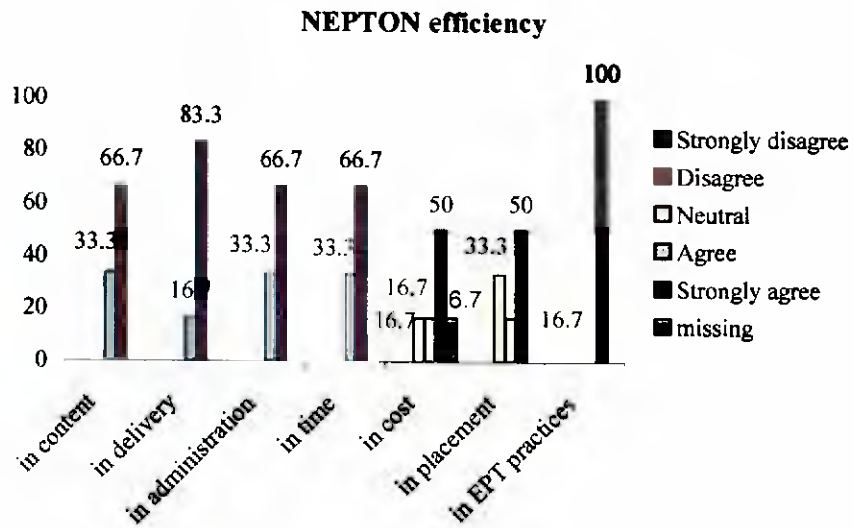
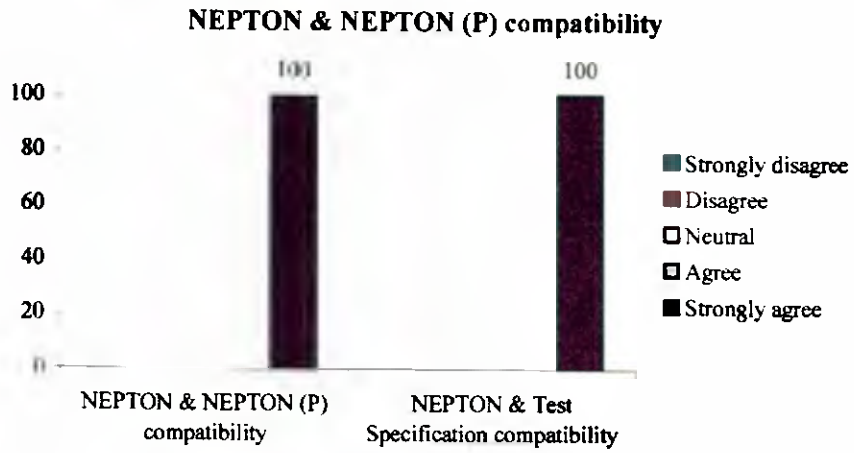


Figure 6.13. Experts Content validity evaluation

NEPTON and NEPTON (P)

The study of figure 6.13 (illustration 1) indicates that all experts strongly agreed that NEPTON and NEPTON (P) were compatible with each other. They also strongly agreed that both tests reflected the NEPTON test specifications. In figure 6.13 (graphic 2), it is evident that sixty-six point sixty-seven strongly agree that the NEPTON is more efficient in content, administration and time. Eighty-three point three percent strongly agree and 16.7% agree that the NEPTON is more efficient in delivery mode. Fifty percent strongly agree that it is more efficient in cost and placement. Sixteen point seven percent agree on both, while 16.7% is neutral about the test's efficiency in cost, 6.7% chose not to answer about the cost and 33.3% were neutral about whether NEPTON was more efficient in placement. All experts strongly agreed that the NEPTON test has brought improvement in EPT practices.

In figure 6.13 (illustration 2), sixteen point seven felt neutral, 50% agreed and 33.33% strongly agreed that NEPTON was better in format than the previous EPT. Finally, according to illustration 3, 83.3% of the experts strongly agreed and 16.7% agreed that NEPTON had brought change, and improvement and all of them agreed that it brought efficiency in Intercollege EPT practices.

Open-ended questions

The experts were asked to also answer some open-ended questions. They did not like the following and felt some aspects needed improvement. Some of these observations came from only one person whereas others came from more than one. Here is a summary of their answers: too many grammatical items tested prepositions; some NEPTON and NEPTON (P) signs and a few passages (particularly the texts testing vocabulary) were culturally biased; some item random selection tested similar grammatical points; occasionally there was ambiguity as to which answer was correct in some items; the experts would have preferred to see some types of written task in electronic form (rather than providing a written task in pen-and-paper format as it is currently done).

The signs are going to be redone by a graphic designer to eliminate any ambiguity or cultural bias. The random selection has already been dealt with. An electronic format of a written component is within the plans we have in the future. Improvement in content is an ongoing process.

The experts were asked to comment on what they liked most about NEPTON and NEPTON (P). Again, some of these observations came from only one person whereas others came from more than one. This is what they said: sufficient variety in text types, interesting topics; overall reflection of the culture of the target language (Britain, Australia, USA); quick, efficient, clear, user friendly, not at all confusing; clarity, variety, security, speed of questions, formats; NEPTON: ability to review previous (already completed) items; less time is needed than the time provided.

Many of the aims of the research project seemed to satisfy the experts, aspects such as variety in text types, interesting topics, clarity, efficiency, user friendliness, security, format, review ability, and time.

Experts also made some suggestions as to how NEPTON and NEPTON (P) could be improved. These are the points they made: more reading skills questions – possibly writing skills as well (structure and vocabulary are adequate); for vocabulary, concentrate on “universal” words instead of British/American/Australian specific (this was mentioned only by one person, in contrast with the others who commented favourably about the overall reflection of the culture of English in different English-speaking countries.); better visuals of signs – bigger maybe; more exercises on tenses.

The experts also made some further comments: “We appreciate the huge effort put into this project”; “It’s a great alternative to the existing placement test system”; “Excellent work done”.

Conclusions from the experts test evaluation

The experts' evaluation (from their answers in both the questionnaire and the open-ended questions) indicates that in general the NEPTON test offers improvement, change and efficiency in the English placement practices at Intercollege. It is also inferred that it reflects the test specifications, and current theories and practices in L2 testing. Their suggestions for improvement positively contribute to its further improvement.

6.2.3. External, construct validity

The construct validity of a test is "the extent to which the test may be said to measure a theoretical construct or trait." (Anastasi 1982, p.144) In the case of the NEPTON the agreement of experts in the field, regarding the content validity of the test also lent credence to the inferred degree of construct validity.

English instructors at Intercollege were also asked to give their feedback about the NEPTON. They were happy to use it and replace the old EPT, however, they wanted to keep the written component, even though this is not offered electronically for the time being.

6.2.4. Conclusion

The NEPTON test overall performance suggested an acceptable reliability of the items at each level. The English language level and class distribution was acceptable for both educational and administrative purposes. Many factors, which could influence reliability such as general, individual, situational, item analysis, inter-item consistency, and correlation with the hand-written component were analysed and they also suggested acceptable reliability. The test was also evaluated for its validity: internal (face and content) and external (construct). It was found to be satisfactorily valid. However, it is in the intentions of this project to include in its continuous process of change and improvement, a continuous process of reliability and validity testing.

Chapter seven

Conclusions and Recommendations

7. Introduction

In this chapter, I will evaluate the project findings and report the main conclusions and recommendations. In order to measure whether the objectives of the project have been met, we have to examine its aims, as set out in the second chapter. The primary objectives of this research project were to:

- (a) Explore a case study through a combination of work based learning and applied linguistics research theory in order to establish the needs, and provide change and improvement in English placement testing practices and to
- (b) Produce evidence of such research.

7.1 Work based learning and applied linguistics research theory combined

Middlesex University work based programme provided me with the opportunity to learn, as a lifelong learner, through equal and open access to high quality learning opportunities, in line with the European Council “Lisbon strategy” (2000). Through the work based learning, required by the nature of my programme of study (Doctorate of Professional Studies) which is a modern way of creating university-level learning in the workplace, I was given the opportunity (a) for academic study grounded within a work context and (b) for enhancing the effectiveness of the Intercollege English placement testing programme. I had the opportunity to be researcher of my practice (Usher 1996). As an insider researcher, I learned to manage work and learning, and manage the influence, reinforcement and complementing of each other. I extended my knowledge in research, and at the same time I was able to develop an English placement test online, a product, which was greatly needed by Intercollege, my workplace. Initially, through work based learning, I was able to work out the problems of my case study, the existing English placement-testing programme. Then I combined the work based learning philosophy with that of applied linguistics research discipline in the area of L2 testing online and developed the *New English Placement Test Online* (NEPTON) to meet the current needs of the college and its students. I studied the theories and practices in L2 testing online and I established the needs of Intercollege in English language placement

testing. As a result, I gained knowledge, skills and abilities in: reviewing the Intercollege English language programme and developing L2 level descriptors; establishing a test design tailored to the local and unique Intercollege needs; developing the test specifications; creating an item bank; moderating the test items; field-testing them in pen-and-paper form; going through item analysis; working together with the IT expert in designing the in-house developed testing electronic environment; trialling the test electronically and then analysing the data to see the effectiveness of the testing programme, its validity and reliability. I learned to establish the NEPTON test reliability using different ways: conducting item analysis and establishing the facility value and the discrimination index of each item; calculating its split-half reliability index; creating a large item bank, going through a moderation process, developing discrete items, using objective scoring, developing test specifications, designing clear computer screen layout, ensuring uniform and non distractive situational factors; comparing the NEPTON electronic results with those of the hand written writing task, and analysing test-taker computer familiarity. I also learned how to establish the NEPTON test validity: I collected and analysed data in order to establish the test's internal face and content validity, and its external construct validity. I gained knowledge, skills and abilities in working out an algorithm for the item presentation and randomization and an algorithm for the cut-off points, and in doing different types of statistical analysis. I worked out a way to develop a computer-based test incorporating smart elements of a computer adaptive test (CAT), and in some instances made it work better than a CAT. For example, in the NEPTON test, the test-takers can skip around or go back to change an answer. And go back and review their answers at the end of the test if they wish to do so, whereas in a CAT test they cannot. As a result of that, I also had to find an alternative way of analyzing the items of such a test. At the same time, I also gained knowledge, skills and abilities in how to lead such a high stake project, work with more than 2200 participants (such as test-takers, faculty, experts, information technology personnel, and administrators), deal with unexpected complications of the project, find solutions to problems, make decisions, take responsibilities, negotiate, present and discuss the project progress and results. The investigation of the case study of changing and improving the Intercollege English placement testing practices as an

insider researcher through the research tradition of applied linguistics generated the following new knowledge: a method of investigating and developing a L2 testing programme; a paradigm of randomised item selection for a computer based test; a paradigm for arriving at cut-off points; an in-house, tailor-made testing electronic environment, and finally the NEPTON test itself, developed for specific needs and tailored to a specific setting and student profile. The reflective aspect of work based learning helped me develop a more systematic awareness of research processes, achievements, and learning.

Moreover, the NEPTON research project developed more local expertise, which could continue to contribute to future locally-based research. Many of my colleagues extended their skills, knowledge and abilities through this project: the Information Technology expert who is in charge of the Information Technology Centre at Intercollege, and everything related to technology, learned how to develop an electronic environment for English Language online testing purposes, based on the specific test requirements and design I developed. My assistant, with extensive knowledge in teaching and testing English and some knowledge in statistics, extended this knowledge by contributing greatly in the formation of the algorithms through discussion and by conducting the iterations with another colleague. He also improved his skills and abilities by helping in the organization and the running of the field test of the pen-and-paper format and then the electronic trialling of the test in the three campuses. All English language lecturers invigilated the field test and thus had an opportunity to have a first-hand experience of what it is to participate in such important projects. More than forty percent of the English language lecturers participated in the test moderation process and developed skills, knowledge and abilities related to such processes. All of them also participated in the English language programme review. This helped them reflect on their practices by acknowledging improvements needed in both the programme content and their practices and contribute to improving the English language programme at Intercollege by participating in the review process and offering suggestions. It also gave them an opportunity as a whole group to improve their knowledge in current theories in L2 teaching, and more specifically in language level descriptors. Other colleagues

developed skills, knowledge and abilities in editing and checking the test against the test specifications. The colleagues involved in the marking of the written task had the opportunity to refresh their knowledge in writing task criteria and put that knowledge into practice. Administrators had the opportunity to improve in their area, by learning how to administer such a test electronically and not manually. They went through a learning process but in the end they welcomed the change because it made their job easier and more efficient. Students also developed skills and confidence in using technology for their English language placement. This locally-based research project reinforced and contributed to the local research in L2 testing online. It offered a case for the use of alternative ways of learning, and recognizing different settings of learning other than the traditional academic learning. In the long term, the continuous improvement and change of the NEPTON project, based on new and changing needs and theories and practices, could continuously contribute to lifelong learning, not only for me, as the insider researcher, but also for my colleagues and other people concerned with English online testing in general in Cyprus and elsewhere. The project has already been presented at seminars in Cyprus and conferences overseas and has received very positive feedback.

7.2 Research project evidence

The major purpose of this research project was to research the existing English placement practices at Intercollege with the purpose of bringing change and improvement. As a result of this research, a New English Placement Test ONLINE (NEPTON) was planned, developed, trialled and evaluated. The questions raised therefore are: has the NEPTON test brought change? Has it brought improvement? Is the NEPTON test more efficient in content, placement, delivery, administration, time and cost? Is the NEPTON test reliable? Is the NEPTON test valid? Is it implemented and accepted?

The previous Intercollege English Placement Testing programme lacked reliability and validity. It had not been systematically developed. It was developed in an *ad hoc* way,

vaguely depending on needs at different times. Although it worked fairly well, there was a general feeling by many stakeholders concerned that it did not accurately place students, it was time consuming, labour intensive and possible wrong placement cost students money and time. It was never analysed to establish its effectiveness.

The NEPTON research project addressed all those issues. It tried to cater for the specific placement needs of a business / social sciences / professional studies oriented college, fit and reflect the specific English language programme courses, descriptors and six language levels. It was developed as an inhouse tailor-made test so that it would directly meet the specific placement test needs of Intercollege; this also meant that Intercollege would not have to pay for royalties on a continuous basis to external agencies and would have the flexibility to improve and change according to future needs. It was based on sound theory and practice incorporating both work based learning and applied linguistics research discipline. It is a *New English Placement Test Online* (NEPTON) which is now used by Intercollege as its new English placement test.

7.3 Conclusions

The NEPTON research project has definitely brought change and improvement in both theory and practice: in theory, it offered knowledge of a case study based on combined work based learning and applied linguistics research theory – both applied to English online testing. In practice, it provided change and improvement in the existing Intercollege English placement practices.

On the theoretical side, this research project gave valuable insights into a case study based on a combination of work based learning and applied linguistics (L2 online testing) research. It provided opportunities to undertake research at work, combining research with work requirements, undertake a project needed for the work place and endorsed by the employers, contribute to a work need, solve a work problem in a scientific way, practice balancing work and research requirements and constraints, managing work needs and research work, leading a research project, cooperating and communicating with faculty, staff, students and other experts, managing worker /

researcher role and subjectivity derived from that. It provided an opportunity to blend theory and practice, combine research with a practical problem and deal with a case study. As a result, it provided a model for the development and the implementation of an English Placement Test Online at tertiary college level based on work based learning and applied linguistics research approach.

On the practical side, the results of this research project provide a new English Placement Test programme for Intercollege students: reviewed, informed by current theory and practice in L2 testing, enriched by current modes of delivery, implemented and evaluated. The aim was to improve current English Placement Practices at Intercollege, thus contribute to English language programme change for the specific needs of this institution.

The NEPTON test was based on sound research, and current theories and practices in L2 testing, compared with the existing English placement test. Second, it catered for the particular needs of Intercollege students, something that many generic well-known English placement tests (TOEFL, IELTS, QPT), as well as the existing Intercollege EPT do not. In other words, it was based on and reflected the particular Intercollege case, the English programme and its language level descriptors. Third, it offered flexibility in the mode of delivery. Although it was meant to be delivered online, it was also offered in pen-and-paper (although as it was proven most test-takers preferred the electronic test). In addition to the booklet given to candidates prior to taking the test so that they familiarize themselves with the test format and activity types, the NEPTON test also included an electronic tutorial which provided hands-on familiarity with those aspects of the test as well as some practice with the basic computer skills needed for taking the test. Although it did not distance itself greatly from the types of activities in the existing English placement test (and that was deliberately done so that novelty was smoothly accepted by stakeholders), the NEPTON test provided a wider range of activity types which takes advantage of the interactive nature of online delivery: it included multiple choice, and dropdown menu selections in sentence-based, text-based, and sign related activity types. These covered writing in a discrete way (structure and vocabulary), and

reading comprehension. Apart from the improvement in the delivery mode and activity types, it was also clearer as to what each item assesses, compared to the existing EPT and it consisted of less items in total for each test (54 compared to 64 items). The previous EPT was used in only four different versions, using the same 64 questions in different order. In the NEPTON test, the questions were randomly chosen from a large item bank and each test was unique. This also ensured test security. The content was rich in topics, settings, vocabulary, and authentic or authentic like text types, reflecting the various language levels covered in the English language programme and informed by CEF and ALTE descriptors. Due to the discrete nature of each item, scoring was objective and computed automatically, without the need of manual markers. This made the test more efficient in marking, human resources, time and cost. The process of item writing, moderation, iteration, field-testing and trialling, item analysis and reliability and validity check (based on current theories and practices) ensured more efficient testing and better student placement. Finally, the NEPTON test offered opportunities for continuous test improvement and change offered by technology.

7.4 Reflection

Thinking back, I can note two major contributions of this research project: its contribution to research in general and its contribution to Intercollege, my work organization.

- (a) The NEPTON project contribution to research in general:
 - (i) An alternative slide paradigm for test item randomization
 - (ii) An alternative paradigm for a cut-off point system
 - (iii) An in-house developed online L2 testing environment design
 - (iv) A 'smart' computer based test (CBT)
 - (v) A case study of applied linguistics research conducted by an insider researcher
- (b) The NEPTON project contribution to Intercollege, my work organization:
 - (i) The NEPTON test

I can also note a few areas in which I expanded my ways of learning. These are:

- (a) Learning to research as an insider researcher
- (b) Learning through problem-solving
- (c) Developing leading and team working skills: project leader, manager, negotiator, co-operator, problem-solver, decision-maker, curriculum writer, presenter, researcher, and worker.
- (d) Extending L2 expertise and research skills: online testing development processes, quantitative analysis, and use of data analysis and presentation software programmes.

7.5 Directions for future research

Upto a total of fifteen hundred students will use the NEPTON test, the practical end product of this research project, annually in the three Intercollege campuses in Cyprus. For both theoretical and practical reasons, the NEPTON test will continue to be researched, analysed, developed and improved. For example:

- (a) The test is still being monitored for possible flaws or weaknesses that may still exist in the system.
- (b) Improvement will be made to the tutorial preceding the test, so that it can be used independently online or on CD ROM as a sampler.
- (c) Editing and item analysis will continue to take place when the NEPTON test is re-used, and new cut-off points may need to be worked out as a result of the item analysis for continuous improvement (for example, curriculum, different type of incoming students).
- (d) Discarding, improving or adding items to the item bank will be an ongoing process.
- (e) Provision will be included for separate, campus-based access to results by administrators, faculty and other people concerned as the NEPTON expands to be used in the three campuses.

- (f) More work will be done in improving the criteria for the subjective, hand-written component of the test and the in-servicing training of markers of this component.
- (g) More work will be done in the improvement of the final scoring report. For example, a system will be worked out whereby the results of the electronic component of the test and those of the handwritten component will be automatically calculated to arrive at the final placement. At the moment, they are manually calculated.
- (h) Further investigation in ways to include the written component of the test online: the possibility of electronic assessment of writing, even by de facto or surrogate processes will also be investigated for more efficiency in delivery, marking and time; organize ways through which test-takers, willing to use the computer for their written task, would be able to do so, for example, write it in a 'writing box' or send it via email.
- (i) Further investigation of ways to include listening comprehension tasks online, for example incorporate sound files and use earphones.
- (j) Further investigation of ways to include speaking tasks, for example have test-takers participate in pre-planned situations by recording their input online.
- (k) Development of unsecured component for test-taker practice at any time before the test.
- (l) Continuous improvement of the NEPTON test, and technology wise.

The research project evolved around some aspects of English language placement that are of discrete nature (structure, vocabulary and reading comprehension) although every effort was made to incorporate authentic language use as much as possible. More work and research will be carried out in electronic testing for more authentic and real-life type tasks such as filling in a form, writing a letter or a report, participating in a phone call conversation, reading or listening to some information, for example being informed about a company and then disseminating that information to the relevant body accordingly, either orally or in writing. Although there are still many technological

limitations, which still do not permit more authentic L2 testing online, with the continuous developments in technology and the easier and cheaper access to it, I hope to have the opportunity to embark on such an endeavour.

On another positive note, I believe that the NEPTON research project offers many aspects of relateability and transferability, in other words it can be related to other, similar cases and can be adjusted to the needs of a different institution. For example, its algorithm can change according to a different number of levels, and the cut off points can be adjusted. Its content can also change accordingly. Its characteristics are flexible and adaptable to different contexts and needs. It has already been presented in various international conferences and meetings, and people from different educational institutions from various European countries such as the Netherlands, Lithuania, Greece, Hungary and the United Kingdom, and even from Canada and Japan have already expressed interest in it.

The NEPTON project Report: 34,642 words
(Project Report requirements: 17,500-35,000 words)

References

- About The Computer-Based TOEFL® Test.*
<<http://www.toefl.org/educator/edabtcbtfl.html>> (accessed 05 July 2002).
- A Brief History of Language Teaching Methodology,*
<<http://www.hh.se/shl/Engelska/history.htm>> (accessed 05 July 2002).
- A Common European Framework of Reference for Languages. Learning, Teaching, Assessment,*
<http://www.culture2.coe.int/portfolio/documents_intro/common_framework.html>
(accessed 03 May 2003).
- ALTE (Association of Language Testers in Europe) UCLES,* <<http://www.alte.org/>>
(accessed 03 May 2003).
- Alderson, J.C. (1978a) *A study of the cloze procedure with native and non-native speakers of English*, PhD thesis, University of Edinburgh.
- Alderson, J.C. and Hughes, A. (eds.) (1981) 'Issues in Language Testing', in *ELT Documents III*. British Council.
- Alderson, J.C. Clapham, C. and Wall, D. (1995) *Language Test Construction and Evaluation*. New York: Cambridge University Press.
- Alderson, J. C. (1987) *Innovation in language testing: Can the micro-computer help?* Special Report No. 1: Language Testing Update. Lancaster, UK: University of Lancaster.
- Anastasi, A. (1982) *Psychological testing* (5th ed.) New York: Macmillan.
- Armsby, P. (2000) 'Methodologies of Work Based Learning' in Portwood, D. and Costley, C. *Work Based Learning and the University: New Perspectives and Practices*. Birmingham: SEDA Paper 109.
- Asher, J. (1977) *Learning Another Language Through Actions: The Complete Teacher's Guidebook*. Los Gatos, CA: Sly Oaks Productions
- Bachman, L. F. (1990) sixth impression 2003) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press
- Bachman, L. F. (1999) *Language Testing at the Turn of the Century*. American Association for Applied linguistics newsletter, Vol. 20 No. 2: Testing.

- Balchman, L. F. Carr, N. Kamei, G. Kim, M. Llosa, L. Sawaki, Y. Shin, S. Sohn, S-O. Vongpumivitch, V. Wang, L. Xi, X. and Yessis, D. (2000) *Developing a web-based language placement examination system*. Poster session presented at the 22nd Annual Language Testing Research Colloquium, Vancouver, BC, Canada.
- Bassey, M. (1981) 'Pedagogic Research: on the relative merits of search for generalisation and study of single events' *Oxford Review of Education Vol. 7 no 1* pp.73-93.
- Bell, J. (1999) *Doing your research project*. (3rd ed.) Milton Keynes: Open University Press.
- Blackboard*, academic online learning environment, <<http://www.blackboard.com/>> (accessed 21/04/2002).
- Board of Studies. (1995) *Curriculum and Standards Framework, Languages Other Than English (LOTE)*. Victoria (Australia): Board of Studies.
- Board of Studies (1997) *Curriculum Course Advice*. Victoria (Australia).
- Boud, D. (2001) Knowledge at work: issues of learning, in Boud, D. and Solomon, N. *Work based Learning: A New Higher Education?* Buckingham: SRHE and Open University Press, pp. 34-43.
- Bowden, J. and Marton, F (1998) *The University of Learning: Beyond Quality and Competence in Higher Education*. London: Kogan Page.
- Breen, M. (1984c) 'Process in Syllabus Design', in Brumfit, C. (ed.) *General English Syllabus Design: Curriculum and Syllabus Design for the General English Classroom*. ELT Documents, 118. Oxford: British Council and Pergamon Press.
- Brindley, G. (1997) *Assessment and the Language Teacher: Trends and Transitions*, The Language Teacher Online, <<http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/97/sep/brindley.html>> (accessed 06 June 2002).
- Brockbank, A. and McGill, I. (1998) *Facilitating Reflective Learning in Higher Education*. The Society for Research in Higher Education. Oxford University Press.
- Brown J.D. *Norm-referenced item analysis (item facility and item discrimination)*, Shiken: JALT Testing & Evaluation SIG Newsletter, Vol. 7 No. 2. Jun. 2003. (p. 16 - 19), <http://www.jalt.org/test/bro_18.htm> (accessed 14 February 2005)
- Brown, J. D. (1997) *Computers in Language Testing: Present Research and some Future Directions*. Language Learning & Technology, Vol. 1, No. 1, pp.44-59. <<http://lt.msu.edu/vol1num1/brown/default.html>> (accessed 19 June 2002).

- Brumfit, C.J. and Johnson, K. (1979) *The Communicative Approach to language teaching*. London: Oxford University Press.
- Brumfit, C.J. (1984) *Communicative Methodology in Language Teaching: The roles of Fluency and Accuracy*. Cambridge: Cambridge University Press.
- Canale, M. and Swain, M. (1980) 'The promise and threat of computerized adaptive assessment of reading comprehension', in C.W. Stansfield (ed.), *Technology and language testing*.
- Canale, M. (1986) 'Theoretical bases of communicative approaches to second-language teaching and testing', in *Applied Linguistics*, vol. 1, pp. 1-47.
- Carlson, R. (1994) 'Computer-adaptive testing: A shift in the evaluation paradigm', in *Journal of Educational Technology Systems*, 22, pp. 213-224.
- Carson, G. R. (2003) 'Education: Testing in the 21st Century'. *Focus on Spartanburg Business*. Volume 26, No. 1; Spartanburg Area Chamber of Commerce.
- Gattegno, C. (1972) *Teaching Foreign Languages in Schools: The Silent Way*, New York: Educational Solutions.
- Council of Europe. (2001) *Common European Framework of Reference for Languages, Learning, Teaching and Assessment*. Cambridge
- Chalhoub-Deville M. (2001) 'Language Testing and Technology: Past and Future', in *Language Learning Technology*, Vol. 5, No. 2, pp. 95-98, <<http://llt.msu.edu/vol5num2/emerging/default.html>> (accessed 19 June 2002).
- Chalhoub-Deville, M. (1996) 'Constructing an operational framework for a CAT of L2 reading proficiency', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency* pp. 1-3. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- Clark, J. L. (1987) *Curriculum Renewal in School Foreign Language Learning*. Oxford: Oxford University Press.
- Cohen, A. (1984) Fourth ACROLT meeting on language testing. *TESOL Newsletter*, 18, 23.
- Cohen, L. and Manion, L (2000) *Research Methods in Education*. 5th Ed. London: Routledge.
- Council of Europe, <<http://culture.coe.fr/lang/index.html>> (accessed 03 May 2003).

- Cruz, J. 'Epistemology', in *Nature Encyclopedia of Cognitive Science* article, accessed 31 December 2003)
<http://www.williams.edu/philosophy/fourth_layer/faculty_pages/jcruz/epistemology.pdf>
- Curran, C. (1976) *Counselling-Learning in Second Languages*. Apple River, IL: Apple River Press.
- Curriculum Corporation. (1993) *A National Statement on Languages Other than English (LOTE) for Australian Schools*. Australia. Victoria: Curriculum Corporation.
- Davey, Lynn (1991) The application of case study evaluations. *Practical Assessment, Research & Evaluation*, 2(9), <<http://PAREonline.net/getvn.asp?v=2&n=9>> (accessed 14 December 2003).
- Davis, M. and Chisholm, C. (2003) 'The Integration of work-based learning with lifelace learning – An approach to a new paradigm' in *Work Based Learning Opportunities For Lifelong Learning*. Conference Proceedings, Work Based Learning Network of the Universities Association for continuing education and Intercollege, Cyprus, Intercollege.
- Denscombe, M. (1998) *The Good Research Guide*. Buckingham: Open University Press
- de Jong, J. (1996) 'Towards integrated learning and testing using structured item banks for CAT', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency*, pp. 38-41. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- Divgi, D. R. (1986) Does the Rasch model really work for multiple-choice items? Not if you look closely. *Journal of Educational Measurement*, 23.
- Dubin, F. and Olshain E. (1987) *Course Design, Developing Programs and Material for Language Learning*. Cambridge: Cambridge University Press.
- Dunkel, P. (ed.) (1991) *Computer-Assisted Language Learning and Testing*. New York: Newbury House.
- Dunkel, P. (1996) 'Checking the utility and appropriacy of the content and measurement models used to develop L2 listening comprehension CATS: Implications for further development of comprehension CATs.' in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency*, pp. 27-37. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.

- Dunkel, P. A. (1997) 'Computer-adaptive Testing of Listening Comprehension: A Blueprint for CAT Development', In *The Language Teacher Online*, 21(10). 20/3/00, <<http://langue.hyper.chubu.ac.jp/jalt/pub/tlt/97/oct/dunkel.htm>> (accessed 19 June 2002).
- Dunkel, P. (1999) 'Considerations in Developing or using second/foreign language proficiency computer-adaptive tests'. *Language Learning & Technology*, Vol. 2, No. 2, pp. 77-93, <http://www.multingles.net/docs/tai_dunkel.htm> (accessed 22 July 2003)
- Eignor, D., Way, W. Stocking, M. & Steffen, M. (1997). *Case studies in computer adaptive test design through simulation* (TOEFL Research Report #93-56).
- Princeton, NJ: Educational Testing Service, < <ftp://ftp.ets.org/pub/toefl/275756.pdf>> (accessed 19 October 2003).
- Felix, U. (1998) *Virtual Language Learning. Finding the gems amongst the pebbles*. Language Australia Ltd.
- Fink, A. (1995) 'What Is a Survey? When Do You Use One?' in *The Survey Handbook*. Sage Publications
- Finocchiaro, M. and Brumfit, C. J. (1983) *The Functional-Notional Approach. From Theory to Practice*. Oxford: Oxford University Press.
- Lord, F. (1980) *Applications of Item Response Theory to Practical Testing Problems*. Educational Testing Service.
- Fulcher, G. (1999) Ethics in Language Testing, <<http://taesig.8m.com/news1.html>> (accessed 20 October 2003).
- Fulcher, G. (2000) 'Computers in Language Testing' in Nrett, P. and Motteram, G. (eds.), *A Special Interest in Computers*. Manchester: IATEFL, Publications, pp. 93-107, <<http://www.dundee.ac.uk/languagestudies/ltest/ltrfile/Computers.htm>> (accessed 26 September 2003).
- Fulcher, G. *Resource In Language Testing WWW Page*, <<http://www.surrey.ac.uk/ELI/ltr.htm>> <<http://www.dundee.ac.uk/languagestudies/ltest/ltr.html>> (accessed 17 July 2002).
- Funk, A. (1995) *The Survey Handbook*. Sage Publications.
- Gault, I. and Papadima-Sophocleous S. (2004) *Voilà 1* Teacher Resource Book and Audio CD. Brisbane: John Wiley & Sons, Australia, Jacaranda.
- Gault, I. and Papadima-Sophocleous S. (2004) *Voilà 2* Teacher Resource Book and Audio CD. Brisbane: John Wiley & Sons, Australia, Jacaranda.

- General Certificate of Education (GCE),
 <[http://en.wikipedia.org/wiki/General Certificate of Education](http://en.wikipedia.org/wiki/General_Certificate_of_Education)> (accessed 14 December 2003).
- Gillham, B (2000) *Case Study Research Methods*. London Continuum
- Godwin-Jones, R. (1996) 'Grating Language Learning Material. for the World Wide Web', In Warschauer. M. (ed.), *Telecollaboration in Foreign Language Learning*. pp.3-17. Honolulu, University of Hawaii, Second Language Teaching and Curriculum Center.
- Godwin-Jones, B. (2001) 'Language Testing Tools and Technologies', in *Language Learning Technology*, Vol. 5, No. 2, pp. 8-12,
 <<http://llt.msu.edu/vol5num2/emerging/default.html>> (accessed 19 June 2002).
- Goldstein, H. (1979) Consequences of using the Rasch model for educational measurement. *British Educational Research Journal*, 5.
- Griffiths, J. et al. including Sophocleous, S. (2000) *Ça Alors 2 Teacher's manual Assessment Task Samples*. Melbourne: CIS-Heinemann.
- Grist, S. (1989) Computerized Adaptive Tests. ERIC Digest No. 107. ERIC Clearinghouse on Tests Measurement and Evaluation Washington DC, American Institutes for Research Washington DC,
 <<http://www.ericdigests.org/pre-9213/tests.htm>> (accessed 11 February 04).
- Harmer, J. (1991) *The Practice of English Language Teaching*. UK: Longman. Cambridge University Press.
- Heaton, J.B. (1988) *Writing English Language Tests*. UK: Longman
- Heinz, C. Taraborelli, D. (2003) *Web Epistemology*, <<http://web.epistemology.free.fr/>> (accessed 21 December 2003).
- Henning, G. Johnson, P. Boutin, A. and Rice, H. R. (1994) 'Automated assembly of pre-equated language proficiency tests.' *Language Testing*, 15, pp. 15-28.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. New York: Newbury House.
- Henning, G. (1989) 'Does the Rasch model really work for multiple-choice items? Take another look'. A response to Divgi. *Journal of Educational Measurement*, 26.
- Higgins, J. (1983) Can computers teach? *CALICO Journal*, 1 (2), pp. 4-6.
- Hot Potatoes, Hot-Baked Potatoes*, <<http://www.halfbakedsoftware.com/>> (accessed 29 December 2003).

- Hudson, D. *Applied Linguistics*. <<http://www.phon.ucl.ac.uk/home/dick/AL.html> > (accessed 04 March 2005)
- Hughes, A. (1989) *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Heylighen, F. (1993) 'Epistemology, introduction' in *Principia Cybernetica Web* (Principia Cybernetica, Brussels), <<http://pespmc1.vub.ac.be/EPISTEMI.html>> (accessed 31 December 2003).
- Houle, C. (1980) *Continuing Learning in the Professions*, San Francisco: Jossey-Bass.
- Intercollege *Placement Test: Marking Guidelines and Sample Placement*
- Intercollege *English Syllabi: BENG 50, BENG 80, BENG 90, BENG 100, ENGL 100, ENGL 101*
- Intercollege *Course Outlines: BENG 50, BENG 80, BENG 90, BENG 100, ENGL 100, ENGL 101*
- Intercollege existing *English Placement Test*, pen-and-paper format
- Intercollege prospectus, March 2004
- International English Language Testing System (IELTS)*, <<http://www.ieltsonline.com/>> (accessed 05 August 2002).
- Item Analysis*, Scoring Office, Michigan State University <<http://www.msu.edu/dept/soweb/itanhand.html>> (accessed 14 February 2005)
- Jarvis, P. (1987) *Adult Learning in the Social Context*, London: Croom Helm.
- Johnson, K. and Porter, D. (ed.), (1983). *Perspectives in Communicative Language Teaching*. London: Academic Press.
- Johnson, K. (1982) *Communicative Syllabus Design and Methodology*. Oxford: Pergamon.
- Johnson, R.K. (1989) *The Second Language Curriculum*. Cambridge: Cambridge University Press.
- Johnston, C. (1996) 'Computerized testing on a large network: Issues for today and tomorrow', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency* pp. 7-10. Symposium conducted at the Center for Advanced Research on Language Acquisition, University of Minnesota, Bloomington, MN.

- Richards, J.C. and Rodgers, T.S. (2001) *Approaches and Methods in Language Teaching*, 2nd Edition Cambridge University Press.
- Kaya-Carton, E. Carton, A. and Dandonoli, P. (1991) 'Developing a computer-adaptive test of French reading proficiency', in P. Dunkel (ed.), *Computer-assisted language learning and testing: Research issues and practice* pp. 259-284. New York: Newbury House.
- Kelly, R. (1978) *On the construct validation of comprehension tests: an exercise in applied linguistics*. PhD thesis, University of Queensland.
- Kenning, M.M. and Kenning, M.J. (1990) *Computers and Language Learning: Current Theory and Practice*. New York, Horwood.
- Kenning, M.J. and Kenning, M.M. (1984) *An introduction to computer assisted language teaching*, Oxford University Press.
- Kitao, S. and Kitao, K. (1996) 'Testing Communicative Competence' *The Internet TESL Journal*, Vol. 2, No. 5, <<http://iteslj.org/Articles/Kitao-Testing.html>> (accessed 26 September 2003).
- Kitao, S. and Kitao, K. (2002) *Test Specifications*. TESL-L Electronic Discussion Forum for teachers of English as a second or foreign language, <<http://ilc2.doshisha.ac.jp/users/kkitao/library/article/test/design2.htm#specification>> (accessed 23 October 2003).
- Kitao, S. and Kitao, K. *Language Testing on WWW*, <<http://ilc2.doshisha.ac.jp/users/kkitao/onlinc/www/kitao/int-test.htm>> (accessed 23 October 2003).
- Kitao, S. and Kitao, K. *Testing Grammar*, <<http://www.aitech.ac.jp/~iteslj/Articles/Kitao-TestingGrammar.html.m9>> (accessed 23 October 2003).
- Kitao, S. and Kitao, K. *The History of English Teaching Methodology*, <<http://www.ling.lancs.ac.uk/staff/visitors/kenji/kitao/tesl-his.htm>> (accessed 23 October 2003).
- Klein, P. D. (1998) 'Epistemology' In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge, <<http://www.rep.routledge.com/article/P059>> (accessed 31 December 2003).
- Kolb, D. (1996) 'On Experiential Learning' In *The Encyclopedia of Informal Education*, <<http://www.infed.org/biblio/b-explrn.htm>> (accessed 31 December 2003).
- Krashen, S.D. (1982) *Principles and Practices in Second Language Acquisition*. Oxford: Pergamon Press.

- Krashen, S.D. (1982) *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
- Larsen-Freeman, D. (1986) *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press.
- Larson, J. (1996) 'Content considerations for testing reading proficiency via computerized-adaptive tests', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency* pp. 24-26. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- Larson, J. W. and Madsen, H. (1985) 'Computerized adaptive language testing: Moving beyond computer-assisted testing.' *CALICO Journal*, 2, p.p. 32-36, 43.
- Lemke, J. L. 'Analysing Verbal Data: Principles, Methods, and Problems', in Tobin, K. and Fraser, B., (eds.) *International Handbook of Science Education* (Kluwer) <<http://academic.brooklyn.cuny.edu/education/jlemke/papers/handbook.html>> (accessed 31 December 2003).
- Leedy, P. D. (1993). *Practical Research, Planning and Design*. USA: MacMillan Publishing Company
- Levy, M. (1997) *Computer-Assisted Language learning, Context and Conceptualization*. Oxford, Clarendon Paperbacks.
- Livesey, C. *Theory and Methods*, <<http://www.sociology.org.uk/primary2.doc>> (accessed 20 October 2003).
- Linacre, J. M. (1996) 'Constructing a reading strength profile with computer adaptive testing', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency* pp. 42-45. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- Lightbown, P. and Spada, N. (1993) *How Languages are Learned*. Oxford: Oxford University Press.
- Littlewood, W. (1986) *Communicative Language Teaching: New Directions in Language Teaching*. Cambridge University Press.
- Lo Bianco, J. (1987) *National Policy on Languages*. Canberra: Commonwealth Department of Education, Australian Government Publishing Service.
- Lo Bianco, J. (1989) *Languages Action Plan*. Victoria: Ministry of Education.

- Lord, F.M. (1980) *Applications of item response theory to practical testing problems*. Hillsdale NJ: Lawrence Erlbaum
- Language Other Than English Curriculum and Standards Framework II*, (2000) Victoria, Board of Studies
- Luecht, R. (1996) 'The practical utility of Rasch measurement models', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency*. pp. 53-60. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- Madsen, H. and Larson, J.W. (1986) Computerized Rasch analysis of item bias in ESL tests', in C.W. Stansfield (ed.) *Technology and language testing*. Washington, DC: TESOL.
- Madsen, H. (1991) 'Computer-adaptive testing of listening and reading comprehension', in P. Dunkel (ed.), *Computer-assisted language learning and testing* pp. 237-257. New York: Newbury House.
- Making a Placement Test for TESL or TEFL,
<<http://www.btinternet.com/~ted.power/esl0820.html>> (accessed 26 October 2003).
- Methods of Language Testing & Assessment,
<<http://www.btinternet.com/~ted.power/esl0704.html>> (accessed 26 October 2003).
- Ministry of Education. (1987) *The LOTE Framework P-10*. Victoria: School Division.
- Ministry of Education. (1990) *LOTE Framework Years 7-10 Course Development Guide*. Draft for Consultation.
- Ministry of Education. (1990) *VCE LOTE: Sample Units*, National Policy on Languages Project. Victoria: State Multicultural Education Coordination Unit.
- Moller, A.D. (1982b) *A study in the validation of proficiency tests of English as a Foreign Language*. PhD thesis, University of Edinburgh.
- Morrison, S, *Second Language Teaching Methodologies*, ERIC Clearinghouse on Languages and Linguistics, <<http://www.cal.org/ericcll/faqs/rgos/methods.html>> (accessed 20 August 2003).
- Morrow, K. (1979) 'Communicative language testing: revolution or evolution?', in Brumfit, CJ and Johnson K (eds.) *The Communicative Approach to Language Teaching*. Oxford: Oxford University Press, pp. 143-157.
- Munby, J. (1978) *Communicative Syllabus Design*. Cambridge University Press. Fourth Printing.

- Nagata, N. (1998) 'Input vs. output practice in educational software for second language acquisition.' *Language Learning & Technology*, 1 (2), pp. 23-40. Retrieved October 15, 1998 from the World Wide Web, <<http://polyglot.cal.msu.edu/llt/vol1num2/article1/default.html>> (21 accessed October 2003).
- N'oon, M. and Blyten, P. (1997) *The realities of work* Mc Milan
- Nunan, D. (1985a) *Language Teaching Course Design Trends and Issues*. Adelaide: National Curriculum Resource Centre.
- Nunan, D. (1987) *Task Design: a perspective on language teaching*. Adelaide: National Curriculum Resource Centre.
- Nunan, D. (1988) *The Learner-Centred Curriculum*. Cambridge: Cambridge University Press.
- Nunan, D. (1989) *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.
- Nunan, D. (1991) *Language Teaching Methodology*. Hertfordshire: Phoenix ELT.
- Nunan, D. (1995) *Research Methods in Language Learning*. Cambridge University Press. USA.
- Oller, J. (1979) *Language Testing at School: A Pragmatic Approach*. Longman: London.
- O'Shea, T. and Self, J. (1983) *Learning and teaching with Computers*, Brighton, The Harvester Press.
- Overview of methodologies for language teaching*,
< <http://simsim.rug.ac.be/staff/elke/recpast/recpast.html> > (accessed 21 May 2003).
- Papadima-Sophocleous, S. et al. (2004) *The COBLE (committee of BENG level English) Report, Review of Intercollege English Programme*. Draft. Cyprus. Intercollege.
- Papadima-Sophocleous, S. (2002) *Voilà 1*, student textbook. Brisbane. John Wiley & Sons, Australia. Jacaranda Publications.
- Papadima-Sophocleous, S. (2002) *Voilà 1*, student workbook. Brisbane. John Wiley & Sons, Australia. Jacaranda Publications.
- Papadima-Sophocleous, S. (2002) *Voilà 2*, student textbook. Brisbane. John Wiley & Sons, Australia. Jacaranda Publications.

- Papadima-Sophocleous, S. (2002) *Voilà 2*, student workbook. Brisbane. John Wiley & Sons, Australia. Jacaranda Publications.
- Papadima-Sophocleous, S. (2002), *Ελληνικά, γιατί όχι*; Textbook. Learning material for Greek as an L2 for Adolescent learners. "Paedia Omogenon" programme and the Greek Ministry of Education.
- Papadima-Sophocleous, S.(2002), *Ελληνικά, γιατί όχι*; Workbook. Learning material for Greek as an L2 for Adolescent learners. "Paedia Omogenon" programme and the Greek Ministry of Education.
- Papadima-Sophocleous, S. (2000) 'Postgraduate Studies online: postgraduate diploma in Educational Studies, specialisation in Teaching Second Language and Computer Assisted Language Learning; Planning, and development.' In Antonopoulos, N. et al. *The teaching of Greek as a second/foreign language, principles-problems-potential*. Salonica: Centre of Greek Language, Department of promoting Greek Language.
- Papadima-Sophocleous, S. (2002) *Voilà 1 Interactive*. Australia: John Wiley Sons Australia, Ltd. Jacaranda. (Windows edition CD-ROM)
- Papadima-Sophocleous, S. (2002) *Voilà 2 Interactive*. Australia: John Wiley Sons Australia, Ltd. Jacaranda. (Windows edition CD-ROM)
- Papadima-Sophocleous, S. (2002) *Voilà*. Australia: John Wiley Sons Australia, Ltd. Jacaranda. Website. <<http://www.jaconline.com.au/languages/index.html>> (accessed 19 June 2002).
- Papadima-Sophocleous, S. (1999-2000) Online Postgraduate Diploma in Educational Studies. LOTE: Greek. RMIT University.
- Papadima-Sophocleous, S. (1999-2000) "French-Australian Links". Eltham High School Web Page of Collaborative Project with Schools from Australia and France.
- Papadima-Sophocleous, S. (1996) Sister school program Templestowe Primary School of Greek Community of Melbourne and Victoria, Australia and 1st Experimental Primary School of Ionian, Greece, 49 minute video production of the school life of students in Australia, sent to the students in Greece.
- Papadima-Sophocleous, S. (1994) *Fabien, Un Jeune Stéphanois*, a ten-minute video, teaching suggestions and student activity booklet, on the life of Fabien, a French teenager.
- Patrikis, P. (1997) The evolution of computer technology in foreign language teaching And learning, in Debski, R. and others, (eds.) *Language Learning through social computing*. Melbourne, ALAA and Horwood Language Centre.

- Portwood, D. and Costley, C. (2000) *Work Based Learning and the University: New Perspectives and Practices*. Birmingham: SEDA Paper 109.
- Quia, <<http://www.quia.com/tutoract.htm>> (accessed 16 February 2000).
- Quick Placement Test (2001) UK. University of Cambridge, Oxford University Press. (Windows edition CD-ROM.
- Question Mark Perception Question Mark Computing Ltd. (Windows edition CD-ROM) sampler.
- Rea, P.M. (1978) 'Assessing Language as Communication', *MAL. Journal*, New Series No. 3, University of Birmingham.
- Richards, J. C. and Rodgers, T.S. (2001) *Approaches and Methods in Language Teaching* (2nd ed.). Cambridge: Cambridge University Press.
- Richards, J. C. and Renandya, W. A. (2002) *Methodology in Language Teaching. An Anthology of Current Practice*. Cambridge University Press.
- Richards, J.C. (2001) *Curriculum Development in Language Teaching*. UK: Cambridge University Press.
- Richards, J.C. (ed.) (1988) *Teaching in Action: Case Studies from Second Language Classrooms*.
- Richards, J.C. and Rodgers, T. S. (1986) *Approaches and Methods in Language Teaching*. Cambridge University Press.
- Richards, J.C. (1985) *The Context of Language Teaching*. Cambridge: Cambridge University Press.
- Richards, Jack C. (1984) 'Language curriculum development'. *RELC Journal* 15, no. 1: pp. 1-29
- Rodgers, T. S. *Language Teaching Methodology*, <<http://www.cal.org/ericcll/digest/rogers.html>> (accessed 21 May).
- Roever, C. (2001) 'Web-Based Language Testing', in *Language Learning Technology*, Vol. 5, No. 2, pp. 84-84, <<http://lt.msu.edu/vol5num2/emerging/default.html>> (accessed 19 June 2002)
- Salkind, N. J. (2000) *Statistics for People who (think they) hate statistics*. California: Sage Publications.
- Savignon, S.J. (1972) *Communicative competence: An experiment in foreign language teaching*. Philadelphia: The Center for Curriculum Development.

- Scarino, A. Vale D. and McKay P. (1988) *The Australian Language Levels Guidelines* (ALL), books 1-4, Curriculum Development Centre. A.C.T.
- Schon, D. (1987) *Educating the Reflective practitioner*. San Francisco: Jossey Bass.
- Sedunary, M. et al. including Sophocleous. S. (1998) *Ça Alors 1 Teacher's manual Assessment Task Samples*. Melbourne: CIS-Heinemann.
- Skehan, P. (1988) *Language testing, part 1: state of the art article*. *Language Teaching*, 21/4, 211-218
- Slowey, M. (1995) *Implementing Change from within Universities and Colleges*. London: Kogan Page Ltd. 19.95 L
- Smith, F. (1983) 'The promise and threat of microcomputers for language learner', in J. Handscombe et al. (eds.) *On TESOL '83*. Washington, DC. TESOL. (ERIC Document Reproduction Service No. ED 275 138.
- Sophocleous, S. (1992-1994) *Γεια σου* material for Greek as an L2 for Directorate of Employment, Education and Training, Canberra, Australia (DEET), Directorate of Secondary Education, Victoria, Australia, (DSE) and Modern Greek Teachers Association of Victoria, Australia (MGTAV) project.
- Sophocleous, S. (1992) *Principles and Practices in Second Language Learning Assessment*. M. Ed. Prelim dissertation. La Trobe University.
- Special Connections, Item Analysis <<http://www.specialconnections.ku.edu/cgi-bin/cgiwrap/speconn/main.php?cat=assessment§ion=main&subsection=qualitytest/item>> (accessed 14 February 2005)
- Spolsky, B. (1968) *Language testing: The problem of validation*. *TESOL Quarterly*, 2, 88-94.
- Stern, H.H. (1983) *Fundamental Concepts of Language Teaching*. Oxford : Oxford University Press.
- Stevenson, J. and Gross, S. (1991) 'Use of a computerized adaptive testing model for ESOL/bilingual entry/exit decision making', in P. Dunkel (Ed.), *Computer-assisted language learning and testing* pp. 223-235. New York: Newbury House.
- Taylor, C. Jamieson, J. Eignor, D. and Kirsch, I. (1998) *The relationship between computer familiarity and performance on computer-based TOEFL test tasks* (TOEFL Research Reports 61). March. Princeton, NJ: Educational Testing Service, <<http://www.ets.org/ell/research/toeflresearch.html>> (accessed 13 March 2004).

Tella, S. (1996) 'Foreign Languages and Modern Technology: Harmony or Hell?' In Warschauer, M. (Ed.), *Telecollaboration in Foreign Language Learning*. pp.3-17. Honolulu, University of Hawaii, Second Language Teaching and Curriculum Center.

Test Scoring Statistics Guide, Interpreting the Reports,
<<http://helpdesk.kent.edu/howto/testscoring/stat/>> (accessed 14 February 2005)

Thanasoulas, D. (2002) *The changing winds and shifting sands of the history of English Language Teaching*. englishclub.com 1997-2002,
<http://teachers.englishclub.com/articles/The_changing_winds.htm> (accessed 22 June 2003).

TOEFL Sampler (1998) USA: Educational Testing Service. (Windows edition CD-ROM).

TOEFL® Testing Program, <<http://www.toefl.org/tstgprog01.cfm>> (accessed 31 March 2001).

Tung, P. (1986) 'Computerized adaptive testing: Implications for language test developers', In C.W. Sandfield (Ed.), *Technology and language testing*: Washington, DC: TESOL.

Ur, P. (1997) *A Course in Language Teaching, Practice and theory*. Cambridge Teacher Training and Development. Series Editors: Marion Williams and Tony Wright. Cambridge University Press.

Usher, R. (1996) 'A critique of the neglected epistemological assumptions of educational research' in Scott, D. and Usher, R. (eds.) *Understanding Educational Research*. London: Routledge.

Vallantin, P. et al. including Sophocleous, S. (2000) *Ça Alors 2 Workbook*. Melbourne: CIS-Heinemann.

van Maanen J. (1995) 'An End to Innocence, The Ethnography of Ethnography', in *Representation in Ethnography*. Sage publications.

Victorian Certificate of Education (VCE) Curriculum Material. Pool, Languages Other than English (LOTE) GREEK, Education Shop, Department of School Education, Ministry of Education and Training, Victoria, (1991), member of Greek teacher group.

Wallace, M. J. (1991) *Training Foreign Language Teachers, A reflective approach*. Cambridge Teacher Training and Development. Series Editors: Marion Williams and Tony Wright. Cambridge University Press.

- Warschauer, M. (ed.) (1996) *Telecollaboration in Foreign Language Learning*. pp.3-17. Honolulu, University of Hawaii, Second Language Teaching and Curriculum Center.
- WebAssign*. <<http://www.webassign.net/>> (accessed 08 September 2001).
- WebCT*, <<http://about.webct.com/>> (accessed 08 September 2001).
- Weir, C.J. (1990) *Communicative Language Testing*. UK: Prentice Hall International (UK) Ltd.
- Weir, C.J. (1995) *Understanding and Developing Language Tests*. Great Britain: Phoenix ELT.
- Weiss, D. (1996) 'A perspective on computerized second language testing', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency*. Symposium conducted at the Center for Advanced Research on Language Acquisition of the University of Minnesota, Bloomington, MN.
- West, R. (1995) 'Computer-inspired testing: Multiple-matching.' *Language Testing Update*. 17, pp. 44-46.
- Wyatt, D.H. (1983) 'Three Major Approaches to Developing Computer-Assisted Language Learning Material. for Microcomputers', *CALICO Journal*, 1/2: 34-8.
- Wyatt, D.H. (1984a) (ed.) *Computer-Assisted Language Instruction: Special Issue of System*. Oxford, Pergamon.
- Yalden, J. (1983) *The Communicative Syllabus: evolution, design and implementation*. Oxford: Pergamon Press.
- Yalden, J. (1987) *Principles of Course Design for Language Teaching*. Cambridge: Cambridge University Press.
- Yule, G. (1994) *The Study of Language*. Cambridge: Cambridge University Press.
- Yao, T.-C. (1995) 'A computer-adaptive test for reading Chinese (CATRC): A preliminary report.' *Journal of the Chinese Language Teachers Association*, 30, pp. 75-85.
- Yin, R. M. (1989) *Case study research, design and methods*. California: Sage
- Yin, R. M. (1993) *Applications of case study research*. California: Sage

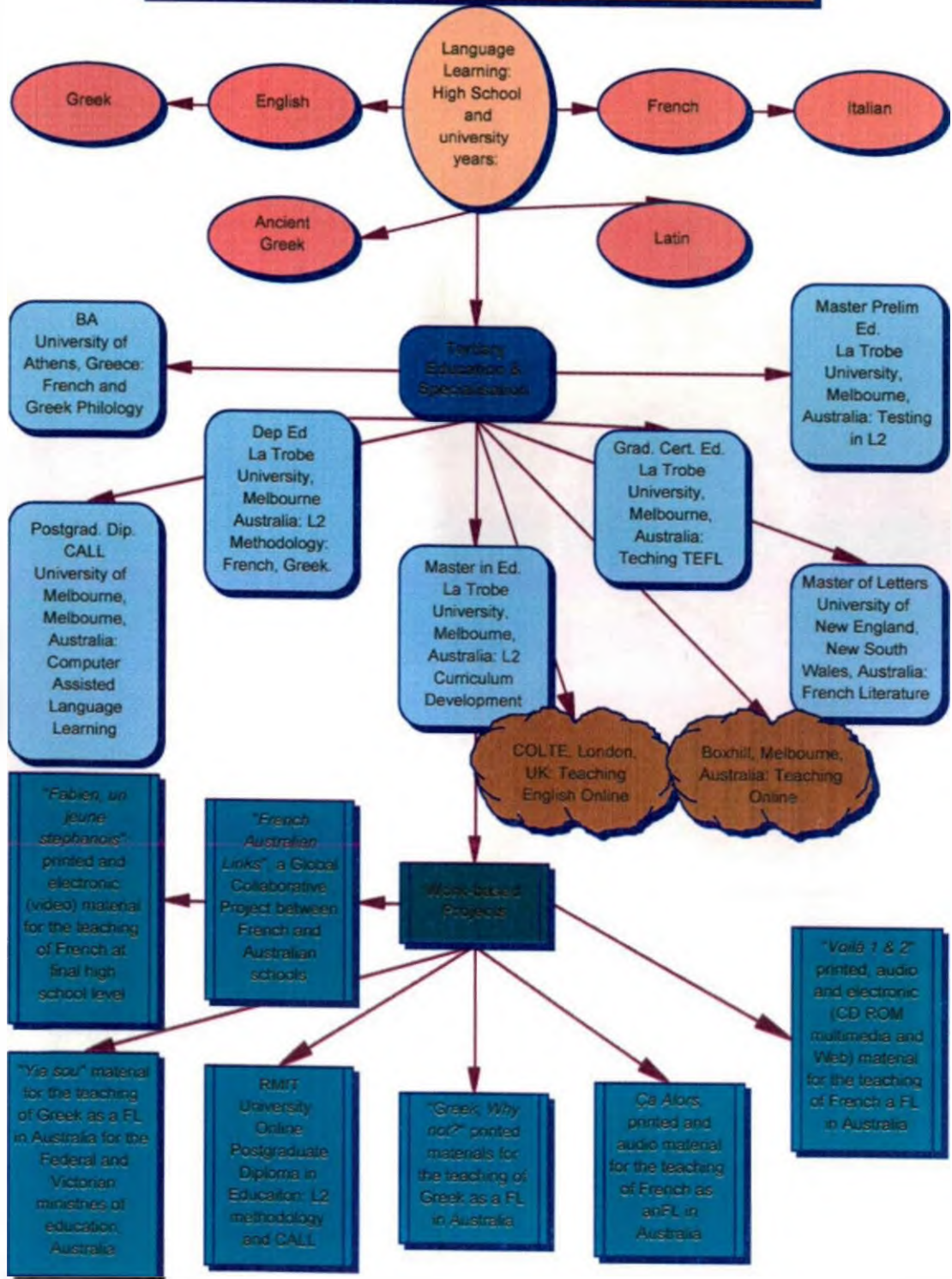
Yoes, M. (1996) 'Exploring new item-types for computerized testing: New possibilities and challenges', in M. Chalhoub-Deville (Chair), *Issues in computer-adaptive testing of second language reading proficiency*. Symposium conducted at the Center for Advanced Research on Language Acquisition, University of Minnesota, Bloomington, MN.

Appendices

Appendix 1

Review of previous learning

Review of Previous Learning



Appendix 2

Collated data from British universities

Collated Data from British universities

Institution	Course Name	Content	Prerequisite	Duration
1. U of St Andrews	Chinese Access Course	English, writing and foundation subjects	High	1 academic year
2. Leeds Metropolitan U	Foundation	English, study skills, IT, Culture, foundation major	IELTS 4.0	1 academic year
3. Edinburgh Telford Coll	English Language Foundation	English language	IELTS <5	1 academic year
4. U of Huddersfield	Foundation Course	English, culture, foundation major	IELTS 4.5	1 academic year
5. U of Manchester	General English	English language	Low (IELTS <5)	44 weeks
U of Manchester	Foundation Course	English, writing, study skills,	IELTS 5 TOEFL 500	10 weeks
U of Manchester	Professional Course	English, culture	IELTS 6 TOEFL 550	12 weeks
6. U of Wales, Lampeter	English Language Courses	English, foundation major	IELTS <6	1 academic year
7. Sussex Downs Coll.	International Foundation Courses	English, study skills, foundation major	IELTS 5 TOEFL 500	1 academic year
8. U of Luton	International Foundation Courses	English, skills, foundation major	TOEFL 510	1 academic year
9. U of Buckingham	EAP Foundation Program	English language, culture	IELTS 4	1 calendar year
10. Trinity Coll, Wales	International Foundation Courses	English language, skills, foundation major	IELTS 5 TOEFL 500	1 academic year
11. Buckinghamshire Chilterns, U Coll	English Language Foundation Course	English language, culture	IELTS 4 TOEFL 400	1 academic year
12. U of Exeter	Foundation Program	English, foundation major	IELTS 5 TOEFL 510	1 academic year
13. Middlesex U	International Foundation Program	English, skills, foundation majors	IELTS 4.5 TOEFL 450	1 academic year
14. Oxford Brookes U	University Foundation Course	English, skills, culture	IELTS 4.5	1 academic year
15. London Metropolitan U (London Guildhall U)	Diploma in Foundation Studies	English, culture, foundation major	IELTS 4.5 TOEFL 475	1 academic year

Appendix 3

Profile of University or College English Placement Tests

Profile of University or College English Placement Tests

Name of Institution	<i>1 Norwalk Community College</i>	<i>2 San Francisco State University</i>
Kind of test takers	All entering students not having previous transfer credits or NCC credits in English	Non-native speakers of English who wish or need to take ESL classes
Name of Test - Purpose	Placement tests http://www.nctc.comnet.edu/placement.htm	http://www.sfsu.edu/~testing/eslpt.html
What they test	Reading comprehension, grammar and sentence skills	Writing, grammar.
Duration	Not timed	Maximum time allowed: 3 hours
Mode of Delivery	Taken on a computer	Not mentioned
Kinds of activities	Multiple-choice questions	Summary & composition writing, multiple choice grammar questions

Name of Institution	<i>3 Columbia University</i>	<i>4 The California State University</i>
Kind of test takers	All non-native speakers	All entering undergraduates, with the exception of those who present proof of one of the following:
Name of Test - Purpose	American Language Program (ALP) placement test: http://www.ce.columbia.edu/tesol/admission.cfm	The CSU English Placement Test http://www.ets.org/csu/ept.html
What they test	Part I is similar to TOEFL. Part II includes a test of writing skills.	Reading and writing skills
Duration	Not mentioned	
Mode of Delivery	Not mentioned	
Kinds of activities	Not mentioned	

Name of Institution	5 University of Cincinnati	6 Pennsylvania College of Technology
Kind of test takers		
Name of Test - Purpose	University College EPT will determine which entry-point course is right for each student. Students can place into one of four courses, which will become their first English course. http://www.uc.edu/englishplacement/ucollege_entry.asp	Placement Test http://www.pct.edu/advise/placement_testing/english.htm
What they test	1) Reading Test & 2) writing of an Essay, a Quick Start, or an On-line English Essay session.	Writing
Duration		
Mode of Delivery		
Kinds of activities	Reading comprehension activities: type not mentioned. - Essay writing	Write a short essay you will be given a choice between two subject matters

Name of Institution	7 Oxford and Cambridge Unis: QPT: Quick Placement Test	8 TOEFL: Test of English as a Foreign Language
Kind of test takers	Any L2 English student	Admission purposes: more than 4,200 two and 4-year colleges & universities in the USA & Canada TOEFL is also used by institutions in other countries where English is the language of instruction. In addition, government agencies, scholarship programmes, & other agencies use TOEFL scores to evaluate English proficiency.
Name of Test - Purpose	Placement Test http://www.oupchina.com.hk/what/elect_arch02.asp#new2	Evaluate the English proficiency of people whose native language is not English
What they test	Grammar, vocabulary, reading and listening	Listening, structure, reading, writing
Duration	Electronic: 15 minutes	Writing: 30 minutes
Mode of Delivery	Pen-and-paper OR computer-based programme electronic adaptive placement test: CD ROM	CB TOEFL test: Listening & structure: computer-adaptive; Structure: incomplete sentences, correct sentence. Reading; Writing: on the computer or handwritten.
Kinds of activities	Multiple-choice questions	<u>Structure question types:</u> 1. Fill in the blank 2. Click on an underlined word or phrase <u>Reading comprehension question types:</u> 1. Select one of four choices 2. Click on a word, phrase or sentence. 3. Add a sentence. <u>Essay writing</u> on an assigned topic

Name of Institution	9 IELTS: <i>International English Language Testing System</i>): Testing: Jointly managed by the University of Cambridge Local Examinations Syndicate, the British Council and IDP Education Australia	10 <i>The Michigan</i>
Kind of test takers	Internationally recognised assessment of a candidate's ability to study or train in the medium of English.	Advanced secondary school students; adult non-native speakers of English interested in college or university study.
Name of Test - Purpose	Academic version is used for admission to undergraduate and postgraduate studies	The Michigan Test of English Language Proficiency (MTELP)
What they test	Speaking, Listening, Reading and Writing	Grammar, vocabulary, reading comprehension
Duration		Time: 75 minutes
Mode of Delivery		Scoring is by punched stencil- on scannable answer sheets.
Kinds of activities		100 item multiple-choice test: grammar items in a conversational format, vocabulary items requiring selection of a synonym or completion of a sentence, and reading passages followed by comprehension questions.

Name of Institution	11 ACT Educational Services Division	12 <i>PEAK English the online Interactive English School</i>
Kind of test takers	ESL teachers.	English placement test
Name of Test - Purpose	ESL Placement Test adaptive testing http://www.act.org/esl/descript.html	http://www.peakenglish.com/placement_test/placementTest.jsp
What they test	Grammar/Usage, Reading, Listening	Usage, Grammar, Reading Comprehension, Listening Comprehension
Duration	Not mentioned	Not mentioned
Mode of Delivery	Taken on a computer	Taken online
Kinds of activities	Multiple-choice questions	40 questions divided into the four skill categories, kind of questions not mentioned

Appendix 4

Profile of other electronic assessment devices used in L2 testing

Profile of other electronic assessment devices used in L2 testing

Name of Institution	<i>Question Mark:</i> Courseware Developer with the School of Biological sciences, University of Sydney.	<i>Hot Potatoes</i> The <i>Hot Potatoes</i> is a six applications suite for the construction of exercises and tests
Kind of test takers		
Name of Test - Purpose	<i>Question Mark</i> is a set of three applications for the construction and delivery of tests or surveys (<i>QM Designer</i> and <i>QM Presenter</i>), and for the reporting of user performance or responses (<i>QM Reporter</i>). http://www.questionmark.com/deu/	http://web.uvic.ca/hrd/halfbaked/ It includes six applications, enabling you to create interactive multiple-choice, short-answer, jumbled-sentence, crossword, matching/ordering and gap-fill exercises for the World Wide Web. <i>Hot Potatoes</i> is not freeware, but it is free of charge for those working for publicly funded non-profit-making educational institutions, which make their pages available on the web. Other users must pay for a licence. Check out the <i>Hot Potatoes</i> licensing terms and pricing on the Half-Baked Software Website .
Mode of Delivery	Online	Online
Kinds of activities	<i>Explanation, (information);</i> Multiple choice (MC) (MC; <i>Likert scale (Multiple choice); True/False (MC);</i> Yes/no (MC); Multiple response (Multiple response); Matrix (MC); Matching (Multiple choice); Pull down list (MC); Ranking (MC); Select a blank (MC); Fill in Blanks (Gap - fill); Essay (Essay); Text match (Gap - fill); Numeric (Numeric); Drag and Drop (D&D) (D&D); Hotspot (Hotspot); Flash (Animation).	Interactive multiple-choice, short-answer, jumbled-sentence, crossword, matching/ordering and gap-fill exercises

Name of Institution	<p><i>Quia</i> It provides technology resources to educators</p> <p>http://www.quia.com/findout.html</p>	<p><i>WebAssign</i> http://www.campussource.de/org/software/webassign/</p> <p>It is a web-based system that integrates written assignments into a virtual university. It comprehensively supports authoring of assignments, download of assignments onto the students' PC, upload of solutions, distribution of solutions over the correctors, assessment of solutions and finally returning the assessed solutions together with sample solutions to the students.</p>
Name of Test - Purpose	<p><i>Quia</i> provides a wide variety of educational services, including:</p> <ul style="list-style-type: none"> • A collection of shared online activities and quizzes in more than 150 categories • Templates for creating 16 different types of online activities* • Tools for creating online quizzes with up to eight question types • Quiz administration and reporting tools • Class Web pages • Online roster and grade book • Calendars and schedules • Online surveys 	<p><i>WebAssign</i> provides a flexible platform for different kinds of assignments and correction modes*:</p>
Mode of Delivery	Online	Online
Kinds of activities	<p>*Flashcards, matching, concentration (memory), word search, battleship, challenge board, columns, cloze exercises, hangman, jumbled words, ordered list, patterns, picture perfect, pop-ups, rags to riches (a quiz-show style trivia game), and scavenger hunt</p>	<p>*Multiple-choice questions, numeric tasks, and matching text with automated assessment and immediate feedback, open questions - where the correctors' comments are delivered in personal web pages for the students, and programming tasks - where an automated testing facility provides immediate feedback (giving the students the possibility to test and improve their solutions step by step); the last solution is assessed by a human corrector and delivered in a personal web page.</p>

Appendix 5

Profile of Assessment Electronic Authoring Tools as part of learning online environments

Profile of Assessment Electronic Authoring Tools as part of learning online environments

Name of Institution	<i>WebCT</i> is a course management online system, with a complete set of teaching and learning tools for course preparation, delivery, and management. http://www.webct.com/	The <i>Blackboard Learning System</i> is an online course management system http://www.blackboard.com/products/academic/ls/index.htm
Mode of Delivery	Online	Online
Kinds of activities	It includes quizzes and testing devices	It includes assessment and testing devices

Appendix 6

Text type inventory

Text type inventory

Text types used at each level in the test.

BENG 50	Account, Letter (introductory, holiday), Profile, Dialogue (party, weekend, studies, Conversation (food, drink, restaurant, Weather Report, Postcard, Email, Daily schedule/routine, Survey, Signs, etc.
BENG 80	Account, Letter, Letter of application, Dialogue, Conversation, Profile, Film reviews, Telephone message, Leaflet information, School report, Friendship corner, Diary, Biography: Greek-Cypriot composer, Internet chatting, Signs, etc.
BENG 90	Account, Dialogue, Letter –Recipe, Biography, Article, Website, Book review, Cook book review, Advertisement, Athlete’s portrait, Tourist brochure, etc.
BENG 100	Dialogue, Business letter, Cover letter, Article, Movie review, Portrait, etc.
ENGL 100	Conversation, Biography, Article (Internet, newspaper, magazine), Travel diary, etc.
ENGL 101	Conversation, Article (Internet, newspaper, magazine), Reference letter, Letter of condolence, Letter of complaint, Poll results, Literary extract, etc.

Appendix 7

Grammar inventory

Grammar Inventory

The following grammatical points have been used in the test at the various levels:

BENG 50	-definite article-indefinite article-partitive article-un/countable nouns-possessive pronouns-personal pronouns-demonstratives: this, these- adjectives, comparatives-prepositions: to, at, for, on, in-present simple-verb <i>to do</i> - verb <i>to be</i> -verb <i>to have</i> -present continuous (now)-past simple-past continuous- future simple-modal: can- shall, should, would, might, could-have to, can't, mustn't to, shouldn't -Wh: why, who, what, when, whose, where, which, what- how many, how much-a few, a, any, an, many
BENG 80	-irregular plural of nouns - -'s-present continuous 2-past simple: irregular: see, teach, go, sing, sleep, shake, wake up, ring- <i>used to</i> & time phrases (ago, in, at)-going to-near future-interrogative-interrogative with <i>did</i> -Present Perfect Simple-Present Tense after "if"&"when"-aren't, didn't, haven't-verbs and -time expressions: never, always, sometimes, rarely, ever, just-there is... there are...- Conditional. 1 & 2-any, much, more, many, some, enough-prepositions: in, in front of, between, over, etc.-that-parallelism-comparatives: adjective and adverbs; good; superlative-phrasal verb: go away / out / off / over / for
BENG 90	-ing as a noun -himself, herself -can and present tense -present and -ing - Past continuous with past simple-past simple and past simple-when and past simple -past perfect and simple past-Past Perfect continuous -future / perfect / continuous-future perfect continuous-passive-if and present with future simple - <i>If</i> with past tense and past conditional -ing: I like playing-I need to have-I wish I... past simple -phrasal verbs: run away from / after / against / out of / into; be up / off / over / on / down; go out-direct and indirect speech-which, for which, whose, who, whom, who to, who for, to whom-capable of and + ing-prepositions-regret to-why / because-compound sentences

BENG 100	<p>-irregular plural: knife, shelf, mouse, leaf, radio-past simple-parallelism-phrasal verbs: <u>bring</u> about / up / off / on / upon; <u>come</u> out / up / off / about / down to / up against / over; <u>fall</u> for / about / on / out / through; <u>get</u> at / down to / on/ across / off with / around to / out of / up / out and about / along; <u>go</u> in / in for / on / off / back on / away / over / under; <u>put</u> off / up on / in for / up / across; <u>run</u> to / down / away / up against / onto / on / into; <u>stand</u> by / up / in for / for / up to; <u>look</u> up/forward/into/around/out; <u>take</u> after/off /up/upon / over; <u>turn</u> in/off/down/up/ away; <u>pull</u> up/over/down/off; <u>throw</u> out/away/in /over;-sentence linkers and collocations:-expressions: for that reason, to tell you the truth, even so, by the way, on the other hand, at any rate, to start with, on top of that, by the way, having said that.-prepositions: above/on top of/over/on/up/before/next/aside/ amongst/ between/apart from/instead of/rather/in case of/regardless of-special superlatives: the best/the least-although/though/even/despite-clauses of time: nowadays/in a week/these days/a day-the last...-(five times) a week-worse and worse-prefixes, suffixes</p>
ENG 100	<p>-Phrasal verbs with 3 words: <u>come</u> down with / up with / along with / up to / round to / down against / up against / round against; <u>go</u> up with / down with / along to / round to / along with; <u>get</u> along with / around with / by to / away with / on with / out of / round to; <u>look</u> over to/round to/round on / around to / forward to / out to / down to / around on /. Over on / down on; <u>check</u> up to / round with / up on / round on / out on; <u>face</u> down with / up with / along with / face up to; <u>hold</u> along to/on with/up with/on to/up to; <u>catch</u> up with/along with/on to/up to/around to; <u>make</u> off with/round with/up to/off to/up with; <u>do</u> away to/away from/away with/around to; <u>grow</u> in to/out of/up to/out to/in of; <u>cut</u> up on /round on/down on/in on/out on; <u>put</u> on with/in on/down to/up with/in of;-of whom / whose / who / which / what-prepositions: popular with, prides itself on, on an individual basis, at reasonable prices, a long way from, to be shocked by, to put the blame on; to be envious of; to be robbed of your savings; to be out of action; to feel sorry for someone; to be weak at English; to reduce something to rubble; to be sorry for; to lose something on account of a mistake; to do something with the view of...; a delay to a schedule; to boast about oneself; I'd rather you didn't tell him...</p>
ENG 101	<p>-Prepositions: have confidence in; make notes on; lose grip on your job; prime oneself on coffee; to be on constant stand-by; to be under illusions; tuck T-shirt into the trousers; the castle dates back to the 16th century; the earthquake through her off her balance; to go about one's job; to ran into financial difficulties; To think of going...; to be accused of;-although, despite, -to believe in doing; to be hopeless in playing; to be anxious about; to try drinking; To remember seeing someone; To regret seeing someone; I can't help wondering; he forgot to do;-Prepositions: to play just for fun; to confide in someone; to mull over something; to fall into categories; to fall open;-Expressions: in addition, also, as yet, to make ends meet.</p>

Appendix 8

Vocabulary inventory

Vocabulary Inventory

The following lexical points have been used in the test at the various levels:

BEGN 50	BENG 80	BENG 90
<ul style="list-style-type: none"> -hungry, thirsty, -types of housing -look, see, observe, notice, watch -money, tip, reward, change, coin -family members -types of jobs (part-time, etc.) -types of schools -time expressions, -lots, many, much, plenty, too, etc. -types of work (job, etc.) -plates and cutlery -times of the day -academic words: certificate, degree, skill, quality, study, etc. -transport means -leave, leaf, lift, live, etc. -position prepositions -shops -rooms -furniture -feeling expressions -sports -remember, forget, forgive, etc. -letter, word, sentence, paragraph, essay -bag, back, bug, buck, etc. -whole, hole, etc. 	<ul style="list-style-type: none"> -numbers -leave, live, leaf, life -expressions of time -descriptive adjectives -message, massage, messenger, passage, collage -travel expressions: land, check in, book in, take, inquire -plans, plane, plain, pain, pity -say, tell, explain, advice, mention -chip, cheap, cheat, check -list, catalogue, menu, book, programme, -dessert, desert, dissertation, deserve, desire -ship, sheep, chip, cheap, sheets -plan, plane, plain, trip -situation, location, -post, credit, student, concession, travel card -cheque, check, cheek, chess, chest -Currency, current, exchange, monetary -toys, games -kilo, litre, kilometer, centimeter, dozen -paper, pepper, -look, look at, watch, see -complain, campaign, blame 	<ul style="list-style-type: none"> -shy, timid, embarrassing -actor, author, writer -photographer, photographic, photograph, photos, -expand, extend, enlarge, widen -on foot -memo, memory, memoir, momentum moment -expressions with <i>neck</i> -raise, rise -single, alone, individual, solitary, simple -think, thing, thick, thigh, thief, -want, desire, wish, crave -cinema expressions: admission, concession, booking fee, subtitle, matinee -bore, bored, boring, board, boredom -embarrassing, embarrassed, embarrassment -allows, permits, admits, forbids, considers

BENG 100	ENG 100	ENG 101
<p>-job / occupation / work / employment / profession -employee / assistant / attendant / clerk / officer -team / crew / party / staff / band -aim / goal / intention / ambition / success -manage / achieve / cope / fulfill / succeed -efficient / able / capable / effective / skilful -combine / compose / assemble / connect / accompany -chances / opportunities / possibilities / occasions / conditions -viewing / looking / seeing / observing / regarding -proofs / documents / certificates / evidences / licenses -explore / investigate / look up / search / look for -onlooker / spectacle / witness / audience / viewer -lively / living / vividly / live / vivid -memorise / recall / remind / recognize / remember -view / sighting / scene / vision / image -distant / far / remote / separate -rise / arise / raise / lift / go up -learn / teach / educate / instruct / study -temper / mood / manner / attitude / behaviour -habit / fashion / routine / trend / custom -popular / favourite / famous / notorious / noticeable</p>	<p>-derivative nouns, verbs, and adjectives such as: value, success, advisable, admissible, alcoholic, appreciative, art, benefit, calculation, child, classic, comfort, comprehensive, confident, considerable, cure, decide, dependant, entertain, event, exciting, forge, forget, fortune, fright, grateful, help, host, impressive, informative, interest, irritable, memory, mountain, oblige, respect, satisfied, simplicity, sympathy, terrific, thought, tired, waste. -synonyms of level, floor, storey, etc. -expression: to be out of order / printing / the question / print / publication - exact word: circulation, wildlife, making money, offence, scholarship, correspondent, verdict, the land, edition, evacuated; vulnerable, strain, tackle, claimed; lunacy, cause, endowed with, raise; acute, treating, provision, urged, afford; memoir, acclaimed, encounters, allowed; seek, ground, hurt, raising, reach; set me up, sweeps me off my feet, terrible, picking me up record.</p>	<p>-informal -immobile -irrational -irrelevant -disrespectful -illegible -shrugged -take out -tune in -misbehaved -impolite -fingers / knuckles / hands/ -fluttered, -twiddled -clench -deceitful -misconception -explanation -snowed under -he hasn't the foggiest idea -a stormy relationship -a sunny smile -storms of protests -to shower someone with gifts -to be caught in a hail of bullets -frosty manner -to bring the traffic to a standstill -recruitment -corporate -tempt -incentives -applying -dermatologist -scheduled -incurred -reimburse -launch -blunders -conventions -premise -core -conduct a survey -respondents</p>

Appendix 9

Topic Inventory

Topic Inventories

BEGN 50	BENG 80	BENG 90		
<ul style="list-style-type: none"> -personal profile -family -Studying overseas: in Cyprus -London weather report -my life -Holidays -correspondence -living overseas -food -drink -shopping -pets -sport -travel -relationships -college life -work -outings -computer -the Internet -house -smoking -languages -leisure activities -Christmas -weekend -transport -hotel and tourism -health -meal -time -health -summer holiday -buildings 	<ul style="list-style-type: none"> -house -family -holiday -films -sport -health -school -daily routine -John Lennon -shopping: clothes -sport car washing / parking, -drinks -smoking -trespassing -litter -animal -shops -mobile phone -friends -savings -test -earthquake -courses: Maths, English -professions -savings -daily schedule -working in a hotel -driving -going to college -British Airways -raccs -work -cooking -films -cards (post, credit, student concession, travel) -meal -learning English -friends -news -Charles Dickens 	<ul style="list-style-type: none"> -Indira Gandhi -the ideal woman -Phar Lap's life -Indian cuisine -welcome to the Cook's Garden -Seven Centuries of English Cooking -Body Dome -The Way to Happiness (book) -A place I like Very much: Central Park -Learning Japanese shopping -lunch -tourism -retirement -cinema -meetings -free time -sport -being sick -exam passing -car accident / driving / buying 	<ul style="list-style-type: none"> -relationships -cooking -Elton John -Simpson Desert in Australia -places in town -lecturers -missing the train -A night in -Printing problems -Student Affairs -Advice to New University Students -Moving overseas -Mobile Ringtones or Picture Messages -Chatting on the Internet -Cats and UK Gardens -the kindergarten teacher -Smoking Kills -History of Soccer and Violence -A parent's guide to Teenage Parties -music -going out with friends 	<ul style="list-style-type: none"> -photos -going to college -chores and house work -memo sending -romantic dinner -careful driving -dining -visiting friends -sharing a few home truths -Nana Mouskouri in concert -holiday recommendations -a French friend -a weekend in Athens -Tourist killed in fall -Catby Freeman -Sightseeing Cruises -weekend gardener -I'm hopeless in English -Pocket money -holiday camp

BENG 100		ENG 100			ENG 101		
-movie review	-exams	-family weekend	-penalty points for mobile use	-fishing books	-research papers	-technologies	-teachers
-Will Smith	-smashing a window	-holidays	-wine making	-food company decisions	-hosting	-taxi ride	-letter of condolence
-Depressing Vancouver	-university qualifications	-Bush walking in New Zealand	-the Lost Art of Letter-writing	-Cyprus tourism, hospitality, food, land prices, and scenery	-discases	-the top five things people hate most	-music therapy
-Weather skateboard history	-and improved job prospects	-medical profession	-London First Impressions	-organized events	-events	-about work	-the war against war
-table setting	-men looking after children	-Learning English	-Mike Judge the animator	-research	-conferences	-Painting the family Pet	-correspondence
-scientific experiments	-weather	-car parks	-sickness	-robbing	-fancy-dress	-the Olive and Culture	-Facts about working women
-autumn collection	-having a baby	-schools	-teacher problems	-bombing	-costumes	-morality and responsibility	-space tourism
-radio	-mythology	-sales department	-board of directors	-accounting	-buildings	-in society	-clothing
-sport	-therapies	-leadership	-terrible behaviour	-flat renovation	-books	-note taking	-handicapped
-keeping promises	-religions	-mobile phones and computers	-sport book reading	-scheduled	-magazines	-morning coffee	-medical insurance
-peoples' characters	-work	-job opportunities	-Cyprus company decisions	-professional babysitting	-wildlife	-language learning	-premium bonds
-meetings	-culinary arts	for the young and the inexperienced	-snobbish people	-property boom	-earnings	-health and medicine	-company cars
-shopping	-assignments	-Nelson	-solution to problems	-The Winepress	-war reports	-taking	-biology
-buying a car	-bombing	-football hooligans	-spending money	-You are what you wear	-hurricane	-police	-relationships
-credit cards	-money	-son and eye effect	-relationship	-Men to learn from	-work stress	-life guard	-student fees
-hotel jobs	-car problems	-diet	-computers	-postgraduate studies	-fund raising	-jobs	-job seekers
-student and teacher relationships	-drunks	-village living	-children clothing	-drink driving	-elderly care	-physical appearance	-choose moral
-visiting London	-work regulations	-sun set	-overweight people	-arts	-transforming a book into a film	-castle history	-letter of money
-inheritance	-diary writing	-final examinations	-friends	-marriage	-Mark Twain	-earthquake	-complaint
-shop assistant and customer relationships	-manners	-employee behaviour	-politicians' speeches	-music	-Blind date	-share prices	-netiquette
-commuting	-book reading	-Jack the Ripper	-work	-cars		-Christmas in Spain	-British business
-drink choice	-London house of parliament	-Business proposal				-sport	-travelers love their mobile phones
-driving through amber light	-going to college	-cover letter				-shopping	
-printer problems	-thankingiving	-Roddick to play in Athens Olympics					
-sport	-The problem of languages	-My desire to study overseas took me to Cyprus					
	-business letter	-friendly dental care					

Appendix 10

Settings inventory

The following settings are reflected in the test at the various levels:

BEGN 50	-United Kingdom (London, Manchester, Leeds, Dublin, Aberdeen-Scotland, - Cyprus -Australia: Melbourne-United States: New York, -Japan-Russia-at home-a party- at a restaurant-at a cafeteria-bank-friends-cinema-street
BENG 80	-Cyprus-New Caledonia-United States: Los Angeles, New York, South Virginia, -United Kingdom: England, Edinburgh-Australia: Melbourne-New Zealand- Germany-Turkey-France-Bangkok: hotels-Paris -Budapest-magazine-shop-street- car wash-car park-pub-university union building-park-ATM outlet-escalators- loading zone-horse riding trail-hotel-neighbourhood -lecture amphitheatre-high school-restaurant-English class-kitchen-at the dentist-meeting-club-beach- Internet-at the doctor's
BENG 90	-India-Australia-England: London: Central Park; UK gardens-Alaska-New York- Alps
BENG 100	-England-Los Angeles-Rome-New Zealand-India
ENG 100	-U.K.-London-Cyprus-USA (San Francisco, etc.)-Asia-Australia-work
ENG 101	-Spain-Cyprus: Limassol-work

Appendix 11

Intercollege NEPTON Writing Task Sample

INTERCOLLEGE ENGLISH PLACEMENT TEST

FALL SEMESTER 2004

GROUP A

Writing Task

Write an essay of about 120 words on ONE of the following topics:

1. Write a description of yourself as you think your best friend sees you.
2. Write a story that begins, "Nobody was pleased to hear that Aunt Beatrice was coming for a visit."
3. Write a letter to a friend telling him / her how much you are enjoying your holiday abroad.
4. "Mothers with small children should not go out to work." Discuss.

Spend some time planning and revising your essay.

When you finish, count the number of words and write the figure at the bottom.

INTERCOLLEGE ENGLISH PLACEMENT TEST

FALL SEMESTER 2004

GROUP B

Writing Task

Write an essay of about 120 words on ONE of the following topics:

1. Describe the view from your window.
2. Write a story about a broken promise.
3. Write a letter refusing an invitation to a party and explaining why you cannot go.
4. "Animal experimentation is evil and should be banned." Discuss.

Spend some time planning and revising your essay.

When you finish, count the number of words and write the figure at the bottom.

Appendix 12

Criteria for marking the writing task

Criteria for marking the writing task

- (a) Natural use of language (mechanical ability: The ability to correctly use punctuation, spelling, capitalization, etc.)
- (b) Range of vocabulary and structure (grammatical and lexical ability)
- (c) Content organization (Stylistic skills. The ability to use sentences and paragraphs appropriately organizational skills. The ability to organize written work according to the conventions of English, including the order and selection of material)
- (d) Clear communication (Judgments of appropriateness: The ability to make judgments about what is appropriate depending on the task, the purpose of the writing, and the audience)
- (e) Completion of task

Appendix 13

Calculation of final placement

Calculation of final placement

- (a) When NEPTON and the written task are of the same level, e.g. BENG 90 the test-taker is placed in the BENG 90 level course.
- (b) When NEPTON and the written task are different by one level, e.g. BENG 90 / BENG 100, (the test being pro-student) the test-taker is placed in the BENG 100 level course, in other words the two placements are averaged.
- (c) When NEPTON and the written task are different by two levels, e.g. BENG 90 / ENG 100, the test-taker is placed in the BENG 100 level course.
- (d) When NEPTON and the written task are different by three levels, e.g. BENG 90 / ENG 101, the written task is looked at more carefully and decision is taken on that basis. This has proved to be a very rare case.

The final NEPTON score calculates in the above way and converted into an Intercollege Placement English Competence level.

The final placement is the combination of both objective and subjective results. The placement boundaries are determined for the test as a whole...p 169

Appendix 14

Letter to Faculty (a)

Salomi Papadima-Sophocleous

Senior Lecturer, Languages Department
Intercollege, 92 Ayias Phylaxeos Str.
P.O.Box 51604
3507 Limassol, CYPRUS

☎ College: + (357) 25 38 11 80

Office: +357 25 73 31 54

☎ Fax: +357 25 38 69 82

Mobile: + (357) 99 94 89 01

✉ email: s.sophocleous@lim.intercollege.ac.cy

To: Intercollege English Faculty
Subject: New English Placement Test Online (NEPTON)
Sent: 3rd December 2003

Dear colleague,

As you probably know, I have been working on NEPTON for some time now. I have completed the background research (Second Language –L2- learning, testing, electronic L2 testing, English Placement tests, Adaptive testing, statistics, etc.) and am about to complete the writing of test items. NEPTON is an electronic test, which will randomly select each item from a big pool of test items. The pool consists of more than 1400 items, divided in 6 levels (BENG 50, BENG 80, BENG 90, BENG 100, ENGL 100 and ENGL 101) in the following areas:

Structure: Sentence based with a blank and 5 multiple choices.

Vocabulary: Sentence based with a blank and 5 multiple choices.

Structure: Text with five test items and a dropdown menu of 5 choices.

Vocabulary: Text with five test items and a dropdown menu of 5 choices.

Reading Comprehension and visual.: match sign with sign meaning / description (for BENG 50 and BENG 80) and text

Reading Comprehension: Text and multiple-choice questions

So far, I had the pleasure of working closely with Chris Alexander for the content, and Dr Dmitry Apraksin for the technological aspect of NEPTON.

We are now at a point where the developed test items need to be moderated by experts, in other words, to be looked at by people like you for the following purposes:

- (a) Editing
- (b) Confirmation of level
- (c) Any other comment

A number of colleagues have already expressed interest in participating in this NEPTON phase. I believe it is a chance to familiarise yourselves with it and contribute in any possible way with your knowledge, experience and expertise.

I'm therefore sending an open invitation to you all to participate in this phase. If you are interested and have the time, please let me know and I'll send you the material early next week. This process needs to be finished before Christmas (preferably in two weeks). The next phase will be for me to edit the material and incorporate your suggestions, and then trial it with students. Trials need to be done early next year.

Looking forward to hearing from you,

Salomi

Appendix 15

Letter to faculty (b)

Salomi Papadima-Sophocleous

Senior Lecturer, Languages Department
Intercollege, 92 Ayias Phylaxeos Str.
P.O.Box 51604
3507 Limassol, CYPRUS

☎ College: + (357) 25 38 11 80

Office: +357 25 73 31 54

☎ Fax: +357 25 38 69 82

Mobile: + (357) 99 94 89 01

✉ email: s.sophocleous@lim.intercollege.ac.cy

To: Intercollege English Faculty
Subject: New English Placement Test Online (NEPTON)
Sent: 15th February 2004

Dear colleagues,

Thanks to all of you who have contributed to the NEPTON validation process during, I know, a difficult exam/correction and end of semester period (November 2003-January 2004.) 44% of fulltime English faculty from all three Intercollege campuses has contributed to this process. Thanks for the valuable editing, the help in establishing the levels for each test item or activity, and all the extra comments made, contributing to the improvement of the test items, before the trial of NEPTON. I very much appreciated the encouraging words, as well as the critical comments, which are very much needed in a process such as this. Thanks to you all, something like 1400 test items covering all six levels (BENG and ENGL) have been validated, thanks to you.

I would also like to thank those colleagues who expressed interest in the whole project, and would have liked to have participated, but too many commitments prevented them from doing so at the time. If they are still interested and have the time during Spring semester, they would be able to participate in other NEPTON phases.

I would like to take the opportunity to particularly thank a few people:

- (a) Intercollege administration in all campuses for their support;
- (b) Mr. Emiliou Solomou for his continuous support;
- (c) Dr James Leigh for his continuous support, encouragement and advice;
- (d) Chris Alexander for his collaboration with some of the writing, with editing and with his general contribution in various other aspects of the project;
- (e) Dmitry Apraksin for his expertise in the technological side of the NEPTON project and his contribution and involvement in the technical side of NEPTON construction.

I have edited the material and incorporate your suggestions. Shortly, NEPTON will be trialled by both faculty and students, both in pen & paper as well as in electronic form. You are welcome to trial it in any for you like, or both, and give us feedback. If you are interested in the trial phase and have the time, please let me know, so that I can make the necessary arrangements.

Looking forward to hearing from you,
Salomi.

Appendix 16

Test-taker post field-test questionnaire

Test-taker post field-test questionnaire

After exiting from the test, please complete the following questionnaire, by ticking the category that most accurately reflects your view. There is also an opportunity for providing more detailed comments

	Strongly <i>disagree</i>	Disagree	Neutral	Agree	Strongly <i>agree</i>
1. The instructions were clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. I found the activities which required filling the blank in a sentence manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. I found the activities which required choosing the best possible answer from a menu in a text manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I found the reading comprehension multiple-choice texts manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. I found the activities which required matching each part of a text with an appropriate heading manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. There was enough variety of activities.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. The length of the test was appropriate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. I found the topics interesting.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. The activity layout was clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. The visuals were appropriate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. Before the test I felt tensed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. I felt comfortable during the test.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Very <i>long time</i>	Neutral <i>time</i>	Little <i>time</i>	Very <i>little</i>	
13. How much time did you spend on this test?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
14. The amount of time for this test was	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
15. The test items were	<i>Too many</i>	<i>Many</i>	<i>Neutral</i>	<i>Few</i>	<i>Very few</i>
16. Tell what you did <i>not</i> like about NEPTON (if anything)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
.....					
.....					
17. Tell what you liked <i>most</i> about NEPTON (if anything)					
.....					
.....					
18. Tell how NEPTON can be improved.					
.....					
.....					
19. Note any further comments:					
.....					
.....					

Appendix 17

Student pre-test NEPTON questionnaire

Intercollege NEPTON STUDENT PRE-TEST QUESTIONNAIRE

Before taking the NEPTON test, please complete the following questionnaire.

This questionnaire is part of the validation process of the New English Placement Test Online, designed by Intercollege and the Language Faculty. The aim of NEPTON is to improve the academic quality and efficiency of the college English Placement Programme. The aim of the questionnaire is to receive feedback from the students in order to continuously improve it. The questionnaire is entirely anonymous.

Please answer all the questions by ticking the category that most accurately reflects your view. There is also an opportunity for providing more detailed comments.

PART A

1. **Gender** MALE: FEMALE:

2. **AGE on your last birthday (circle)**
 17 18 19 20 21 22 23 24 25 26 27
 28 29 30 31 32 33 34 35 36 37 38
 39 40 Other, please specify:

3. **First language** (e.g. English, Greek, Russian, Chinese, Pakistani, etc.)

4. **Country of origin** (e.g. Cyprus, Russia, China, Cameroon, Belarus, etc.)

5. **Years of English studies (circle)**
 1 2 3 4 5 6 7 8 9 10 11
 12
 If other, please specify:

6. **English Qualifications, if any:** NONE
 IELTS-level
 TOEFL-level
 If other, please specify:

10. **Intended Degree program at Intercollege**

PART B Previous Basic Computer Experience

	<i>Very uncomfortable</i>	<i>Uncomfortable</i>	<i>Neutral</i>	<i>Comfortable</i>	<i>Very comfortable</i>
1. Comfortable with the use of the mouse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Can click	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Can scroll up and down the screen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Comfortable with the use of technology	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<i>Never</i>	<i>Rarely</i>	<i>Neutral</i>	<i>Often</i>	<i>Always</i>
5. Use computer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Use Internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Use Email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Use chart	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 18

Student post-test NEPTON questionnaire

Intercollege NEPTON STUDENT POST-TEST QUESTIONNAIRE

After exiting from the test, please complete the following questionnaire, by ticking the category that most accurately reflects your view. There is also an opportunity for providing more detailed comments

	Strongly <i>disagree</i>	Disagree	Neutral	Agree	Strongly <i>agree</i>
1. The Tutorial helped me.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Although I had some previous computer experience, the Tutorial helped me	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The instructions were clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. I found the structure/vocabulary (sentence-based) section manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. I found the structure/vocabulary (text and dropdown menu selection) section manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. I found the reading comprehension sign section manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. I found the reading comprehension (text and multiple-choice) section manageable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. I found the topics of the electronic part of the test interesting.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. There was enough variety of activities.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. The length of the electronic test was appropriate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. The computer screen was clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. The activity layout was clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. The visuals were appropriate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. Before the electronic test I felt tensed.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. I felt comfortable during the electronic test.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. I found the topics of the written part interesting.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. The length of the written test was appropriate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Very <i>long</i>	Long	Neutral	Little	Very <i>little</i>
18. How much time did you spend on this test as a whole (electronic AND written)?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. The amount of time for this test was	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. The test items were	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. Tell what you did <i>not</i> like about NEPTON (if anything)					
22. Tell what you liked <i>most</i> about NEPTON (if anything)					
23. Tell how NEPTON can be improved.					
24. Note any further comments:					

Appendix 19

Experts NEPTON Test Specification questionnaire

EXPERTS NEPTON (& NEPTON Paper)PRE-TEST QUESTIONNAIRE

After studying the NEPTON Test Specifications and the NEPTON / NEPTON (P) tests, please complete the following questionnaire.

This questionnaire is part of the content validation process of the New English Placement Test Online, designed for Intercollege and the Language Department. The aim of NEPTON is to improve the academic quality and efficiency of the college English Placement Programme. The aim of the questionnaire is to receive feedback from English Language experts in order to establish its content validity. The questionnaire is entirely anonymous. Please answer all the questions of this questionnaire by ticking the category that most accurately reflects your view AFTER looking at the *NEPTON Test Specifications* and *the NEPTON / NEPTON (P)*.

PART A

- | | | |
|--------------------------------------|-----------|-------------|
| 1. Gender | MALE..... | FEMALE..... |
| 2. First language: | | |
| 3. Other languages: | | |
| 4. Country of Origin: | | |
| 5. Years you have taught each level: | | |
| BENG 50 | | |
| BENG 80 | | |
| BENG 90 | | |
| BENG 100 | | |
| ENGL 100 | | |
| ENGL101 | | |

PART B Previous Basic Computer Experience

	<i>Very uncomfortable</i>	<i>Uncomfortable</i>	<i>Neutral</i>	<i>Comfortable</i>	<i>Very comfortable</i>
1. Comfortable with the use of the mouse	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Can click	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Can scroll up and down the screen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Comfortable with the use of technology	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	<i>Never</i>	<i>Rarely</i>	<i>Neutral</i>	<i>Often</i>	<i>Always</i>
5. Use computer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Use Internet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Use Email	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Use chat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

EXPERTS NEPTON (& NEPTON Paper)POST-TEST QUESTIONNAIRE

This questionnaire is part of the validation process of the New English Placement Test Online, designed for Intercollege and the Language Department. The aim of NEPTON is to improve the academic quality and efficiency of the college English Placement Programme. The aim of the questionnaire is to receive feedback from English Language experts in order to establish its content validity. The questionnaire is entirely anonymous. Take / look at both the NEPTON and NEPTON (P) and compare it with the *NEPTON Test Specifications*. Then answer all the questions by ticking the category that most accurately reflects your view. There is also an opportunity for providing more detailed comments.

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1. NEPTON meets the requirements of an English placement test for the needs of Intercollege students.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. NEPTON accurately represents the content of Intercollege English Language Programme at the various levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The grammatical points represent the structure content of the respective levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The lexical points represent the vocabulary content of the respective levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. The reading comprehension sections represent the level of understanding of reading texts of the respective levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. The topics are appropriate for the respective levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. The topics are interesting.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. The instructions are clear.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. The test reflects students' ethnic background.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. The test reflects settings test-takers would come across in their studies.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. The test reflects settings and contexts in English-speaking settings.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. The test reflects the setting where test-takers will be studying	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. There is sufficient variety in text types, appropriate for the respective levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. NEPTON test items generally target the appropriate English level.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 15. The kinds of activities are appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 16. The activity types cover variety of learning styles. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 17. The length of the test is appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 18. The time given to the test-takers to take the test is appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 19. There is time flexibility to cater for test-takers' differences in time needs. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 20. NEPTON secures test and item security and item disclosure | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 21. Item selection criteria are appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 22. The levels of performance criteria are appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 23. The scoring system is satisfactory. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 24. The test results are appropriately reported to the test-takers. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 25. The NEPTON test booklet helps test-takers familiarise themselves with the test; the test format, the test layout, and the test items. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

NEPTON

- | | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 26. The Tutorial helps test-takers familiarise themselves with the computer skills they need to take the NEPTON test. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 27. The NEPTON tutorial helps students even though they may have some previous experience with computers. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 28. The Tutorial helps students familiarise themselves with the NEPTON test format. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 29. The Tutorial helps students familiarise themselves with the NEPTON test layout. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 30. The Tutorial helps students familiarise themselves with the NEPTON test items. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 31. The computer screen is clear. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 32. The activity layout is clear. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 33. The visuals are appropriate. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

34. Students are given opportunities to control the test
(Review items, going backwards and towards, etc.)

NEPTON (P)

35. The pen-and-paper test layout is clear.

36. The activity layout is clear.

37. The visuals are appropriate.

38. There is satisfactory provision of test familiarity:
(a) Test format,

(b) Types of activities.

39. Students are given opportunities to control the test
(Review items, going backwards and towards, etc.)

NEPTON and NEPTON (P)

40. NEPTON and NEPTON (P) are compatible.

41. The Test is compatible with the Test Specifications.

42. NEPTON is more efficient in content than
the previous EPT.

43. NEPTON is more efficient in delivery than
the previous EPT.

44. NEPTON is more efficient in administration
than the previous EPT.

45. NEPTON is more efficient in time than the previous EPT

46. NEPTON would be more efficient in cost than
the previous EPT.

47. NEPTON would be more efficient in placement
than the previous EPT.

48. NEPTON is better in format than the previous EPT.

49. NEPTON has brought change in EPT practices.

50. NEPTON has brought improvement in EPT practices.

51. NEPTON has brought efficiency in EPT practices.

Open-ended:

52. Tell what you did *not* like about NEPTON / NEPTON (P) (if anything)

.....
.....
.....
.....
.....

53. Tell what you liked *most* about NEPTON / NEPTON (P) (if anything)

.....
.....
.....
.....
.....

54. Tell how NEPTON / NEPTON (P) can be improved. Mention specific areas, if possible.

.....
.....
.....
.....
.....
.....
.....
.....
.....

55. Note any further comments

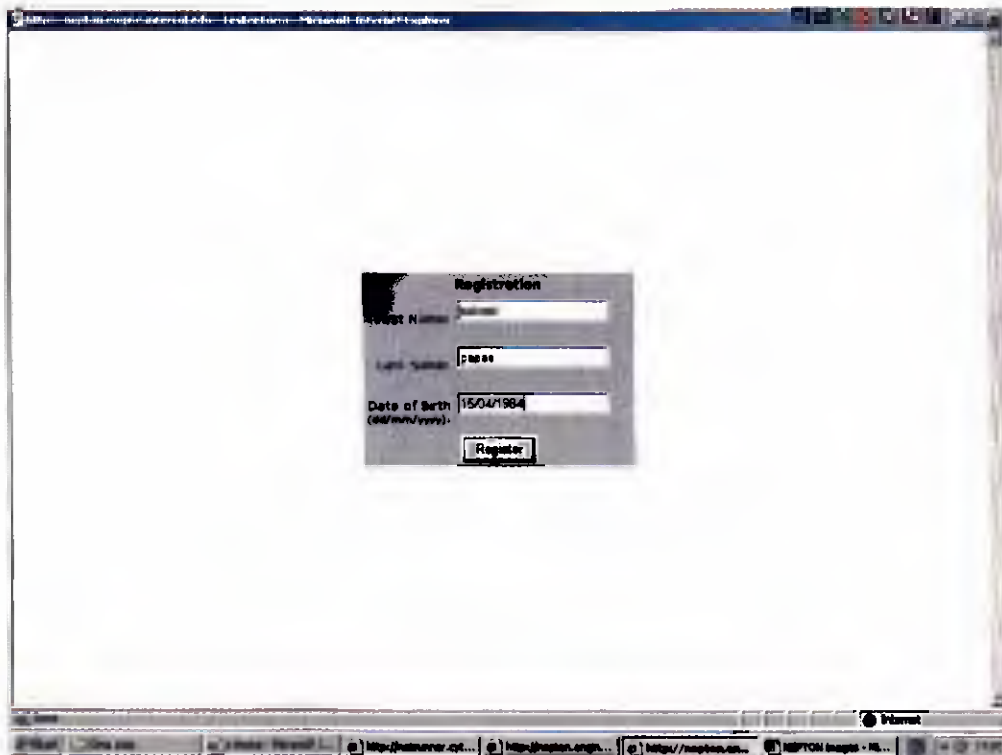
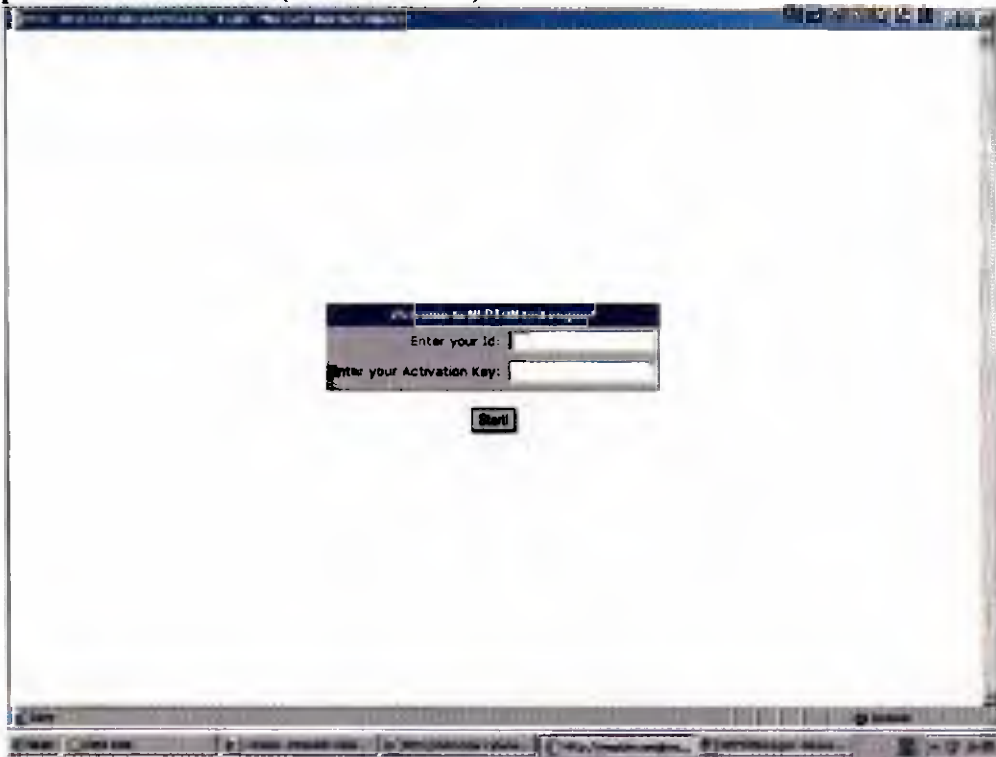
.....
.....
.....
.....
.....

Thank you
for accepting to contribute to the content validation of
NEPTON / NEPTON (P)

Appendix 20

Printed samples of the NEPTON (electronic test) and activities

Samples of the NEPTON (electronic test) and activities



http://nepton.mcgill.ca/... NEPTON Engine - Microsoft Internet Explorer

Questions: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

Choose an answer

She TV all night.

- looked
- saw
- observed
- noticed
- watched

papas, salomi Remaining Time: 79 min. Finish Test


Done Internet

Start Drive data i-choice - Microsoft L... http://netrunner.ct... http://nepton.engh... http://nepton.en... NEPTON Image - M... 17:03

http://nepton.mcgill.ca/... NEPTON Engine - Microsoft Internet Explorer

Questions: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33

Where would you see this notice?



- In a night club
- In a hospital
- at a petrol station

nepton, salomi Remaining Time: 79 min. Finish Test

Start Drive data i-choice - Microsoft L... http://netrunner.ct... http://nepton.engh... http://nepton.en... NEPTON Image - M... 17:03

Questions: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

Letter to a Friend

Read the passage about Letter to a Friend and choose the correct answer from the dropdown menu selection.

Dear Sean,

It was good to see you in Manchester last week. Unfortunately I had a lot of work every day, so I didn't have much time to see you more. I didn't have the time to visit many of the interesting places [21] are famous in Manchester either. However, I [22] many useful contacts so I hope my company can do a lot of business in the United Kingdom in the future. I know you don't have much free time during your visits in Cyprus, but please try to come and see us next time you are here. There are a lot of interesting places in Cyprus, too. Do you have many things to do during your next visit? You [23] come from the beginning of summer because the weather is terrific and when you [24] with your work we can go to the beach. My [25] organises bush walking and he can take us to some beautiful [26]

Looking forward to seeing you then.

Love,
Petros.

Answers: [21] [22] [23] [24] [25] [26]

Answers dropdown menu:
 wife brother
 wife's brother
 wives brother
 wife's brother
 wife brothers

Remaining Time: 77 min. Finish Test


Questions: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35

4,000 years of wine making

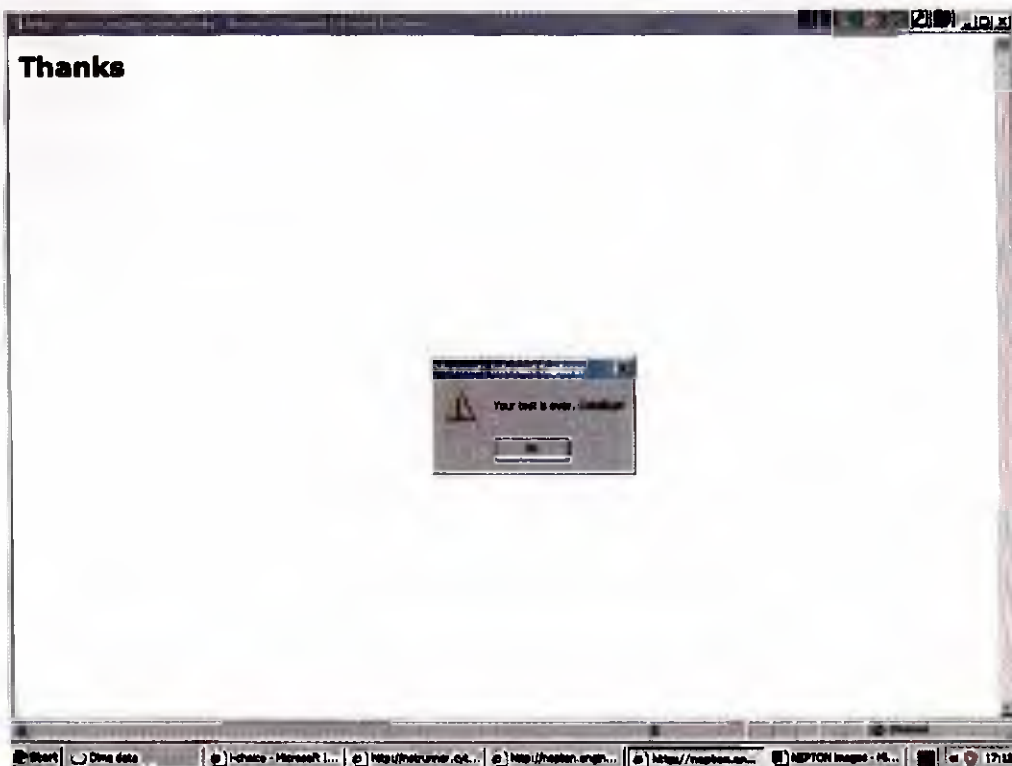
Read the text and then answer the multiple choice questions.

CYPRUS is an island of incomparable ancient sites and a wealth of romantic legends. It also occupies a unique place in the history of wines and the development of the wine trade. In fact our wine traditions go back over 4,000 years, virtually to the dawn of civilisation. One of our wines - **Commandaria** - is acknowledged to be the oldest 'in the world' and probably the first-ever to be given an 'Appellation of Origin.'

From the very beginning, we have shared our rich heritage with other nations. We supplied the Pharaohs of Egypt, and our wines were in great demand from the ancient Greeks and Romans. Recent excavations in Paphos have yielded old coins with a representation of a vine on one side. These date from the 3rd century BC and indicate that even in those days, wine was a major source of the Island's wealth. An old Cypriot chalice of the 6th century BC recommends its user to "Be happy and drink well", signifying the quality of wine in the life of Cyprus. And mosaics discovered in ancient sites on the Island bear witness to the importance of Cyprus grapes and the wines they produced.

The reason for our 40 long centuries of success is the area coloured green on the map. It lies to the northern slopes of the Troodos mountains. 

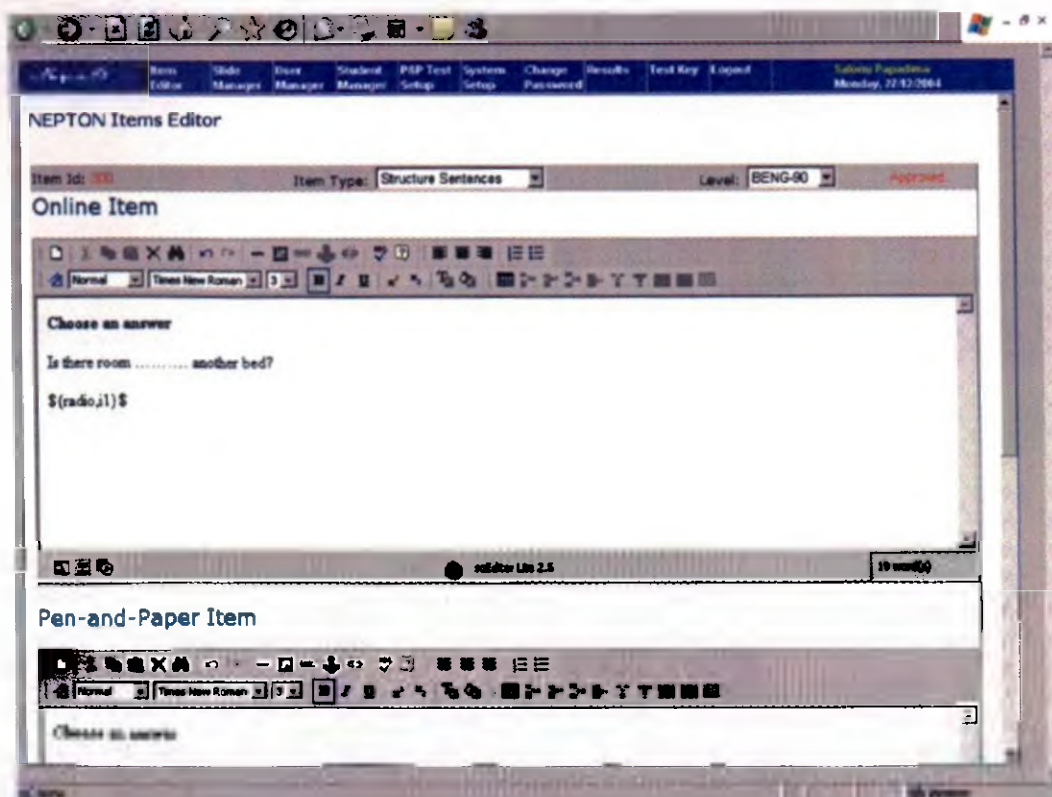
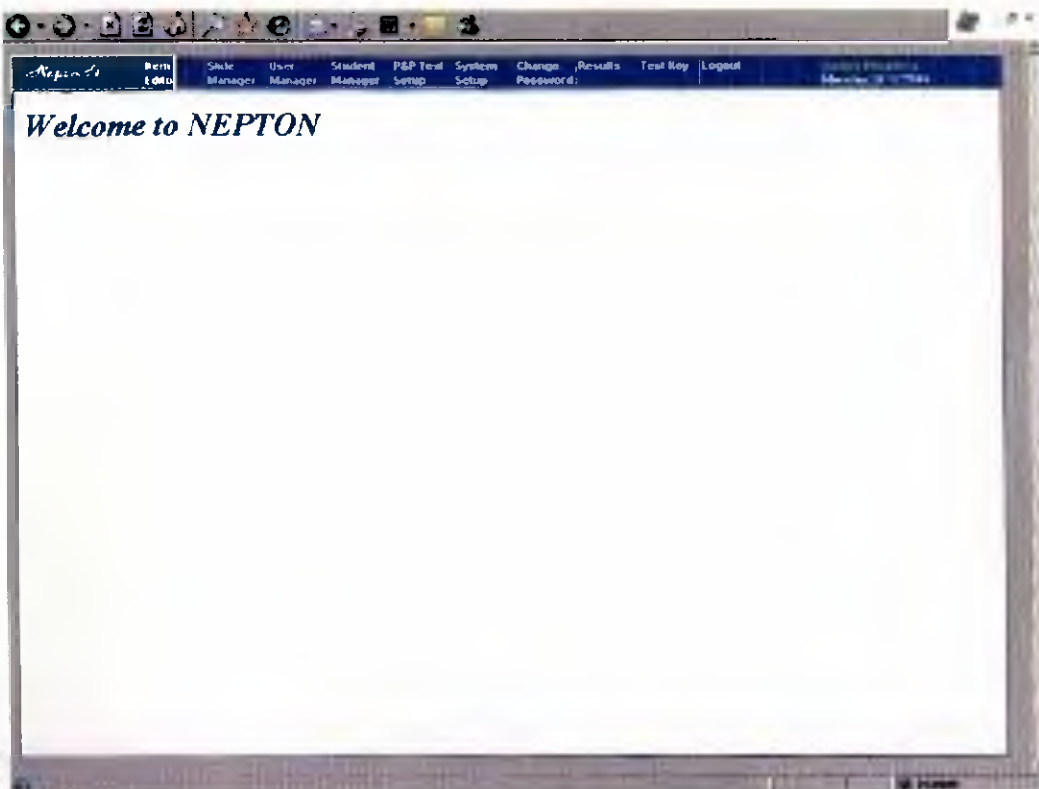
- The history of Cyprus is as long as
 - the history of wines and wine trade.
 - the history of ancient sites.
 - the history of the Island of Commandaria.
 - ancient romantic legends.
 - the 'appellation of Origin'.
- Pharaohs, Greeks and Romans
 - shared their wine with the Ancient Cypriots.
 - bought wine from the Ancient Cypriots.
 - sold their wine to Ancient Cypriots.
 - supplied the Ancient Cypriots with wine.
 - sold wine in Paphos.
- Mosaics discovered in ancient sites reveal the great importance of
 - coins in Ancient Cyprus.
 - vines and wines in Cyprus.
 - old Cypriot chalices in the 6th century BC.
 - recent excavations in Paphos.
 - of the island's wealth in the 3rd century BC.

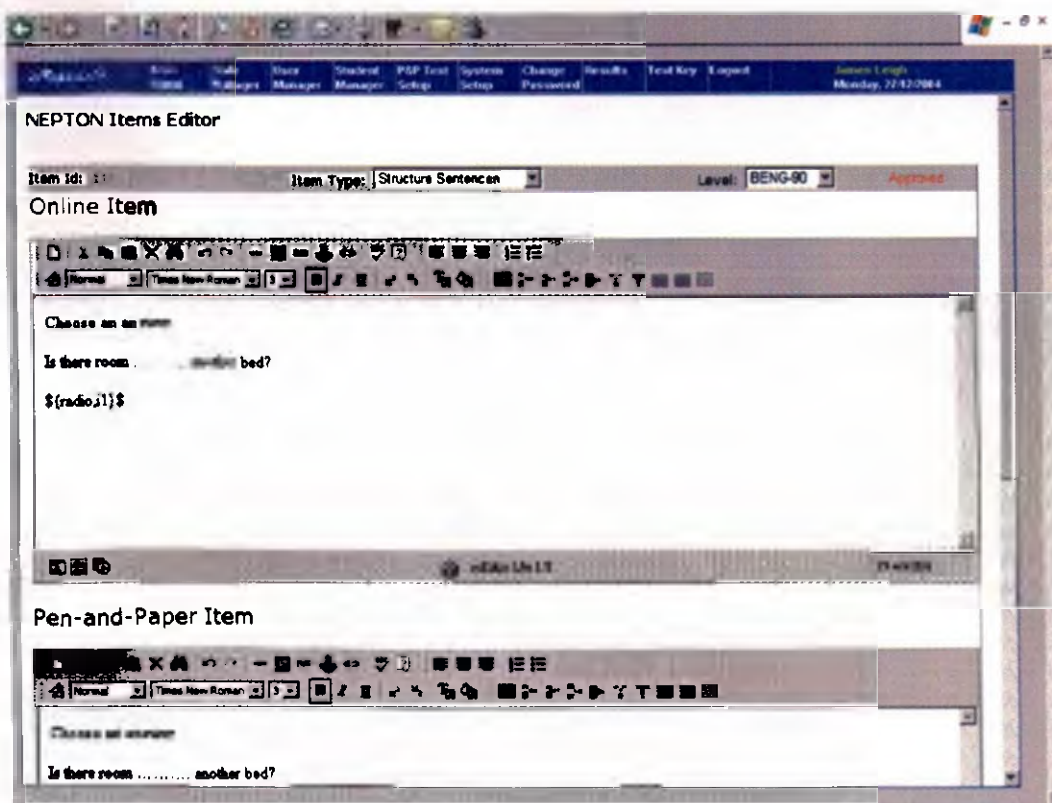
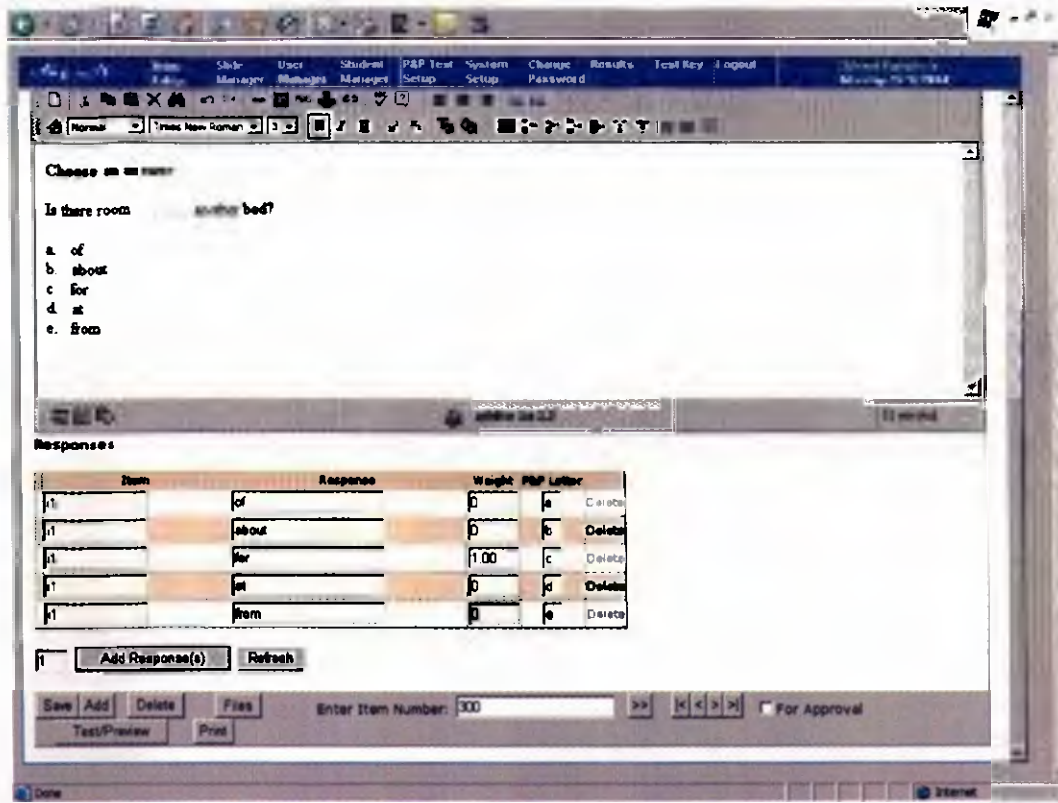


Appendix 21

Samples of the NEPTON (electronic test) editor environment

Samples of the NEPTON (electronic test) editor environment





CTD ID	Name	Country	DOB	Age	Language 1	Language 2	Language 3	Time 1	Time 2	Time 3
CTD40714	WEGDEL	GRECE	17/09/1984	20/09/2004	ENGL-101	ENGL-100	ENGL-101	08:50:34	14:00:00	15:02:09
CTD40101	MARIE-CLAIRE	GUIONNET	19/10/1983	20/08/2004	ENGL-101	ENGL-101	ENGL-101	14:40:00	15:02:09	00:22:08
CTD40743	GULSUN	GULER	24/10/1986	20/09/2004	ENGL-101	ENGL-100	ENGL-101	11:30:16	12:16:45	00:46:29
CTD40800	CHRISTODOULOS	HADJICHRISTODOULOU	11/01/1985	22/09/2004	ENGL-101	ENGL-100	ENGL-100	14:32:28	15:21:24	00:48:55
CTD40383	ELENI	HADJIGIANNAKI	07/12/1983	20/09/2004	ENGL-101	BENGL-100	BENGL-100	08:55:27	09:20:30	00:25:02
CTD40511	TASOS	HADJIPAVLIS	07/08/1980	14/09/2004	ENGL-101	ENGL-100	ENGL-100	15:30:43	16:08:52	00:26:09
CTD40991	PRADDEEP SINGH	HANDA	24/03/1986	07/10/2004	ENGL-101	BENGL-100	ENGL-100	14:23:03	15:22:36	00:59:32
CTD40807	Nadia	Hawatmah	23/02/1984	22/09/2004	ENGL-101	ENGL-100	ENGL-101	14:22:15	14:57:24	00:26:08
CTD40765	CHARITHEA	HERODOTOU	20/01/1987	20/09/2004	ENGL-101	BENGL-100	ENGL-100	11:32:22	12:00:28	00:28:05
CTD40598	KYRIAKOS	IOANNIDES	27/07/1984	20/09/2004	ENGL-101	ENGL-100	ENGL-101	08:33:59	09:27:22	00:53:22
CTD40711	IRINI	IOANNOU	23/01/1987	20/09/2004	ENGL-101	BENGL-100	ENGL-100	08:35:14	09:32:32	00:57:17
CTD41038	senha	laneti	27/09/1987	14/10/2004	ENGL-101	BENGL-90	ENGL-100	14:25:03	15:07:36	00:42:32
CTD40715	ELENA ATHENA	KARAOLI	31/07/1986	20/09/2004	ENGL-101	ENGL-100	ENGL-101	08:50:49	09:57:40	01:06:51
CTD40420	DUSAN	KARIC	25/02/1987	07/09/2004	ENGL-101	BENGL-100	ENGL-101	14:27:51	14:43:52	00:16:01
CTD40795	ELENI	KARIOLI	04/06/2004	22/09/2004	ENGL-101	ENGL-100	ENGL-101	08:36:24	09:24:03	00:47:38
CTD40301	CHRISOVALANTI	KATTOU	04/05/1986	01/09/2004	ENGL-101	ENGL-101	ENGL-101	08:30:08	09:18:04	00:47:58
CTD40476	LILIT	KAZANJIAN	27/07/1986	10/09/2004	ENGL-101	ENGL-100	ENGL-101	08:29:14	09:09:45	00:40:30
CTD40159	NICHOLAS	KAZANTZIS	15/01/1984	26/08/2004	ENGL-101			08:01:09	08:34:51	00:33:41

Rank	Time	Result	Count
1	408	Correct	1
2	420	Correct	1
3	421	Wrong	1
4	442	Correct	1
5	445	Correct	1
6	448	Correct	1
7	449	Correct	1
8	449	Correct	1
9	561	Correct	1
10	161	Correct	2
11	162	Correct	2
12	163	Correct	2
13	164	Correct	2
14	196	Correct	2
15	196	Correct	2
16	198	Correct	2
17	198	Correct	2
18	985	Not Attempted	3
19	268	Correct	3
20	269	Correct	3
21	270	Wrong	3
22	271	Wrong	3
23	311	Correct	3
24	311	Correct	3
25	311	Correct	3
26	311	Correct	3
27	311	Correct	3
28	217	Correct	4
29	218	Wrong	4
30	218	Correct	4
31	221	Correct	4
32	222	Correct	4
33	758	Wrong	4
34	758	Correct	4
35	758	Correct	4
36	758	Correct	4

Cut-off Points

Level	Points
BENG-50	5
BENG-80	4
BENG-90	4
BENG-100	5
ENGL-100	4
ENGL-101	5

Levels

Code	Description	Order	
BENG-50	BENG-50	1	Delete
BENG-80	BENG-80	2	Delete
BENG-90	BENG-90	3	Delete
BENG-100	BENG-100	4	Delete
ENGL-100	ENGL-100	5	Delete
ENGL-101	ENGL-101	6	Delete

Question Types

Code	Description
RC	Reading Comprehension
S	Signs
SK	Skimming/Scanning
SS	Structure Sentences
ST	Structure Text Cloze
V	Vocabulary Sentence
VT	Vocabulary Text cloze
W	Writing

Test-taker answer history

#	Question Id	Answer	Slide #
1	256	Correct	1
2	257	Correct	1
3	258	Correct	1
4	259	Correct	1
5	264	Correct	1
6	264	Correct	1
7	264	Correct	1
8	264	Correct	1
9	663	Correct	1
10	325	Correct	2
11	326	Correct	2
12	327	Correct	2
13	383	Correct	2
14	383	Wrong	2
15	383	Correct	2
16	383	Correct	2
17	383	Correct	2
18	692	Correct	2
19	109	Correct	3
20	110	Correct	3
21	111	Correct	3
22	112	Correct	3
23	113	Correct	3
24	722	Wrong	3
25	722	Correct	3
26	722	Wrong	3
27	722	Correct	3

28	76	Wrong	4
29	81	Correct	4
30	84	Correct	4
31	85	Wrong	4
32	103	Correct	4
33	103	Wrong	4
34	103	Correct	4
35	103	Correct	4
36	103	Wrong	4
37	509	Wrong	5
38	510	Wrong	5
39	511	Correct	5
40	512	Correct	5
41	513	Correct	5
42	518	Correct	5
43	518	Correct	5
44	518	Correct	5
45	518	Wrong	5
46	544	Correct	6
47	545	Correct	6
48	546	Correct	6
49	547	Correct	6
50	555	Correct	6
51	555	Correct	6
52	555	Correct	6
53	555	Wrong	6
54	555	Correct	6

