

**The Evaluation of English Listening Courses
at Taiwanese Universities:
An Exploratory Study**

Mu-hsuan Chou

Doctor of Philosophy

University of York
Department of Educational Studies
February 2008

ABSTRACT

In the early 2000s, the Taiwanese Ministry of Education (MOE) began to promote an English and/or bilingual environment in higher education, whereby English should be used for academic communication and as preparation for future careers. This implied the need for communicative/task-based language training courses and tests, and as a result, courses in English listening have become compulsory in the majority of Taiwanese universities. In 2002, the MOE decided that the English courses would form part of national university evaluations. However, in 2006, IELTS and TOEFL test data showed that Taiwanese students' English listening scores were lower than their reading and speaking scores. Listening thus seems to be a particular problem, but to date there has been little research on how listening courses are taught or assessed.

The present thesis is an exploratory study focusing on evaluating the teaching and assessment (both mid-term and final exams) of university English listening courses. The study reports on four case studies carried out at two Taiwanese universities to investigate how far the annual assessments matched up with MOE guidelines, using qualitative research methods (classroom observations, semi-structured interviews, and a document survey), plus one quantitative technique (questionnaire surveys). 112 English-major students in their second and third years of study and their four teachers participated in the study. The data were collected between 2005 and 2007.

The main findings were: (1) using reading and grammar items to test listening skills led to problems with establishing students' listening ability, and assessing whether they had reached the skill levels which they were expected to achieve, (2) only a few features of communicative or task-based instruction were discovered in either the teaching or the tests, (3) there was a large discrepancy between the exams in both difficulty level and test contents, which again made it impossible to compare the students' skill levels or progress within or between universities. The conclusion is that English support courses need a more transparent and comparable system of evaluation if the aims of a bilingual teaching environment are to be met.

Keywords: programme evaluation, task-based instruction, teaching listening, testing listening, higher education, Taiwan

CONTENTS

Title	i
Abstract.....	ii
List of Contents.....	iii
List of Figures and Tables.....	x
Acknowledgement.	xvi
Declaration.....	xvii
Conventions and Abbreviations.....	xviii

Introduction.....	1
--------------------------	----------

Chapter One Background to English Listening Courses in Taiwanese Universities

1.1 Introduction.....	6
1.2 Educational Reforms in Taiwanese Higher Education..	7
1.3 The Teaching Evaluation Programme in Taiwanese Universities.....	8
1.4 English Listening Courses in Taiwanese Universities: Background and Problem.....	13
1.5 Evaluation in Foreign Language Classroom.....	16
1.6 Quality Assurance in Higher Education	18
1.7 Listening to English as a Foreign Language	19
1.8 Teaching English Listening in the Foreign Language Classroom	24
1.8.1 Teaching Objectives.....	24
1.8.2 Approaches to Teaching Listening.....	25
1.8.3 Task-based Instruction (TBI) in Teaching Listening	31
1.8.4 Current Implementation of Task-based Instruction in Taiwan.....	38
1.9 Summary	39
Notes to Chapter One.....	41

Chapter Two English Listening Assessment and Washback Effects

2.1 Introduction.....	43
2.2 The Nature of Classroom Assessment... ..	44
2.3 Three Main Approaches to Assessing Listening.....	47
2.3.1 Discrete-point Testing.....	47
2.3.2 Integrative Testing.... ..	49

2.3.3	Communicative Language Testing...	55
2.4	Validity of Test Contents.....	57
2.5	Text Characteristics in Listening Tests..	63
2.6	Task Characteristics in Listening Tests.....	65
2.6.1	Test Setting	66
2.6.2	Test Rubric	66
2.6.3	Response Format.....	67
2.6.4	Topical Knowledge.....	67
2.7	Washback Effects and Test Consequences.....	68
2.8	Summary	72

Chapter Three An Overview of Research Design and Methodology

3.1	Introduction.....	74
3.2	The Purpose of the Research and Research Questions.....	74
3.3	Research Design: Main Study.....	75
3.4	Research Design: Pilot Study	77
3.5	The Ethical Issues in the Study.....	78
3.6	Data Analysis.	82
3.7	Validating Case Studies.....	82
3.8	The Use of Questionnaire Surveys	84
3.9	The Use of Semi-Structured Interviews	87
3.10	The Use of Structured Classroom Observation....	90
3.11	The Importance of Piloting.....	92
3.12	The Design of the Two Questionnaires.....	94
3.12.1	Questionnaire for the Mid-Term Examination.....	94
3.12.2	Questionnaire for the Final Examination.....	103
3.13	The Design of the Interview Questions.	103
3.13.1	Interview Questions for the Mid-term Exam.....	104
3.13.2	Interview Questions for the Final Exam....	105
3.14	The Design of the Classroom Observation Checklist... ..	106
3.15	Summary	109

Chapter Four The Results of the Pilot Study

4.1	Introduction to Pilot Study.....	111
4.2	Pilot Study: Procedural Overview.....	111
4.3	Pilot Study: The Classroom Observation before the Mid-term Exam.....	112
4.3.1	Time Sampling of Classroom Events before the Mid-term Exam... ..	112

4.3.2	Classroom Observation of Task-based Instruction before the Mid-term Exam.....	113
4.4	Pilot Study: Questionnaire Survey on the Mid-term Exam... ..	115
4.5	Pilot Study: Mid-term Interviews with the Teacher.....	116
4.6	Pilot Study: The Classroom Observation after the Mid-term Exam.....	119
4.6.1	Classroom Observation of Task-based Instruction after the Mid-term Exam.....	120
4.7	Pilot Study: Questionnaire Survey on the Final Exam.. ..	122
4.8	Pilot Study: Final Interview with the Teacher..... ..	122
4.9	Summary	124
	Notes to Chapter Four.....	126

Chapter Five Data Analysis – Case 1

5.1	Introduction.....	127
5.2	Case 1: Background to the Listening Course.....	128
5.3	Case 1: Classroom Observation before the Mid-term Exam.....	129
5.3.1	Case 1: First Observation.....	130
5.3.2	Case 1: Second Observation..... ..	131
5.3.3	Case 1: Third Observation	133
5.4	Case 1: Questionnaire Survey on the Mid-term Exam.. ..	135
5.4.1	Case 1: The Teaching Materials and the Course before the Mid-term Exam	136
5.4.2	Case 1: Students' General Preferences about the English Listening Classes	137
5.4.3	Case 1: Questions about the Mid-term Exam... ..	139
5.4.4	Case 1: Students' General Comments on the Mid-term Exam.. ..	142
5.5	Case 1: Mid-term Interview with the Teacher..... ..	143
5.6	Case 1: Classroom Observation after the Mid-term Exam.....	148
5.6.1	Case 1: Fourth Observation..... ..	148
5.6.2	Case 1: Fifth Observation.. ..	149
5.6.3	Case 1: Final Observation.....	150
5.7	Case 1: Questionnaire Survey on the Final Exam	152
5.7.1	Case 1: The Teaching Materials and the Course after the Mid-term Exam.....	153
5.7.2	Case 1: Students' General Preferences about English Listening Classes.....	154
5.7.3	Case 1: Questions about the Final Exam... ..	155
5.7.4	Case 1: Students' General Comments on the Final Exam	158

5.8	Case 1: Final Interview with the Teacher.....	159
5.9	Summary.....	161
	Notes to Chapter Five... ..	164

Chapter Six Data Analysis – Case 2

6.1	Introduction.....	165
6.2	Case 2: Background to the Listening Course	165
6.3	Case 2: Classroom Observation before the Mid-term Exam	167
6.3.1	Case 2: First Observation.....	167
6.3.2	Case 2: Second Observation.....	169
6.3.3	Case 2: Third Observation	170
6.4	Case 2: Questionnaire Survey on the Mid-term Exam..	172
6.4.1	Case 2: The Teaching Materials and the Course before the Mid-term Exam..	173
6.4.2	Case 2: Students’ General Preferences about the English Listening Classes.....	174
6.4.3	Case 2: Questions about the Mid-term Exam...	176
6.4.4	Case 2: Students’ General Comments on the Mid-term Exam..	179
6.5	Case 2: Mid-term Interview with the Teacher.....	180
6.6	Case 2: Classroom Observation after the Mid-term Exam...	185
6.6.1	Case 2: Fourth Observation.....	185
6.6.2	Case 2: Fifth Observation..	186
6.6.3	Case 2: Final Observation.....	187
6.7	Case 2: Questionnaire Survey on the Final Exam.	189
6.7.1	Case 2: The Teaching Materials and the Course after the Mid-term Exam..	189
6.7.2	Case 2: Students’ General Preferences about English Listening Classes.....	190
6.7.3	Case 2: Questions about the Final Exam...	191
6.7.4	Case 2: Students’ General Comments on the Final Exam	194
6.8	Case 2: Final Interview with the Teacher.....	195
6.9	Summary	197

Chapter Seven Data Analysis – Case 3

7.1	Introduction.....	202
7.2	Case 3: Background to the Listening Course	203
7.3	Case 3: Classroom Observation before the Mid-term Exam	204

7.3.1	Case 3: First Observation.....	204
7.3.2	Case 3: Second Observation.....	206
7.3.3	Case 3: Third Observation	207
7.4	Case 3: Questionnaire Survey on the Mid-term Exam..	209
7.4.1	Case 3: The Teaching Materials and the Course before the Mid-term Exam..	210
7.4.2	Case 3: Students' General Preferences about the English Listening Classes.....	210
7.4.3	Case 3: Questions about the Mid-term Exam... ..	212
7.4.4	Case 3: Students' General Comments on the Mid-term Exam..	215
7.5	Case 3: Mid-term Interview with the Teacher.....	215
7.6	Case 3: Classroom Observation after the Mid-term Exam...	219
7.6.1	Case 3: Fourth Observation.....	220
7.6.2	Case 3: Fifth Observation..	220
7.6.3	Case 3: Final Observation.....	221
7.7	Case 3: Questionnaire Survey on the Final Exam	222
7.7.1	Case 3: The Teaching Materials and the Course after the Mid-term Exam..	223
7.7.2	Case 3: Students' General Preferences about English Listening Classes.....	224
7.7.3	Case 3: Questions about the Final Exam... ..	225
7.7.4	Case 3: Students' General Comments on the Final Exam	228
7.8	Case 3: Final Interview with the Teacher.....	228
7.9	Summary	231
	Notes to Chapter Seven.....	234

Chapter Eight Data Analysis – Case 4

8.1	Introduction.....	235
8.2	Case 4: Background to the Listening Course	235
8.3	Case 4: Classroom Observation before the Mid-term Exam	236
8.3.1	Case 4: First Observation..	237
8.3.2	Case 4: Second Observation.....	238
8.3.3	Case 4: Third Observation	240
8.4	Case 4: Questionnaire Survey on the Mid-term Exam..	241
8.4.1	Case 4: The Teaching Materials and the Course before the Mid-term Exam..	243
8.4.2	Case 4: Students' General Preferences about the English Listening Classes.....	244

8.4.3	Case 4: Questions about the Mid-term Exam.....	245
8.4.4	Case 4: Students' General Comments on the Mid-term Exam..	248
8.5	Case 4: Mid-term Interview with the Teacher.....	248
8.6	Case 4: Classroom Observation after the Mid-term Exam...	252
8.6.1	Case 4: Fourth Observation.....	252
8.6.2	Case 4: Fifth Observation..	253
8.6.3	Case 4: Final Observation.....	253
8.7	Case 4: Questionnaire Survey on the Final Exam	255
8.7.1	Case 4: The Teaching Materials and the Course after the Mid-term Exam..	256
8.7.2	Case 4: Students' General Preferences about English Listening Classes.....	256
8.7.3	Case 4: Questions about the Final Exam... ..	257
8.7.4	Case 4: Students' General Comments on the Final Exam	260
8.8	Case 4: Final Interview with the Teacher.....	260
8.9	Summary	262
	Notes to Chapter Eight.....	266

Chapter Nine Discussion

9.1	Introduction.....	267
9.2	The Implementation of the Task-based Instructions.....	268
9.3	The Test Objectives.....	272
9.4	The Communicative Language Testing Approaches....	275
9.5	The Difficulty the Students Encountered in the Two Exams	275
9.6	The Washback of Test Results on the Teaching.....	278
9.7	Summary	280

Chapter Ten Conclusions and Suggestions

10.1	General Aims of the Study.....	282
10.2	Implications for Instruction and Curriculum Development.....	284
10.3	Implications for In-class Listening Materials	286
10.4	Implications for Assessment Methods... ..	288
10.5	Implications for Policy and Practice.....	290
10.6	The Limitations of the Study.....	292

Appendices	
Appendix A.....	295
Appendix B.....	315
Appendix C.....	335
Appendix D.....	342
Appendix E.....	363
Appendix F.....	388
Appendix G.....	411
Glossary	429
References and Bibliography	431

FIGURES AND TABLES

Tables

1.1	English Listening Courses in Taiwanese Universities.....	13
3.1	Main Study Design.....	76
4.1	Pilot: Occurrence of Task-based Instruction Features before the Mid-term Exam.....	114
4.2	Pilot: Translation of Mid-term Interview Data (Questions 1 to 7).....	117
4.3	Pilot: Occurrence of Task-based Instruction Features after the Mid-term Exam.....	121
4.4	Pilot: Translation of Final Interview (Question 1 to 4).....	123
5.1	Case 1: Occurrence of Task-based Instruction Features before the Mid-term Exam.....	134
5.2	Case 1: Mid-term Survey – Gender of the Respondents.....	136
5.3	Case 1: Mid-term Survey – Perceived Difficulty of In-class Listening Materials.....	137
5.4	Case 1: Mid-term Survey – Frequency of Reasons for Material Difficulty.....	137
5.5	Case 1: Mid-term Survey – Preferred Mode of Answering.....	138
5.6	Case 1: Mid-term Survey – Preferred Type of Speech.....	138
5.7	Case 1: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	139
5.8	Case 1: Mid-term Survey – Performance in the Mid-term Exam.....	139
5.9	Case 1: Mid-term Survey – Topics of Test tasks.....	140
5.10	Case 1: Mid-term Survey – Linguistic Features of the Listening Extracts (1).....	140
5.11	Case 1: Mid-term Survey – Linguistic Features of the Listening Extracts (2).....	141
5.12	Case 1: Mid-term Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam.....	142
5.13	Case 1: Mid-term Survey – Interview with the Teacher (a).....	143
5.14	Case 1: Mid-term Survey – Interview with the Teacher (b).....	144
5.15	Case 1: Mid-term Survey – Interview with the Teacher (c).....	144
5.16	Case 1: Mid-term Survey – Interview with the Teacher (d).....	145
5.17	Case 1: Mid-term Survey – Interview with the Teacher (e).....	146
5.18	Case 1: Mid-term Survey – Interview with the Teacher (f).....	147
5.19	Case 1: Mid-term Survey – Interview with the Teacher (g).....	147

5.20	Case 1: Occurrence of Task-based Instruction Features after the Mid-term Exam.....	151
5.21	Case 1: Final Survey – Perceived Difficulty of In-class Materials.....	154
5.22	Case 1: Final Survey – Frequency of Reasons for Material Difficulty.....	154
5.23	Case 1: Final Survey – Preferred Mode of Answering	154
5.24	Case 1: Final Survey – Preferred Type of Speech.....	154
5.25	Case 1: Final Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	155
5.26	Case 1: Final Survey – Performance in the Final Exam.....	155
5.27	Case 1: Final Survey – Topics of Test tasks	156
5.28	Case 1: Final Survey – Linguistic Features of the Listening Extracts (1).....	156
5.29	Case 1: Final Survey – Linguistic Features of the Listening Extracts (2).....	157
5.30	Case 1: Final Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam.. ..	158
5.31	Case 1: Final Survey – Interview with the Teacher (a).....	159
5.32	Case 1: Final Survey – Interview with the Teacher (b).....	159
5.33	Case 1: Final Survey – Interview with the Teacher (c).....	160
5.34	Case 1: Final Survey – Interview with the Teacher (d).....	161
6.1	Case 2: Occurrence of Task-based Instruction Features before the Mid-term Exam.....	171
6.2	Case 2: Mid-term Survey – Gender of the Respondents.....	173
6.3	Case 2: Mid-term Survey – Perceived Difficulty of In-class Listening Materials	173
6.4	Case 2: Mid-term Survey – Frequency of Reasons for Material Difficulty.....	173
6.5	Case 2: Mid-term Survey – Preferred Mode of Answering	175
6.6	Case 2: Mid-term Survey – Preferred Type of Speech	175
6.7	Case 2: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	175
6.8	Case 2: Mid-term Survey – Performance in the Mid-term Exam.....	176
6.9	Case 2: Mid-term Survey – Topics of Test tasks.....	177
6.10	Case 2: Mid-term Survey – Linguistic Features of the Listening Extracts (1).....	177
6.11	Case 2: Mid-term Survey – Linguistic Features of the Listening Extracts (2).....	178
6.12	Case 2: Mid-term Survey – Easiest vs. Most Difficult Type of	

	Comprehension Question in the Mid-term Exam	179
6.13	Case 2: Mid-term Survey – Interview with the Teacher (a)	180
6.14	Case 2: Mid-term Survey – Interview with the Teacher (b)	182
6.15	Case 2: Mid-term Survey – Interview with the Teacher (c)	182
6.16	Case 2: Mid-term Survey – Interview with the Teacher (d)	183
6.17	Case 2: Mid-term Survey – Interview with the Teacher (e)	183
6.18	Case 2: Mid-term Survey – Interview with the Teacher (f)	184
6.19	Case 2: Mid-term Survey – Interview with the Teacher (g)	184
6.20	Case 2: Occurrence of Task-based Instruction Features after the Mid-term Exam.....	188
6.21	Case 2: Final Survey – Perceived Difficulty of In-class Materials.....	190
6.22	Case 2: Final Survey – Frequency of Reasons for Material Difficulty.....	190
6.23	Case 2: Final Survey – Preferred Mode of Answering	190
6.24	Case 2: Final Survey – Preferred Type of Speech.....	190
6.25	Case 2: Final Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	191
6.26	Case 2: Final Survey – Performance in the Final Exam.....	191
6.27	Case 2: Final Survey – Topics of Test tasks.....	192
6.28	Case 2: Final Survey – Linguistic Features of the Listening Extracts (1).....	192
6.29	Case 2: Final Survey – Linguistic Features of the Listening Extracts (2).....	193
6.30	Case 2: Final Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam.. ..	194
6.31	Case 2: Final Survey – Interview with the Teacher (a).....	195
6.32	Case 2: Final Survey – Interview with the Teacher (b).....	196
6.33	Case 2: Final Survey – Interview with the Teacher (c).....	196
6.34	Case 2: Final Survey – Interview with the Teacher (d).....	197
7.1	Case 3: Occurrence of Task-based Instruction Features before the Mid-term Exam.....	208
7.2	Case 3: Mid-term Survey – Gender of the Respondents.....	209
7.3	Case 3: Mid-term Survey – Perceived Difficulty of In-class Listening Materials.....	210
7.4	Case 3: Mid-term Survey – Frequency of Reasons for Material Difficulty.....	210
7.5	Case 3: Mid-term Survey – Preferred Mode of Answering.....	211
7.6	Case 3: Mid-term Survey – Preferred Type of Speech	211
7.7	Case 3: Mid-term Survey – Preferred vs. Least Preferred Type of	

	Comprehension Question in Class.....	211
7.8	Case 3: Mid-term Survey – Performance in the Mid-term Exam.....	212
7.9	Case 3: Mid-term Survey – Topics of Test tasks.....	212
7.10	Case 3: Mid-term Survey – Linguistic Features of the Listening Extracts (1).....	213
7.11	Case 3: Mid-term Survey – Linguistic Features of the Listening Extracts (2).....	213
7.12	Case 3: Mid-term Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam	215
7.13	Case 3: Mid-term Survey – Interview with the Teacher (a)	215
7.14	Case 3: Mid-term Survey – Interview with the Teacher (b)	216
7.15	Case 3: Mid-term Survey – Interview with the Teacher (c)	217
7.16	Case 3: Mid-term Survey – Interview with the Teacher (d)	217
7.17	Case 3: Mid-term Survey – Interview with the Teacher (e)	218
7.18	Case 3: Mid-term Survey – Interview with the Teacher (f)	219
7.19	Case 3: Mid-term Survey – Interview with the Teacher (g)	219
7.20	Case 3: Occurrence of Task-based Instruction Features after the Mid-term Exam.....	221
7.21	Case 3: Final Survey – Perceived Difficulty of In-class Materials.....	224
7.22	Case 3: Final Survey – Frequency of Reasons for Material Difficulty.....	224
7.23	Case 3: Final Survey – Preferred Mode of Answering	224
7.24	Case 3: Final Survey – Preferred Type of Speech.....	224
7.25	Case 3: Final Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	225
7.26	Case 3: Final Survey – Performance in the Final Exam.....	225
7.27	Case 3: Final Survey – Topics of Test tasks	226
7.28	Case 3: Final Survey – Linguistic Features of the Listening Extracts (1).....	226
7.29	Case 3: Final Survey – Linguistic Features of the Listening Extracts (2).....	227
7.30	Case 3: Final Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam..	228
7.31	Case 3: Final Survey – Interview with the Teacher (a).....	229
7.32	Case 3: Final Survey – Interview with the Teacher (b).....	230
7.33	Case 3: Final Survey – Interview with the Teacher (c).....	230
7.34	Case 3: Final Survey – Interview with the Teacher (d).....	231
8.1	Case 4: Occurrence of Task-based Instruction Features before the Mid-term Exam...	241

8.2	Case 4: Mid-term Survey – Gender of the Respondents.....	243
8.3	Case 4: Mid-term Survey – Perceived Difficulty of In-class Listening Materials.....	243
8.4	Case 4: Mid-term Survey – Frequency of Reasons for Material Difficulty.....	243
8.5	Case 4: Mid-term Survey – Preferred Mode of Answering.....	244
8.6	Case 4: Mid-term Survey – Preferred Type of Speech	244
8.7	Case 4: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	245
8.8	Case 4: Mid-term Survey – Performance in the Mid-term Exam.....	245
8.9	Case 4: Mid-term Survey – Topics of Test tasks.....	246
8.10	Case 4: Mid-term Survey – Linguistic Features of the Listening Extracts (1).....	247
8.11	Case 4: Mid-term Survey – Linguistic Features of the Listening Extracts (2).....	247
8.12	Case 4: Mid-term Survey – Interview with the Teacher (a)	249
8.13	Case 4: Mid-term Survey – Interview with the Teacher (b)	249
8.14	Case 4: Mid-term Survey – Interview with the Teacher (c)	249
8.15	Case 4: Mid-term Survey – Interview with the Teacher (d)	250
8.16	Case 4: Mid-term Survey – Interview with the Teacher (e)	251
8.17	Case 4: Mid-term Survey – Interview with the Teacher (f)	251
8.18	Case 4: Occurrence of Task-based Instruction Features after the Mid-term Exam.....	254
8.19	Case 4: Final Survey – Perceived Difficulty of In-class Materials.....	256
8.20	Case 4: Final Survey – Frequency of Reasons for Material Difficulty.....	256
8.21	Case 4: Final Survey – Preferred Mode of Answering	256
8.22	Case 4: Final Survey – Preferred Type of Speech.....	256
8.23	Case 4: Final Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class.....	257
8.24	Case 4: Final Survey – Performance in the Final Exam.....	257
8.25	Case 4: Final Survey – Topics of Test tasks.....	258
8.26	Case 4: Final Survey – Linguistic Features of the Listening Extracts (1).....	259
8.27	Case 4: Final Survey – Linguistic Features of the Listening Extracts (2).....	259
8.28	Case 4: Final Survey – Interview with the Teacher (a).....	261
8.29	Case 4: Final Survey – Interview with the Teacher (b).....	261
8.30	Case 4: Final Survey – Interview with the Teacher (c).....	262

9.1	Course Instructors in the Main Study.....	267
9.2	Comparison of TBI Characteristics between Four Teachers.....	270
9.3	Comparison of the Teaching Objectives of the Four Teachers... ..	274
9.4	Difficulties in Comprehending the Listening Contents in the Two Exams.....	276

Figures

1.1	Evaluation Relationship between Minister of Education, University and HEEACT.....	10
1.2	Framework for Task-based Instruction.. ..	34
5.1	Case 1: The Layout of the Classroom.....	129
6.1	Case 2: The Layout of the Classroom.....	167
7.1	Case 3: The Layout of the Classroom.....	204
8.1	Case 4: The Layout of the Classroom.....	237

ACKNOWLEDGEMENTS

In the past three and half years, I received much help from my teachers and friends in the UK and Taiwan. Without their kind and generous help, this tough job could not be completed. In addition to thanking my mother and sister for their support and encouragement from Taiwan, I would like to thank the following people:

Dr. Graham Low (University of York) for his long-term supervision and insightful discussions on my topic.

Dr. Chris Kyriacou (University of York) for his kind help and practical suggestions on my research design and data analysis.

Dr. Emma Marsden (University of York) for her critical ideas about my data analysis method.

Sharon Tsau, Shu-fang Ni, Hsin-rong Tsai, Lan-hsin Jiao, Shu-chen Tuan, Ching-Yi Tien, Stephano De Caro for bridging connections with the course instructors and for collecting data.

The anonymous 153 students for participating in my research and providing me with very useful information.

Jing Sheng (University of Lancaster) and **Yun Wang** (University of Leeds) for their supply of academic journal articles.

Alex May for proofreading and suggestions.

There are too many people I need to thank – my family, my teachers, and my friends – for their company throughout this lonely long journey.

DECLARATION

I declare that this thesis is based on my own work. All published references are cited, and official governmental websites are included. So far until the time of submission, none of any chapters in this thesis is published.

CONVENTIONS and ABBREVIATIONS

- “√ x”** “√ x” is used in the observation list, which means only the TBI characteristics were found among some students, not all of them.
- “A”** “A” means “agree” in Tables.
- “bonus item/mark”** An item is scored when students answered it correctly, but the item is not scored if they answered it incorrectly. In other words, a bonus item adds an extra mark to the total score, but no deduction from the total score is made if students fail to answer it.
- “D”** “D” means “disagree” in Tables.
- “HEEACT”** Higher Education Evaluation and Accreditation Council of Taiwan
- “look-up-and-say”** When the students read the listening extracts in the textbook, they need to “speak” or “say” the conversations rather than simply read them.
- “MOE”** Ministry of Education
- “Neither A nor D”** “Neither A nor D” means “neither agree nor disagree” in Tables.
- “SA”** “SA” means “strongly agree” in Tables.
- “SD”** “SD” means “strongly disagree” in Tables.
- “(strongly) dis/agree”** This is used to report the result of Likert type items in the two questionnaires. Where the number of students is small, it is more useful to aggregate “strongly agree” and “agree” as “(strongly) agree”. The same also applies to “strongly disagree” and

“disagree” as “(strongly) disagree”.

“(very) dis/satisfy”

It is used to report the result of the Smiley Face in the two questionnaires. The reason for the aggregation is the same as with “(strongly) dis/agreed” described above.

Introduction

Evaluating language courses in Taiwanese universities has recently become a matter of considerable concern. Since 2002, globalisation has become a key evaluation issue, particularly for social sciences and the humanities in Taiwanese higher education (MOE, 2002; Chen, 2007). In order to achieve globalisation, the Ministry of Education (MOE) has encouraged university teachers to use English to teach specialised subjects in class; this policy, to be effective, implies some form of communicative or task-based teaching needs to be implemented. It also assumes that university students will have sufficient aural and oral skills to understand lessons and to express their opinions in discussions.

In order to establish whether the policy is being implemented, evaluation procedures are needed. In 2002, the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT) was created to evaluate academic departments. Evaluation in higher education now plays an important role in quality assurance and quality enhancement in teaching, learning, and research. In Taiwan, university academic departments are evaluated by a committee in the HEEACT, consisting of professors, administrators and teachers working in different universities. The evaluation takes place every five years, and each evaluation lasts four months – an academic term. The main evaluation criteria include departmental teaching goals, curriculum design, the qualification of full-time teachers, and alumni performance. However, the evaluation system focuses on general departmental policy and curriculum rather than taking a close look at what really happens in the language classroom. In other words, how language courses are taught and assessed in the classroom remains unexplored. It appears that there is a mismatch between the MOE requirement and the

HEEACT scheme. If the evaluation at classroom level shows problems, then this would have policy implications for making the two dovetail more effectively.

English listening courses are compulsory now in the majority of universities in Taiwan and are designed to provide students with opportunities to develop their English listening ability. However, in 2006, IELTS and TOEFL test data showed that Taiwanese students' listening and writing scores were lower than their reading and speaking scores (IELTS, 2006; ETS, 2006). The problem is thus that although English listening courses are compulsory in universities, the students' listening performance in tests appears to remain poorer than their reading and speaking performance. As the manner in which teaching and testing are implemented is closely related to the quality of teaching and learning, it is important to establish how both are implemented and influence each other, at a practical level, in Taiwanese universities. Thus, the aim in writing the present thesis was to look at how the university students taking English listening courses are taught and tested, and how far the teaching and assessment methods in English listening classrooms in Taiwanese universities involve interactive listening and speaking between teachers and students. Evaluating the classroom assessment in the present study involved identifying whether the English listening programmes in a sample of Taiwanese universities had their intended effects, to compare the teaching and assessment methodologies across different groups of students, and to provide recommendations regarding enhancing the quality of English listening courses at university level. The ultimate aim, at a general level, is to explore and evaluate how far communicative approaches are employed in teaching and testing English listening courses, and to look for connections between the courses, the university evaluation programme, and government policy.

As an exploratory survey, I conducted interviews with ten teachers who

taught English listening and speaking courses in eight universities in Taiwan, inquiring about current assessment methods regarding the listening and speaking courses in these institutions from April to October 2005. I also carried out an Internet-based survey of English listening and speaking courses at Taiwanese universities in 2006. I discovered that, firstly, English listening and speaking courses were indeed compulsory in the majority of Taiwanese universities, and university teachers were free to decide in-class teaching materials and assessment methods. Next, students were required to take and pass two examinations – mid-term and final – and to participate in class in order to earn the necessary credits. In the interviews, interestingly, I was told that students in the same year of study were divided into two or more “groups”, each with a different teacher, so that the students could have more opportunity to speak English in class. When students in the same year of study are divided into smaller groups led by different teachers, it is possible, even likely, that the teaching, teaching materials, and assessment methods will be diverse. Since the purpose of dividing students into groups was to provide them with more opportunities to engage in in-class tasks, the present thesis also examines whether students felt “smaller group” teaching encouraged them to speak more English and to interact with each other in class.

In order to explore and compare the similarity and the differences between groups at a detailed level, a case study approach was adopted. Four groups of students and their course instructors from two universities (two groups from University A and another two from University B) in Taiwan were chosen and agreed to participate. There are, thus, there are four cases in the main study. Although English listening courses are compulsory for students from all subjects, the participants selected for the pilot and formal case studies were all English language major students, because the English listening courses were established

and run for several years.

The data collection methods comprised classroom observations, questionnaires, interviews, examination of in-class teaching materials or textbooks, collecting teacher's syllabus notes, and the students' academic marks. Six classroom observations were carried out for each of the four cases, with three observations before the mid-term exam and the remaining three after it, to see how far communicative approaches were implemented in the listening classes, and to see if the results of the mid-term exam impacted on the teaching afterwards. A mid-term and a final questionnaire were designed, piloted, and administered (by me) in order to understand what kind of problems influenced students' listening comprehension in the two exams. Face-to-face interviews were conducted with the four course instructors after both the mid-term and the final exams, to investigate how they decided on the content and the difficulty of the two exams, to what extent the teaching and test objectives were clearly specified in class, and whether test items had been piloted. Document surveys of in-class materials, textbooks, and the teachers' syllabuses were also used to contextualise the situation and to help interpret the information from the classroom observations and interviews. The data analyses mainly involved a qualitative approach, with support where relevant from descriptive statistical analyses, particularly for the questionnaire results. The research procedure involved observing the three classes *before* the mid-term exam, distributing the questionnaires *after* the exam, and finally interviewing the teachers *after* students' mid-term exam marks were known.

The thesis is divided into ten chapters. Chapter One examines Taiwan's recent policy towards educational evaluation, gives background information about English listening courses at Taiwanese universities, and reviews the teaching of

English listening, including its objectives, approaches, and classroom interactions. Chapter Two looks at the nature of classroom assessment, approaches to testing listening, the importance of validity in testing procedures, test and task characteristics of listening assessments, and washback effects. Chapter Three gives an overview of the research design, the methods used, and associated ethical issues, reviews the relevant literature on validating case studies, and discusses the advantages and disadvantages of using questionnaires, interviews, and classroom observations. Chapter Four gives the results of the pilot study and lists the amendments made for data collection in the main study (Appendix A presents the pilot version of the questionnaires, interview questions, and observation schedules, and Appendix B presents the revised versions). Chapters Five and Six examine the data from the two cases at University A, while Chapters Seven and Eight analyse the data from the two cases at University B. Chapter Nine is a discussion chapter and pulls together the findings from Chapters Five to Eight. The thesis ends with Chapter Ten, where I consider the implications of the findings for policy and teaching practice at university level in the light of current government educational policy, and make a number of suggestions for future research.

Chapter One

Background to English Listening Courses in Taiwanese Universities

1.1 Introduction

Since language is a means of communication, developing an appropriate balance between the four English language skills is an important aspect of language teaching and learning. In Taiwan, where Chinese is the native and official language and English for almost all Taiwanese is a foreign language, the communicative language environment for listening to English is very limited. Since the 1990s, the educational reforms of Taiwanese higher education and the impact of joining the World Trade Organisation have led to an increase in the number of universities, a need for an English teaching and learning environment, and the establishment of the higher education evaluation system. One aspect of this is the need to develop university students' comprehension and communication skills in English, and to this end an English listening module has become compulsory for students in over 80% of universities. This chapter begins by introducing the educational reforms in Taiwanese higher education, followed by a discussion of the current situation of educational evaluation. Thirdly, I examine how English listening is organised, taught and tested in Taiwanese universities based on the preliminary interviews with ten university English listening teachers in Taiwan. Next, the nature of listening to English as a foreign language and the teaching of English listening in foreign language classrooms are examined; the discussion includes teaching objectives, approaches to teaching listening, and classroom interactions.

1.2 Educational Reforms in Taiwanese Higher Education

Educational policy change is normally caused by external system events such as changes in economic and political conditions that affect actors' belief systems (Sabateir and Jenkins-Smith, 1993; see Bleiklie, 2001: 24). This was very much the case with higher education changes in Taiwan in the 1990s. Before the 1990s, Taiwanese universities were directly run and governed by the Kuomintang (KMT)¹ party who monopolised the administration, development, and funding of higher education (Lo, 2004). There were many restrictions to higher education due to the regulation of the Martial Law, such that very few educational activities or development were promoted. With the relaxing of the Martial Law in the early 1990s, the KMT initiated a series of reforms to make the educational system more open and innovative (Lo, 2004: 78). "Liberalisation" was the central idea to introduce autonomy and flexibility to Taiwanese higher educational institutions, and the removal of unnecessary political controls on education was a major concern for higher education reform. Lo (ibid: 79) notes that higher education in Taiwan has successfully freed academic institutions from governmental control to run as "independent agents". As independent non-government administrators increased, more universities, particularly private ones, were established to provide more tertiary education. However, the Ministry of Education (MOE) considered that the development of higher education should focus not only on increasing the quantity of programmes and universities, but also on ensuring the quality of teaching (MOE, 2008). The development of educational reforms in Taiwanese higher education includes: (1) the liberalisation of university governance, and more importantly for present purposes, self-evaluation, (2) the mobilisation of non-government sectors in higher education provision and, (3) the establishment of a quality evaluation system for higher education (Council on Education

Reform, 1996; see Lo, 2004: 79). However, the evaluation system was not fully-fledged and nation-wide until the 2000s.

Since Taiwan entered the World Trade Organization (WTO) in 2002, the change of economic situation has impacted on its higher education policy. Chen (2003) discovered that the resulting economic change in Taiwan has had a great influence on teaching, student recruitment, school operation, and the quality of teachers in higher education. Che (2004) pointed out that the impact from the WTO led to several problems: (1) an imbalance between the *quality* and the *quantity* of the tertiary institutions, (2) a deficiency of educational resources and a poor English learning environment, (3) the internationalisation of higher education, (4) high tuition fees, and (5) a university financial crisis. Specifically, the private universities were affected more than the national ones, and the universities in the southern part of Taiwan more than those in the northern part (Chen, 2003). However, there is another worrying situation, namely that the shrinking population² in Taiwan cannot meet the rapid increase in places in tertiary institutions, and this has created a serious imbalance between supply and demand in the higher education market. Although the liberalisation in higher education has provided more learning opportunities, the control of both quality and quantity remains a major problem to be resolved.

1.3 The Teaching Evaluation Programme in Taiwanese Universities

After a comparison of evaluation schemes used in higher education systems in developed countries, the MOE in Taiwan held a symposium to discuss the importance and feasibility of establishing a designated organization to undertake higher education evaluation affairs in 2002 (HEEACT, 2008). In the same year, the MOE decided that aspects of internationalisation and English globalisation in

English courses would form part of university evaluations for the purpose of making Taiwanese universities competitive globally (MOE, 2002). In 2004, Minister Du, in the *Report of the Education and Cultural Committee* (Du, 2004), broadened this by recommending that all university courses should be taught in English, or bilingually, to create an English teaching environment:

In order to enhance our students' English ability, to strengthen their English ability in understanding their specialised subjects in Taiwanese universities, and to attract foreign students to study in Taiwan, the Ministry of Education strongly encourages all universities in Taiwan to teach part of the specialised programmes fully in "English" or "bilingually in two languages" (Chinese and English).

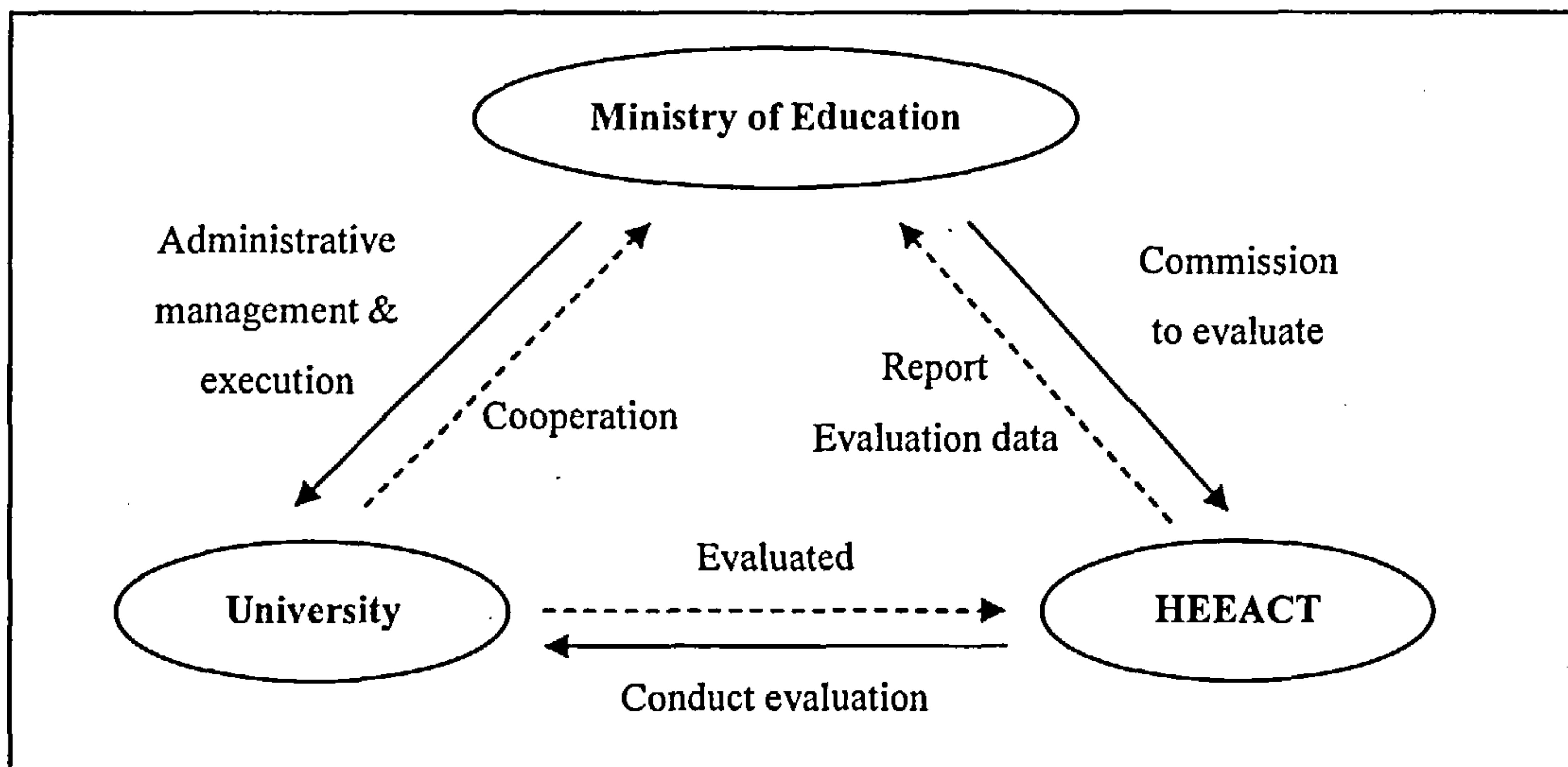
Du (2004) (<http://www.edu.tw/>)

The problem with this official pronouncement was that it was somewhat vague. The range of skills that universities needed to add to their programmes was not detailed, nor were the elements (skills or modules) that would form part of the evaluations. In addition, the MOE claimed that courses taught in English would be included in the university evaluation, but I was unable to find any relevant literature regarding the inclusion of language issues in this evaluation.

In 2005, all universities and colleges in Taiwan were asked to fund the establishment of the Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT, 2008). This teaching evaluation system aims at "assuring academic programmes provide a sound and qualified learning environment" (HEEACT, 2008). The job of the HEEACT is (1) to investigate the teaching situation and quality in academic departments, (2) to conduct on-site visits, (3) to assist each university in establishing self-evaluation mechanisms, (4) to strengthen academic excellence and distinctiveness, and (5) to provide the evaluation results to the government as a reference for future educational policies. That is to say, the HEEACT is commissioned to evaluate university programmes on the behalf of the

MOE. Each university is evaluated every five years, with the evaluations lasting a semester (16 to 18 weeks). The academic programmes are evaluated by an HEEACT committee, including professors, administrators and teachers working in different Taiwanese universities; each university is evaluated by five HEEACT evaluators. Figure 1.1 illustrates the relationship between the MOE, university, and the HEEACT in university programme evaluation.

Figure 1.1 Evaluation Relationship between Ministry of Education, University and HEEACT



Source: HEEACT, 2008

The criteria for evaluation cover the characteristics of the programmes, curriculum design, student learning and extracurricular activities, research development, and employability of graduates (Wang, 2007: 8). Thus, evaluations are divided into five areas (HEEACT, 2008):

- (1) Goals, features, and improvement
- (2) Curriculum design
- (3) Students' opinions and student affairs
- (4) Research and professional performance
- (5) Alumni performance

The results of the evaluations are reported to the MOE and published in Chinese on the HEEACT website (HEEACT, 2007). However, for present purposes, the results are problematic, or of limited usefulness, in several ways.

Firstly, the process of evaluation is slow, in the sense that it takes five years to complete a cycle. Secondly, the results of the evaluation are fairly general; they cover administrative management (department features and goals, curriculum design, and the academic qualification of teachers), rather than comment on how successfully programmes comply with the MOE requirements, or the real classroom situation; how teachers teach students, how students learn, and their interactions in class remain unknown. In addition, the language issue is only reported indirectly, for example, via students' opinions of the department or occasionally of a certain course. The last problem is that classroom assessment is closely related to teaching, but the HEEACT evaluation does not take this into consideration. Evaluating the administrative management of a department provides administrators with suggestions about curriculum design and policy, whereas the provision of information about what works, or what needs changing with respect to teaching and assessment, requires some input at classroom and learner level. The overall result is that the MOE requirements and the five-year evaluations do not dovetail well, and it is almost impossible to establish whether, or how far, the MOE requirements are in fact being met.

The HEEACT serves as an external evaluation organisation, but before academic programmes are evaluated by them, self-evaluation within an institution takes place (Liu, 2007; Wu and Chang, 2007: 10). In 2007, there were 109 universities, excluding colleges, in Taiwan (Table 1.1, Section 1.4). Official statistics show that, in 2007, there were 72 universities in Taiwan implementing self-evaluation systems, though of these, only 34 (48%) clearly specified their evaluation items and criteria (Wu and Chang, 2007). The aim of establishing self-evaluation in each university is to allow institutions to evaluate the teaching, to decide either to continue or terminate a teaching contract with a teacher, and to

determine whether or not to reward a teacher academically (ibid.). From Brenna and Shah's point of view (2000: 56), the purpose of a self-evaluation is to enable the institution to "provide appropriate, relevant and up-to-date information about itself". However, Wu and Chang (2007) suggest that it is precisely because each university in Taiwan is allowed to decide what to include in their evaluation for their own programmes that the criteria are often unclear, and the focus seems to be on decision-making (e.g. whether to continue or to terminate a teaching contract). The idea of individual self-evaluation in Taiwanese universities is relatively recent, and is not yet fully developed, as less than half of these universities clearly and specifically employed it at the time of writing. Also, it is unclear if the results of this self-evaluation are published.

According to the IELTS test data for 2006, Taiwanese students' English listening and writing skills were poorer than their reading and speaking skills (IELTS, 2006).³ The published TOEFL test data for 2005 and 2006 also supported this claim that Taiwanese students' English listening skills were poorer than their reading and writing skills (ETS, 2006).⁴ If Taiwanese university students want to be competitive internationally, their English listening skills need to be strengthened. Moreover, in the situation where university teaching in English, or in English and Chinese, is now officially promoted, English listening courses have become particularly important, in the sense that students are now expected to listen or even to use English in their regular classrooms. Thus, English listening courses need to become part of the evaluation at least within each institution, in order to ensure the quality of teaching and learning of English listening skills.

One of the main purposes of running English listening courses is thus to train students to comprehend the sort of English language which they are very likely to

hear in other university courses. Compared with the vagueness about what is taught in English classes at university level, the MOE in Taiwan was clearer about its policy on teaching listening and speaking skills in secondary schools. Wang (2002: 132) notes that the MOE published new curricula for English teaching in secondary schools that clearly demanded communication-based teaching and which guided materials development and classroom practice in English classes. Thus the MOE required that each textbook (one per term) should have colourful pictures and short daily-life dialogues, and that lessons should be arranged according to themes and functions of communication where English speaking and listening skills are the focus of teaching (ibid.). Although the guidelines for university teaching are less clear, we may assume that here too the methods for teaching students to use English in class for listening or for interacting with the teacher (or with each other) will need to be communicative or task-based. However, how far they currently achieve this remains unknown.

1.4 English Listening Courses in Taiwanese Universities: Background and Problem

In 2006, I conducted an Internet check of the number of the universities in Taiwan and how many of them run English listening courses. Table 1.1 shows that there were one hundred and nine universities in Taiwan which included 41 national universities, and 68 private ones in 2006. The titles of English listening modules varied from university to university, but in general, two titles were used the most: *English Listening* and *English Listening and Speaking Practice*.

Table 1.1 English Listening Courses in Taiwanese Universities

University	Total Number	No. with <i>English Listening</i> or <i>English Listening and Speaking</i> Courses
National	41	33
Private	68	57
Total	109	90 (approx. 83%)

Source: Internet check on 3 December 2007. Colleges were not included.

From April to October 2005, I conducted preliminary interviews with ten teachers who taught English listening courses in eight universities in Taiwan, to inquire about current assessment practices regarding the listening courses in their institutions. The ten teachers were from four national universities and six private ones. I asked them four questions regarding their teaching situations. Each interview was accordingly short, lasting approximately fifteen minutes (see Appendix C.1). Firstly, I asked the teachers how they usually evaluated students' English listening ability in the classroom. All ten claimed that according to the university policy regulations, students' final scores had to comprise two marks from two examinations – a mid-term and a final examination; in addition, the teachers also had to evaluate students' in-class performance and added this to the exam marks to create a final composite score. Six teachers said that they tended to use in-class participation, such as doing class-based exercises, or quizzes for in-class performance assessment; the remaining four said that in-class performance usually consisted of a combination of attendance and homework. Secondly, I asked them who decided teaching materials and assessment methods. They all said that they could determine the teaching materials and assessment methods themselves. Thirdly, they were asked what types of listening extracts or passages they used in class. All ten stated that "conversations" were mostly used, but seven also said "news broadcasts". "Academic lectures" were employed by just four teachers. Finally, I asked if they taught all the students from one year of study in a single class. Eight teachers told me that they were responsible for teaching a single group of students, though they were aware that there were other teachers teaching the same course. However, when asked if they knew why their students were divided into separate groups taught by different teachers, they all said that this was arranged by the department. Interestingly, there is no government official

documentation regarding this division of groups into two classes; it appears that a university is free to determine how classes are organised. Four teachers assumed that it was possible that using smaller groups provided their students with more opportunity to participate actively in class. Four teachers knew some of the other teachers who also taught the English listening courses, while the rest did not. None of these eight teachers knew what in-class materials the other teachers used for their English listening classes. The remaining two teachers claimed that they did not divide students into groups, because of problems of teacher availability.

According to the preliminary interviews, English listening modules are organised such that within a university, different instructors using different in-class materials teach different groups of students, but the title of the course is the same for all students. A problem therefore arises, as noted above, when English teachers within a university come from different educational backgrounds, bringing diverse personalities and their own teaching styles. Each English instructor chooses in-class materials to teach their students, and designs tests, or uses test questions from previous exams, to assess their students based on their own individual preferences. There are probably considerable variations in teaching materials and test methods within and between institutions, but the nature and extent of the variation remains very unclear and there are no published reports or research studies that I could find that have explored this variation. In addition, this freedom may lead to problems of quality control, since a listening course which is divided into two groups may be taught by two different instructors using very different in-class materials and assessment methods. The result is that it is extremely difficult to ensure a valid assessment system to test students' listening abilities and interpret their test scores within a university, let alone between universities.

1.5 Evaluation in Foreign Language Classrooms

Evaluation is defined in various ways. Cronbach *et al.* (1980: 14) defined an evaluation as a systematic examination of events occurring in, and consequent on, a contemporary programme which is conducted to assist in improving this programme (see Bennett, 2003: 5). Davies *et al.* (1999: 56) consider that evaluation in a language education programme is carried out “to provide information about the programme to stakeholders (e.g. sponsors, teachers or parents), and to make decisions about the future of the programme”. In a foreign language classroom, the process of evaluation normally involves collecting information about teaching, learning, and assessments from a specific programme, and making decisions on improving or innovating the teaching and/or the assessment. Rea-Dickins (1990; 1994), surveying recent work on evaluation, suggests that there are three main purposes to evaluation: *accountability*, *development* of curriculum and course monitoring, and *awareness raising* for staff training. “Accountability” means that evaluations are carried out to benefit the different requirements of administrators, funders, institutions, or individual course instructors. In the context of university evaluation in Taiwan, “accountability” in individual institutions relates to the MOE and HEEACT evaluation projects with an emphasis on quality assurance, on teaching, and with decisions taken concerning teacher promotion.

In the past, students’ academic records were considered to be a key indicator for evaluating whether or not the teaching and testing were effective or met certain agreed criteria. However, Rea-Dickins (1994: 73) argues that the purpose of evaluation should not simply focus on judging the target objectives of a particular programme, but also take into account the quality of teaching and the “development” of both the curriculum and the staff. As Kiely (1998: 78) notes,

evaluation findings should “provide information on the technologies of teaching”, such as classroom tasks, materials, tests, or homework, thus helping to plan for the future. This implies that close attention needs to be paid by teachers when examining their usage of in-class materials and tasks. In the present thesis, the teaching materials were decided by the course instructors, so it was necessary needed to interview them in order to be able to report their opinions on the textbooks/materials they had used in class, and what they thought about the students’ reactions to them. In addition, Rea-Dickins and Germaine (1998: 12) consider that evaluation is expected to provide information regarding decision making (e.g. whether or not the teaching materials should be changed), planning (e.g. what will be done to improve the teaching), action and change (e.g. how existing teaching will be changed), based on the particular programme. What an evaluation can do to improve the curriculum is strongly related to the criteria set for inspections and the goals of the programme, otherwise the evaluation cannot provide appropriate judgements.

Although bureaucracy in Taiwan exerts a strong external force on university programme evaluation, course instructors can evaluate the lessons they teach by themselves. Roger Ellis (1997) suggests that evaluating teaching materials can be done by teachers either before and/or after the teaching – i.e. evaluation can be *predictive* or *retrospective*. Predictive evaluation helps course instructors to choose materials or textbooks that suit their teaching objectives. Retrospective evaluation, on the other hand, provides teachers with information regarding whether the materials chosen materials functioned effectively for students, and can serve as “a means of testing the validity of a predictive evaluation” (ibid: 37). By using retrospective evaluation, Roger Ellis (ibid: 37) emphasises that teachers will be able to know “whether it is worthwhile using the materials again, which

activities work and which do not, and how to modify the materials to make them more effective for future use.” He suggests that teachers can carry out a “micro-evaluation” to examine one particular task that they are interested in via students’ attitudes towards the task, or via their actual performance outcomes. This idea assumes that the feedback from “micro-evaluation” provides teachers with information to improve their teaching, classroom activities, and materials by themselves.

Rea-Dickins (1994: 76; see also Parsons and Davidson, 1989: 4) considers that the ultimate goal of evaluation is not simply to include the *antecedents* (i.e. needs analysis), *processes* (i.e. implementation of teaching), and *products* (i.e. learning outcomes), but also to raise teacher’s awareness of staff development and training needs. However, an awareness of the need for staff training is not part of the academic evaluation in Taiwan; staff research publications or conference talks will improve evaluation results, but this will not be related to staff training projects (Wu and Chang, 2007).

1.6 Quality Assurance in Higher Education

In Section 1.5, it was pointed out that quality enhancement and inter-institutional competitiveness are two significant aims of university evaluation in Taiwan. Roger Ellis (1993: 7) suggests that quality assurance for university teaching should include external examination, course validation, professional commitment, peer review, and the collection of documents giving statements of aims and objectives bearing upon students’ learning, regulations and procedures for admission, details of course contents, and examination results. Ellis makes the important point that quality assurance should be a central concern of staff,

especially where teachers are responsible for teaching, assessing, and/or examining their own teaching.

However, quality assessment can improve the policies or the institutions while affecting existing teaching systems. Brennan and Shah (2000: 13) note that quality assessment has an impact both at *institutional* level (e.g. the individual, the course, the department, the institution, or the national system) and *mechanism* level (e.g. policies and structures). They argue that although the impact is treated as the extent of “presumed improvement or enhancement” in much of the literature on quality assessment, it may challenge existing academic values and conceptions (ibid.). When it comes to decision-making, drastic demands for change in institutions or internal mechanisms may lead to an imbalance between administrators, teachers, curriculum, materials, and assessment methods. Institutions need to strike a balance between improving the programme, or alternatively raising awareness of the need for staff training, and dealing with changes affecting teachers, teaching, and policy. In short, the justification for looking at the English listening modules is that listening ability, even if not yet a part of HEEACT evaluation, now contributes, albeit to a small extent, to the quality of programmes that are evaluated, and ultimately to the quality and ranking of the entire university.

1.7 Listening to English as a Foreign Language

Processing spoken information involves phonetic, phonological, prosodic, lexical, syntactic, semantic, pragmatic, and sociolinguistic levels. In spoken discourse, listeners of English first hear sounds, then interpret them, and finally hypothesise the correct or intended meaning of words and phrases. Rost (1990; see Ellis, 2003: 39) suggests that listening involves “interpretation”, because listeners are

involved repeatedly in hypothesis-testing and formulating inferences, not merely decoding that which is said. It is also common for learners to draw on their knowledge of the world – their schemata⁵ – to help them interpret the listening contents.

In addition to background knowledge, listeners can also make predictions in terms of collocations, idioms, and proverbs which are commonly used. Ellis (2003: 45) notes that interactive listening also refers to the social processes of collaboration that listeners enter into to ensure a degree of convergence between their schematic world and that of the speaker. That is to say, listeners and speakers in an interactive communication need to reach a mutual understanding of discourse (interactions are discussed further in the following sections). However, Ur (1984: 17) argues that listeners will also fail to recognise many words they *have* learned but are not yet sufficiently familiar with when they occur within the swift stream of speech. In other words, the fact that students find the vocabulary in a spoken utterance difficult to understand may not mean that they have never learned the words before, but it is possible that they have never heard the words in spoken discourse or in that particular context. Moreover, even though students can recognise the sounds of lexis, it is still problematic for them to find correct words or make predictions due to a lack of background knowledge. In a dialogue of colloquial language, listeners may not understand every word in spoken discourse. Different pronunciations of known words, or colloquial reductions in the pronunciation of collocations also can cause difficulties. A rapid delivery is a further characteristic of informal discourse which makes it hard for foreign language learners to understand listening passages. In such a situation, listening to the sounds when said quickly in an unemphasised position in a sentence and juxtaposed with other words may affect the perception of pronunciation (ibid: 17).

This implies that there may be a gap between the words students have learned in English listening classes, and the number that they are able to successfully recognise in authentic spoken discourse.

One of the assumptions of communicative language teaching (CLT) is that teachers will use authentic texts which are not designed with contrived or simplified language aimed at learners in a classroom. Bachman and Palmer (1996: 23) define *authenticity* as the degree of correspondence between the characteristics of a given language test task and the features of a Target Language Use (TLU) task. Gilmore (2007: 98) agrees with Morrow's (1977: 13) definition of authenticity, considering that "an authentic text is stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort". Breen (1985: 61), however, states that authenticity needs to be concerned with texts, the learner's own interpretation of texts, tasks, and the actual situation of a language classroom. He argues that listening extracts are viewed as authentic when learners can develop authentic interpretations towards the discourses (Breen, 1985). Buck (2001) sees an authentic task as a task that shares characteristics with target-language use tasks. However, the classroom is an environment for learning in the sense that activities or tasks may not always duplicate events in real-life situation. Breen (1985: 66) accordingly suggests that it is better to use tasks that "generate authentic language learning behaviour which would involve communication and meta-communication during and about learning". From Breen's perspective (1985), it is important that verbal and non-verbal interaction is required in the language classroom, where the students can publicly solve the problems and share overall learning process together as "socially motivated and socially sustained activities". In short, listening

discourses, tasks, and classrooms are considered authentic when students develop authentic interpretations of discourses through interactions with other students.

However, Ur (1984: 22) notes that many listening comprehension exercises used in the classroom are based on formal spoken prose which is read aloud by the teacher or on tape. While Hedge (2000: 68) argues that contemporary listening course-books involve authentic discourses such as radio talks, news items, travel news, weather forecasts, airport and station announcements, or interviews, Ur (1984: 23) points out that it can be very difficult technically to plan and administer stretches of spontaneous speech, whether live or recorded, and that there are two main drawbacks of using recordings of authentic unrehearsed discourse.

First, being authentic, the speech used in such recordings is ungraded and the language is often very difficult, suitable only for the highest levels. Second, anyone who has listened to recordings of natural conversations knows how difficult they are to understand; without seeing the speakers it is very hard even for a native listener to disentangle the thread of the discourse, identify the different voices and cope with frequent overlaps. (Ur, 1984: 23)

There is thus a potential problem in CLT of engaging students in authentic and communicative listening materials where they are expected to understand the texts without encountering the appropriate contexts. The listening materials students come into contact with in the classroom may be different from those encountered in real target language situations they encounter, unless teachers use authentic listening texts and students can understand them to a reasonable degree. Ur (ibid.), therefore, suggests that “students may learn best from listening to speech which, while not entirely authentic, is an approximation to the real thing, and is planned to take into account the learners’ level of ability and particular difficulties”. Alternatively, using video-tapes might be a more effective way for

students to understand authentic discourses, since students can see facial expressions or gestures from videotapes instead of merely listening to the speech. For example, MacDonald, Badger, and White (2000), in their research on the authenticity of academic English listening extracts, supported the idea that it was easier for undergraduate students to understand listening extracts on videotapes than on audiotapes, because videotapes allow access to paralinguistic features. In addition, an audio-visual stimulus can help students to situate themselves in different real-life English social contexts. However, the quality of playing equipment, audio, and audio-visual tapes has to be ensured, so that the clarity of spoken discourse is not impaired. In the real-life situation of Taiwanese university lectures, however, the speakers are usually visible to listeners, in the sense that listeners are able to see facial expressions, or body language to aid their understanding. What happens in listening classes is unknown; it is possible that teachers use audio-visual tapes to support students' listening input, but this needs to be established.

It has been argued that using discourse and tasks which are similar to those in the real world helps to predict a candidate's ability to communicate in real-life situations (Alderson and Banerjee, 2002: 98). Spence-Brown (2001) carried out a study asking learners of Japanese at an Australian university to complete an authentic interview task by using the language skills they had learned from their Japanese courses to engage in authentic conversations with Japanese native speakers. He discovered that authenticity was important with respect to implementation as well as task design, since those students who focused on the assessment outcome of the interview tasks, showed a lack of engagement with the task as interaction.

Bachman and Palmer (1996: 23) claim that authenticity is important in considering candidates' test performances, because they consider authenticity to increase the ability to generalise from the test scores to real-life situations, and to be more indicative of a candidate's actual linguistic ability (see also Alderson and Banerjee, 2002: 99; Davies et. al, 1999: 50). Although Bachman and Palmer regarded authenticity as an important element in testing listening, Lewkowicz (1997, 2000), in her study of using authentic discourses to test the English of learners in Hong Kong, discovered that the learners cared more about their familiarity with the task type than the authenticity of the discourses provided whilst taking English listening tests. The results of Lewkowicz's study show that using authentic discourses to test foreign language students, in the Hong Kong case, appeared to have no influence on the test scores. It is possible that the practical situation of using the target language can be perceived as less important than the need to pass the tests. In the situation of Taiwanese universities, how often authentic listening extracts and tasks are implemented in the classroom is unclear.

1.8 Teaching English Listening in the Foreign Language Classroom

Teaching English listening in the foreign language classroom involves (1) the objectives or goals of the lessons, tasks, or activities (i.e. what teachers intend to teach their students), (2) teaching approaches (i.e. how teachers teach students), and (3) classroom interaction (i.e. how students interact with each other and with the teacher).

1.8.1 Teaching Objectives

Teaching is effective when there are clear objectives related to appropriate teaching materials, methods, assessments, and students' needs. There are three types of objectives in education: global, educational, and instructional (Krathwohl and Payne, 1971; cited from Banks, 2005: 6). *Global objectives* refer to stated standards and objectives which encompass curriculum guidelines in a broad sense (Banks, 2005: 6); for example, "students will listen to different types of listening materials and practice speaking skills". *Educational objectives* in teaching serve as "an intermediate level of specification for curriculum and instructional decision" (ibid: 7); for example, "students will improve their listening and speaking skills". *Instructional objectives* are specific to particular classroom tasks or activities, for example, "students will describe in their own words a listening passage they have listened to".

Banks (2005: 7) also highlights the fact that classroom instructional objectives are directly related to the purpose, goals, lesson plan, and assessment outcomes for a particular course. Teaching can be more effective if teachers plan lessons on the basis of students' learning needs and goals. Parrott (1982: 5) claims that without goals, students' achievements become random and accidental rather than controlled and predictable. However, it may be difficult for all students in a class to reach a unanimous agreement on the topics which they wish to learn, and teachers have to strive for a balance between different learning needs. Determining learning needs does help teachers to decide teaching approaches, relevant learning activities and tests.

1.8.2 Approaches to Teaching Listening

Both Mendelsohn (1998: 81) and Hedge (2000: 228) note that there has been a shift in listening instruction from the audiolingual approach to communicative

approaches over the past 50 years. Audiolingual teaching of listening provided only restricted practice of scripted dialogues, and had the main aim of presenting and practising language forms. Rivers (1981: 41-43) and Jones (2002: 178) state that the audio-lingual method had the following features: firstly, it put great emphasis on listening and speaking in terms of correct pronunciation and intonation. Students were encouraged to listen to native speakers' utterances and repeat them (Rivers, 1981; Jones, 2002). Secondly, "mimicry-memorisation" and "structural pattern drilling" of decontextualised words and sentences were also promoted, to provide learners with "automatic control" of the framework of the target language without paying attention to the forms (Rivers, 1981). In addition, contemporary colloquial language was taught. As structural linguists rejected the notion of a universal grammatical system for all languages, they considered each language had unique interrelationships with its culture. In addition, the focus on tight instructional control, exercises and practice, and avoidance of error in audiolingual teaching assumed that students could understand the authentic language they heard outside the classroom (Wesche and Skehan, 2002: 209). In other words, audiolingual listening teaching put the emphasis on teaching through a restricted practice of pattern drills, vocabulary, and pronunciation. Rivers (1981: 47) argues that if students are trained to make variations on language patterns without being given a clear idea of how the target language is used appropriately while performing it, they may not understand the possibilities and limitations of the operations they are performing. Also, research reveals that students who speak fluently and accurately in practicing pattern drills may fail to transfer such "gains" to actual communicative language use (Cohen, Larson-Freeman, and Tarone, 1991; Jones, 2002: 180). The deficiencies of audiolingualism imply the importance of keeping learners active in the classroom by reducing the amount of

teacher-talk and of highly controlled drills (Allwright, 1984: 156); otherwise drilling can foster students' pronunciation rather than real use of language.

Teaching listening skills covers aurally discriminating between different words (minimal pairs), segmentation and identification of words in continuous speech (sentence level), extrapolation of unrecognised words (guessing the spellings of difficult-to-recognise cognates), anticipation of listening contents, and identification of information which is related to the tasks (Ur, 1984: 35-46; Field, 1998: 114). In short, the skills involved in teaching listening include breaking the listening texts into word and/or sentence units, and teaching the students how to identify them morphosyntactically; this essentially requires a bottom-up processing approach⁵ to decoding linguistic input rapidly and accurately, where input refers to the written or spoken language to which students are exposed (Davies et al., 1999: 83), and mapping the input against these expectations to confirm consistencies or to refute implausible interpretations (ibid.) at a detailed level. However, Nunan (1991: 18) argues that successful listeners are those who can utilise both bottom-up and contextually sensitive top-down processing to interpret what they hear; this implies learners need to use schemata⁶ and contextual information, as well as phonological information, to interpret speech, to create plausible expectations of what they are about to hear, and to understand the contexts at a general level (see Section 1.7; Tsui and Fullilove, 1998: 433; Nunan, 2002: 239; Ellis, 2003: 45). Anderson and Lynch (1988; see Nunan, 1991: 18) view bottom-up processing learners as "tape recorders", who focus on decoding linguistic form, but top-down processing learners as "model builders" who listen for meaningful information and use their schemata to comprehend discourses. There is some evidence showing that skilled listeners are better able to use top-down processing, while less-skilled listeners tend to rely more on bottom-up

processing (Hildyard and Olson, 1982; Shohamy and Inbar, 1991). Indeed the two modes can interact; in Tsui and Fullilove's study of testing Hong Kong university students' listening (1998), it was found that less-skilled listeners, who were weak in bottom-up processing tended to use plenty of contextual support to compensate for the lack of linguistic decoding skills. Tsui and Fullilove accordingly conclude that bottom-up processing and top-down processing are both important in teaching listening. In Taiwanese context, the government requirement that universities teach listening (and other skills) for real life application implies that teachers should focus at least part of their lessons on teaching top-down processing. However, it is unknown how far they do actually focus on either type.

Another aspect of learning to listen in a second language involves learning to use strategies. Strategy teaching for listening is heavily endorsed by researchers like Mendelsohn (2005), Cohen (1998), and Cohen and Macaro (2007). However, proponents of strategies take differing approaches to how teachers and learners should proceed. Mendelsohn (1995: 134), for example, suggests that learners need to be aware of how the language functions first, and then teach the strategies so that learners can use them in tackling the listening tasks they encounter. Flowerdew and Miller (2005), on the other hand, suggest that strategies should be directed toward educating L2 learners to identify their own preferred strategies for listening. To this end, they propose that listening strategies can be explored by having learners check a summary of listening strategies such as that developed by Vandergrift (1997: 392-394). However, Field (1998: 116) argues that current proponents of strategy-based teaching do not make a clear enough distinction between listening strategies which are used for extracting meaning ("communication strategies") and those which are used for the purposes of acquiring new language ("learning strategies"). For instance, students may

resolve a linguistic item but may not learn how to use it, and it is also possible that strategies taught in class could prove difficult to apply in real-life situations. In addition, individual students may use diverse strategies. If teachers just introduce strategies and expect students themselves to work out which are suitable for them, not all students will succeed in finding appropriate ones and the teacher will be left with a serious pedagogical problem. On the other hand, it is likely to be time-consuming for teachers to establish the most appropriate strategies for each individual student. Field (1998: 117) tries to resolve the situation by arguing that skills may be seen as competencies which native listeners process and which non-natives need to acquire in relation to the language they are learning; skill learning thus involves mastering (inter alia) auditory phonetics, word-identification techniques, and the patterns of sentences in the target language. Strategies, by contrast, are better seen as strictly compensatory and as a listener's ability improves, they can be dropped (Field, 1998: 117). Essentially, he argues that compensatory strategies tend to be developed for learning the first language. They remain relatively dormant in later years but can be easily reactivated in an L2 context. The implication is that teachers do not need to spend too much time or effort on strategies, but should focus more on the teaching of skills.

In addition to the effects of skill- and strategy-based teaching, Flowerdew and Miller (2005: 87) note that learning to listening can be influenced by individual variations such as attitude, personal interest, and particularly motivation. It is generally agreed that high motivation is important to learners while learning second or foreign languages (Dörnyei, 2001). In order to encourage students to learn in class, it is necessary to explore students' listening needs (Jordan, 1997), and at the very least to design motivating tasks or activities that encourage them to learn in class. Thus, it is important for teachers to choose

listening materials that can motivate their students and that are appropriate to their level of English proficiency.

Field (2002: 242; 1998: 110) notes that in the late 1960s and early 1970s, listening comprehension lessons tended to follow a three-stage format:

1) Pre-listening Stage

– pre-teaching of new vocabulary in passages;

2) Listening Stage

– extensive listening (followed by general questions in contexts)

– intensive listening (followed by detailed information)

3) Post-listening Stage

– analysis of language in the text

– listen and repeat

Linguistic forms were strongly emphasised. Vocabulary and grammar were consciously taught rather than letting students make inferences from listening passage contexts. Drills and the reproduction of sentences were frequently used in classroom exercises. Field (2002: 245; 1998: 110) goes on to suggest that “a present-day listening lesson” is likely to be very different from the model in 1960s, and to include some or all of the following stages:

1) Pre-listening Stage

– set context and motivation

2) Listening Stage

– extensive listening (followed by questions on context)

– preset task/preset questions

– intensive listening

– check answers

3) Post-listening Stage

– examining functional language

– inferring new vocabulary

In this situation, listeners engage in the listening contents or a discussion of its topics first, and then in listening to contents and doing tasks before vocabulary and grammar features are taught at the end of the listening. The procedure for teaching listening ideally involves the three stages that Field demonstrates.

However, it is unclear what procedures are used for teaching English listening in Taiwanese universities.

1.8.3 Task-based Instruction (TBI) in Teaching Listening

As far back as 1984, Ur suggested that listening exercises are most effective if they are constructed round a *task* (1984: 25). This implies that a form of task-based teaching is the preferable approach to teaching listening. Nunan (1991) strongly supports this idea that tasks are useful in teaching listening, since tasks can be determined differently based on the listener's purposes and needs. Pica, Young, and Doughty (1987; see Ellis 2003: 23) also agree, arguing that tasks can be modified to investigate the relationship between different inputs and students' responses in classroom. Ellis (2003), taking an overview of Second Language Acquisition, further suggests that using tasks to elicit a sample of communicative language can be pedagogically useful, as it can help a teacher analyse learners' use of specific linguistic features. In the situation of teaching listening, the main idea behind a task-based approach to listening is to encourage students to become *active* learners (Brown, 1987; Flowerdew and Miller, 1995: 14). Flowerdew and Miller (1995: 14) note that, with task-based instruction, students are asked to listen to what are described as "authentic" situations and "do something" with the information. They argue that "doing something" in a task does not simply mean reporting or repeating everything in a spoken discourse in a way reminiscent of audiolingual classroom (see Section 1.8.2). A listening task requires students to listen to the discourse, and to use their language skills and background knowledge, as well as to apply top-down and bottom-up processing approaches, and individual compensatory learning strategies to resolve a problem. Also, task-based

instruction involves asking students to reflect on what they do and how well they do it in class. This allows task-based listeners to reflect on their learning processes and this criticality plus the active engagement are held to trigger “deep processing” which in turn leads to retention in long-term memory and to learning (Hulstijn, 2001). Flowerdew and Miller (1995: 14) thus consider the process students employ in finding a successful outcome to a task is pedagogically more important than being able to understand every single word in the discourse. In the Taiwanese university context, the requirement for an English-speaking environment and the ability to operate in real-world contexts would seem to imply that teachers need to employ if not a pure task-based approach, at least the key assumptions of TBI.

Given the above, it becomes important to establish what the term “task” is considered to mean and to just what the key assumptions are. There are now many definitions of pedagogical tasks in the relevant literature. Richards *et al.* (1986: 289), for example, consider a pedagogical task as an activity or action which learners carry out enabling them to develop their understanding of a language, and which “may or may not involve the production of language” (e.g. drawing a picture while listening to a tape). Breen (1987: 23), taking a different view, assumes that “tasks” refer to a range of activities which have “the overall purposes of facilitating language learning”, such as group problem-solving, simulations, or decision-making. Long and Crookes (1992: 43), based on their ideas of second language acquisition (SLA), claim that pedagogical tasks provide a vehicle for the presentation of appropriate target language samples to learners – input which they will inevitably reshape via application of general cognitive processing – and for the delivery of comprehension and production opportunities of negotiation. Long and Crookes (*ibid.*) argue that tasks not only facilitate

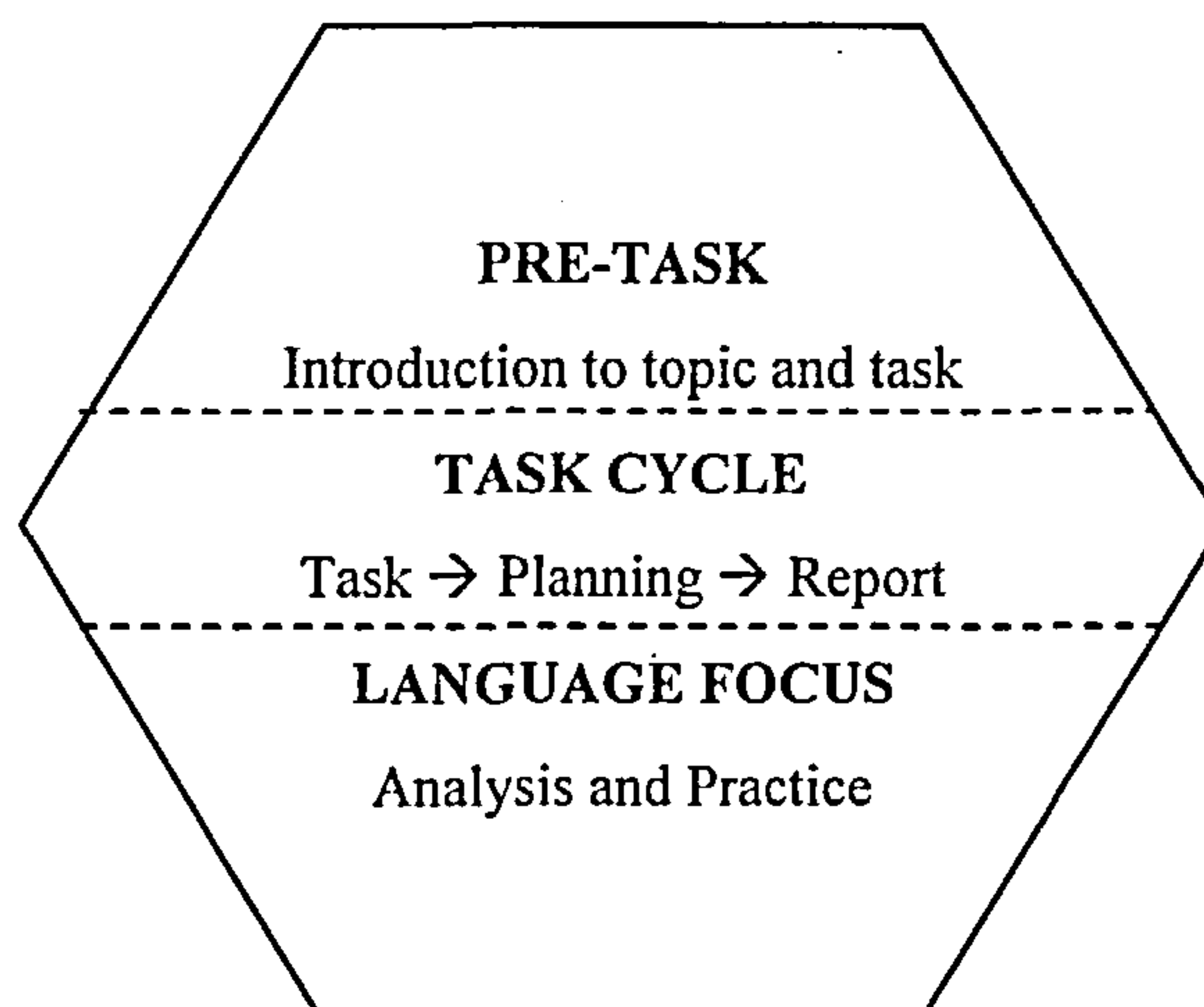
acquisition for pedagogical purposes but also have non-pedagogical outcomes, including buying foods, making reservations, or seeing a doctor. Ellis (2003: 16) and Nunan (2004: 4) define a pedagogical task as classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language, in order to express meaning rather than manipulate form. Students are required to do goal-oriented listening activities or exercises in response to what they hear in the classroom (rather than in the outside world) demonstrating their understanding (J. Willis, 1996: 53; Nunan, 1989: 6). The purpose of having communicative goals is that students can practice exchanging information, ideas, opinions, and attitudes via interpersonal communication; for example, sociocultural tasks require students to understand everyday life patterns in the target language speech community.

As in CLT, tasks in task-based instruction should focus on meaning, relate to real-world activities, assess outcomes, and encourage pair or group work (Skehan, 1996; Nunan, 1989; Wesche and Skehan 2002: 217). J. Willis (1996: 53) suggests that students should be exposed to “a rich but comprehensible input of real language”; which means that the listening input (i.e. in-class materials) should consist of authentic and spontaneous language use. In addition to real input, students need to be provided with opportunities to speak the language (*ibid.*), in the sense that students need to use the target language to express ideas, exchange information, and interact spontaneously with other students. As both CLT and TBI assume that teaching is more acceptable, meaningful, and encouraging to learners when learners are able to perform communicative functions, J. Willis (1996: 54) notes that while doing tasks, students are focusing on meaning what they say, and exchanging information for real purposes which “replicate(s) features of language use outside the classroom”. Thus, the success of

implementing a task is judged by whether or not students communicate successfully (Willis and Willis, 2007: 5). Similarly, Savignon (2005: 636), in her pioneering research on adult classroom second language acquisition, characterises communicative competence as “the ability of classroom language learners to interact with other speakers, to make meaning, as distinguished from their ability to recite dialogues or to perform in discrete-point tests of grammatical knowledge”. However, tasks that do not include authentic and spontaneous use of language for speaking practice and for display of listening inputs do not count as tasks (J. Willis, 1996: 54). Also, role-plays and reproductions at syntactic levels do not count as tasks, since there is no real exchange of meanings involved. In addition to the meaning exchange in tasks, Willis and Willis (2007: 11) suggest that a task should incorporate activities that help promote students’ interest and interactions. In the Taiwanese university context, it is simply unknown how far task-based instruction is implemented in the English listening classes. It would be particularly useful to know how far authentic language and materials are used in the listening classrooms, whether or not the listening tasks involve oral interactions between groups or pairs, and if the focus meaning is primarily on forms.

While implementing TBI, J. Willis (1996: 53) summarises task-based instruction with the following framework (Figure 1.2):

Figure 1.2 Framework for Task-based Instruction (Willis, 1996: 53)



At the “Pre-task” stage, teachers introduce listening exercises or tasks to the students, and help them understand the objectives of the task. In the listening situation, students are given time to listen to extracts before doing the task properly. Next, at the “Task Cycle” stage, students resolve the task in groups or in pairs by using available language and knowledge; spontaneous and fluent talk in oral discussions is recommended. The teacher needs to be physically active and walk around and monitor the students’ discussions. Students are then asked to prepare for reporting the findings of their discussions by using accurate and organised language either in oral or written form. While reporting the findings, the teacher can ask students to compare findings from different groups or pairs. After discussing or sharing the task problem that the learners have solved, the teacher moves on to the “Language Focus” stage, where the grammar or forms of language relating to the topic of the task are analysed and explained, either by the teacher or by the students themselves. Teachers may give feedback to help students improve their performance in class (or their learning) by using quizzes or tests to assess whether the learning outcomes have been met. Wiggins (1998: 43) suggests that constant isolated drill work and testing without concurrent effective feedback means that answers are isolated from actual effects, causes, and purposes in terms of the students’ learning outcomes. Ur (1984: 28) too notes that students should be given immediate feedback on their performance of the task, particularly in listening and speaking tasks. Unlike receiving feedback in reading and writing assignments where students can reread the text, students should receive immediate feedback before they forget what they have listened to or spoken. Moreover, Wiggins (1998: 46) suggests that feedback should be “highly specific, directly

revealing or highly descriptive of what actually resulted” and closely related to their individual task performance, rather than providing general praise such as a “good job” or “excellent”.

In addition to the feedback on tasks or exercises, students should have an opportunity to reflect on the language they have learned and how well they are doing at the end of the task (Nunan, 2004: 37). Developing learners’ awareness of the need to reflect on their learning is not always easy in the language classroom, since it depends on how and what the teacher does to encourage and to engage their students in reflection, and/or employ learning strategies. Listening tasks that involve students in making predictions of language use, in monitoring and evaluating the learning process, and in solving problems can also help students “develop metacognitive knowledge that is critical for the development of self-regulated listening” (Vandergrift, 2007: 197). In addition to using language and in-class materials that replicate real-life situations, it is important to know whether reflection periods are given in the Taiwanese classes to students before the end of each lesson.

Nunan (2004: 42) specifies that goals may not always be explicitly stated, although they can usually be inferred from the task itself. Students do, however, have to perceive task objectives as intended. If the intentions are unclear, or subject to misinterpretation, students are unlikely to learn what the teacher wants them to learn (Malamah-Thomas, 1987: 41). Kumaravadivelu (1991), investigating the relationship between teachers’ intentions and students’ interpretation of tasks, uncovered serious mismatches between how students interpreted task objectives and what the teacher asked them to do (see Ellis, 2003: 40). One result of such mismatches was that students sometimes completed the tasks in a wrong way when they misinterpreted the teacher’s intentions. The

general conclusion is that it may be difficult for teachers to assess students' learning outcomes based on performances which do not meet the teaching and learning goals.

Teachers using task-based teaching need to ensure that the levels of tasks are neither too difficult nor too easy, and also to ensure that students do in fact achieve the goals in each task. Using tasks in teaching English listening aims at strengthening students' English ability in terms of understanding, digesting, and resolving tasks they complete. Such task-based teaching also expects students to be able to apply what they learn in the classroom to actual target environments. In addition, tasks that involve a lot of reading (such as reading questions with long and/or complex sentences) or writing (such as taking notes) lead to difficulties where students have to concentrate on listening while interpreting the questions and writing down correct notes. Such tasks are no longer just listening exercises, but reading or writing ones, and this may threaten the achievability of the goal. Ur (1984: 26) emphasises that if teachers want to concentrate on aural comprehension tasks, it is best to base the task on easily grasped visual materials (pictures, diagrams, grids, or maps) and quick simple responses such as physical movements, ticking-off, or one-word answers. Picture-based tasks are suitable for learners at different ages. It is important to keep the pictures informative and adequately detailed, not involving irrelevant information that may confuse listeners. Diagrams can also be used since the advantage of diagrams as the basis for task-centred activities derives from the fact that diagrams can be designed to convey a large number of facts clearly and quickly without necessitating a heavy reading load (ibid: 31). Even so, diagrams should not be unnecessarily complicated. Any writing on a diagram should be kept to a minimum; reading and

listening interaction takes much more time and students will probably miss information in the listening passage.

To summarise, key or essential task-based instruction features in a foreign language classroom include at least (1) a problem-solving task(s) for students to do in class, (2) opportunities for students to speak English in class, (3) findings reported by students in pairs or groups, (4) the major focus on being the meaning and then on the form, (5) reflection periods for students, and (6) authentic in-class materials. It is important to investigate how far task-based instruction, based on the six characteristics, is implemented in Taiwanese universities.

1.8.4 Current Implementation of Task-based Instruction in Taiwan

The general notion of TBI has been discussed in 1.8.3; current research on TBI in Taiwan focuses very much on teaching pupils in primary and secondary schools (Lun, 2004; Chiang-Fan, 2004; Lin, 2004; Tseng, 2006)². Lun (2004), in his research on vocational secondary school students' perceptions of TBI in English classes, found that TBI had positive effects on language development, including all four skills and the ability to self-monitor. Secondly, TBI provided the students with an enjoyable learning atmosphere and with opportunities to develop their positive learning attitudes, independent thinking, creativity, and self-esteem. Next, TBI helped the students develop interpersonal skills and communicative skills. The findings of Lun's study suggest that TBI is a potential alternative to teacher-centred teaching in the Taiwanese context and that it facilitated the secondary students' English learning.

With respect to using TBI in primary schools, Lin (2004) and Tseng (2006) found that task-based activities improved learners' social skills, their linguistic performance on all four skills, and their affective development. In Chiang-Fan's

(2005) studies, the results suggest that task-based activities enhanced pupils' motivation and attitudes towards learning English. However, Chiang-Fan (ibid.) argues that time limitations and the lack of professional task-based teaching materials made it more difficult to implement TBI, so she suggests that teachers should adjust their teaching approaches based on students' needs and preferences to make their teaching more effective. In short, based on the recent studies carried out in primary and secondary schools in Taiwan (ibid.), using "tasks" to teach pupils had a positive influence on students' performance, English skills, and communicative ability. I could find no equivalent Taiwanese research on TBI at university level after the promotion of all-English (or bilingual) teaching in any language courses. Also, the teaching approaches that the teachers used in university English listening classes remain unknown.

1.9 Summary

In the early 1990s, the educational reforms of Taiwanese higher education allowed individual academic institutions more autonomy to mobilise teaching and learning within their universities, and also tried to establish an evaluation system. After Taiwan joined the WTO in 2002, the resulting economic changes had a great influence on higher education; there was a need for more universities, more teaching resources, and the creation of an English environment, and this spurred the development of a national evaluation organisation – HEEACT. At the same time, the MOE in Taiwan strongly promoted the use of English in classroom teaching at university level, and included English courses in its evaluation schemes in 2002. English listening ability is now accordingly an essential language skill for Taiwanese university students, and more than 80% of Taiwanese universities include English listening modules as compulsory and

assessed training courses. One key aspect of working in English in regular classes is the ability to listen and understand English. However, the 2006 IELTS and TOEFL test scores appear to show that the students listening skills were lower than their other language skills (see Section 1.3). It is important to investigate how general English listening modules are taught and assessed, and whether the courses are taught and assessed based on curriculum and test objectives agreed by the institutions. Since it is considered necessary to create an English learning environment to confront the socio-economical change in higher education, task-based instruction would appear to be an effective way to teach listening (see Section 1.8.3). Thus, it is essential to examine how far communicative or task-based instruction is implemented in teaching and testing English listening.

Notes to Chapter One

¹ KMT was the ruling party in Taiwan from year 1949 to 2000.

² The population in Taiwan decreased from 3.45 million people in 1985 to 2.29 million people in 2007 (Kuo, 2005; National Statistics in Taiwan, <http://www.stat.gov.tw/>).

³ IELTS mean band score by most frequent Asian countries or regions of origin (Academic) – Test-taker performance 2006. Test scores in 2007 were unavailable at the time of writing, January 2008.

Country	Listening	Reading	Writing	Speaking	Overall
Bangladesh	5.58	5.38	5.38	5.62	5.55
China	5.47	5.80	5.23	5.39	5.53
Germany	7.44	7.23	6.75	7.26	7.23
Hong Kong	6.70	6.75	5.91	6.06	6.42
India	6.30	5.82	5.79	6.10	6.07
Indonesia	6.10	6.27	5.43	5.83	5.97
Iran	6.04	5.96	5.81	6.31	6.09
Japan	5.87	5.86	5.33	5.80	5.78
Korea	5.87	5.87	5.36	5.72	5.77
Malaysia	6.93	6.85	6.13	6.41	6.64
Nepal	6.34	5.79	5.71	5.88	5.99
Nigeria	5.65	5.84	6.22	6.93	6.22
Pakistan	5.83	5.58	5.49	5.86	5.75
Philippines	6.68	6.27	6.18	6.74	6.53
Russia	6.49	6.48	5.98	6.68	6.47
Sri Lanka	6.27	5.97	5.93	6.39	6.21
Taiwan	5.52	5.81	5.23	5.66	5.62
Thailand	5.82	5.89	5.28	5.70	5.74
United Arab Emirates	4.99	5.10	4.86	5.43	5.16
Vietnam	5.59	6.01	5.56	5.70	5.78

On-line Source: <http://www.ielts.org/teachersandresearchers/analysisoftestdata/article382.aspx>

⁴ TOEFL CBT scores by Asian countries between July 2005 and June 2006. Test scores after June 2006 were not published on line at the time of writing – January 2008.

Country	No. of Examinees	Listening	Structure writing	Reading	Total Score Mean
Afghanistan	99	18	19	18	182

Azerbaijan	226	21	22	21	214
Bangladesh	1287	22	22	23	228
Cambodia	74	21	21	20	206
China, People's Republic of	9,017	20	23	22	216
Hong Kong	5947	21	22	22	216
India	72,973	23	24	23	236
Indonesia	4,641	21	21	21	214
Japan	78,635	18	19	20	192
Kazakstan	1,198	22	22	22	217
Korea DPR	4,203	19	19	20	193
Korea ROK	128,445	21	22	23	218
Kyrgyzstan	108	23	23	23	232
Macau	241	19	20	20	196
Malaysia	1,998	23	23	23	232
Mongolia	132	21	20	20	202
Myanmar Burma	138	20	21	21	206
Nepal	5,027	21	23	21	218
Pakistan	4,258	23	25	23	238
Philippines	6,389	24	24	23	238
Singapore	456	26	26	25	255
Sri Lanka	162	24	23	23	234
Taiwan	33,327	19	21	21	206
Thailand	13,162	19	20	21	200
Uzbekistan	675	21	22	22	218
Vietnam	705	19	22	21	207

On-line Source: <http://www.etsliteracy.com/Media/Research/pdf/TOEFL-SUM-0506-CBT.pdf>

⁵ According to Lynch and Mendelsohn (2002: 197), schemata are used to refer to a “package” of prior knowledge and experience that we have in memory and can call on in the process of comprehension. Listeners organise their knowledge of the world by using three types of schemata: (1) general factual knowledge, (2) local factual knowledge, and (3) socio-cultural knowledge to interpret key lexical items (Anderson and Lynch, 1988). Listeners use these schemata to comprehend a discourse in three major ways: interpretation, prediction, and hypothesis testing to recognise key lexical items, to make prediction, and to confirm/disconfirm predictions (Ellis, 2003: 41).

⁶ According to Field (2003: 20-21), bottom-up processing in listening is defined as the process whereby listeners build acoustic features into phonemes, phonemes into syllables, syllables into words, words into syntactic patterns, and connect syntactic patterns into propositional (abstract) meaning. Bottom-up processing is essentially “data-driven” and relies upon linguistic form. Top-down processing, on the other hand, is defined as processing which involves the use of information stored in existing schemata (relating to events, scenarios, or just words) to support hypotheses about meaning and the words used. Top-down processing is “knowledge-driven” in that it relies more on external information.

Chapter Two

English Listening Assessment and Washback Effects

2.1 Introduction

As noted in Chapter One, course instructors in the English listening courses in Taiwanese universities are responsible not only for their teaching, but also for designing and processing tests. In such a situation, it becomes particularly crucial for the classroom teacher to have a clear idea of what will be assessed and how the assessment will be conducted (Cohen, 1994). In addition, it is recognised in the literature that listening comprehension can be influenced by different variables; Rubin (1994: 199), for example, identified five major factors that affect listening: (1) text characteristics (variation in text type, accents, and speech rate); (2) interlocutor characteristics (variation in the speaker's personal characteristics); (3) task characteristics (variation in the type of comprehension questions, topics, and test rubrics); (4) listener characteristics (variation in the listener's language proficiency level, background knowledge); and (5) process characteristics (the listener's interpretation of texts). Lynch (2002: 43) supports the view that listening test scores are affected by the input, the task, and the listener. In language testing, the term "input" is used to denote the information or stimulus material contained in a given test task (Davies et al., 1999: 83). The input can be forms of written language texts, non-linguistic pictures, oral utterances or discourse from tapes or live speakers.

This chapter begins by reviewing the nature of classroom assessment. Next it discusses the three main approaches to listening assessment, such that details of good and problematic practice can be applied to the tests to be examined in Chapters Five to Ten. Thirdly, the validity of assessment is examined, primarily

focussing on content and construct validity. Text and task characteristics are then considered. The chapter ends with a discussion on washback effects on teaching.

2.2 The Nature of Classroom Assessment

Most classroom-based language assessment is likely to be criterion-based. Students are measured with reference to their degree of achievement on the course, rather than evaluated by comparison with the achievement of other students. Brown and Hudson (2002: 31) classify three different types of criterion-referenced evaluations in curriculum assessment: diagnostic testing, progress testing, and achievement testing. While criterion-referenced diagnostic testing is often done at the beginning of a course and progress testing is usually carried out between the beginning and the end of the course, achievement testing is normally done at the end of the course.

The objectives of a curriculum can be partially determined by needs analysis, in that the teaching materials are collected and analysed to satisfy students' learning requirements within particular teaching and learning contexts. Jordan (1997: 56) states that designing a curriculum involves examining needs analyses and establishing goals, which entails the selection, grading, and sequencing of the manageable teaching content and assessment methods. A major advantage of classroom assessment is that it helps to clarify the course objectives. In a similar way to the three teaching objectives – global, educational, and instructional – discussed in Chapter One, Brown and Hudson (2002) define testing objectives in terms of three different aspects – educational, instructional, and performance objectives. *Educational objectives* describe what teachers expect students (a) to be able to do at the end of a course, or (b) to “be able to do in a specific domain of knowledge, skills, or abilities” (Brown and Hudson, 2002: 36); for example,

“students are expected to understand how and why to use specific research methods in individual case study and learn how to analyse the data”. *Instructional objectives* are specific in the sense that students are expected to perform what they learn under a particular course criterion in order to pass it. *Performance objectives* refer to students’ ability to perform in the language, or to use it to accomplish certain tasks in observable ways, such as speaking or writing. Anastasi and Urbina (1997: 115) argue that a well-constructed educational test should cover the objectives of instruction, and not just its subject matter. Hughes (2003: 55) further notes that if classroom tests are based on specific objectives rather than on detailed teaching or textbook contents, they will provide “a truer picture of what has actually been achieved”; he too agrees that teaching and assessment are not separate issues, and teaching objectives should be incorporated into classroom assessment. In the present study, therefore, it will be important to investigate how far test objectives are clearly specified in relation to the curriculum and teaching contents.

Classroom assessment allows teachers to better plan their lessons by taking into account the strengths and weaknesses of their students, and this in turn serves as quality control of programmes, teachers, and the curriculum in the educational system concerned (Muijs and Reynolds, 2001: 185). In other words, the purpose of classroom-based language assessment is not only to measure how much knowledge and skills students have learned from the course, but also to help to maintain the effectiveness and efficiency of the language teaching. The mid-term exams which the Taiwanese students receive as part of their listening modules can be considered as progress testing. So, on the one hand, the mid-term exams can benefit the teachers, allowing them to revise the teaching and the in-class materials for the rest of the programme (i.e. between the mid-term and final exams). That

is to say, teachers may modify existing materials, or create additional new ones, based on students' performance in the mid-term exam, which will help them ensure the quality of teaching and learning. On the other hand, the mid-term exams help the students to review their own learning and approaches to study. The final exam, however, is better seen as an achievement test, to assess students' achievement at the end of the course. However, studies on second language learning argue that there is rarely a one-to-one correspondence between what is taught and what is learned in the language classroom (Ellis, 1990; Larsen-Freeman and Long, 1991; Cohen, 1994: 16). The results of assessment may accordingly be unexpected due to different perceptions or interpretation of teaching and learning goals from teachers and learners.

Students may not learn what is taught or may learn only partially or even incorrectly. Sometimes they learn incorrectly because they pay inadequate attention or because they do not have the proper basis for comprehending the materials – a basis gained, for example, from coming to class regularly, doing the homework, or having exposure to the language out of class. (Cohen, 1994: 16)

Teachers can use either standardised test items developed by professionals or develop tests themselves. Muijs and Reynolds (2001: 189) claim that teacher-made tests are better matched to the curriculum and students' level of progress than standardised tests. Whatever test methods teachers select or develop, test contents should match what is taught, and test items should be reliable and valid for their purposes. Teachers should try to maximise reliability and validity in their tests by using unambiguous items and test instructions or by increasing the number of test items (Gronlund, 1968: 10). Students should also be provided with feedback, and the teacher should recap these items that the students have demonstrated problems with after the test to ensure students'

understanding of the items. Shohamy (2001; see Adair-Hauck *et al.*, 2006: 362) suggests that in the absence of feedback, test takers are utilised by institutions or assessment bodies to meet their testing agenda, and the test takers receive little or no benefit from the tests. Since the test objectives in a criterion-referenced test must be specific, a negative aspect of criterion-referenced tests noted by Henning (1987: 7) is that the test objectives are often too limited, or restricted to certain topics. Harrison (1983) also argues that achievement testing is vulnerable to problems of sampling, since what has been learned in a term cannot all be assessed in one test, yet the test must somehow reflect the content of the whole course. The choice of test tasks which are representative of the teaching and testing objectives is thus important in classroom-based assessments, and something that will need to be examined in the present study.

2.3 Three Main Approaches to Assessing Listening

It was discussed in Chapter One how communicative or task-based instruction has been promoted by the government and found to enhance acquisition at primary and secondary levels. As teaching and assessment are inseparable in classroom-based settings, using communicative teaching implies a use of communicative testing. In the history of language testing, according to Buck (2001: 61), there have been three main methods of testing: discrete-point testing, integrative testing, and communicative language testing.

2.3.1 Discrete-point Testing

Discrete-point testing focuses on testing students' knowledge of the grammatical system, vocabulary and of aspects of pronunciation tested via decontextualised linguistic questions (McNamara, 2000: 14). Listening comprehension is often

viewed very much as a process of recognising the sounds of a language, and Lado (1961: 218) recommends that items such as segmental phonemes, stress, intonation, grammatical structure and vocabulary should be measured by using true/false, multiple-choice questions, and pictures (Buck, 2001: 62). Dichotomous items like true/false questions are easier and quicker to construct and score than multiple-choice questions, in the sense that test-takers simply respond by judging statements as true or false. However, the obvious disadvantage of this format is that test-takers can have a fifty percent chance of getting the questions correct by random guessing (Brindley, 1998: 177; Buck, 2001: 147; Alderson, 2000: 222). Some testing manuals suggest adding a third option of “no information” or “can’t tell from the text” to minimise the possibility of guessing (Carroll and Hall, 1985; Rivers, 1981). Nevertheless, Burger and Doherty (1992: 315) report that the “true-false-not given” format did not work well in an English as a Second Language listening test they developed at the University of Ottawa, because listeners tended to focus on what was said rather than what was not said (see also Brindley, 1998: 178; Buck, 2001: 147). As there is no text for test-takers to refer back to what is not mentioned, listeners have no means of checking “not given” information.

Since discrete-point testing usually uses selected or restricted responses, multiple-choice questions have become the main type of discrete-point testing, involving the selection of one correct answer out of three or four options. However, not all multiple-choice questions test discrete-point knowledge. Buck (2001: 63) argues that constructs underlying multiple-choice test items may be different, even though the format looks the same. In some cases, multiple-choice questions test overall comprehension and even inferences from listening extracts but still remain within the format of multiple-choice tests. Lado (1961: 218),

cautions testers not to use too much context in language tests, suggesting that the context should be enough to resolve any ambiguity as to the problem at hand, but no more. The notions of isolating language forms from a stream of speech and asking students to recognise them correctly from decontextualised listening extracts tend to ignore the fact that language processing involves making inferences and predictions based on contexts.

2.3.2 Integrative Testing

Since the exclusive focus of discrete-point testing is on assessing formal linguistic abilities rather than on assessing language ability with understanding contexts, Oller (1979) suggests that integrative tests should be used to assess learners' capacity to *use* many bits of language at the same time, rather than their ability to *know* about them. Integrative tests require test takers to combine various language skills to answer test tasks, in the sense that test takers need not only to recognise the forms of the language, but also to process the language and understand its usage and function; for example, test takers write down what they hear in listening extracts or answer questions via oral interactions with interlocutors. Oller (1979) sees language use in integrative tests as involving (1) the on-line processing of language in real time (e.g. listening and speaking activities), and (2) a "pragmatic mapping" component (understanding the language and its contexts of usage) (McNamara, 2000: 15). Common integrative testing techniques include listening cloze tests, gap-filling tasks, dictation, and sentence-repetition tasks.

Listening Cloze

A cloze test requires listeners to integrate grammatical, lexical, contextual, and

pragmatic knowledge in order to be able to supply the missing words (McNamara, 2000: 15). Listening cloze tests are easy to construct in the sense of simply deleting words and straightforward to score because the scoring is based on the “exact words” spoken. Lewkowicz (1991: 30) notes that testers should pay attention to the deletions to ensure (a) that they cannot be filled in purely on the basis of general knowledge, and (b) that they are not testing trivial questions likely to be missed by test-takers. Buck (2001: 70) also points out that the most obvious problem with this sort of gap-filling test is that test-takers could treat the passage as a normal cloze test, and fill in the blanks without listening to the passage, in which case “it is no longer a listening test at all, although it may still be a perfectly good test of reading or general language ability”. Alternatively, test-takers may only listen for the missing words and respond based on word recognition without understanding the whole passage. Anderson (1972) argues that listening extracts can often be processed on a perceptual or phonological level without bringing to mind the actual meaning of the words, which cannot be held to constitute comprehension unless evidence of semantic processing of the content is provided (Buck, 2001: 71). In addition, research has shown that cloze tests seem mostly to be measuring the same kinds of things as discrete point tests in terms of grammar and vocabulary (McNamara, 2000: 16), both of which fail to assess students’ communicative ability to *use* the target language. Buck (1992) considers that using *summaries* which are closely related to the listening passages test-takers hear and asking them to fill in blanks is better than simply asking students to reproduce exactly what they hear. Summaries of listening passages encourage test-takers to understand, process, and interpret the meaning of contexts. However, it is not always easy to produce summary tasks from a listening passage, and pre-testing is required (Lewkowicz, 1991: 30); moreover, gap-filling

summaries are subject to several of the same problems as cloze tests – the possibility of responding without understanding.

Dictation

The basic idea of dictation is to ask students to listen to a passage and write down what they hear. Oller (1971: 259) states that dictation tests students' ability to (1) discriminate phonological units, (2) make decisions concerning word boundaries in order to discover sequences of words and phrases, and (3) translate this analysis into a graphemic representation. In other words, dictation is considered to be related to the interpretation of the acoustic signal, phonemic identification, lexical recognition, morphology, syntactic analysis and semantic interpretation (Oakeshott-Taylor, 1977; see Buck, 2001: 74). However, it has also been argued that using dictation to test listening comprehension is very indirect and inadequate, due to a limited sampling of the relevant listening passages (Anderson, 1953: 43). Nevertheless, Oller (1971: 257) argues that dictation is more effective than tests involving multiple-choice, short answers, or fill-in-the-blanks items, since it demands greater "understanding of complex English structures than the more isolative and analytical objective tests". Brindley (1998: 179), however, does not recommend dictation as a surrogate for a listening test, because it involves skills other than listening, such as auditory memory, spelling, and grammatical and lexical knowledge. Weir (1993: 124) also supports this point that "the condition under which this task is conducted only in a very limited sense reflects the normal condition for the spoken language". Testing listening by using dictation demands students' memory of phrases and sentences they hear. Buck (2001: 77) further notes that if segments of passages are very short and do not challenge test-takers, then dictation is probably testing little more than the ability to recognise simple

elements; but if the segments are a little longer, it will be testing understanding on a local, literal, linguistic level, in which case it clearly is an integrative test. In addition, dictation does not seem to require the ability to understand inferred meaning or to relate the literal meaning to a wider communicative context (Buck, 2001: 78), which largely focus on the semantic level. Dictation also requires a good short-term memory, as well as writing ability, making it far more than a test of listening skills.

Short-answer Questions

Short-answer questions are easy to produce if testers are prepared to allow test-takers to construct answers themselves. It is suggested that such questions are particularly suitable for testing the understanding of clearly stated information (Buck, 2001: 138). The justification for short-answer questions is that they are a better method than multiple-choice questions of establishing whether test-takers really understand the listening extracts. Hughes (2002: 145) suggests that short-answer questions can also work well for testing the ability to predict the meaning of unknown words heard from contexts, as the possibilities for guessing are lower than with dichotomous or multiple-choice items. It is further recommended that the answers in short-answer questions should be kept very short, so as not to make serious demands on test-takers' writing skills (Hughes, 2002: 166); if test-takers spend too much time on one question, they may miss the following listening passages and questions. Indeed, Buck (2001: 141) supports the idea of restricting responses to no more than three words. However, due to the subjectivity of scoring short-answer questions, compared with "objectively" scoring multiple-choice items, Brindley (1998: 177) and Buck (2001: 140) both suggest that short-answer questions require a detailed scoring key containing a list

of acceptable responses to reduce the problem of determining what constitutes a correct answer and a sufficient response.

Multiple-choice Questions

Common formats of multiple-choice questions include a question stem, a correct or best key option, and several distractors. Alderson (2000: 211) states that multiple-choice items allow the tester to control the range of possible answers to comprehension questions, and to some extent to control students' answers when responding. Henning (1987: 44) notes that the most difficult task in preparing multiple-choice distractors is to avoid giving response cues. Students who have frequently been exposed to examinations with multiple-choice questions may develop test wiseness; that is, students may choose "the correct options independently of any knowledge of the content field being tested" (Henning, 1987: 44). Vandergrift (2007: 191) notes that the narrow focus on the right answer to comprehension questions does little to help students understand or control the processes leading to comprehension. Alderson (2000: 211) also worries that test-takers' ability to answer multiple-choice questions is different from their language ability, in the sense that students can learn how to answer multiple-choice questions by eliminating improbable distractors. Thus, designing distractors for such questions is a very skilled and time-consuming task, and also requires a pre-testing process. Cohen (1994: 63) contends that teachers often choose distractors on the basis of intuition; however the problem is that a teacher's intuition may not always be accurate enough to "distract" students from the correct answer. Henning (1987: 45) also notes that weak distractors are subject to negative washback in language teaching, i.e. students may learn errors from the examination itself. Increasing the number of distractors appears to

minimise the chance of selecting the correct answer by guesswork alone. However, Shizuka *et al.* (Shizuka, Takeuchi, Yashima, and Yoshizawa, 2006), in their study of three- and four-option English tests for university entrance purposes in Japan, discovered that using three options instead of four did *not* significantly change the mean item difficulty or the mean item discrimination. Their results suggest that using three options would save testers time and effort in writing multiple-choice questions, as well as saving students time when responding. Nevertheless, if teachers take test questions without adaptation from published textbooks, the reliability and validity of the questions and particularly the distractors are called into question. Buck (2001: 146) also suggests that the forms of the correct key and the distractors should not differ in any way structurally and the options should not provide clues to the answer. Moreover, none of the options should be particularly long, otherwise test-takers may spend time reading the options at the expense of listening to the next question, and the result may become a test of reading rather than listening. A further serious difficulty with multiple choice questions is that the tester does not know why the test-takers respond the way they do, in the sense that test-takers may choose the correct answer, but for the wrong reasons (Wu, 1998: 40; Alderson, 2000: 212).

Listening ability can also be assessed by asking students to repeat sentences orally immediately after they hear them. However, Buck (2001: 79) regards this as a test of recognising and repeating sounds, which may not involve processing of the meaning at all. Clearly, if a listening test assesses more language abilities than just listening (e.g. reading or writing), its validity as a listening test may be threatened. For example, asking students to write complex sentences in a listening test assesses writing ability and students' writing abilities may be very diverse. The differences between listening and writing levels may leave too

many unpredictable variables in interpreting scores appropriately and fairly.

2.3.3 Communicative Language Testing

The basic idea underlying communicative teaching is that language is used for the purpose of communication in a particular situation and for a particular purpose; what is important is for a person to know how to use this language appropriately, rather than how grammatically correct they are when they use it (Buck, 2001: 83). Thus, the *use* of the language is more important than *usage* in communication (Widdowson, 1978). Although Oller (1979) claims that cloze and dictation, which are fundamental tests of basic language comprehension, focus on test takers' structural and lexical processing, Morrow (1979: 149) argues that neither cloze nor dictations give any convincing proof of test takers' ability to actually use the language "in ordinary situations", i.e. to use the language to listen, speak, read, and write in real-life situations. Skehan (1988: 215), summarising from Morrow (1979: 149-150) and Canale (1984), suggests that genuine communication in a testing situation involves interaction with more than one participant, who may be unpredictable and creative, and situated in discursal and sociolinguistic contexts, purpose/achievement-oriented, using authentic stimulus, and outcome evaluated. In short, communicative testing values whether test-takers can manipulate the target language and their ability to communicate appropriately. As discussed in Chapter 1.8.3, a prominent feature of communicative language teaching and task-based instruction is the use of authentic texts and tasks in listening classrooms and this also applies to testing situations. However, it is not a simple task to design easily score-able listening tests which correspond closely to real-life situations; Brindley (1998: 174) argues that the difficulty lies in the fact that a great many of the listening tasks people undertake in everyday life (e.g. listening

to radio or television programmes) do not require a specific response – the listeners simply process the information and store it until it is needed or forget it. Widdowson (1978; see Buck, 2001: 85) also argues that although the discourses may be genuine where they are taken from the target-language use situation, “they are not authentic unless the test-taker is required to deal with them in a way that corresponds to their use in the target-language use situation”. Buck (2001: 84) further notes that there are many different communicative topics and situations in language use, and successful performance on one complicated task does not always indicate the ability to perform well on others. In other words, authentic tasks in communicative tests can not comprehensively be used to generalise and predict test-takers’ language ability across target-language use situations.

Brindley (1998: 174) notes that many standardised listening tests tend to focus on non-participative listening tasks which require candidates to listen to pre-recorded texts and respond through activities like ticking boxes, circling alternatives, or writing short answers. He considers that “a good deal of listening, however, happens in the context of oral interaction where listening and speaking ability are closely interconnected”, and a person cannot carry on a conversation effectively if he or she does not understand what the other interlocutor says. Savignon (2005: 640) supports the fact that because communicative language teaching aims at teaching students’ functional language competence, a global and qualitative evaluation of learner achievement is needed, rather than a purely quantitative assessment of discrete linguistic features. In communicative language testing, performance tests are usually used to assess students’ speaking or writing abilities, and there are three main types of performance tests: *direct assessment tests*, *work sample tests*, and *simulation techniques* (Wesche, 1987; see Skehan, 1988: 216). However, Wesche (1987; see Skehan, 1988: 216) suggests

that “such performance tests are more appropriate at higher proficiency levels”. It is sometimes claimed that performance tests establish greater predictive validity because the purpose of such tests is to provide information about likely behaviour in the real world (Davies *et al.*, 1999: 149; Skehan, 1988: 216). However, performance tests are subject to narrowness and are hard to generalise (Weir, 1990), because it is difficult to make inferences from a score until the meaning of the score can really be related to the real-life situation. In addition, testing listening via oral interaction can easily become a test of speaking, since in speaking assessment, one’s listening comprehension is subsumed into one’s speaking ability. In this case, assessing listening ability will not only involve testing listening comprehension but also involve testing speaking ability. That is, assessment methods for testing speaking ability will need to be applied (e.g. rating scales with detailed descriptors). In the contexts of English listening courses at Taiwanese universities, it is unknown how teachers assess students’ listening abilities, and what kind of approaches they adopt, and whether oral assessment is used to assess students’ listening ability.

2.4 Validity of Test Contents

In general, validity in a test refers to whether a test measures what it claims to measure, and does not measure what it claims not to measure. According to Messick (1993: 13), validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment. Validity is thus a multifaceted but unified concept which integrates “considerations of content, criteria and consequences into a comprehensive framework” to support any claim for the validity of scores on a test (Messick,

1995: 742; cited in Weir, 2005: 13). In addition, Messick (1993: 13) emphasises that validity is a matter of degree, not all or none; for example, he points out that a test can be highly valid in content coverage but have a low reliability coefficient. Weir (2005: 12) defines validity as the extent to which a test can be shown to produce data, i.e. test scores, which are an accurate representation of a candidate's level of language knowledge or skills.

Validity can be examined in terms of different foci or facets (rather than types). Content validity means that test content covers a representative sample of the behaviour domain to be measured, and the domain under consideration should be fully described in advance, rather than being defined after the test has been prepared (Anastasi and Urbina, 1997: 114). Weir (1993; 2005: 19), using *context validity* to represent content validity, defines this context validity as the extent to which the choice of tasks in a test is representative of the larger universe of tasks of which the test is assumed to be a sample. However, it is difficult to systematically sample items which are highly representative of real-life situations. Bachman and Palmer (1996: 23) try to overcome this by setting up a dimension they call "test authenticity" and defining it as "the degree of correspondence of the characteristics of a given language test to the features of a target language use task." So, the more authentic the tasks that are used to test students' language ability, the more confidence testers can have in relating students' performance to likely real-life performance.

The term "construct validity" refers to the extent to which performance on tests is consistent with predictions that the tester makes on the basis of a theory of the relevant abilities (Bachman, 1990: 255). Construct validity is a multifaceted and complicated issue; Messick actually integrates content validity into construct validity, by arguing that construct validity has six main aspects (1993; also see

1996: 248): (1) *content* (content relevance and representativeness), (2) *substantive* (item correlations), (3) *structural* (construct domain and structure of assessment), (4) *generalisable* (the extent to which score interpretations can be generalised to the population), (5) *external* (the relationship between assessment scores and non-assessment behaviours), and (6) *consequential* (the consequences of test use).

In the present study, the content, structural, and consequential aspects of construct validity seem particularly relevant to the topic of English listening classroom assessment. Messick (1993: 17) emphasises that construct validity subsumes content relevance and representativeness as well as criterion-relatedness, because such information about the content domain of reference and about specific criterion behaviours predicted by the test scores clearly contributes to score interpretation. Both Messick (1980, 1988) and Bachman (1990: 242) argue that while using test scores for a particular purpose, not only construct validity should be justified, but also the relevance or utility of the particular test purpose and the social consequences of using the test score in this particular way. In other words, in order to justify the interpretations of the test scores, the test constructor has to take construct validity and its implications with reference to test scores into consideration, making the test scores more meaningful to the test purpose(s). Construct validity, thus, almost embraces all forms of validation evidence (Messick, 1993: 17). Chapelle (1999: 258), summarising key differences between past and current conceptions of construct validity, states that construct validity in the past was seen as one of *three types of validity* (i.e. content, criterion-related, and construct). However, validity is currently regarded, following Messick (1993), as a more *unitary* concept with construct validity as central, taking performance differences across different groups of examinees, test times, test settings, examinees, raters, and interlocutor behaviour into

consideration (Chapelle, 1999: 258; Davies et al., 1999: 33).

Messick (1996: 244) emphasises that an assessment is deficient when either construct under-representation or construct-irrelevant variance threaten the validity of the test content. Construct under-representation means that the content is too narrow, and fails to cover important dimensions or facets of the focal construct. Weir (2005: 18) notes that construct under-representation in a test may have an adverse washback effect on the teaching that precedes the test; teachers may simply not teach certain important skills if they are not in the test. Construct-irrelevant variance, on the other hand, is when the content is too broad and the test results contain excess but reliable variance that is irrelevant to the interpreted construct. In short, if a test does not assess that which it is supposed to test, based on the tester's construct, or else if it tests something else, it cannot be considered adequately valid for its purpose. In the Taiwanese context, if there is a marked variation between teachers' desired test constructs and what they actually test, this will inevitably reduce the overall validity of the test. It will thus be important to establish, by observation or interview, how far the teaching and tests match each other and match the students' needs.

Construct validity is also important when making inferences about examinees' performance based on test scores (ibid: 242; Davies et al., 1999: 31). Messick (1996: 251) suggests that validity of test consequences includes "evidence and rationales for evaluating intended and unintended consequences of score interpretation and use in both the short and long-term", particularly those associated with bias and unfairness of score interpretation and test use, which will lead to positive or negative washback (washback issues are discussed in Section 2.7). In addition, the content of the test items can also influence students' attitudes towards tests and the consequences of tests. If the interpretations of the

test consequences are inappropriate, the test content cannot be regarded as representative of the course objectives. For example, if teachers interpret students' test consequences from a certain class as their overall ability in English, this may lead to a misuse of test content and misinterpretations of test consequences. In addition, group differences in a language classroom also influence the validity of a test. Elder (1997: 261) states that group differences may be treated as: (1) a real difference in ability being tested in terms of students' individual language abilities or cultural backgrounds, or (2) the effect of confounding variables within a test with reference to test methods, question types, or background knowledge. However, if the differences between test-takers are not controlled, it may be very difficult to discover whether a test is biased due to test-taker differences or simply the test itself. Pre-test piloting has always been an important aspect of testing, but this last point makes it even more central (Fulcher, 1999: 226). Piloting tests permits item analysis and item revision, where poor items can be rejected, which has the effect of "minimising construct contaminants" and strengthening the empirical domain structure that "underlies the final test form and ultimate score interpretation" (Messick, 1993: 43; Fulcher, 1999: 226). Piloting test items thus helps eliminate construct under-representation and construct-irrelevant variance. However, in the case of the English listening test practices in Taiwanese universities, it is (in my experience at least) very rare that test items are piloted before testing students. Whether it is possible in the case of the modules observed in this study for piloting to be undertaken before students are tested will need to be checked with the course instructors. Failure to pilot tests would leave inappropriate items undetected and content validity might be reduced. Again, whether this happens in practice needs to be checked, as part of the research.

There is also the related question of test fairness. Kunnan (2000: 3) proposes that there are three main concerns of test fairness: validity, access, and justice. Validity in this context concerns whether the interpretation of a test score has *equal construct validity* (and reliability) for different test takers with respect to variables such as gender, ethnicity, field of specialisation, native language, or cultural background (ibid: 3). In the present study, the students in Taiwanese universities can be assumed to share the same native language, cultural background, and field of specialisation (since only English and foreign language major students will be chosen, see Chapter Three), so the possibility of unfairness may not be a problem. In addition, it would be unfair if some individuals in the same testing group received biased scoring. The access aspect of fairness concerns whether test takers access the test equipment equally. If test takers use different equipment or are tested in different environments, this can lead to unfairness. Justice, on the other hand, concerns whether test scores are used in terms of *societal equality* without bias towards test-taker groups (ibid: 4). In other words, test takers would not be discriminated against by their ethnicity, gender, or native language when their test scores are used for societal purposes (e.g. job interviews). In the context of Taiwanese universities, justice with respect to test fairness lies in how students' listening ability is interpreted by society based on their academic listening scores, something which is beyond the scope of the present study. To put it another way, test fairness emphasises the idea that test takers operate in a system that allows each testee to perform to the best of his/her ability and to be marked accordingly. The test is regarded as valid and appropriate when the tasks are topical and content related, and the use of the test score is justified.

2.5 Text Characteristics in Listening Tests

Listening comprehensibility can be affected by the differing input of listening texts (i.e. different formats or stimulus materials). It will be recalled from Section 2.1 that among the key variables that influence listening are the *nature of the input* (e.g. speech rate, length of text, lexical and syntactic characteristics, accent, register, propositional density, amount of redundancy), the *nature of the assessment task* (e.g. amount of context, clarity of instructions, output required), and *individual listener factors* (e.g. memory, topic-relevant knowledge, motivation) (Brindley, 1998: 175; Lynch, 2002: 43). In the present study, the influences of accent, speech rate, length of text, lexical and syntactic characteristics, text type, test rubric, outputs, and topic-relevant knowledge on students' listening comprehension appear to be highly relevant to published ideas of task-based teaching, learning, and testing that the Taiwanese government expect will be investigated.

Different accents of English may cause comprehension problems to a certain degree even to native speakers. Derwing and Munro (2005: 379) note that a foreign accent is a complex aspect of language that affects speakers and listeners in both perception and production, and consequently in social interaction at many levels. Understanding a foreign accent has three aspects: (a) *intelligibility* (the extent to which the speaker's intended utterance is actually understood by a listener); (b) *comprehensibility* (the listener's perception of the degree of difficulty encountered when trying to understand an utterance); and (c) *accentedness* (how much an foreign accent differs from the variety of English commonly spoken in the community) (Munro and Derwing, 1995; Derwing and Munro, 2005: 385). In order to help students understand "clear" English, the language that speakers use in published listening materials is usually Standard UK or American English

without a “strong” regional accent. However, students may still have problems understanding listening texts in a test if they are not familiar with the accents. Weir (2005: 81) suggests that the bottom line is that speakers should have clear accessible pronunciation and intonation.

In addition to accent, speech rate can also markedly affect students’ listening comprehension in a test. Buck (2001: 38) argues that listener perceptions that speech is too fast are often due to a lack of processing automaticity, so as listeners get better, and as they learn to process the language more automatically, they feel that speech seems to become slower. This implies that it is possible to improve input processing automaticity by frequently listening to English at a relatively fast speed. Current research on speech rate suggests that the mean rate for British conversation is around 210 words per minute and for lectures 140 (Tauroza and Allison, 1990). Whether students are able to follow the speed of the listening extracts used in class relates to their level of English. In my experience, Taiwanese students are more frequently exposed to short conversations than short talks in English listening courses. However, in terms of short conversations which last less than one minute, it may be difficult to judge whether the speaker speaks too fast or not.

Text type can also affect students’ listening comprehension. It is assumed that most prewritten edited monologues, such as news broadcasts, are more syntactically complex, less redundant, lexically denser, and use fewer pauses and repetitions than unplanned texts, and as such are potentially more difficult to understand (Rubin, 1994: 204; Shohamy and Inbar, 1991: 28). Shohamy and Inbar (ibid: 30), in their research on 150 US EFL secondary learners, found that “*news broadcast*” was the most difficult category for the learners to understand, followed by “*lecture*”; “*dialogue*” proved to be the least difficult text type.

Specifically, they also found that understanding certain types of text appeared to be related to the level of familiarity of the students with the genre. That is to say, limited exposure to certain genres may hinder comprehension (ibid.). In contrast to Shohamy and Inbar's findings, Read (2002) found that scripted monologues were significantly easier than unscripted discussions, because interlocutors in a discussion frequently share a degree of background knowledge, which may not be the case in a test. All three discourse types are common in English listening tests, teaching materials, and in-class activities in Taiwanese universities, according to the teachers in the preliminary interviews (see Chapter 1.4); however, different teachers may put the emphasis on practicing certain type(s) of discourse. If a majority of students find it hard to comprehend a certain type of discourse, but teachers still put an emphasis on it, regardless of the students' level of English, this may decrease students' willingness to learn. The conclusion is that some indication is needed of how speech rate and genre characteristics are matched in both classes and tests, and of how the students perceive the situation.

2.6 Task Characteristics in Listening Tests

Undertaking an English as a Foreign Language listening test requires intense concentration on the listening passages, and test-takers can become very stressed, particularly when they are only allowed to listen to the passages once. For this reason, Brindley (1998: 176) suggests that it is important to minimise the possible effects of extraneous factors such as test presentation and administration on candidates' performance. Just as linguistic features may influence comprehension of listening to a large extent, so task characteristics may also affect not only students' understanding of task content, but also their performance on tasks. The framework of task characteristics which was developed by Buck (2001: 117),

based on Bachman (1990); and Bachman and Palmer (1996: 49), consists of four main categories: (1) the setting, (2) the test rubric, (3) the input (e.g. topical knowledge), and (4) the expected responses. Buck suggests that all four components need to be taken into consideration in listening test tasks.

2.6.1 Test Setting

The physical characteristics of the test setting are considered important in language testing (Buck, 2001; Bachman and Palmer, 1996), as it is essential to ensure a good quality acoustic environment and minimise background hiss or other noises from the recording equipment, so that students are not disturbed. Similarly, the environment outside the test setting (i.e. the classroom) needs to be kept quiet, or at least students need to be acoustically isolated from it.

2.6.2 Test Rubric

Test rubric includes the information about the test instructions, the duration of the test as a whole and of the individual tasks, and about how the language of testees' responses will be evaluated, or scored (Bachman and Palmer, 1996: 50). It is important for students to understand the instructions both clearly and identically, before carrying out tasks or answering questions. Instructions should be given in the native language to minimise the possibility of confusion; however, this is clearly not a feasible option in cases where the test candidates do not share the same native language (Buck, 2001: 119; Brindley, 1998: 176). If the target language is used, one should try to ensure that the language is easier than the level the test aims to measure (Buck, 2001: 119). Brindley (1998: 176) also makes the important point that preparing clear examples of each new item type with students before giving them a test also helps them understand the instructions. In addition,

the allotment of test time and the length of tests in classroom assessment should not be too long or too short; and there should be enough time for students to finish all tasks. Scoring methods also need to be made explicit in the test rubric; it is important that students know the relative value of each task in the test as it can help them to structure their time and effort (Buck, 2001: 122). Teachers can also constrain the responses available to students, so the responses become easier for them to score.

2.6.3 Response Format

The format of the expected response can vary considerably. Common types of listening comprehension question which are frequently used to elicit students' responses in listening tests are: short-answer, multiple-choice, true/false, cloze, and dictation questions (Buck, 2000; Brindley, 1998). These four types of question have been discussed in Sections 2.3.1 and 2.3.2.

2.6.4 Topical Knowledge

The topical characteristics of a task relate to certain types of background or domain, such as cultural, academic, political, or technical topics. In the context of listening tests, when test-takers process listening extracts, they integrate the new information from the content into their existing schemata (Alderson, 2000: 33). Formal and content schemata are considered to have important effects on test-takers' performance. Formal schemata refer to knowledge of language and linguistic conventions, including text organisation and genre types (Carrell, 1983). Alderson (2000: 34) classifies content schemata into *background knowledge*, which may or may not be relevant to the content of listening passages, and *subject-matter knowledge*, which is directly relevant to extract content or topic.

Current research evidence suggests that background knowledge is important in listening comprehension, and that it does affect test performance (Chiang and Dunkel, 1992; Long, 1990; Schmidt-Rinehart, 1994; Buck, 2001; Chang and Read, 2006). Chang and Read (2006), investigating the effect of providing topical knowledge to university students in English listening tests, discovered that low language proficiency students benefit most from being provided with topical knowledge, because prior study of the listening topics allowed those learners to compensate for their more limited language knowledge. Chiang and Dunkel (1992), also investigating Taiwanese undergraduate EFL students' listening comprehension, found that their Taiwanese EFL listeners scored higher on listening comprehension tests when they listened to a lecture on a familiar topic than when they listened to one on an unfamiliar topic. In a criterion-based classroom assessment, the topics taught in class need to be related to the topics tested; however, this is something that needs to be checked in Taiwanese English listening programmes.

2.7 Washback Effects and Test Consequences

Validity is an important factor not only in deciding test content relevance, but also in interpreting test consequences. The term “washback” (or “backwash”) is frequently used to refer to the effects of tests on teaching and learning (Wall, 1997: 291). Alderson and Wall (1993) state that evidence of washback is typically sought in terms of behavioural and attitudinal changes in teachers and learners, which are associated with tests. It is believed that the consequences of tests influence educational processes in various ways; Taylor (2005) points out that students' attitudes or opinions towards test content and test method can also influence teachers' teaching methods and lesson contents:

One common assumption is that teachers will be influenced by the knowledge that their students are planning to take a certain test and will afterwards adapt their teaching methodology and lesson content to reflect the test's demands. (Taylor, 2005: 154)

Washback is usually perceived as either negative or positive. Negative washback typically occurs when a test's content or format is based on a narrow definition of language ability, and so constrains the teaching/learning context (Messick, 1996: 242). In other words, teachers emphasise in their teaching what will be tested in the test, rather than teaching for understanding of a topical knowledge domain. Positive washback, on the other hand, results when a good test encourages positive teaching and learning practices. Alderson and Wall (1993) and Messick (1996) all note that in some educational contexts, direct testing or assessment facilitates positive consequences for teaching and learning, where authenticity and directness imply realistic simulations or criterion samples. In other words, communicative or direct tests which provide students with practice in performing the target language can be the most beneficial for improving students' individual language ability.

However, Messick (1996: 242) notes that good or bad educational practices, which are separate from the quality of a test, can also influence test washback; this becomes a particular problem where a poor test is associated with positive effects and a good test with negative effects. In other words, how teachers conceive of the test can affect the adjustments they make to their teaching, which in turn impacts on good or bad educational practices. In Alderson and Wall's study (1993: 120), the concept of washback is explored by establishing several possible "Washback Hypotheses". Briefly, the "Washback Hypotheses" centres on two main propositions – *a test will influence teaching* and *a test will influence*

learning – and the key questions are how this happens, in what ways, how quickly, in what degree and depth, and what consequences will follow. Specifically, they hypothesise that *a test will influence what and how teachers teach* (Alderson and Wall, 1993: 120). Nevertheless, Messick (1996: 242) argues that a test might influence *what* is taught but not *how* it is taught, might influence *teacher* behaviours but not *learner* behaviour, or might influence both “*what* is taught” and “*teacher* behaviour” with little or no improvement in skills. Hamp-Lyons (1997: 300) suggests that classroom investigation and observation may help to uncover changes in teaching and learning. Tests therefore should have an impact on teaching and learning, and this will need to be checked, to the extent that it is possible, in the case of Taiwan, exploring whether the results of the tests influence the teaching, the curriculum, the test content, or test method.

However, studying washback effects is not the only approach to exploring the impact of the variation of English listening test practices on validity; Messick (*ibid.*) suggests that one can instead turn to the test properties likely to produce washback – namely, authenticity and directness – and ask how they might influence the impact of variations in test content on validity. For the present study, it seems reasonable to hypothesise that if teachers use direct tests, students’ listening ability will be more likely to improve than if they use indirect tests.

Washback can be affected by a range of different factors in a test. Watanabe (2004: 22), in a review of Alderson and Hamp-Lyons (1996), Brown (1997), Cheng (2004), Shohamy et al. (1996), and Wall (1997), concludes that the process of washback seems to be mediated by five factors:

The factors may include the following: test factors (e.g., test methods, test contents, skills tested, purpose of the test, decisions that will be made on the basis of the test results, etc.); prestige factors (e.g., stakes of the test, status of the test within the entire system, etc.); personal factors (e.g., teachers' educational background, their beliefs about the best methods of teaching/learning, etc.); micro-context factors (e.g., the school setting in which the test preparation is being carried out); and macro-context factors, that is, the society where the test is used.

(Watanabe, 2004: 22)

Washback clearly needs to be included in the present study. The English listening modules to be observed in Taiwanese universities are all compulsory courses and students are required to pass the exams in order to pass the course. In such a situation, the two exams that the students have to take have high stakes, because the outcomes of the tests are very likely to affect students' future study. For example, the results may restrict them from taking advanced related English courses, or even stop them from graduating. Since test consequences may influence teaching and learning to certain extent, test factors such as test content and test methods may well influence washback. Buck (1988), in his analysis of the washback potential of various listening tasks in Japanese university entrance examinations, discovered that many accepted testing procedures such as noise tests, listening cloze, tape-recorded sentence repetition, and dictation, fell short on providing the testees with communicative tasks which invited beneficial washback, except for cloze tests based on summaries of listening passages, which appeared to be able to promote students' understanding of spoken English (cited in Bailey, 1996: 265). Wall (1996) states that washback effects may take much longer to be visible than we typically are prepared to wait before judging the influence of a testing innovation; and although washback effects may occur, outsiders may fail to recognise it (Hamp-Lyons, 1997: 298). If a course finishes before washback

effects occur, it is possible that no effect can be detected. In the context of Taiwanese university listening courses, the results of the mid-term test may affect the teaching of the second half of the course, the design of the final exam, and the students' performance on the final exam. There has been no research to date that I can find exploring the relation between the two tests.

2.8 Summary

Test objectives are an important consideration when designing test contents because objectives help teachers to determine what they want to know about students' listening ability or skills that they have learned in class. However, specifying a test objective in an English listening classroom assessment does not always guarantee a valid test or interpretation of students' English listening ability. Students' listening ability can be elicited in terms of different types of listening comprehension question, and while choosing the types of question, teachers need to take the pros and cons of each type into consideration and establish the most appropriate way to elicit their students' listening ability. The test contents and items must be highly relevant to the lessons that the teacher delivers. The ways in which English listening is taught may influence the ways students learn and consequently influence how tests/examinations are designed and students' test performance. If there is inconsistency between the test constructs, test contents, test methods, and scoring methods, this could threaten the validity of the test objectives.

There is no relevant literature to date regarding Taiwanese university students' expectations of either teaching or assessment of listening. There is also no guarantee that either of the tests used by teachers will actually measure listening skills or allow them (or me) to make valid inferences about what students

have learned. In these circumstances, it seems reasonable to restrict the aims of the present study to investigate, firstly, how far the test objectives are clearly specified in relation to the curriculum and teaching contents; secondly, how far task-based instruction is implemented in English listening classes; thirdly, what the problems are that influence students' listening comprehension in the mid-term and final exams; and lastly how far the results of the mid-term exam have discoverable washback effects on the teaching that follows it. The extent to which there are variations in teaching and assessment methods between teachers of listening courses, or how far testing impacts on teaching and learning, are unknown; the focus of the study will accordingly be on teachers' aims and opinions regarding their teaching and assessment methods, and on discovering whether students are satisfied with the teaching and the assessment of their listening courses. The study will also explore students' expectations and opinions towards the two tests.

Chapter Three

An Overview of Research Design and Methodology

3.1 Introduction

This chapter discusses the research design of the study, considers a number of important methodological questions, and describes the preliminary designs of the surveys. The first part discusses how the study was designed and implemented, based on the research questions presented in the previous two chapters. There is also a consideration of the main ethical issues involved. The second part reviews the advantages and disadvantages of the methods used, namely questionnaires, interviews, and classroom observation schedules. The third part covers the designs of the questionnaires, interview questions, and classroom observation checklists for the pilot work in the study.

3.2 The Purpose of the Research and Research Questions

To recap, the focus of the research was to explore how far the teaching, learning, and assessment of English listening courses in Taiwanese universities influence each other. At a general level, the research question in the study was – *Are the general listening test practices within and between Taiwanese universities similar or markedly different?* Thus, the study focused on three areas: (1) teaching and testing English listening comprehension, where teachers were test constructors, (2) teachers' and students' opinions towards tests, and (3) test consequences. Specifically, it aimed at investigating five things:

- (1) *How far is task-based instruction implemented in English listening classes?*

(2) *How far are the test objectives clearly specified in relation to the curriculum and teaching contents?*

(3) *How far are communicative language testing approaches applied in the mid-term and final listening exams?*

(4) *What kind of problems will influence students' listening comprehension in the two listening exams?*

(5) *How far do the results of the mid-term exam have washback effects on teaching?*

In Chapter Two I explained that the differences in test methods between discrete-point, integrative, and communicative testing. The test methods of the listening courses needed to be checked and compared, to see what type of test methods were used in the mid-term and final exams, and whether communicative language testing produced positive washback in teaching and learning.

3.3 Research Design: Main Study

The main study was designed to examine four groups of students from two different universities in Taiwan. A case study approach was used to investigate the same research questions across different groups of samples.

Participants

In the absence of an agreed ranking scheme for Taiwanese universities, a mixture of six national and private institutions was contacted and three of the private ones agreed to participate. Of these, two were randomly selected in the main study and the third was used for the pilot. For the main study, each university provided two groups of students (i.e. two classes) in the same year of study, taking the same title of English listening module – either *English Listening* or *English Listening and*

Speaking Practice, depending on the university. The students were “English” or “foreign language” major students. One reason for choosing language majors was because it was known that all of the courses run in the English departments at both universities were taught in English (thereby meeting the MOE’s proposal for an English or bilingual teaching environment); English listening courses would thus be important for training students both in order to understand their regular classes better, and when it came to gaining employment. The other reason was that these English listening courses were established and had run for several years – nine years at University A, and six years at University B – so there was less likelihood of temporary effects due to a recent innovation. Students from Cases 1 and 2 were in their second year of study, while students from Cases 3 and 4 were in their third year. Each group of students was taught by a different course instructor, who was able to decide on the teaching materials and examination content. Two examinations – mid-term and final – were administered at both universities, and it was compulsory for the four groups of students to take them for graduation. In total, there were 112 respondents and four different course instructors from four different classrooms at the Department of English Language (Cases 1 and 2) and Department of Foreign Languages (Cases 3 and 4) at the two universities taking part in the main study (Table 3.1). Participants in the four classes were not randomised because the purpose of the study was to evaluate existing English listening programmes rather than to do experiments.

Table 3.1 Main Study Design

University	University A		University B	
Course Title	English Listening		English Listening and Speaking	
Student Major	English Language		Foreign Languages	
Group	Case 1	Case 2	Case 3	Case 4
Course Instructor	Dr. N	Miss T	Dr. C	Dr. D
Exams	A Mid-term Exam + A Final Exam			

Research Methods

Two questionnaires were designed and group-administered to survey the opinions of the four groups of students towards (a) the listening courses and (b) the two examinations. Face-to-face interviews with the course instructors were conducted after the marks of the two examinations had been made available. Classroom observations were used to investigate the in-class interaction between teachers and students and to establish whether the in-class situation made any difference before and after the mid-term exam. Three observations were carried out before the mid-term exam and the other three afterwards. In addition to the classroom observations, two examinations in each case were audited by the researcher. Observations of classroom activities and examinations were tape recorded with the prior permission of course instructors (see Section 3.5). Relevant documentary data such as course textbooks, in-class supplementary materials, and students' examination marks were collected from the course instructors.

3.4 Research Design: Pilot Study

A pilot study, which was separate from the four main case studies, was undertaken at a different university in Taiwan and a group of 41 university students taking an English listening course participated in it. The backgrounds of the participants and the educational environment in the pilot study were all similar to those in the main study, in order to make the samples in the pilot representative of the participants in the main study.

The two questionnaires and interview questions were designed originally in English and translated into Chinese versions by the researcher. All Chinese translations were cross-checked by another person who understands both Chinese and English. Two problems with respect to the precision of nouns in Chinese

translation were noted. One was the translation of “type of speech” in Question 6, Part Two; “speech” in Chinese was replaced by “expression” in order to approximate better the intended meaning. The other was the translation of “is” in Question 11, Part Three; it was suggested that “is” in Chinese should not be translated literally and its occurrence made the sentence hard to follow. As all questionnaire respondents in the present study were Chinese speakers, the purpose of using Chinese versions with respondents was to help them understand the wording by using their native language. As for the English course instructors in the interviews, questions were presented in English with a Chinese translation included. Translation of the interview questions was also cross-checked by the same Chinese speaker who can speak English. No translation problems arose. Because it was possible that the English teachers would include some whose first language was Chinese but who understood English and some whose first language was English, the interviewees could choose to answer in either language. The classroom observation checklist was used by the researcher only in an English version and no translation of English into Chinese was made. The questionnaires, interview questions, and classroom observation checklists were all piloted; the pilot testing and the results will be discussed in Chapter Four.

3.5 The Ethical Issues in the Study

Ethical issues are rightly regarded as an important aspect of doing educational research. De Vaus (2002: 59) illustrates five important ethical responsibilities towards survey participants: (1) voluntary participation, (2) informed consent, (3) chance of causing harm, (4) confidentiality/anonymity, and (5) maintaining privacy. Voluntary participation means that respondents should not be forced to participate in a survey. Compulsory participation might increase the response rate,

but it could undermine the quality of the responses if respondents are not willing to participate. It is the researchers' responsibility to obtain consent from those who are personally involved in the study and to explain the purposes of the research to them. In this study, the students were all observed. It was then explained to them what the study involved and they were asked orally whether or not they were prepared to participate further, by answering the questionnaire. All agreed.

Informed consent is often used in science and social science experiments, interviews, or questionnaires; participants are informed about a range of matters concerning the purposes of the study or any foreseeable risks or discomfort that might arise during the survey before being asked to agree. Several studies on informed consent procedures in the late 1970s suggest that conventional surveys had advocated keeping the introduction short, in order not to lose the respondents' interest or attention; a general explanation of purpose was preferable to a more detailed one, which might antagonise some people (Blumberg, Fuller, and Hare, 1974; Singer 1978a, 1978b; National Research Council 1979; Reamer 1979; Singer and Frankel, 1982; Singer 1984). De Vaus (2002: 60) supports the view that providing too much technical research information to participants may discourage participation, distort responses, and undermine the validity of the findings. Asking people to sign consent forms can sometimes be problematic, as it may make them unnecessarily defensive and suspicious about the study. On the other hand, Singer (1993: 365), in her later survey of using informed consent to research interviewees' response rate and quality, discovered that information about consent had no perceptible effect on response rate or quality. In other words, there was no point in asking participants to sign a consent form unless sensitive or private topics were involved. The participants in the present research were neither

asked to provide any bodily samples (e.g. blood) nor asked to take part in any tests or experiments (e.g. language tests or randomised controlled trials). They were simply asked to provide information regarding their English listening experiences and give their opinions, and they could choose at several points in the project either to participate or not. Although informed consent plays an important role in gaining permission to collect data, the situation in Taiwan is somewhat different from that in the UK or USA. In the present thesis, I contacted the head of the department and the course instructors in the three private universities from April to October 2005, to explain the aims of the study. The head of the department, the course instructor, and the students all consented to participate in or facilitate the research. General research information was given to the head of a department or school in both universities, before contacting the course instructors (Appendix C.2). A letter of permission was sent to the five course instructors in the pilot and main studies in order to gain access to their classrooms (see Appendix C.3). All of them, including the head of the department and the teachers gave me permission either by telephone or emails, to observe and analyse their teaching and assessment procedures. The description of the questionnaire survey that was given to the students in both questionnaires can be found in Appendix B.

It is often considered best that all participants in a research study should be anonymous and their data kept confidential. Indeed, Grinyer (2002) highlights the fact that anonymity for respondents/participants is assumed to be an integral feature of ethical research. The introduction of the UK Data Protection Act (1998) emphasises that the consideration of anonymity and privacy is no longer simply a matter of ethics; it can also involve legal implications. The fundamental principle of the Act is the protection of the rights of individuals in respect of personal data held about them by a data controller, including academic researchers (*ibid*, see

Grinyer, 2002). In other words, the researchers are responsible for protecting their respondents' identities while presenting and/or publishing their materials. However, at the time of writing in 2007 there was no equivalent law protecting the confidentiality of survey data in Taiwan. In addition to anonymity, assuring confidentiality helps improve the quality and honesty of responses, and has often been found to encourage participation in a study (De Vaus, 2002: 62). One way of achieving confidentiality is to avoid having a third party (e.g. a teacher) administer surveys (e.g. to students). Grbich (1999; see Grinyer, 2002) suggests that respondents should also be told how confidentiality and anonymity will be maintained before surveys. In this case, I followed the rules of the UK Data Protection Act to ensure the anonymity and confidentiality of all participants. The university students did not give their names when I collected their responses, so that they cannot be identified by themselves or by others. Pseudonyms were adopted for staff and students alike and no information from one group was disclosed to the other.

In addition, asking sensitive questions in the survey, such as "Were students satisfied with the way the teacher taught in class?" and "Were students satisfied with the assessment methods the teacher used in this course?" need to be agreed in advance with the teacher, and the five teachers in both pilot and main study agreed before the survey. It was also considered ethical that feedback was provided upon request by course instructors after data analysis, though in the event none was requested. In order to eliminate concerns over emotional responses resultant from the findings, the universities and the teachers were anonymous. The textbooks concerned need to be named, but they are used, to the best of my knowledge, at several universities in Taiwan. In addition, I have tried to avoid couching my report of the results as any form of personal attack.

3.6 Data Analysis

The data collected in the present study comprised classroom observations, interviews, questionnaires, teachers' syllabuses, mid-term and final exam marks, and in-class teaching materials. Photocopies of the published textbooks were requested and permitted. SPSS (Statistics Package for Social Science) 11 was used for analysing the data from the questionnaires and the data presented below are basic descriptive statistics – frequencies and percentages. Data from the classroom observations, interviews, teachers' syllabuses, students' marks, and in-class teaching materials are also analysed and presented via verbal descriptions or tables.

3.7 Validating Case Studies

As case studies generally relate to qualitative approaches to research, according to Patton (2002: 447), case studies involve organising the data by specific cases for in-depth study and comparisons, with the result that well-structured case studies are holistic and context sensitive. Opie (2004: 74) points out that a case study focuses on the interactions of a single instance in an enclosed system that can range from a single person to a department within a school. Case studies can penetrate situations in ways that are not always susceptible to numerical analysis (Cohen, Manion, and Morrison, 2000: 181). In other words, case studies can provide insights into observed effects in real contexts where unanticipated events or uncontrolled variables can be embraced.

Like other research methods, case studies still require checks for reliability and validity. This can be difficult; given the differences and uniqueness of various situations, a case study may prove inconsistent with other case studies or be unable to demonstrate a positivist view of generalisation so that reliability and

validity cannot be guaranteed (ibid: 184; Guba and Lincoln, 1987: 148). Subjective personal judgement and bias may occur when there are not sufficient cases for cross-checking. Naturalists like Guba and Lincoln (1987: 147; Lincoln and Guba, 1985), argue that the aim of naturalism is to develop an idiographic body of knowledge which is a series of “working hypotheses” that describe the individual case; generalisations are not possible since human behaviour is never time or context free. Naturalists suggest using criteria for “trustworthiness” to replace the positivistic concepts of reliability and validity. Lincoln and Guba (1985; Guba and Lincoln, 1987: 150), for example, proposed the use of four criteria in case studies: credibility, transferability, dependability, and confirmability. A number of techniques were developed by them, to achieve these four criteria, including prolonged engagement, persistent observation, triangulation, referential adequacy (checking findings against archive data), or checking collected data with respondents. Viewpoints on triangulation vary. In general, triangulation in a research study refers to the combination of different research methods to answer the research questions. Most sources explain that triangulation in fact requires only a minimum of two vantage points or datasets, usually qualitative and quantitative approaches, to examine a third phenomenon or test hypothesis (Gorard and Taylor, 2004: 43; Aldridge and Levine, 2001: 59). In the present study, both quantitative (i.e. questionnaires) and qualitative (i.e. interviews, observations, documentation) approaches were adopted in order to compensate for any weaknesses in one method through the strengths of another. Triangulation via diverse methods also enhances the trustworthiness of an analysis (Perlesz and Lindsay 2003; ibid.). Cross-checking data with the interviewee was also used to ensure credibility of the case studies.

Case data consist of all the information one has about each case: interview data, observations, the documentary data (e.g. programme records or files), and impressions and statements of others about the case (Patton, 2002: 449). That is, the data tend to be qualitative information. However, Patton (*ibid.*) goes on to demonstrate that quantitative data can be part of a qualitative case study; thus case data can include programme documents, statistical profiles, programme reports, interviews and observations. In the present case studies, the units of data collection comprised questionnaire surveys, interviews, classroom observations, and relevant documents.

3.8 The Use of Questionnaire Surveys

One of the purposes of the study was to measure students' opinions about and reactions to the course and the exams. Opinions are used as the means for measuring attitudes where opinions are verbal expressions of attitudes. Questionnaires are regularly used to explore opinions, as they are reasonably quick to administer and the questions are on the surface at least consistent. The main advantage of conducting a self- and group-administered questionnaire survey is that it is easy to get information from a large number of respondents very quickly, thus saving both respondents' and researchers' time, particularly when the questionnaire consists largely of closed questions (Gillham; 2000; Oppenheim, 1992; Munn and Drever, 2004; Dörnyei, 2003).

One of the conspicuous features of a questionnaire is that the researcher determines the questions to ask with respect to the research purposes, and a range of possible answers may also be provided. Analysing answers to closed questions is relatively more straightforward than open ones. However, questionnaires are not without their problems. Gillham (2000: 2) indicates that the researcher has

already decided the possible answers in a closed questionnaire; all one can find out is which answers are selected. This makes it systematic for the researcher to analyse the collected data while saving time and money, but the opportunity for discovering unexpected answers is much reduced owing to the constraint of preselected answers. Munn and Drever (2004: 5) also note that the information collected by closed questions tends to describe rather than explain why things are the way they are, so that the information may be superficial. In a self-completion survey, descriptive information can be superficial in the sense that it cannot provide the researcher with detailed reasons and information as to why the participants chose certain options. Denscombe (2003: 156) also argues that using closed questions, where the answers are established by the researcher, may not exactly reflect the respondents' true feelings if their opinions happen to be complicated, or they may not fit exactly into the range of options supplied. That is, closed questions are prone to over- or underestimating respondents' real opinions. Open questions, on the other hand, allow respondents the opportunity to express their own viewpoints and feelings without being restricted by preselected answers. Open questions may, however, leave the researcher with enormous amounts of raw data which can be very time-consuming and difficult to analyse. It was considered that using a questionnaire survey in the present study would increase the efficiency of gathering a large amount of data from four groups of students simultaneously, compared with face-to-face interviews. In order to discover answers beyond predetermined ones in the present study, open options were added to most questions.

In a questionnaire, all respondents are presented with the same questions which have been standardised at the piloting stage and there should be no intervention or negotiation from any interviewers, thereby avoiding interviewer

bias. The advantage of using standardised questions is that the stimulus presented to all respondents is strictly under the researcher's control (Munn and Drever, 2004: 4). As participants in a questionnaire survey receive the same questions, De Vaus (2002: 96) states that a question is considered unreliable if it fails to achieve consistent responses. In other words, the reliability of a questionnaire can be increased by eliminating ambiguous or vague wording in its questions. Thus, Payne (1951), Oppenheim (1992: 128-130), Cohen, Manion, and Morrison (2000: 248-249), and De Vaus (2001: 97-99) argue that the designer of a questionnaire should avoid complex language, questions which the respondents may not understand, double-barrelled questions, questions that use negatives and particularly double negatives, leading questions, and "dead giveaway" questions which use absolute, all-inclusive or exclusive words, such as "all", "every", "nothing." Using specific and understandable language while taking respondents' educational and cultural backgrounds into consideration makes it easier for them to interpret and answer the questions. Although the standardised questions are the same for all participants, their interpretation of each question is nevertheless beyond the researcher's control. Low (1999: 505) notes that respondents tend to use the clues provided by the way questions are framed and sequenced in order to help them decide what the researcher is asking; questions are treated as related in topic unless they are told otherwise. Clark and Schober (1992: 27) add that interpretations will often be idiosyncratic, because vague words allow a latitude of specific interpretation within which respondents can presume the question to operate. In this study, wordings in the two questionnaires were examined before distribution, in both English and Chinese, in order to avoid different interpretations of questions.

Gillham (2000: 13) demonstrates that not only do people often tend not to take questionnaires seriously, in that their answers may be frivolous, but that it can be difficult to check the seriousness or honesty of answers. As the extent of interviewee seriousness is difficult to control, there might be too much subjectivity and bias if only a questionnaire survey were to be used in the present study. Hence, listening to the voices of the teachers and making unobtrusive observations of classroom interaction were used to complement the constraints of a questionnaire survey.

Moreover, anonymity for respondents/participants is assumed to be an important ethical dimension of much research (Grinyer, 2002). Munn and Drever (2004: 3) also state that people are less likely to be frank if they are interviewed than if they are able to provide information anonymously, and an anonymous questionnaire survey makes respondents feel freer and safer about giving more personal information. Thus, the questionnaire respondents were all anonymous in the present study.

3.9 The Use of Semi-Structured Interviews

The reason for using interviews was that interviews can, if handled well, yield rich insights into respondents' experiences, opinions, aspirations, and feelings towards predetermined questions or specific topics. Different types of interview can be employed, based on how an interviewer structures the research questions and mode of answering. In the present study, semi-structured interviews were adopted.

Interviews are usually divided into three types: structured, unstructured, and semi-structured. Questions in a structured interview are clear and specific, incorporating pre-coded items or options for interviewees to answer, and they are

also normally sequenced and may be given out in advance. The interviewer can ask interviewees to give direct answers to each question without wasting time on unrelated issues: a solution which is quick and cost effective when it comes to analysing the data. Though structured interviews limit interviewees' answers to predetermined questions, detailed information concerning interviewees' feelings and opinions towards the answers cannot be explored. In contrast, interviewees are allowed much more freedom of speech and thought in unstructured conversations. Interviewers are not required to explain or justify the discussion and it is not necessary for them to prepare well for the interview. As interviewees are free to discuss their ideas with the interviewer without being restricted to a certain issue, the information gathered by the interviewer may be too vast and unrelated to the issue, on which very little factual, useful, or detailed information is in fact provided.

Suchman and Jordan (1990; 1992), comparing structured interviews with unstructured interaction in face-to-face surveys, suggest that it can be difficult to deal with many questions by structured interviewing. Ordinary conversation is locally controlled by speakers mutually agreeing on a topic, being sensitive to the content of current talk and accommodating to different topics. Unstructured conversation affords sufficient occasions to discover differences in world-views but rarely provides sufficient leeway to accommodate differences (Suchman and Jordan, 1992: 254). Interviewers using structured interviews, in contrast, are trained not to redesign questions based either on information acquired in previous responses or on the observable circumstances (*ibid*: 244) and to rely on predetermined questions and predetermined sequences. However, prohibiting redesign of questions in a structured interview can result in serious validity problems, since the interviewer is not allowed to make inferences based on

information gathered, and corrections of questions' premises cannot be made. In other words, apparent mismatches between precoded responses and respondents' real opinions are neither negotiated through interaction nor acknowledged (ibid: 254). Validity for interviews ideally requires that all participants, including the interviewer and the interviewees have a mutual understanding of what questions mean, and the analyst knows how to interpret the data correctly.

Compared with structured interviews and unstructured conversations, semi-structured interviews allow the interviewees to express their concerns, feelings, and opinions relatively freely with regard to specific research topics. May (1997: 111) notes that interviewers can seek *clarification* and *elaboration* of questions based on the answers given via a semi-structured interview. In other words, interviewers can also encourage respondents to provide more detailed reasons by probing related questions based on their answers. The advantage of using semi-structured interviews lies in the fact that the interviewees have freedom to express their opinions in depth when necessary, while the event as a whole stays under the control of the interviewer. It was recognised that using semi-structured interviews in the present study would be more appropriate, as they combined the advantages of structured and unstructured interviews. Answers from structured interviews would be likely to under- or overestimate the reactions of respondents. Unstructured conversational interactions allow different viewpoints and sufficient leeway from respondents, but can lead to irrelevant information with regard to the research questions.

However, Bell (2005: 157) argues that interviews are a highly subjective technique and therefore there is *always* the danger of bias. Miles and Huberman (1994: 253-4) suggest that interviewers tend to differentially overweight information they believe in or depend on, by concentrating on part of the gathered

data rather than all of it. In addition to the bias from interviews, the characteristics of the respondents and the substantive content of the questions are also sources of bias. Respondents may misunderstand the questions. Hence, interview techniques and questions should be carefully planned so as not to make the questions offensive or biased. Triangulation of other research approaches in the present study, for example, quantitative surveys and documentation, can reduce bias of a purely interview approach.

3.10 The Use of Structured Classroom Observation

“Classroom observation” is an approach by which researchers observe and record what really goes on in classrooms in a systematic way. To some extent what happens in the classroom can be accessed by interviewing a course instructor or chatting with his or her students. However, people may not actually do what they say. Observation enables researchers to understand the context of programmes, to see things that might otherwise be missed, to discover things that participants might not freely talk about in situations like interviews, and to move beyond perception-based data (Cohen, Manion, and Morrison, 2000: 305). Classroom observation is thus important where what in fact happens in a classroom can be observed. In the present research, conducting classroom observation was considered useful in three ways (Blythman at al., 1989):

- *As a means of identifying the characteristics of the learner:* different types of students can be detected, such as spontaneous or passive students and teaching can be modified to meet their needs.
- *As a means of on-going monitoring and detection of learning problems as they arise:* teaching and learning are evaluated to discover and prevent any unbalanced interaction between teaching and learning.

- *As a means of assessing students' assimilation of learning materials and resources:* examining how students react to the textbooks or additional in-class material is also related to how the teachers manage students' learning.

Observation is best when it is a kind of “unobtrusive measure” which does not involve direct elicitation of information from research subjects, and where researchers do not seek to manipulate the situation by posing questions (Webb *et al.*, 1966; Adler and Adler, 1994: 378). Although researchers themselves can pose as a participant in order to obtain valuable data, subjective, impressionistic, and idiosyncratic judgements as well as a lack of precise quantifiable measures can result in biased interpretations of observation (Bell, 2005: 187). Since it is difficult to obtain permission to conduct participant observation in Taiwanese universities, non-participant observation was considered more appropriate in this case.

To the extent that observation can really be considered to be an unobtrusive technique, one justification for using it lies in the methodological weakness of questionnaires and interviews. Lee (2000: 2), concluding from Webb *et al.* (1966: 1), suggests that interviews and questionnaires tend to influence respondents because respondents commonly try to manage impressions of themselves in order to maintain their standing in the eyes of an interviewer. Besides, the characteristics of interviews and questionnaires can, under different circumstances, affect the answers respondents provide to sensitive topics or questions (Lee, 2000: 3). An interviewer's intention in asking questions in interviews or questionnaire surveys determines in part the wordings they use. Speakers and their addressees tend to take it for granted that the addressees recognise what the speakers mean by what they say and accept their perspectives (Clark and Schober, 1992: 16). However, Bradburn *et al.* (1979; see Lee, 2000: 2)

found a tendency for survey respondents to over-report socially desirable behaviours when interviewed using less anonymous methods. In the case of questionnaire surveys, respondents may also behave in a socially desirable way to give what they regard as “correct” answers and to create or maintain a socially positive image. The less face-to-face contact there is with the respondents, the less pressure they face.

Schuman and Presser (1981; see Lee, 2000: 3) point out that survey questions are almost never asked in isolation, but as part of a flow of questions; respondents can adjust their responses to a later question in order to make it consistent with the answers they have provided previously. In the present case, it could lead to a biased result if respondents answered questions in the light of previous answers rather than their actual situation. That is, it is possible that respondents might answer questions in terms of what the researcher expects them to answer, rather than stating what they actually think. Observing the actual situation in the classroom can compensate for the inevitable presuppositions made in questionnaire and interview questions.

Observation helps the researcher to uncover the actual events that happen in the classroom. A highly structured observation also takes considerable time to prepare but the data analysis is fairly rapid (Cohen, Manion, and Morrison, 2000: 305). In the present study, using a structured observation in classroom situations is suitable because it lets the researcher pay attention to the lecture without spending time writing down every event.

3.11 The Importance of Piloting

The term “pilot study” (also called “feasibility” study) refers to mini versions of a full-scale study, and involves the specific pre-testing of a particular research

instrument, such as a questionnaire or interview schedule, which is regarded as a preparation for a major study before the formal stage of data collection (van Teijlingen and Hundley, 2002; Baker, 1994: 182-183). Piloting is essential in this study since it helps the researcher understand how long it takes to answer the questionnaire, discover any major omissions or problems, and establish if the wording or instructions of survey questions are unclear or ambiguous.

One of the advantages of conducting a pilot study is that it can give advance warning about where the main research project could fail, where research protocols may not be followed, or whether proposed methods or instruments are inappropriate or just too complicated (van Teijlingen and Hundley, 2002). The purposes of conducting the pilot work in the present study were to assess the feasibility of the full-scale survey and instrument, assessing whether the research protocol was realistic and workable, and identifying logistical problems which may occur.

However, pilot studies have several limitations. These include the possibility of making inaccurate predictions or assumptions on the basis of pilot data, and problems arising from contamination. If data from the pilot study are reused in the main study, errors in the pilot data may contaminate those in the main study. The resulting problems of inconsistent data and participants may lead to inaccurate and flawed data and thus threaten the quality, credibility, reliability, and validity of the main study. Van Teijlingen and Hundley (*ibid.*) argue that the findings from a pilot study cannot guarantee success in the main study because they do not have a proper statistical foundation and are nearly always based on small numbers of participants. In short, accomplishing a pilot study does not promise success in the main study, but it does increase the likelihood of reliability and validity in the

survey. Participants in the present pilot study and in the formal data collection were separate, and the data from the pilot study was not reused.

3.12 The Design of the Two Questionnaires

Two questionnaires – one for the mid-term examination and the other for the final – were designed for this study. Both questionnaires were divided into four parts: the first part surveys the relationship between the course and students' listening needs; the second part is about their general preferences concerning the English listening course and the in-class tasks; the third part investigates students' opinions about the two exams; part four consists of general comments about the course and the test. Closed questions were primarily used in the questionnaire, but open options were included in most cases. The content and layout of the mid-term questionnaire and the final one were nearly the same except for two questions; the layout of the mid-term questionnaire was as follows. The version of the questionnaire displayed below is the one used for piloting (Appendix A.1); the modified questionnaire after piloting is given in the next section (Appendix B.1).

3.12.1 Questionnaire for the Mid-Term Examination

**Questionnaire on Taiwanese University Students' Attitudes towards the English
Listening Course and the Mid-term Test**

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. This questionnaire is intended to investigate the relationship between the English listening course and university students' listening needs, their general preferences concerning the course and the in-class tasks, their opinions about mid-term exam, and their general comments about the course and the test. This questionnaire is for academic research only, and your answers will not be revealed other than when reporting the work.

The beginning of the questionnaire included the title of the questionnaire, the name of the researcher, and the purpose of the research, including how the information gathered would be used. It was considered ethical to provide the

respondents with the background information about the research and the questionnaire. All data collected in the present study were kept confidential and anonymous.

PERSONAL INFORMATION

1. Department: _____
2. Year: _____
3. Gender: M F

Respondents were asked to provide their basic personal information, such as department, the year of their study, and gender in Questions 1, 2, and 3.

Part One – About the course before the mid-term exam

4. Do you think the contents of the in-class English listening materials have been difficult to understand? Please put a tick (✓)

1 2 3 4 5

never : ___ : ___ : ___ : ___ : ___ : always

If "1" NEVER, go on to Q5

If you tick 2 to 5, what in particular has been hard to understand? Please tick (✓) the appropriate box(es).

- Topical content Vocabulary Accent Speech rate
- The use of colloquial language Sound quality
- Text type (e.g. *news broadcasts*, *lecture*, and *consultative dialogue*); if you tick this box, please specify which of the three has given you most difficulty in understanding the content? _____

5. Are the course contents relevant to your listening needs?
 - Yes, please explain. _____
 - No, please explain. _____

Two questions in Part One aimed at investigating university students' opinions towards the English listening course before the mid-term. Question 4 asked the respondents to say how far they perceived the contents of the in-class listening materials too difficult to comprehend. Seven options regarding topical content, vocabulary, accent, speech rate, the use of colloquial language, sound quality, and text type were developed from Question 4 to act as follow up

responses about the reasons why they thought the aspects of course contents had been hard to understand. Respondents who reckoned the text types to be hard to understand were asked to specify which of the three had given them most difficulty. Question 5 asked the respondents whether the course they had taken was relevant to their listening needs. If the course content did not meet the students' listening needs, this would indicate inconsistency between the course objectives and the students' needs and threaten the validity of the test content.

Part Two – Your general preferences about English listening classes

6. In which mode of answering in English listening comprehension classes do you think you perform better?

Writing

Speaking

Please explain. _____

7. Which type of speech in English listening comprehension classes do you think you can understand better?

Monologue (only one person speaks in the listening passage)

Conversations/dialogues (more than two people interaction)

Please explain. _____

8. What type of comprehension question do you prefer most for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

9. What type of comprehension question do you prefer least for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

Questions 6, 7, 8, and 9 in Part Two were designed to elicit students' general preferences about the English listening course they were currently taking.

Different modes of answering English listening comprehension questions might influence students' performance on their mid-term and final exams and then "wash back" to affect their motivation towards the lessons. Writing and speaking were viewed as two basic modes of output by which students gave their answers in common listening tests. Different types of speech – monologue or conversations/dialogues – could also affect students' comprehension and their reaction to the two types of speech might vary owing to the text components. This object was to find out whether students were in favour of certain modes of answering and types of speech, as it was assumed that students' preferences for a particular mode of answering or type of speech in listening tasks, might well affect their performance on the two examinations. In addition, different test methods could also influence students' performance in the examinations. It was hypothesised that one reason why students might prefer certain types of comprehension question was that they could gain more marks by answering that kind of question. On the other hand, a type of question which respondents were not capable of answering could cause them to lose marks more easily. Five common types of comprehension question – short-answer, multiple-choice, true or false, cloze, and dictation question – were accordingly included in Questions 8 and 9 to discover which of them were preferred most and least.

Part Three – About the mid-term test

10. By and large, were you satisfied with your performance on this test? Please circle O the number which best describes your feeling.



Part Three contained seven primary questions (Question 10 to 15) regarding students' opinions and feelings towards their performance on the mid-term exam.

Question 10 used a Smiley Face Assessment Scale (SFAS) to evaluate the respondents' satisfaction with their performance on the mid-term test by circling one of the five faces. The scale was devised by Mortimore *et al.* to measure children's preferences and attitudes towards various aspects of school life (see Mortimore *et al.*, 1986; Davies and Brember, 1994: 448; Macklin and Machleit, 1989: 253). Because of the age of the children, a verbal scale like "strongly agree", "agree" etc. was not used, but they developed a set of faces with expressions which changed in five steps from very happy to very sad (Davies and Brember, 1994: 448). As Davies and Brember (*ibid.*: 447) obtained reliability indices of .71 and .87, it was decided to use them here with young adults. Using the cartoon smiley face scale was intended to give a general indication of students' satisfaction with the exams before moving to specific questions later on.

11. Listening passages can be difficult for many reasons. IN THIS TEST, here are ten possible reasons. Can you say how far each of these proved difficult? Please circle (O) the appropriate number.

1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree

<u>IN THIS TEST,</u>	SD				SA
a. The topics of the test tasks were difficult.	1	2	3	4	5
b. The topics of the test tasks were representative of the curriculum taught in class.	1	2	3	4	5
c. The test tasks were harder than those used in class.	1	2	3	4	5
d. The accent was too difficult to understand.	1	2	3	4	5
e. The vocabulary was difficult.	1	2	3	4	5
f. The sentences were too complicated to understand.	1	2	3	4	5
g. It was hard to understand what speaker(s) said because they spoke fast.	1	2	3	4	5
h. It was hard to understand monologue speech.	1	2	3	4	5
i. It was hard to understand conversations/dialogues.	1	2	3	4	5

Question 11 aimed at discovering how far the topics in the test were representative of the curriculum taught in class, and whether the test tasks were

harder than the exercises students did in class. In addition, Question 11 also explored how far the linguistic features, such as topical content, accent, vocabulary, the complexity of sentences, and the speed of a speaker(s), worried the students when they heard the texts in the exam. In other words, the question explored how respondents perceived the phonological, morphological, syntactic, and sociolinguistic knowledge involved in the listening texts as difficulties they worried about, and how far the test contents were perceived as relevant to the course taught in class. In order to do this, a five-point scale with a midpoint was used. The midpoint thus functioned slightly differently from its use in traditional attitude questionnaires, in that there was no validation problem if the midpoint frequencies were high. Indeed a high figure lent greater credence to the “Agree or Disagree” figures (i.e. they were less contaminated.).

From social psychologists’ views, attitudes can be measured by a quantitative method, such as attitude scales (Thurstone, 1928: 128; Likert, 1932; Hogg and Vaughan, 2005: 180). Respondents are asked to indicate whether they agree or disagree with a series of statements that reflects particular attitudes or opinions. In his survey of reliability of attitude scales, Edwards (1957; see Shaw, 1966: 615) states that the reliability of five-point bipolar Likert scales lies between .72 and .94 and can be considered as a reliable device for measuring attitudes. In the present study, five-point bipolar verbal scales were used in Question 11. Respondents indicated their intensity of agreement or disagreement from “strongly disagree”, “disagree”, “neither agree nor disagree”, “agree” to “strongly agree” based on the statements given. However, Low (1999: 504) argues that Likert questionnaire responses are problematic as they are frequently associated with incongruities of various sorts. Incongruity refers to problems of coherence between the question stem and the rating expression (*ibid.*), which makes it difficult either for

respondents to answer or for the questionnaire designer to interpret a response. Low (ibid.) discovered that it can be effective to use a Think Aloud (TA) approach to explore reactions to questionnaire items involving incongruities, but this was/is very time consuming to do so a TA approach was not employed in the present study.

12. IN THIS TEST, was colloquial language used?

Yes

No

If YES, did you think that it was difficult to understand colloquial language?

Yes

No

In addition to the influence of linguistic features on test performance, Question 12 investigated whether it was difficult for students to understand colloquial language in the mid-term exam.

13. Did you find any of the following to be a problem when you took the mid-term test?

Please tick (✓) the appropriate box(es).

1. The quality of the recording and/or the visual equipment

a. The quality was so poor that I couldn't hear the text clearly.

b. The background noise was too loud.

c. The volume of the tape recorder or audio-visual equipment was too small.

d. The quality of the recording was clear, and I can hear the texts clearly.

e. Other: _____

2. Testing time

a. Time was too limited to answer all the questions properly.

b. The testing time was sufficient for me to answer all the questions properly.

c. Other: _____

3. The test/task instructions

a. The instructions of each test section were not clear.

b. The instructions were too complicated.

c. The test/task instructions were clear.

d. Other: _____

4. The length of the listening texts

a. The texts in the test were longer than I was used to listening to in class.

b. Longer texts increased the difficulty of understanding.

c. The length of texts in the test was similar to what I was used to listening to in class.

d. Other: _____

While taking an English listening exam, different task characteristics need to be taken into account. On the basis of Buck's classification of task characteristics (2001), students' performance on the tests might be mainly influenced by characteristics of the setting, the test rubric, and the input. Consequently, examining students' opinions towards (1) the task features regarding the quality of the recording and/or the visual equipment, (2) testing time, (3) the test/task instructions, and (4) the length of the listening texts were viewed as an essential part of judging how a test was designed. Respondents were asked to tick the options provided. An open option was left at the end of each item for respondents to fill in if they found other task characteristics to be a problem.

14. What type of comprehension question did you find easiest in this test? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
 True/false questions Cloze questions

15. What type of comprehension question did you find most difficult in the listening test tasks? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
 True/false questions Cloze questions

As mentioned in the discussion on Questions 8 and 9 in Part Two, Questions 14 and 15 aimed to discover whether respondents found it easy or difficult to answer test questions using particular test methods. It was hypothesised that students might find a certain type of question difficult to answer, and this might influence their performance on the tests.

Part Four – General Comments

16. (a). Have you learned what you expected to learn so far from this class?

Yes

No, please explain the reason(s): _____

(b). Were you satisfied with the way the teacher taught so far in class?

Yes

No, please explain the reason(s): _____

(c). Were you satisfied with the assessment method the teacher used in the mid-term exam?

Yes

No, please explain the reason(s): _____

Thank you very much for taking time to finish this questionnaire!

It is very much appreciated!

Part Four surveyed students' general comments towards the English listening course and the test. Question 16a asked students if they had learned what they expected to learn so far from the course. Questions 16b and 16c asked if students were satisfied with the way the teacher taught in class, and the assessment the teacher used. It was important to find out, from the students' perspectives, whether the course was designed, taught, and assessed in a way which the students accepted. In other words, finding out whether the course was taught and tested based on the teacher's curriculum plan or the students' learning needs was considered relevant to a classroom assessment situation. Blank spaces were provided for respondents to explain the reasons why they chose the negative answer NO. It was discussed in Section 3.5 that sensitive questions (i.e. Questions 16b and 16c) were permitted by the five teachers in both pilot and main study before they consented to the survey. A brief thank you for completing the questionnaire was included at the end of the questionnaire.

3.12.2 Questionnaire for the Final Examination

The content and layout of the final questionnaire was almost the same as the mid-term one, but a few modifications of wording were made. The wording “mid-term” was replaced by “final” in Parts Two and Three; the wording “before the mid-term exam” was replaced by “after the mid-term exam” in Question 4. The phrase “so far” in Questions 16a and 16b was replaced by “after the mid-term exam”, and the wording “in the mid-term exam” in Question 16c was replaced by “in the final exam”.

To summarise, the questionnaires were considered to constitute a method for examining whether a test meets its purposes because students’ feelings towards the two exams might be closely linked to their test results. For example, if a few students did not prepare well for the exam and failed it, they might respond negatively to later questions about test content, test method, and even the teaching. Hence, it was important to conduct interviews with the teachers, who were also the test constructors, to see if the course and test contents were perceived as being as difficult as their students felt.

3.13 The Design of the Interview Questions

The interview questions were designed to explore the course instructors’ own perceptions of course objectives and their beliefs in teaching and testing students. Seven questions investigated teachers’ attitudes towards testing and related to test constructs/objectives, the selection of the test content, the perceived level of difficulty of the test content, and washback effects.

3.13.1 Interview Questions for the Mid-term Exam

There were seven interview questions with the teachers in the present study:

1. Where did the content of the test come from?
2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?
3. What did you expect the students to have learned from your class?
4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them?
5. How did you identify and decide on the difficulty of the content/items in the mid-term test? What were the criteria?
6. Why did you choose the particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?
7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier)?

Question 1 asked each course instructor how he or she decided on the content of the mid-term test. Did they design it by themselves or did they reuse other tests? Pass marks for the mid-term and final exam were asked for in Question 2. Cut-off scores are crucial in classroom achievement tests. Normally 60% is set as the pass mark in the Taiwanese educational system, but different teachers might determine or adjust the cut-off scores in light of their teaching objectives and the results obtained. Question 3 investigated what the teachers expected their students to have learned from their classes. It is possible that the language skills each course instructor expected his or her students to have learned from class depended very much on their teaching and testing purposes. The test objective of the mid-term exam and its relation to the teacher's teaching plan and their students' learning outcomes, including what the teachers think about their achievements,

can be a clue to how they evaluate their assessment methods and teaching contents, and this was asked in Question 4.

It was hypothesised that the difficulty of the content and items in the mid-term test might have a considerable impact on students' performance. Consequently, it was important to investigate how course instructors identified and decided on the difficulty level of a test. Question 5 accordingly examined what kind of criteria the teachers used: for example, the students' listening performance in class, the amount of vocabulary they knew, or the level of the course book/materials. In addition to the level of difficulty of the test, Question 6 explored the reasons why the teachers chose particular type(s) of comprehension question to test their students. It is possible that the teachers believed that different test methods might favour particular types of student, or they might believe the students would perform better on certain types of comprehension question. Question 7 asked if the result of the mid-term exam would impact on the teaching and testing of English listening in the following second half semester. In other words, would the perceived difficulty in the contents of teaching materials be increased or decreased and would the final test be designed to be more difficult or easier?

3.13.2 Interview Questions for the Final Exam

Four questions, which were similar to the questions in the mid-term interview, were asked again, but in the context of the final exam.

1. Where did the content of the test come from?
2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them?
3. How did you identify and decide on the difficulty of the content/items in the final test? What were the criteria?
4. Why did you choose the particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

To summarise, face-to-face interviews with the teachers were intended to balance and compensate for students' own points of view on the exams. However, it is possible that the teachers' and students' accounts did not reflect what went on in class. Unobtrusive observation in classroom settings was thus adopted to observe actual classroom interaction between teachers and students.

3.14 The Design of the Classroom Observation Checklist

The aim of conducting the classroom observations was to explore the classroom communication and interaction between the course instructor and the students observed by the researcher. A checklist was designed and used in the present study. Items 1 and 2 in the checklist, which were adopted from part of a Laptop Learning Classroom Observation Form on the Internet (http://www.mcmel.org/MLLS/eval/Observation_Checklist_v4.pdf), were slightly modified to meet the purpose of the study – the interaction between teacher and student in the English Listening classes in Taiwanese universities. Item 3 consisted of several essential characteristics of task-based instruction in language classrooms selected by me. Whether or not Items 1 and 2 from the Internet were valid for their purpose was unknown, so a pilot study was conducted to validate the modified checklist for the main study.

The beginning of the checklist consisted of a brief statement of the goal of the observation, the case number, the number of observation, and the date of observation.

<u>Classroom Observation Checklist</u>	Case No.: _____
<p>The purposes of the Classroom Observation Checklist are to investigate teachers' and students' in-class behaviours as well as the in-class atmosphere, including the interaction between teachers and students and its relation to the curriculum objectives.</p>	
No. of Observation: _____	Date: _____

Time sampling is used to record what was happening at each moment (Cohen, Manion, and Morrison, 2000: 305). In the present case, it was known that English listening lessons tended to last about 100 minutes with a 10 minute break in the middle. Class time was accordingly broken into segments of five minutes. The role of a teacher can determine how actively students are involved in a class, and it can affect their motivation and attitudes towards learning. Item 1 recorded the teacher's role: directing the whole group (telling or lecturing), discussing with the whole group, managing feedback discussion, and facilitating or coaching while students are discussing. The observer needed to circle the appropriate number every five minutes.

Minutes	5	10	15	20	25	30	35	40	45	50
Item 1. Teachers Role										
① Directing whole group (telling, lecturing)	①	①	①	①	①	①	①	①	①	①
② Discussing with whole group	②	②	②	②	②	②	②	②	②	②
③ Managing feedback discussion	③	③	③	③	③	③	③	③	③	③
④ Facilitating / coaching	④	④	④	④	④	④	④	④	④	④

Minutes	55	60	65	70	75	80	85	90	95	100
Item 1. Teachers Role										
① Directing whole group (telling, lecturing)	①	①	①	①	①	①	①	①	①	①
② Discussing with whole group	②	②	②	②	②	②	②	②	②	②
③ Managing feedback discussion	③	③	③	③	③	③	③	③	③	③
④ Facilitating / coaching	④	④	④	④	④	④	④	④	④	④

As the role of an instructor and the organisation of a class can impact on the interaction between the teacher and their students, how students behave in class may in turn reflect the way a teacher teaches. Item 2 assumed most student behaviour in class would involve: paying attention to what the teacher teaches, paying attention to other students' presentation or speech, discussing with the teacher, discussing with each other, or feedback from small group work.

Item 2. Student Involvement	Minutes	5	10	15	20	25	30	35	40	45	50
	① Paying attention to what the teacher teaches		①	①	①	①	①	①	①	①	①
② Paying attention to other students' speech or presentation		②	②	②	②	②	②	②	②	②	②
③ Discussing with the teacher		③	③	③	③	③	③	③	③	③	③
④ Discussing with each other		④	④	④	④	④	④	④	④	④	④
⑤ Doing listening tasks		⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥ Feedback from small groups		⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥

Item 2. Student Involvement	Minutes	55	60	65	70	75	80	85	90	95	100
	① Paying attention to what the teacher teaches		①	①	①	①	①	①	①	①	①
② Paying attention to other students' speech or presentation		②	②	②	②	②	②	②	②	②	②
③ Discussing with the teacher		③	③	③	③	③	③	③	③	③	③
④ Discussing with each other		④	④	④	④	④	④	④	④	④	④
⑤ Doing listening tasks		⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥ Feedback from small groups		⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥

Item 3 concerned several common features of task-based instruction, which was defined and discussed in Chapter 1.8.3, in foreign language classrooms. Feature 1 looked at whether there was at least a problem-solving task for students to do in class. Feature 2 investigated how many opportunities were given to students to speak English in class. Opportunities for speaking English include sentence practice and/or reproductions of listening extracts, group discussion and/or teacher-student interaction for exchanging information and opinions, problem-solving tasks, and games. Feature 3 asked whether students reported findings in groups after problem solving or not. Feature 4 looked at whether the

teachers focused on the understanding of meaning before teaching grammar. In addition, it was important to know if students were given opportunities to reflect on what they have learned and how well they were doing (Feature 5).

Item 3. TBI Characteristics	Obs. No.	Observed or not	Comments
1. There is at least one problem-solving task for students to do in class.			
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.			
3. Students report findings of a task to class, in groups or pairs, after problem solving.			
4. The major focus of teaching is on the meaning and then on the form.			
5. Students were given opportunities to reflect on what they have learned and how well they were doing (i.e. reflection period).			

3.15 Summary

There were two different universities participating in the present study. Each university provided two groups of students (i.e. two classes), and an English listening course; each group of students was led by a teacher. That is, there were four case studies from two universities involved in the present study. The purposes of the present study was to investigate how far task-based instruction was implemented in English listening classes, how far test objectives were clearly specified in the mid-term and final exams, how far communicative language testing approaches were used in the two exams, and how far the results of the two exams had washback effect impact on teaching and learning.

Using any single research approach is likely to be biased; using a range of research methods, however, helps to increase the trustworthiness of the processes of analysing the data. As case studies are usually very context-bound, the decision was made to adopt some form of triangulation in this study. Questionnaires were regarded as efficient for collecting large amounts of data from groups of students in a limited time, and analyses of large amount of data could be undertaken relatively easily. Semi-structured interviews would allow both the interviewer and the teachers freedom to express their opinions beyond predetermined answers; inferences could be made and tracked in the light of their previous answers. Observations in the classroom would help to track events without involving the teachers' and students' personal perspectives. In addition, documentary data, such as in-class teaching materials, test contents, students' examination marks, and official university evaluation programme database were included in the analysis.

Chapter Four

The Results of the Pilot Study

4.1 Introduction to the Pilot Study

The purpose of the pilot study was to test (1) the usefulness of the research design, (2) the appropriateness, validity, and practicality of the three research techniques, and (3) the feasibility of the research questions. The results of the pilot study consist of analyses of the observation, questionnaire, and interview data and these are used to amend items for the main study. The pilot test version is included in Appendix A, and the modified one for the final data collection in Appendix B.

4.2 Pilot Study: Procedural Overview

A group of 41 university students and one course instructor from a private university in Taiwan were selected for the pilot test. The students all majored in English language, and the sample participants in the pilot study were similar to the participants selected from those in the main study in terms of the major and the year of study. The students in the pilot study were all aged eighteen, and comprised thirty-four female and seven male students. Two lessons – one before and another one after the mid-term exam – were observed. The procedure involved observing the classroom *before* the exams, distributing the questionnaires *after* the exams, and finally interviewing the teachers *after* students' exam marks were known. I shall analyse the classroom observation first, then the questionnaire survey, and finally the interview results.

4.3 Pilot Study: The Classroom Observation before the Mid-term Exam

The piloting of the classroom observation checklist was carried out by observing two lessons given to the same class. The first observation was made before the mid-term exam on 3 November 2005, and the second after the mid-term exam (before the final exam) on 15 December 2005. The three primary purposes of piloting the checklist were to find out: (a) whether or not the statements of “presumed” classroom situations applied to the “real” classroom situations observed, (b) whether I would be able to apply the categories on the checklist in real time, and (c) the answerability of research question 1 (*How far is task-based instruction implemented in English listening classes?*). The class started at 1:10pm, lasted for 100 minutes with a 10 minute break in the middle, and ended at 3:00pm every Thursday. I attempted to act as an unobtrusive observer who recorded every event happening in class by filling out the classroom observation checklist and using a digital tape recorder. The evidence suggests I was reasonably successful, as there was little evidence of the students being distracted by me or the recorder; two or three students looked at me for a second or two, and then focused on the class. The observation was carried out with the permission of the course instructor; the letter of permission is included in Appendix C.3.

4.3.1 Time Sampling of Classroom Events before the Mid-term Exam

I noted and classified the class activities every five minutes (see Items 1 and 2; Appendix A.7). However, I found it hard to record the events that actually occurred in the classroom, because events did not always happen every five minutes, some of them had finished after just two or three minutes. Though it was hard to use the schedule while observing the classroom, it was decided to keep the

full schedule for the second observation, and to think about modifying or discarding it after that (see Section 4.6).

4.3.2 Classroom Observation of Task-based Instruction before the Mid-term Exam

The class was taught in English; Chinese was used once or twice only for clarifying ideas. At the beginning of the lesson, the teacher checked if the students had done the homework assigned and discussed the quiz they had taken the previous week. She reminded students of their responsibility to finish the homework in order to pass the course. She then discussed several difficult questions where students had lost marks on the quiz, and checked that her students understood the questions at the end of the discussion.

The teacher then turned to focus on the lesson in the textbook. Students firstly listened to the texts and were asked to spell the key vocabulary and practice the pronunciation in the texts they had heard. The listening questions were related to general social and cultural issues, for example, sightseeing, sports, shopping, and invitations. Students then did the exercises by filling in missing vocabulary which they had heard and reading the answers out together to the teacher. Students were not nominated individually to answer the teacher's questions, but answered all together. Two or three students asked questions to inquire about the usage of certain vocabulary or grammatical sentences, or confirm what the teacher had taught. The teacher tended to discuss the listening contents and ask students do the task simultaneously. She tried to use English to explain vocabulary and meaning in the texts, but while encountering new vocabulary, such as "moustache", she used both English and Chinese to describe it. The teacher said that they would see an English movie in class after the mid-term exam. As the end of class time approached, the teacher assigned the homework, reminded the

students to work harder for the upcoming mid-term exam, and then finished the lesson.

In this lesson, no tasks that required the students to solve problems were used. The opportunities for the students to speak English were primarily focused on spelling vocabulary and answering the exercise questions in the textbook. The students reported answers or spelled vocabulary all together; no group or pair discussions were observed. The major focus of the teaching was not on the meaning, but on the form, because the teacher put the emphasis on correct vocabulary spelling and pronunciation rather than on the understanding of listening contexts. At the end of the lesson, there was no reflection period for the students. The description, which was a post-observation summary, was recorded immediately after the lesson ended. I found no difficulty in using the task-based instruction checklist (Table 4.1), and this appeared to provide exactly the quality and amount of data needed to answer research question 1.

Table 4.1 Pilot: Occurrence of Task-based Instruction Features before the Mid-term Exam

3. Task-based Instruction Features	Obs. No.	Observed or not	Comments
1. There is at least one problem-solving task for students to do in class.	1 st	x	No tasks were used in class. The students were told to answer the questions regarding the listening passages in the textbook.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st	✓ x	The majority of the students spoke English when being asked to answer the questions; only two of them asked questions.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	x	Group or pair discussions were not found, and the students reported answers all together.
4. The major focus of teaching is on the meaning and then on the form.	1 st	x	The major focus of teaching was on the form – vocabulary, pronunciation, and spelling.
5. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	x	No reflection period was found.

4.4 Pilot Study: Questionnaire Survey on the Mid-term Exam

The first survey was conducted right after students took their mid-term exam on 10 November 2005. The forty one respondents were all in their second year of study in the Department of Foreign Languages. The purposes of piloting the questionnaire were to see if I could apply the questionnaire to collecting information from a group of students in the time available, and to examine the answerability of research question 4 (*What kind of problems will influence students' listening comprehension in the two listening exams?*).

The questionnaire was group-administered to the students who completed the questionnaire in the classroom in which they had their listening classes and exams. The teacher was not present while they answered it. The questionnaire used for the pilot test was translated into a Chinese version and the translation was checked by a native speaker of Chinese who also understood English. Two problems with precision of nouns in the Chinese translation were suggested (see Chapter 3.4). I acted as an observer, in case there were any unclear questions or any instructions that needed to be clarified. The students were additionally asked to write down anything they felt was ambiguous or which confused them at the end of each item in the questionnaire. The students did not report any particular difficulties in answering any of the questions. Everyone had finished well within the 30 minutes anticipated.

According to the data collected from the respondents, the topics, accents, vocabulary, and speech rate were perceived as the main problems that influenced their listening comprehension while taking the mid-term exam. Although neither the respondents nor I found any problems with Parts One, Two, or Four of the mid-term questionnaire, I found two problems with items in Part Three. In Question 10 (Part Three), the students were asked to report their general

satisfaction with their performance in the exam by circling a “smiley face” number. As they were in effect asked to infer a scale of satisfaction/dissatisfaction and map this onto the smiley faces, it was difficult to be absolutely certain of how the ratings had been interpreted. It was decided that, for the main study, the oral descriptors “Very satisfied”, “Satisfied”, “Neither Satisfied nor dissatisfied”, “Dissatisfied”, “Very dissatisfied” would be provided (see Appendix B.1 and B.3), so that students would have a clearer sense of the intended meaning of the five faces. However, it was felt that this did not make the faces redundant, as the visual stimulus provided variety on the answer sheet and seemed to retain students’ concentration.

In Question 13(1), a problem was detected while interpreting the phrase “background noise”. As the question did not clearly specify whether it referred to background noise from the recording equipment or noise from the environment, it was not possible to distinguish what kind of background noise was referred to. Thus, it was decided to modify the expression “background noise” for the main study by specifying “noise outside the testing environment” (see Appendix B.1 and B.3). I did not find any problems in reporting and analysing the data in the remaining items in Question 13, which meant that the questions were usable for the main study. The data collected from the questionnaire was adequate to answer research Question 4, in the sense that the topics, accents, vocabulary, and fast speech rate in the listening extracts were perceived as the three main problems that influenced students’ listening comprehension in this exam.

4.5 Pilot Study: Mid-term Interviews with the Teacher

The interview with the instructor was conducted the day after the mid-term exam on 11th November 2005 in the teacher’s private office in the college, which was a

quiet location well suited to face-to-face interviewing. The purposes of piloting the interview questions were to investigate whether interview was an appropriate approach for obtaining information from the teacher, and to answer the research questions 2 (*How far are the test objectives clearly specified in relation to the curriculum and teaching contents?*) and 3 (*How far are communicative language testing approaches applied in the two listening exams?*). The interview questions were presented both in English and Chinese and the teacher could choose to answer in either language. Because she had completed a master's degree from a university in the UK, it was assumed that she could understand the interview questions in English. The teacher nevertheless chose to answer in her native language – Chinese. The interview was tape recorded with her permission, transcribed into Chinese and then translated into English (see Appendix C.4). Table 4.2 is the translation of teacher's interview. The translation was examined by a Chinese speaker who could also speak English and a native English speaker. The main suggestions were simply to make the verb tenses consistent, which was done. The interview lasted approximately twenty minutes.

Table 4.2 Pilot: Translation of Mid-term Interview Data (Questions 1 to 7)

Question 1. Where did the content of the test come from?
Teacher (T): I reused other tests. Because textbook publishers provided practice tests as supplements to their coursebooks, I used the practice tests as the content of the mid-term exam.
Interviewer (I): Was it possible that students accessed the practice tests before the exam?
T: No, it was impossible, since those (practice tests) were only for the teacher, it was impossible for students to obtain the test contents.
I: Did you test their speaking ability in the mid-term exam?
T: No.
I: Could you tell me why you did not want to test their speaking ability?
T: Well....I think that they needed to polish their listening skills first before moving on to the speaking skills, so I'd rather focus on training their listening first.

Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?
T: The cut-off score was 60 for the two exams. The mid-term exam counted as 30% and the final one counted as 40% of the total final score. The other 30% was for in-class coursework.

Question 3. What did you expect the students to have learned from your class?
--

T: The course was for basic English listening practice at beginner level, so I put more emphasis on the pronunciation and correct spelling of vocabulary in the listening texts. I found that the level of students' English ability did not meet the level it should in their present year of study (the second year of a five-year programme), so understanding correct pronunciation and spelling were important for them at this time! They would, I hoped, learn the vocabulary by pronouncing it. Hence, when you audited my class you would have seen that I asked students to practice the pronunciations and spellings of the vocabulary a lot. Besides, the majority of the students were lazy about studying. If you didn't push them to memorise the vocabulary, they would not do so. Because of this, there was a chance that their marks would be terrible, and then the dean would then put pressure on us (teachers); that's why I tended to focus on the basic memorisation of vocabulary and spelling.

Question 4. What are the mid-term testing objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Do you think you have achieved them?

T: I hoped that the students would be able to memorise the vocabulary they read in the textbook because the vocabulary was pretty basic. If they could not memorise the basic words, it would be more difficult for them to understand advanced listening texts. The marks in the mid-term exam tended to be lower than I had expected. Most of the listening contents in this exam were taught in class, but the way questions were asked was different from the way they had practiced in class; the results were not good.

I: The only difference was the ways in which the questions were asked?

T: Yes! In fact the contents of the listening were almost the same.

Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?

T: I decided the level of difficulty based on that in the textbook. When I chose the test content, I considered whether the difficulty of the test content was similar to that taught in class. It was likely that their marks would be lower if I used questions that were too difficult and this might discourage them.

Question 6. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?

T: In fact I did not specifically choose any types of comprehension question. On the one hand, test time was limited, and listening only once was insufficient for them. In this case, asking them to write a lot of words as an answer would have taken too much time; that's why I used questions with "options" for answering, as well as a few short-answer questions. It was also easier for me to score by using questions with options for answering.

I: Listening contents are usually heard only once in an exam, like TOEFL and IELTS, so why did you allow the students to listen more than once?

T: If they merely heard the content once, my experience was that they usually performed very poorly.

Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)

T: I felt that the students did not perform as well as I expected, so I will demand more as regards practicing the pronunciation and spelling of vocabulary. I won't increase the in-class material or the level of difficulty. Because they could not handle such basic listening contents, how can I increase the difficulty of the material? As for the test content in the final exam, the level difficulty will be similar to that in the mid-term one. I hope that their English ability will increase, so I will maintain the level of difficulty in the final exam.

As can be seen from Table 4.2 above, the data was sufficient to allow answers to Questions 2 and 3, in the sense that the teacher clearly focused on vocabulary and grammar practice in class, and vocabulary memorisation in the mid-term exam; no communicative language testing approaches were used. The table also shows that she gave no evidence of finding problems or difficulties in responding to any of the questions. I was accordingly satisfied with the quality of the information the teacher provided since she had answered all the questions and probes and had not contradicted herself. My influence on the answers appeared to be minimal or non-existent (see above data), and my role in giving feedback to the teacher's answers was purely to clarify her answer or to ask for further details; I asked no leading questions.

4.6 Pilot Study: The Classroom Observation after the Mid-term Exam

The second observation was carried out after the mid-term exam on 15th December 2005 with the same participants in the same classroom as in the first observation. This class again lasted 100 minutes and was also observed every five minutes and analysed in terms of the events that happened. While using the time sampling checklist for the two observations, I found one problem. Items 1 to 2 divided the class organisation, the role of the teacher, and student involvement into separate and different events (Appendix A.7); however, in addition to the problem detected in 4.3.1 before the mid-term exam, it was noted that two, or more than two, events could take place simultaneously within one item. For example, from minute 75 to minute 90, the students completed listening tasks while paying attention to what the teacher taught, and discussed things with her. When the three events happened in the same time segments, it was difficult to analyse and interpret the data. However, the TBI characteristics in Item 3 and the lesson description covered the

events in Items 1 and 2, so I found that using Item 3 and the lesson description was sufficient to describe the implementation of TBI in the listening classroom (see Section 4.3.2 and 4.6.1). Thus, Items 1 and 2 were discarded due to the difficulty in encoding, decoding, and analysing them.

4.6.1 Classroom Observation of Task-based Instruction after the Mid-term Exam

At the beginning of the class, the teacher checked the homework which she had assigned the previous week. The teacher then turned to the textbook to continue the lesson she had not finished the previous week. Students were told to do the exercises in Lesson 11 by listening to the contents of the tapes and read out the answers together. The teacher taught the grammar in the exercise and asked whether the students had understood it or not. None responded with questions. Consequently, the teacher started the next lesson. All students listened to the content once and the teacher explained the conversations to them. The students concentrated on the lecture and were asked to pronounce and spell out the new vocabulary that they heard in the conversation.

After listening to the contents twice and practicing the new vocabulary, the teacher asked the students to do a task in the textbook. The teacher nominated a student to be the representative to do the task in front of the white board. Then the teacher told the students to record the news report she read by drawing pictures on a piece of paper (the representative student drew on the white board). The content of the news report involved descriptions of three bank robbers, and the students were asked to listen to descriptions of the robbers' appearances from their teacher and then draw their pictures. After the teacher had read the report, she examined the representative student's drawing and gave the correct answers to all the

students. The class had to finish in 100 minutes; none of the students at any point asked any questions.

In this second lesson observed, the task that was assigned to the students had a specific goal – drawing the characteristics of these robbers. The students did the task individually and did not report their results to the teacher, that is, they were not given a chance to discuss their answers in English. In addition, the students did not appear to ask questions actively. In this lesson, the teacher again focused more on the vocabulary and pronunciation than on the meaning of the listening passages. Similar to the case in the lesson observed before the mid-term, the students were not given any opportunity to reflect on what they had learned or on how well they had done.

In theories of task-based instruction, using authentic in-class materials is regarded as a way to involve learners in authentic language activities in real-life situations (see Chapter 1.5.3). With hindsight, “Authentic texts which reflect a real-life situation were used” should have been included in the pilot. However, the situation was rectified for the main study. It was also recognised that listening scripts would need to be examined for the main study in order to cross-check their authenticity. Apart from adding this item, I did, as before, not find any problems with using the TBI checklist (Table 4.3); it thus appeared to be feasible to use the TBI checklist for the main study.

4.3 Pilot: Occurrence of Task-based Instruction Features after the Mid-term Exam

3. Task-based Instruction Features	Obs. No.	Observed or not	Comments
1. There is at least a problem-solving task for students to do in class.	2 nd	✓	A problem-solving task was found in the drawing-picture task.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	2 nd	✗	The students did not speak English in class.

3. Students report findings of a task to class, in groups or pairs, after problem solving.	2 nd	✗	Although a task was assigned to the students, they were not told to report the findings.
4. The major focus of teaching is on the meaning and then on the form.	2 nd	✗	The teacher again put emphasis on practicing the vocabulary and spelling.
5. Students were given opportunities to reflect on what they have learned and how well they were doing (i.e. reflection period).	2 nd	✗	No reflection period was found.

4.7 Pilot Study: Questionnaire Survey on the Final Exam

The second survey was conducted immediately after students had taken their final exam on 18 January 2006. The procedure for the questionnaire administration was the same as for the mid-term one: the teacher was not present, only I remained with the students. This time, everyone had finished after twenty minutes. Again, the respondents did not report any problems with Parts One, Two or Four; and, I found no problems in analysing the data in them. As discussed earlier, in Section 4.4, it was still difficult to interpret the smiley scale, as different interpretations of the faces were possible. This reinforced the decision to add verbal descriptors for the main study.

4.8 Pilot Study: Final Interview with the Teacher

The course instructor, who was interviewed after the mid-term exam, was interviewed again after the final exam on 20 January 2006 in the same room as before. The teacher answered the interview questions in Chinese, and the interview was recorded with her permission, transcribed and translated into English (Appendix C.5). Table 4.4 is the translation of teacher's interview after the final exam. The translation was examined by a Chinese speaker who could also speak English and a native English speaker; suggestions regarding the translation were again to make the verb tenses more consistent. It took approximately fifteen to twenty minutes to get through the four questions.

Table 4.4 Pilot: Translation of Final Interview (Question 1 to 4)

<p>Question 1. Where did the content of the test come from?</p> <p>T: I reused other test items for most of the test content, just like the listening test items I used in the mid-term exam. There was an extra point for a dictation test in the last test item. It was not compulsory for students to answer, but there would be a point for a correct answer, but no minus point for wrong answers.</p> <p>I: Where did the dictation question come from?</p> <p>T: It was from a small paragraph in the textbook which had been taught in class. It was an easy test item. As long as they worked hard in and after class, the students should have gained point on this item.</p>

<p>Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them?</p> <p>T: I still expected that the students could memorise the vocabulary and its pronunciation they had heard and the basic grammar sentences in class. The mean of their final marks was slightly higher than the mid-term one, yet more failed, although this was not as high as I expected. Although I increased the amount of exercises done in class, it seemed that the students did not review them after class. What I can say was I have tried my best.</p> <p>I: What were the marks you expected them to obtain in order to achieve your teaching objective?</p> <p>T: I felt that they should reach at least 75 to 80 because the test was really not that difficult.</p>

<p>Question 3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria?</p> <p>T: The difficulty of the final exam was decided based on the level of the textbook and the students' performance in the mid-term exam. Although they did not perform well in the mid-term exam, the reason why I added a bonus point in the final exam was to enhance their marks. But as they often failed to spell correctly, they lost marks.</p>
--

<p>Question 4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p> <p>T: There were no big differences in the types of comprehension question between the mid-term and the final exam. The only difference was the short answer question for dictation, including filling in missing vocabulary and phrases in blanks. I expected that they would be able not only to understand the vocabulary but also spell it out correctly.</p>
--

As can be seen from Table 4.4 above, the results of the mid-term exam did not change the way the teacher taught after the mid-term exam or the difficulty of the final exam, and this was exactly the quality and amount of data needed to answer research question 5. Again, the teacher also gave no evidence of finding problems or difficulties in responding to the interview questions. I was also satisfied with the results of the final interview; examination of the transcript suggested that my influence on the answers again appeared to be minimal or non-existent.

4.9 Summary

The purposes of conducting the pilot study were to examine whether the research questions were useable, and to test the appropriateness, validity, and practicality of the three research techniques in the research design. Essentially the three methods held up well and between them generated enough data of the right quality to answer all four research questions. The respondents did not report any problems in answering the questionnaire, but I did modify and rephrase two questions in the two questionnaires to make it easier for the students to answer; no fatigue was reported by any students after completing the two questionnaires.

As for the observation checklist, I discarded two questions on the form to optimise the match with the actual classroom situations I observed. Essentially, I also found almost all the research points and checklist items were useable and answerable. The procedure of administering the classroom observation form, the questionnaires and the interviews was also found to be practical, in the sense that I, as the observer, was able to operate the systems and the participants were at all times helpful and did not react adversely.

Although the primary aims of the pilot study were methodological, there were several interesting points about the content of the answers which are worth discussing briefly. First and foremost is the fact that there were indeed problems designing appropriate listening tests, and although the teacher tried to link the mid-term and final tests, based on her perception of the students' strengths and weaknesses, the criteria for success seemed unrelated to communicative uses of language (spelling and word identification). It appeared that traditional teacher-centred language teaching *was* used in the case of the pilot study. Secondly, although the teacher was concerned about the students' apparent failure to learn anything, or progress, no attempt was made to examine the test scores, or to

discover and to react to the problems that the students consistently raised in the questionnaire. The results showed that there might be a large mismatch between teaching and assessing English listening, and a mismatch between the actual implementation of communicative English teaching and the policy promoted by the government, at least in this particular Taiwanese university. These results clearly validate and justify the general research aim of how far the teaching, learning, and assessment of English listening courses in Taiwanese universities influence each other. The next four chapters discuss the findings of the four cases from the two universities in Taiwan.

Notes to Chapter Four

¹ A bonus item gives extra score to the total score, but no deduction of the total score is made when students fail to answer it.

Chapter Five

Data Analysis – Case 1

5.1 Introduction

Chapters Five and Six include the analyses of two groups of students from University A. University A is a private university in central Taiwan which includes five colleges – Engineering, Management, Design and Arts, Biotechnology and Bioresources, and Foreign Languages. It operates a four-year system, in the sense that it normally takes four years to complete an undergraduate degree. In the Cases 1 and 2, there were 64 students in total from the Department of English Language taking the English Listening course which was titled “*English Listening*”. The 64 students were in their second year of study; they were divided into two separate classes and were taught by two different teachers. The department provided two teachers for this listening course, and the students could choose the class led by either teacher. The first case from University A is analysed in this chapter while the second is examined in Chapter Six. Firstly, the background, including the aims of the listening course, in-class teaching materials, and exercises, is introduced in Section 5.2. Secondly, the teaching approaches or methods the teacher used in the English listening class are examined, to explore how far task-based instructions were implemented before the mid-term exam in Section 5.3. Thirdly, the students’ opinions towards the mid-term exam are presented in Section 5.4. Finally, the teacher’s perspectives on teaching and testing objectives, in-class materials, and test results are examined in the light of the test contents and teaching materials in Section 5.5. The classroom observations *after* the mid-term exam are compared in Section 5.6.

The discussion of the questionnaire and interview *after* the final exam are compared in Sections 5.7 and 5.8.

5.2 Case 1: Background to the Listening Course

There were 31 students and one course instructor in this listening class (Case 1). The teacher (Dr. N) obtained her PhD in Teaching and Leadership at a University in the USA and had three year's experience of teaching students at university level. The students in her class were second-year undergraduate students who specialised in English Language. According to Teacher N's syllabus, the course in this semester was designed to

improve undergraduate students' listening ability at intermediate level to understand English in "general and academic" situations. Listening comprehension would be achieved through various exercises inside and outside the classroom. The primary emphasis was on listening comprehension and the secondary one was on oral presentation. (Extract from Dr. N's syllabus notes)

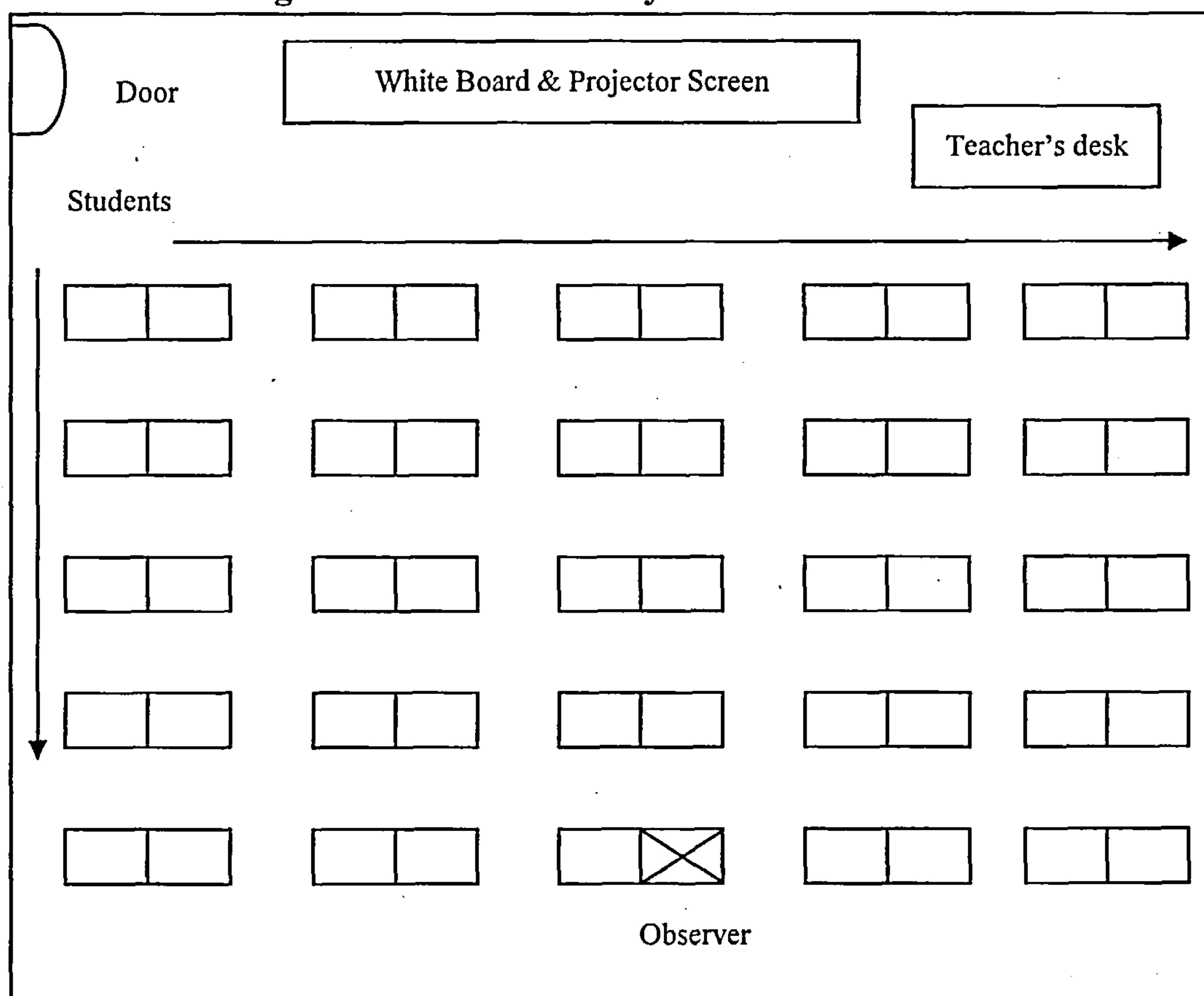
It is clear from the above extract that Dr. N expected that her students would be able to understand English in both general and academic situations. She used two textbooks in class. According to the interview with Dr. N in Section 5.5 and her course description, one textbook (*Impact Listening 3*) was aimed at listening to various daily life contexts, in the sense that general social and cultural aspects in everyday situations were included (Appendix D.1). According to the Introduction of *Impact Listening 3* (2001: 3-4), the listening extracts in the textbook, which were "drawn from or based on authentic conversations" and were "based on unscripted recordings", aimed at providing listeners with a lively variety of tasks. The other textbook (*Mosaic 1*) concerned different academic topics, where lectures were the main listening contents (Appendix D.2). Three

observations were carried out before the mid-term from 10:20am to 12:10pm on the 16th, 23rd, and 30th of October 2006. Researching the first case lasted 12 weeks from the 16th of October 2006 to the 10th of January 2007. Each lesson lasted 100 minutes with a ten-minute break in the middle.

5.3 Case 1: Classroom Observation before the Mid-term Exam

The students had their lessons in an audio-visual classroom which was specifically designed for English listening classes. Each student was equipped with a personal headset and a microphone (Figure 5.1). Exams also took place in this classroom.

Figure 5.1 Case 1: The Layout of the Classroom



5.3.1 Case 1: First Observation

The class was first observed on the 16th of October 2006. Unit 3 of *Impact Listening 3* and Chapter 2 of *Mosaic 1* were the main activities in this lesson. At the beginning of Unit 3 in *Impact Listening 3*, Dr. N asked the students to do the vocabulary exercises in the textbook, and she checked the answer with the students by listening to the CD. Before the students listened to a passage, she told the students to listen to the main idea first; she then played the passage again and asked the students to do the exercise in the textbook. After doing the first part of exercises, Dr. N asked the students whether they knew the answers or not. Almost all of the students seemed eager to answer the questions (possibly as those who answered correctly could get an extra point added to their participation mark). Next, Dr. N played the CD again and asked the students to do another exercise which asked for detailed information from the passages. Dr. N explained new vocabulary or useful phrases that the students encountered. In the second half of the class, Dr. N went on teaching Chapter 2 in *Mosaic 1*, which she had not finished the previous week. In order to remind the students of the content they had listened to, she played the CD and asked the students to do the exercises in the textbook.

In this first observation, neither pair nor group discussion took place and the students worked either alone, or as a whole class. They answered questions individually and concisely when nominated. The students remained silent in class if no questions were asked. Though Dr. N tried to give all students equal opportunities to answer questions by not nominating the same student twice, it nevertheless appeared that the students did not all receive equal or sufficient opportunities to practice their English with each other, or with the teacher in class. Dr. N taught and explained ideas in English in this first observation, and the

students appeared to understand the lecture and took notes when necessary. In class, Dr. N emphasised that it was more important to understand the main ideas in the listening passages than the detailed information. As for the authenticity of the listening materials, the listening extracts in *Impact Listening 3* tended to be scripted in the sense that the sentences were too fluent without any broken sentences or false starts in the conversations (Appendix D.3). In such a situation, the listening extracts used in class were not considered authentic compared with language used in real-life situations.

5.3.2 Case 1: Second Observation

The second observation was carried out on the 23rd of October 2006. Unit 4 of *Impact Listening 3* and student presentations were the main activities in the lesson. In the first half of the lesson, Unit 4 of *Impact Listening 3* was taught. The students did the vocabulary exercises in the textbook first, and listened to the CD for general information and did the exercises; Dr. N checked the answers with the students. The students then listened to the CD again for detailed information and did more exercises. They were asked to answer questions individually – by raising their hands, and Dr. N nominated one student to answer each question. Students who answered correctly could again gain an extra credit. The procedure for teaching *Impact 3* in the second observation was similar to that seen in the first.

After the break, two groups of students presented a radio talk show. There were four students in a group and they were asked to listen to different kinds of radio talk show, such as ICRT (International Community Radio Taipei), outside the classroom and to produce a live talk show by themselves. The students had to present in English and scripts were not allowed. The topics of the students' radio

talk shows ranged broadly, from pop music, short story telling, news, weather reports, and movie discussions, to interviews with super stars. One group discussed a movie they had seen the previous week; the content included the reasons why they had decided to see the movie and how they thought about it after seeing it. The four students all presented their opinions on the movie and the presentation lasted for 15 to 20 minutes. After the presentation, the presenters asked the audience three questions related to it. The listeners who answered the questions received as before extra credits. The other group of students presented an interview with a famous Taiwanese female singer. One student acted as the female singer and the other three asked her questions about her new album and her future development. This presentation also lasted around 15 to 20 minutes. They too asked the audience three questions regarding the interview. Dr. N was in charge of the opening and the ending of the presentations; she did not ask questions during the presentation.

In the two observations, the students were asked to do the exercise question in the two textbooks. No problem-solving tasks were found in this lesson. Compared with the lesson in the first observation, more students (though still not all of them) had more opportunities and a longer time to practice their English via presentation in class. Compared with the conversations in the two textbooks, the target language was more authentic in the sense that the presenters had to produce their own language to interact with the audiences. The presenters and the audience looked happy and involved in the presentation. However, the listeners answered the questions individually to the presenters or the teacher, but not in groups or pairs. Dr. N again told the students to listen for the main ideas in the passages and then she would teach new vocabulary and useful phrases.

5.3.3 Case 1: Third Observation

The third observation was carried out on the 30th of October 2006. In this last observation before the mid-term, Unit 5 of *Impact Listening 3* and Chapter 3 of *Mosaic 1* were taught, and the students watched an American situation comedy called “Seinfeld” (Series 1). At the beginning of the lesson, Dr. N taught Unit 5 of *Impact Listening 3*. The procedure for teaching Unit 5 was similar to that used in the previous two units (see Sections 5.3.1 and 5.3.2). After finishing Unit 5, Dr. N started teaching Chapter 3 of *Mosaic 1*. She played the CD initially and asked the students to listen for general information. She told the students to preview the exercises in the textbook and played the CD again. The students were then asked to do the exercises in the textbook, after which they took a ten-minute break. In the second half of the class, Dr. N discussed the answers to the exercises, but only five or six students raised their hands. Dr. N did not finish the whole of Chapter 3, but left the rest till after the mid-term exam. Finally, she played “Seinfeld” (episode 4) to the students. The episode was shown with English subtitles.

No problem-solving tasks were found in any of the three observations. However, the interaction between the teacher and the students in the lessons was limited to the “question-and-answer” mode when the students did the exercises in the two textbooks. In other words, the students simply answered a question when Dr. N asked it. Interactions became more frequent when the students presented the talk show. While doing the exercises in *Mosaic 1*, the students who answered the questions were not as interactive as when they answered the questions in *Impact Listening 3*. Although the listening extracts in the two textbooks were less authentic than speech in real-life situations, the activities, such as the

presentations and the comedy, appeared to be reasonably authentic. Table 5.1 summarises the features of task-based instruction observed in Dr. N's three classes before the mid-term exam.

Table 5.1 Case 1: Occurrence of Task-based Instruction Features before the Mid-term Exam

TBI Characteristics	Obs. No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	1 st	✗	Students were asked to do the exercise question in the two textbooks. No problem-solving tasks were found in this lesson.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
2. There are many opportunities for students to practice English orally.	1 st	✓✗	Interaction between the teacher and the students was limited to students' answers to the exercises in the textbook. Answers were also limited to vocabulary, phrases, and a few sentences.
	2 nd	✓✗	Two pairs of the students presented a live radio talk show. They were not allowed to bring any scripts with them. At the end of the presentation, the pairs asked the audience three to four questions related to their radio show. Other students answered their questions.
	3 rd	✓✗	The same as 1 st observation.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	✗	Students reported their answers individually when Dr. N asked.
	2 nd	✗	Students reported their answers when Dr. N asked or when the presenters asked.
	3 rd	✗	The same as 1 st observation.
4. Authentic texts which reflect a real-life situation are used.	1 st	✗	Although the publisher claimed that the listening passages were based on unscripted recordings in which were authentic, the language in conversations were too fluent without any broken sentences or false starts (Appendix D.3 and Section 5.3.1). The situational contexts were close to real-life situations, but the language and the speed were not as authentic as native speakers' speaking.
	2 nd	✓✗	The same as above. The students' presentations appeared to be more authentic.
	3 rd	✓✗	The conversations in the "Seinfeld" episodes were produced by American actors, which is more

			authentic than the conversations in the two textbooks.
5. The major focus of teaching is on the meaning, and then on the form.	1 st	✓	Dr. N asked students to listen to main ideas first; vocabulary and grammar were taught after they listened to the passages once.
	2 nd	✓	The same as above.
	3 rd	✓	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	✗	The students were not given opportunities to reflect on what they had learned in class.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.

5.4 Case 1: Questionnaire Survey on the Mid-term Exam

The students took the mid-term exam in the same classroom in which they had had the English listening lesson on the 6th of November 2006. The test comprised 40 questions, consisting of multiple-choice and cloze tests (Appendix D.4). In this mid-term exam, the usage of the cloze test might have saved time for the teacher as regards constructing and scoring, but it was not a test which assessed students' oral communication skills (McNamara, 2000). Buck (2001) also worries that test-takers may fill out the missing words by "reading" rather than "listening", so they would not concentrate on listening. In addition, designing good multiple-choice questions is tough in terms of creating appropriate distractors, piloting the procedure, and there is a high possibility of guessing, although the scoring is more objective owing to a restriction of possible answers. In this mid-term exam, Dr. N used the multiple-choice questions that had already been designed and (presumably) piloted by the textbook publisher, and she assumed that she did not need to re-test those questions (see Section 5.5). However, the textbook publisher did not specify whether the test questions in the teacher's manual were designed and piloted on students similar to those in the present study. This lead to a problem of whether the published test questions in a

textbook could really test students' listening skills in class. In addition, the mid-term exam was paper-based; speaking skills were not directly tested.

In the mid-term exam, Dr. N played the CD twice, so that the students could listen to the questions twice. However, more than one-third of the students asked the teacher to play the CD for a third time. The exam lasted approximately one hour. After the exam and after Dr. N had left the classroom, the questionnaire was group-administered to the students, who completed it in the classroom in which they had just taken the exam. Twenty-eight students were female, and only three were male; there were no missing responses (Table 5.2). Though the questionnaire had been piloted before, I still acted as an observer in case there were any ambiguous question wordings or instructions. However, in the event, the students did not have any questions or report any problems with filling out the questionnaire, and it took the group approximately twenty minutes to finish it, as expected.

Table 5.2 Case 1: Mid-term Survey – Gender of the Respondents (Q3)

Gender	Frequency	Percent
Male	3	9.7%
Female	28	90.3%
Total	31	100%

5.4.1 Case 1: The Teaching Materials and the Course before the Mid-term Exam

In Question 4, only two students felt that the contents of the in-class teaching materials were “frequently” difficult (Table 5.3). In short, there were 27 students who considered the contents difficult to some extent, of whom 24 claimed that the speech rate in the listening materials made them hard to understand (Table 5.4). The use of colloquial language, the accent and the text type in the in-class materials were also perceived as hard by more than ten students. Text type in the

listening extracts was regarded as a problem for ten students; all ten said that the most difficult text type to listen to was the academic lecture in was *Mosaic 1*.

TABLE 5.3 Case 1: Mid-term Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	4	12.9%
Rarely	13	41.9%
Sometimes	12	38.7%
Frequently	2	6.5%
Always	0	0%
Total	31	100%

TABLE 5.4 Case 1: Mid-term Survey – Frequency of Reasons for Material Difficulty (Q4)

Factors	Frequency
Topical content	1
Vocabulary	9
Accent	12
Speech rate	24
The use of colloquial language	17
Sound quality	3
Text type	10

In Question 5, all except two students agreed that the course contents so far were relevant to their listening needs. Two thirds of them stated that the topics in *Impact Listening 3* were easy to understand, because the topics were related to situations in everyday life. The others said that they did not have particular listening needs in relation to the course contents; they simply wanted to improve their listening ability in general. On the other hand, two students thought the course contents were irrelevant to their listening needs; Student 6 expected that the course would be focused on General English Proficiency Tests¹ (GEPT), TOEFL or TOEIC, while Student 11 considered the course was too easy and did not help her improve her English listening ability.

5.4.2 Case 1: Students' General Preferences about the English Listening Classes

Nineteen students (61.3%) believed that they performed better in pen and paper exams of English listening than in oral exams (Table 5.5; Question 6). Approximately half of them said that pen and paper exams allowed them more time to think about the answers; this would stop them losing marks. The others were afraid that they would “lose face” if they gave wrong answers in front of other students, so they preferred written exams. However, more than one third

preferred oral exams because they believed that they could train their listening and speaking abilities at the same time. In addition, 25 students (80.6%) thought that they understood better when listening to “conversations/dialogues” than monologues (Table 5.6; Question 7). One third of the students (25.8%) claimed that the topics in conversational type speech tended to be interesting. The other half of the students (29%) said that they could guess the information from the second speaker if they missed the information from the first speaker in a conversation.

TABLE 5.5 Case 1: Mid-term Survey – Preferred Mode of Answering (Q6)

Mode	Frequency	Percent
Writing	19	61.3%
Speaking	12	38.7%
Total	31	100%

TABLE 5.6 Case 1: Mid-term Survey – Preferred Type of Speech (Q7)

Types of Speech	Frequency	Percent
Monologue	6	19.4%
Conversations	25	80.6%
Total	31	100%

Twenty-two students (71%) preferred “multiple-choice” questions when taking listening tests (Table 5.7; Question 8). One third said that it was easier and faster to answer this type of question, because they did not need to write down any words. The other half said that it was easy to determine the correct answer based on pre-selected answers in multiple-choice questions. However, the type of comprehension question which nearly two thirds of the students preferred least was “short-answer” questions (Question 9), because choosing pre-selected answers in the multiple-choice question decreased the probability that they would misspell vocabulary and lose marks. The other a third also reported that short-answer questions usually accounted for a high percentage of the total marks; if they missed the information conveyed in the discourses, they could easily lose marks.

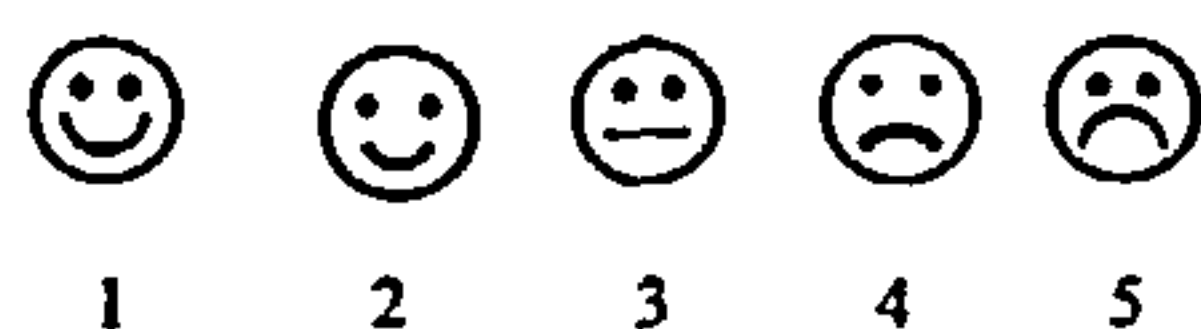
Table 5.7 Case 1: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class

Types of Questions	Preferred Question (Q8)		Least Preferred Question (Q9)	
	Frequency	Percent	Frequency	Percent
Short-answer	1	9.7%	20	64.5%
True/false	4	12.9%	2	6.5%
Multiple-choice	22	71%	2	6.5%
Cloze	3	9.7%	0	0%
Dictation	1	3.2%	7	22.6%
Total	31	100%	31	100%

5.4.3 Case 1: Questions about the Mid-term Exam

In the mid-term exam, nearly half of the students (48.4%) were “(very) satisfied”² with their performance (Table 5.8). It appears that nearly half of the students were happy with the results of their performance. However, the other twelve students (38.7%) were neither satisfied nor dissatisfied with their performance. Only a small number were “(very) dissatisfied”.

TABLE 5.8 Case 1: Mid-term Survey – Performance in the Mid-term Exam (Q10)



Satisfaction with the Mid-term Exam	Frequency	Percent
Face 1 – Very satisfied	2	6.5%
Face 2 – Satisfied	13	41.9%
Face 3 – Neither Satisfied nor dissatisfied	12	38.7%
Face 4 – Dissatisfied	3	9.7%
Face 5 – Very dissatisfied	1	3.2%
Total	31	100%

Just two students (6.5%) “agreed” that the topics of the test tasks were difficult, and that the test tasks were harder than those used in class (Table 5.9; Question 11-a and 11-c); the test topics thus appeared to be comprehensible for the majority of the students. In addition, more than half of the students (51.6%) “(strongly) agreed” that the topics of the test tasks were representative of what had been taught in class (Table 5.9; Question 11-b) which assumed that the difficulty of the tasks in the exam and the textbook were similar.

Table 5.9 Case 1: Mid-term Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q11-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q11-b)		Test Tasks Were Harder than Those in Class (Q11-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	8	25.8%	1	3.2%	7	22.6%
D	8	25.8%	7	22.6%	14	45.2%
Neither A nor D	13	41.9%	7	22.6%	8	25.8%
A	2	6.5%	12	38.7%	2	6.5%
SA	0	0%	4	12.9%	0	0%
Total	31	100%	31	100%	31	100%

Fewer than five students (16.2%) “(strongly) agreed” that the accent in the listening passages was too strong to understand (Table 5.10; Question 11-d). Fewer than three “agreed” that the vocabulary was too hard, and that the sentences were too complicated to understand (Table 5.10; Question 11-e and 11-f). Thus, the majority of the students did not regard the accent, vocabulary, or the sentences as problems while listening to the extracts in the mid-term exam.

Table 5.10 Case 1: Mid-term Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q11-d)		Vocabulary was too Difficult to Understand (Q11-e)		Sentences were too Complicated to Understand (Q11-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	6	19.4%	4	12.9%	7	22.6%
D	12	38.7%	18	58.1%	16	51.6%
Neither A nor D	8	25.8%	7	22.6%	7	22.6%
A	3	9.7%	2	6.5%	1	3.2%
SA	2	6.5%	0	0%	0	0%
Total	31	100%	31	100%	31	100%

Nevertheless, more than half of the students (51.6%) “(strongly) agreed” that it was hard to understand what speakers said because they spoke fast (Table 5.11; Question 11-g). In this exam, both monologue and conversational speech were included. Understanding conversational speech appeared to be easier than comprehending monologues, since only two students (6.5%) “agreed” that the conversations were hard to understand, while more than a quarter (25.8%)

“(strongly) agreed” that it was hard to understand monologue (Table 5.11; Question 11-h and 11-i).

Table 5.11 Case 1: Mid-term Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q11-g)		Monologue Speech Hard to Understand (Q11-h)		Conversations Hard to Understand (Q11-i)	
	Frequency	%	Frequency	%	Frequency	%
SD	1	3.2%	1	3.2%	5	16.1%
D	4	12.9%	5	16.1%	13	41.9%
Neither A nor D	10	32.3%	17	54.8%	11	35.5%
A	13	41.9%	7	22.6%	2	6.5%
SA	3	9.7%	1	3.2%	0	0%
Total	31	100%	31	100%	31	100%

In Question 12, eighteen students (58.1%) reported that colloquial language was used in the mid-term exam, but only eight out of the eighteen students said that they found it hard to understand. These corresponded to the students’ responses in Table 5.10, where more than 70% of students regarded the vocabulary used in the exam as not difficult to understand. Again, this evidence supports the fact that the language used in the mid-term exam was generally highly comprehensible.

As regards the test characteristics in Question 13, eighteen students (58.1%) reported that there were no problems with the quality of the recording equipment (Appendix D.6). However, another six (19.4%) claimed that the background noise outside the testing environment was too loud. Testing time was sufficient for fourteen students (45.2%), but twelve (38.7%) did feel that time was too limited to answer all the questions properly. Another five students said that the time between questions was too short, and that they did not have sufficient time to digest the questions. In the situation of the test instructions, nineteen students (61.3%) reported that the instructions were clear, but another ten (32.3%) complained that the instructions for the individual test sections were not clear.

More than two thirds (67.7%) said that the length of the listening texts was similar to what they had heard in class. Twenty-one students (67.7%) found that the most difficult questions in the mid-term exam were the multiple-choice questions (Table 5.12; Question 14). Cloze questions, on the other hand, were considered the easiest type of question (Table 5.12; Question 15). The exam did not contain short-answer, true/false, and dictation questions.

Table 5.12 Case 1: Mid-term Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam

Types of Question	Easiest Question (Q14)		Most Difficult Question (Q15)	
	Frequency	Percent	Frequency	Percent
Short-answer	0	0%	0	0%
True/false	0	0%	0	0%
Multiple-choice	13	41.9%	21	67.7%
Cloze	18	58.1%	10	32.3%
Dictation	0	0%	0	0%
Total	31	100%	31	100%

5.4.4 Case 1: Students' General Comments on the Mid-term Exam

Twenty seven students (87.1%) out of the 31 reported that they had learned what they expected to learn thus far from the classes (Question 16-a). The remaining four students said that what the teacher taught was too easy for them, and they hoped that Dr. N would choose more advanced English listening materials. All but one student were satisfied with the way the teacher taught in class (Question 16-b); Student 23 pointed out that the teacher spoke unclearly and this made it hard to understand the session. Twenty seven students (87.1%) were satisfied with the assessment method the teacher used (Question 16-c), but the other three students felt that the test items were almost the same as those they did in the textbook, which were not challenging at all. The remaining student said that not all the types of question were covered in the exam.

5.5 Case 1: Mid-term Interview with the Teacher

The face-to-face interview with the teacher was conducted on the 13th of November 2006, one week after the mid-term exam, when the students' marks were known. The interview took place in the teacher's private office in the University and lasted approximately thirty minutes. The interview questions were presented in both English and Chinese; Dr. N chose to answer in her native language – Chinese. The interview was tape recorded with her permission and transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in Appendix D.7. The translation was double-checked by a Chinese speaker who could speak both Chinese and English and a native English speaker.

Table 5.13 Case 1: Mid-term Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Dr. N (N): The first part came from the textbook.</p> <p>Interviewer (I): Were those exercises in the textbook?</p> <p>N: The questions were taken from the teacher's manual.</p> <p>I: Were those more difficult than the ones in the textbook?</p> <p>N: Not really, the conversations in the teacher's manual were the same as those in the textbook, but the questions were different. The remaining questions were from the textbook, but I designed different questions.</p> <p>I: Did the types of question change?</p> <p>N: Basically no, I simply asked different questions, but if students paid attention to the lessons, they should have no problems.</p> <p>I: Did you pilot the test items, I mean run a small test, before you test your students in this exam?</p> <p>N: Did I have to? Maybe the book publisher had tested those questions.</p>

According to Dr. N's statement, the topics of the test tasks were all based on the two textbooks; she simply changed the questions (Table 5.13). Dr. N assumed that the students would have fewer problems in answering the questions. It will be recalled that more than half of the students, too, agreed that the test tasks were not hard and the test was indeed representative of what had been taught in class (Table 5.10).

Table 5.14 Case 1: Mid-term Survey – Interview with the Teacher (b)

<p>Question 2. What are the cut-off scores for the mid-term and final examination tests? What percentage does each of the two tests count for in the total final score of the course?</p>
<p>N: “60” is the passing score. Mid-term scores count for 25% of the total marks; final scores count for another 25%.</p>
<p>I: During my observations, you did not give the students quizzes directly in class but asked them to complete quizzes afterwards on the Internet. So why did you test your students on the Internet?</p>
<p>N: The quizzes were from an adult English learning website in California³. There were many topics which were similar to those we learned in class. They had to listen and complete all lessons in the category of “School” before the mid-term exam, and the categories of “Going Places” after the mid-term. Students did different types of question after they listened to a topic. After finishing the questions, there were be scores on the screen and they had to send their scores directly to me via the Internet. It saved more class time for the teaching.</p>
<p>I: Did you find any problems with using this kind of on-line quiz?</p>
<p>N: There have been no technical problems with using the website at all so far.</p>
<p>I: Was it possible for students to cheat because they did not take the quiz in the classroom?</p>
<p>N: Well....of course I could not prevent them from cheating....but if they cheated, their listening ability would not improve. So they had to be responsible for themselves.</p>

In Question 2, Dr. N also took participation and quiz scores into consideration for the final total mark, in addition to the two marks from the mid-term and final exams (Table 5.14). The students were asked to take quizzes by themselves after the class in order to save more class time for teaching. However, the danger of asking students to test themselves privately after the class could have led to cheating and which may have threatened the reliability and validity of the quizzes. In other words, the quiz scores could be meaningless from the point of measuring improvement in students’ listening ability. Moreover, I found that the students had many opportunities to test themselves again and again on that website before submission. That is to say, if students were dissatisfied with the results of a quiz, they could retest themselves before submitting the results to the teacher. It was highly likely that the students with lower English listening ability received the same or even higher scores via repeated tests; this was unfair since the quiz scores accounted for part of the total mark.

Table 5.15 Case 1: Mid-term Survey – Interview with the Teacher (c)

<p>Question 3. What did you expect the students to have learned from your class?</p>
<p>N: I use two textbooks. One is for “general” listening, which means conversations we use in everyday life. The other is for “academic” listening. I usually teach everyday English conversations for the first hour, and academic listening for the second.</p>

I: Did the academic listening take the form of a “lecture”?
N: Yes, the lecture in the textbook was quite long.
I: So did students find listening to it problematic?
N: Of course, they found it difficult to understand.
I: Did you think the textbook you used matched students’ English level?
N: I think the “general” book is OK, but the “academic” one is too hard for them.

In Question 3, Dr. N expected that her students would learn not only the English used in everyday life but also English for academic purposes (Table 5.15). She was aware of the situation that her students understood everyday English better than academic English; however, she still hoped that the students would try to listen to academic lectures. According to the classroom observations before the mid-term (See 5.3.1 and 5.3.3), it seemed that the students did not interact as actively when they listened to *Mosaic 1* as they did when they listened to *Impact Listening 3*. Ten students (Table 5.4) reported that the topics of the academic lectures were sometimes hard to understand.

Table 5.16 Case 1: Mid-term Survey – Interview with the Teacher (d)

Question 4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and students’ learning outcomes? Did you think you have achieved them?
N: I did ask them to review and practice the units I taught every week. The purpose of the test was to see whether they had practiced the listening contents after the class and whether they understood what I taught in class.
I: Did all students have a copy of the listening CD?
N: Yes, a CD was included in the “general” textbook; they have a copy of the CD of “academic” listening.
I: What did you think of the students’ performances in the mid-term test? Did you think you achieved your test objective?
N: I think the scores are a little lower than I expected. I think they should have performed better because the test questions were all from the textbooks. It was possible that they didn’t review and practice after the class.
I: I noticed that you did not test students’ speaking ability in the mid-term exam. Could you let me know the reason?
N: I don’t think that testing speaking was necessary for this course, since the title of the course was *English Listening*. But I gave students opportunities to speak English in class, I mean, the presentations, so I think that it was enough for this course.
I: So...will you use any speaking test for the final exam?
N: No, I won’t.

The test objective which Dr. N wanted to achieve was to test whether her students had learned what had been taught in class (Table 5.16; Question 4). The test objective was somewhat vague because the teacher did not specify what was included in the mid-term test based on her teaching. Nearly half of the students

were “(very) satisfied” with their performance in the mid-term exam (Table 5.8), unlike the teacher who was not very satisfied with the test results (the mean of the mid-term scores was 74.42). She thought that the students would practice the English outside the classroom, but they did not review it enough. Four students (12.9%) did not pass the mid-term exam; the teacher expected her students would perform better, but they failed to achieve her expectation. She decided not to test students’ speaking skills, because she did not think that it was necessary to test speaking on a listening course.

Table 5.17 Case 1: Mid-term Survey – Interview with the Teacher (e)

<p>Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?</p> <p>N: The test content was based on the textbook and the curriculum progress.</p> <p>I: If the students told you that the in-class teaching materials were too difficult, would you make the test easier?</p> <p>N: So far I haven’t had any reports from the students regarding the difficulty of the materials. But as they felt that “academic” listening was more difficult, I designed fewer questions regarding “academic” listening. Those were in Part Three.</p> <p>I: Did they score lower in this part?</p> <p>N: Not really. On the contrary, they scored lower in the first part of the test that was from the “general” listening textbook. In fact, I simply slightly changed the types of questions.</p> <p>I: Did you think that they could cope in this listening class with their vocabulary?</p> <p>N: Erm...I am afraid not, because there were some colloquial language usages in the “general” listening textbook. If they didn’t use colloquial language frequently, they wouldn’t understand it. But I didn’t focus on testing the vocabulary; I tested their understanding of the listening passages.</p> <p>I: Did this mean that they didn’t find the vocabulary difficult?</p> <p>N: Yes.</p>
--

Since the students felt that it was harder to understand the “academic” listening part than the “general” one, Dr. N tested fewer items regarding academic topics in order to decrease the difficulty level (Table 5.17; Question 5). Surprisingly, the students did not perform worse on the items regarding academic topics but on topics about everyday life. One possibility was that the students thought they had completely learned what had been taught in *Impact Listening 3*, but in fact they had not. It was also possible that the students had not prepared well enough for the exam. In addition, although the majority of the students also reported that the vocabulary was not a problem for them in the mid-term exam

(Table 5.10), Dr. N still felt that the students' vocabulary was too limited to cope with the colloquial language.

Table 5.18 Case 1: Mid-term Survey – Interview with the Teacher (f)

Question 6. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?
N: I personally felt that they would score higher on cloze questions because they could write down what they heard. However it turned out not necessarily so; they did not score higher in this part. Only more than half of the students got full marks in this type of question, but others lost marks for unknown reasons.

In Question 6, Dr. N thought that it was easier to answer cloze questions than multiple-choice and true/false questions (Table 5.18). However, she did not know why the students lost marks in the cloze test. One possibility is that that some students had learned the words but failed to recognise them while listening to a stream of speech, so they lost marks.

Table 5.19 Case 1: Mid-term Survey – Interview with the Teacher (g)

Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)
N: Maybe I would give them more time to listen to the “academic” listening textbook in class, but they seemed to fall asleep when listening to it. On the other hand, they found the “general” textbook more interesting, so I wouldn't amend difficulties in this part.
I: How about the final exam?
N: I think I wouldn't adjust the difficulties of the test questions; difficulties would be similar to those in the mid-term exam. In fact, this mid-term did distinguish those who studied harder from those who didn't. The harder they studied, the higher scores they received. But there were still some students who didn't study.
I: What did you think the level of your students in this class?
N: I think in general there were not many differences between them, but there were some students whose English was poorer and they didn't study hard.
I: Did students feel that the teaching materials were too easy for them?
N: I didn't hear that they felt the materials were too easy.

From the interview with Dr. N, she appeared to have felt that the reason why the students did not perform well did not lie in the difficulties of the test tasks but in the students themselves (Table 5.19; Question 7). That explains why Dr. N did not wish to change the difficulties of the test tasks in the final exam.

To summarise, a number of the features of task-based instruction, such as specific targets of teaching and the focus of teaching being more on the meaning

than form, were observed regularly before the mid-term exam. However, the use of authentic texts was observed in just two lessons. Opportunities for students to speak English and to discuss in groups were limited. In this mid-term exam, the teacher was aware of the students' in-class learning situation, for example, whether they understood the listening extracts in the two textbooks or not. The teacher also tested her students on the basis of the two textbooks. The majority of the students, however, were sensitive to different types of question such that they believed they could perform better on certain types. The teacher claimed, and the majority of the students agreed, that the listening extracts were taken from the two textbooks. However, since Dr. N said that she had changed the questions, it was possible that the majority of students did not completely understand the listening extracts and then lost marks.

5.6 Case 1: Classroom Observation after the Mid-term Exam

Dr. N continued to use the two textbooks in class after the mid-term exam. Another three observations were carried out from 10: 20am and to 12: 10pm on the 27th of November, the 11th and 18th of December 2006.

5.6.1 Case 1: Fourth Observation

The first observation after the mid-term exam was carried out on the 27th of November 2006. Unit 8 of *Impact Listening 3* and students' presentations were the main activities in this lesson. She asked the students to do the vocabulary exercises and then played the CD once. She asked the students to listen for the main idea and then she played the CD twice and asked the students to do the exercises in the textbook. She checked the answers with the students by playing the CD yet again and asked the students to answer each question. The students

appeared eager to answer each of the questions. They appeared to understand both the listening passages and the tasks. In other words, the students practiced their oral English ability by answering the questions in the textbook.

In the second half of the class, two groups of students again presented a radio talk show in front of their peers. The first group of four students discussed a TV commercial they had seen recently; the content included how the most famous female supermodel in Taiwan was chosen to front the commercial for an airline in Taiwan, and how the supermodel felt about shooting the commercial and about the airline. One student acted as the supermodel, another as the talk show host, and the other two as the journalists. The presentation lasted for 15 to 20 minutes. After the presentation, the presenters asked the audience four questions related to their show. As with the presentations observed before the mid-term exam, the audience could receive extra points by answering the questions. The other group of students discussed a UK movie which was adapted from a famous series of books. The students discussed the storyline in the movie as well as the previous three movies in a series. This presentation also lasted around 15 to 20 minutes. The presenters again asked the audience three questions regarding the discussion. Nearly one quarter of the students raised their hands to answer questions from two presentations. However, it was hard to decide whether those who did not raise their hands had understood the presentations or not. Dr. N was in charge of the opening and the ending of the presentations; as before, she did not ask questions during the presentation.

5.6.2 Case 1: Fifth Observation

The second observation after the mid-term exam took place on the 11th of December 2006. The activities the students did in class were to learn Unit 10

(*Impact Listening 3*) and to watch the American situation comedy “Seinfeld” (Episode 5 in Series 1). The procedure for teaching Unit 10 was similar to that used to teach the other units (see Sections 5.3.1, 5.3.2, or 5.6.1). In the second half of the class, the students watched “Seinfeld” (episode 5) with English subtitles. Before playing the episode, Dr. N gave the students four questions regarding the content on a piece of paper, and asked them to answer the questions after watching the comedy. The teacher briefly introduced the episode and then watched it with the students. At this point, the students submitted their answers individually to Dr. N and the class ended. I asked the teacher why she asked questions regarding the comedy this time (after the mid-term exam), but did not do so before the exam. She said that she simply wanted the students to get familiar with aspects of the English, such as the speed and the colloquial language that Americans used. However, by this point she wanted to know how far her students could understand the plot line, so she asked a few questions regarding the episode as a whole.

5.6.3 Case 1: Final Observation

The last observation was carried out on the 18th of December 2006. At the beginning of the class, Dr. N told the students that they would finish Unit 12 of *Impact Listening 3* and Chapter 6 of *Mosaic 1*. Then, she reminded the students to e-mail the results of the on-line quizzes to her before Friday. Unsurprisingly, the procedure for teaching Unit 12 of *Impact Listening 3* was similar to that used on the previous occasions. In the second half of the class, Dr. N continued to teach Chapter 6 of *Mosaic 1*, which she had not finished the previous week. She asked the students to take the transcript of the listening extracts she had given them the previous week, and asked them to read it while listening to the lectures.

However, I noticed that Dr. N did not give her students transcripts of the *Mosaic 1* passages before the mid-term exam. After playing the CD, the students were asked to do the exercises in the textbook. I saw that almost half of the students answered the questions by reading and copying from the transcripts. It was thus hard to know whether the students understood the lectures in Chapter 6 of *Mosaic 1* or not. Dr. N, however, did not respond to their behaviour immediately. The reason why she gave the students transcripts of listening extracts in *Mosaic 1* after the mid-term but not before it emerged in the interview after the final exam (See Section 6.8). Table 5.20 shows the extent of using task-based instruction after the mid-term exam.

Table 5.20 Case 1: Occurrence of Task-based Instruction Features after the Mid-term Exam

TBI Characteristics	Obs. No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	4 th	x	The activities the students did in class were to answer the questions in the textbooks and the students' presentations individually. No problem-solving tasks were found in class.
	5 th	x	The same as above. The students were asked to write down their answers to the questions based on the "Seinfeld".
	6 th	x	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	4 th	✓ x	The students practiced their English only when being asked questions in the textbook or by the presenters.
	5 th	✓ x	The students only spoke English while answering questions in Unit 10. They wrote down the answers to the questions about the comedy.
	6 th	✓ x	The students only spoke English while answering questions in Unit 12 and Chapter 6.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	4 th	x	The students worked individually in class.
	5 th	x	The same as above.
	6 th	x	The same as above.
4. Authentic texts which reflect a real-life situation are used.	4 th	✓ x	The listening extracts in the Impact Listening 3 were not considered authentic (see Appendix D.3; Section 5.3.1). The presentation,

			however, demanded students' oral ability to speak English without scripts.
	5 th	✓ x	The conversations in the textbook were not as authentic as those in the real-life situations; however, the conversations in "Seinfeld" were produced by American actors or scriptwriters which Dr. N considered as authentic.
	6 th	x	The listening contents of the two textbooks were scripted and produced by the publishers, which the teacher did not consider authentic.
5. The major focus of teaching is on the meaning, and then on the form.	4 th	✓	Dr. N asked the students to listen for main ideas first, and then for detailed information.
	5 th	✓	The same as above.
	6 th	✓	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	4 th	x	The students were not given opportunities to reflect on what they had learned in class.
	5 th	x	The same as above.
	6 th	x	The same as above.

To summarise, the interaction between the teacher and the students did not change obviously before or after the mid-term exam. Dr. N still encouraged the students to answer the questions in the textbooks by giving them extra marks. As for the teaching of the two textbooks, the procedure for teaching the units of "general listening" (*Impact Listening 3*) was similar before and after the mid-term exam. The teaching of academic listening, on the other hand, changed. The teacher gave transcripts of listening passages to the students before playing the CD, which might cause the students to focus on reading rather than on listening to the passages. In addition, Dr. N paid more attention to whether the students understood the comedy by asking them questions after the mid-term exam.

5.7 Case 1: Questionnaire Survey on the Final Exam

The students took the final exam on the 10th of January 2007. It consisted of 39 questions, involving multiple-choice, true/false, and cloze tests (Appendix D.5).

Dichotomous test items, such as true/false questions, are subject to a higher possibility of guessing than multiple-choice items. Also, true/false questions test nothing about oral communication skills. Again in the final exam, the test items were not piloted because the teacher assumed the items she used had been tested by the textbook publisher (see Section 5.8). In Case 1, no communicative testing approaches were used in either the mid-term or the final exam, instead, the discrete-point and integrative testing approaches were used.

In this final exam, Dr. N played the CD twice and none of the students asked Dr. N to replay it. The exam lasted approximately 45 minutes. After the exam and Dr. N had left the classroom, and the final questionnaire was group-administered to the students. The questionnaire had been piloted before, but as with the mid-term course questionnaire, I stayed in the classroom in case that there were there were ambiguous instructions or wordings. Again, no questions were reported, and it took the group around 20 minutes to finish.

5.7.1 Case 1: The Teaching Materials and the Course after the Mid-term Exam

Only one student (3.2%) felt that the in-class listening materials after the mid-term were “frequently” difficult (Table 5.21). Another thirteen students thought the materials were “sometimes” difficult. The perceived difficulty of the in-class materials remained consistent throughout the term, because the frequencies in Tables 5.3 and 5.21 were similar. Table 5.22 showed that speech rate was still a main reason why 22 students found listening to the passages hard.

TABLE 5.21 Case 1: Final Survey – Perceived Difficulty of In-class Materials

(Q4)		
	Frequency	Percent
Never	5	16.1%
Rarely	12	38.7%
Sometimes	13	41.9%
Frequently	1	3.2%
Always	0	0%
Total	31	100%

TABLE 5.22 Case 1: Final Survey – Frequency of Reasons for Material Difficulty (Q4)

Factors	Frequency
Topical content	2
Vocabulary	7
Accent	6
Speech rate	22
The use of colloquial language	6
Sound quality	1
Text type	2

5.7.2 Case 1: Students' general preferences about English listening classes

The number of students who preferred written and oral exams remained the same after the mid-term exam (Tables 5.5 and 5.23; Question 5). Similarly, the number of students who preferred monologue and conversations in the final exam was nearly the same as it was for the mid-term exam (Tables 5.6 and 5.24; Question 6). It may thus be concluded that the results of the mid-term exam did not change the students' preferences for the mode of answering or the type of speech involved.

TABLE 5.23 Case 1: Final Survey – Preferred Mode of Answering (Q5)

Mode	Frequency	Percent
Writing	19	61.3%
Speaking	12	38.7%
Total	31	100%

TABLE 5.24 Case 1: Final Survey – Preferred Type of Speech (Q6)

Types of Speech	Frequency	Percent
Monologue	7	22.6%
Conversations	24	77.4%
Total	31	100%

Compared with the figures concerning students' preferred mode of answering before mid-term exam, the proportions of the students who preferring multiple-choice questions decreased from 71% to 51.6% (Tables 5.7 and 5.25). However, the proportion of students who preferred true/false questions increased markedly from 12.9% to 32.3%. Since the exercises in *Impact Listening 3* were similar across the units, as were the exercises in the *Mosaic 1*, it is unlikely that the types of question in the textbooks influenced the students' preferences after

the mid-term. It is more likely that the results of the mid-term exam – where the students found it hard to answer the multiple-choice questions – were at least partly responsible. Multiple-choice questions remained the type of question students preferred most (Tables 5.7 and 5.25).

Table 5.25 Case 1: Final Survey – Preferred vs. Least Preferred Type of Comprehension

Types of Questions	Preferred Question (Q7)		Least Preferred Question (Q8)	
	Frequency	Percent	Frequency	Percent
Short-answer	1	9.7%	20	64.5%
True/false	10	32.3%	0	0%
Multiple-choice	16	51.6%	3	9.7%
Cloze	1	3.2%	3	9.7%
Dictation	3	9.7%	5	16%
Total	31	100%	31	100%

5.7.3 Case 1: Questions about the Final Exam

Sixteen students (51.6%) were “(very) satisfied” with their performance in the final exam (Table 5.26; Question 9). In general, the students’ satisfaction with the two exams was similar (Tables 5.8 and 5.26).

TABLE 5.26 Case 1: Final Survey – Performance in the Final Exam (Q9)



Satisfaction with the Mid-term Exam	Frequency	Percent
Face 1 – Very satisfied	5	16.1%
Face 2 – Satisfied	11	35.5%
Face 3 – Neither Satisfied nor dissatisfied	14	45.2%
Face 4 – Dissatisfied	1	3.2%
Face 5 – Very dissatisfied	0	0%
Total	31	100%

In the final exam, only two students “(strongly) agreed” that the topics were difficult (Table 5.27; Question 10-a), and none of the students “agreed” that the test tasks in the final exam were harder than those used in class (Table 5.27; Question 10-c). Tables 5.27 and 5.9 show that the difficulties of the tasks in the two exams were similar, since the frequencies in the two tables are close. Nearly

half of the students (48.4%) “(strongly) agreed” the test tasks were representative of the curriculum taught in class (Table 5.27; Question 10-b). The topics of the test tasks in the two exams were thus considered representative of the curriculum by approximately half of the students (Tables 5.9 and 5.27).

Table 5.27 Case 1: Final Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q10-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q10-b)		Test Tasks Were Harder than Those in Class (Q10-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	8	25.8%	1	3.2%	8	25.8%
D	11	35.5%	4	12.9%	15	48.4%
Neither A nor D	10	32.3%	11	35.5%	8	25.8%
A	2	6.5%	13	41.9%	0	0%
SA	0	0%	2	6.5%	0	0%
Total	31	100%	31	100%	31	100%

Nearly a quarter of the students (22.6%) “(strongly) agreed” that the speakers’ accent in the final exam listening passages was too strong to understand (Table 5.28; Question 10-d). Just three (9.7%) “(strongly) agreed” that the vocabulary was too difficult, and only two “(strongly) agreed” that the sentences were too complicated to understand (Table 5.28; Question 10-e and 10-f). The vocabulary and the sentences in both two exams were thus not considered a problem by the majority of the students (Table 5.28 and 5.10).

Table 5.28 Case 1: Final Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q10-d)		Vocabulary was too Difficult to Understand (Q10-e)		Sentences were too Complicated to Understand (Q10-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	4	12.9%	2	6.5%	7	22.6%
D	12	38.7%	19	61.3%	13	41.9%
Neither A nor D	8	25.8%	7	22.6%	9	29%
A	7	22.6%	3	9.7%	1	3.2%
SA	0	0%	0	0%	1	3.2%
Total	31	100%	31	100%	31	100%

Thirteen students (42%) “(strongly) agreed” that it was hard to understand what speakers said because they spoke fast (Table 5.29; Question 10-g). It

seemed that the fast speed of speaking in the listening passages was still a problem for approximately half of the students in both exams (Tables 5.29 and 5.11). In the final exam, both monologue speech and conversations were included. Understanding monologue speech was considered a problem by nearly a quarter of the students (22.6%) (Table 5.29; Question 10-h). Compared with the monologues, only four (12.9%) “(strongly) agreed” that it was difficult to understand the conversations (Table 5.29; Question 10-i). The students thus appear to have found it relatively easy to understand the conversational types of text in both exams.

Table 5.29 Case 1: Final Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q10-g)		Monologue Speech Hard to Understand (Q10-h)		Conversations Hard to Understand (Q10-i)	
	Frequency	%	Frequency	%	Frequency	%
SD	2	6.5%	5	16.1%	4	12.9%
D	7	22.6%	3	9.7%	8	25.8%
Neither A nor D	9	29%	16	51.6%	15	48.4%
A	11	35.5%	7	22.6%	4	12.9%
SA	2	6.5%	0	0%	0	0%
Total	31	100%	31	100%	31	100%

In Question 11, five students (16.1%) reported that colloquial language was used in the test, but only one of them found it difficult to understand. In Question 12, twenty seven students (87.1%) reported no problems regarding the quality of the tape recording; only two felt that the volume of the tape recorder was too low (Appendix D.6). The problems of the poor sound quality, the background noise outside the testing environment, and the low volume found in the mid-term exam were thus reduced in the final exam. Twenty six students (83.9%) regarded the testing time as sufficient. The percentage of students (16.1%) who reported that testing time was too limited to answer all the questions properly in the final exam also decreased. As for the clarity of the test/task instructions, twenty seven

(87.1%) felt that the instructions were clear. Only four felt that the instructions were not clear and too complicated; these two problems were thus also reduced in the final exam. All students reported that the length of the texts in the final exam was similar to what they had listened to in class. Twenty students (64.5%) considered that cloze questions were the easiest type of question in the final exam (Table 5.30; Question 13). Sixteen (51.6%) found that true/false questions were the most difficult type of question in the final exam (Table 5.30; Question 14).

Table 5.30 Case 1: Final Survey – Easiest vs. Most Difficult Type of Comprehension

Question in the Mid-term Exam

Types of Questions	Easiest Question (Q13)		Most Difficult Question (Q14)	
	Frequency	Percent	Frequency	Percent
Short-answer	0	0%	0	0%
True/false	3	9.7%	16	51.6%
Multiple-choice	8	25.8%	6	19.4%
Cloze	0	0%	3	9.7%
Dictation	20	64.5%	6	19.4%
Total	31	100%	31	100%

5.7.4 Case 1: Students' General Comments on the Final Exam

All but two students claimed that they had learned what they expected to learn after the mid-term exam (Question 15-a). However, the two students concerned did not specify the reasons why they did not learn what they expected to learn. All of the students but one said that they were satisfied with the way the teacher taught after the mid-term exam (Question 15-b); student 29 said that what the teacher taught was too easy for her. As for the assessment methods, all of the students said they were satisfied with the assessment the teacher used after the mid-term exam (Question 15-c). After finishing the course, twenty nine students out of the thirty one felt that their English listening skills had improved (Question 15-d). Student 14 said that her English listening did not improve because she had not worked very hard on the course. The other student, Student 29, said the teaching materials were too easy.

5.8 Case 1: Final Interview with the Teacher

The final interview was conducted after the final exam on the 15th of January 2007. The face-to-face interview also took place in the teacher's private office in the University and lasted approximately fifteen to twenty minutes. In the final interview, Dr. N again chose to answer in Chinese. The interview was tape recorded with her permission, transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in the Appendix D.8. The translation was again double-checked by a Chinese speaker who could speak both Chinese and English and a native English speaker.

Table 5.31 Case 1: Final Survey – Interview with the Teacher (a)

Question 1. Where did the content of the test come from?
Dr. N (N): It was from the teacher's manual, which was similar to the mid-term exam. But the second part was taken from <i>Mosaic 1</i> .
Interviewer (I): Had the students heard the second part from <i>Mosaic 1</i> in previous class?
N: Yes, that's why they didn't make many mistakes in this part.
I: Did you include what you have taught before the mid-term in the final exam?
N: No. I simply tested what I had taught after the mid-term.
I: Did you pilot the test items in the final exam?
N: No. I think the textbook publisher had piloted the questions.

In the final exam, Dr. N also used the questions from the teacher's manual. Only Part II was taken from a listening passage in *Mosaic 1* (Table 5.31; Question 1). This probably explains why twenty students (64.5%) found cloze questions, which were in Part II, to be the easiest type of question in the final exam, since they had already listened to the passage in class (Table 5.30).

Table 5.32 Case 1: Final Survey – Interview with the Teacher (b)

Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them?
N: Because I didn't test what had been taught before the mid-term but only after it, the test content was different from the mid-term exam.
I: What did you think of their performance this time, compared with their mid-term performance?
N: Erm...I personally felt that they performed a little bit better than the mid-term this time, but not particularly well.

The test tasks in the final exam did not include the teaching contents before the mid-term exam, but only those taught after it. Although it appeared that the students performed better in the final exam, Dr. N was still not very satisfied with their results (Table 5.32; Question 2). The mean of the final scores was 77.29, which was slightly higher than the 74.42 for the mid-term exam. She expected that her students would perform better. Nearly all students reported that they were satisfied with the teaching materials, the teacher and the assessment method, but only half of them were “satisfied” with their test results. The question was why nearly all students were satisfied with the teaching materials, the teacher, and the assessment, but half were *not* satisfied with their test results. On the one hand, it was possible that those students did not work hard in class, so they failed to perform well. On the other hand, there could be problems with the test items that no one had discovered.

Table 5.33 Case 1: Final Survey – Interview with the Teacher (c)

<p>Question 3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria?</p> <p>N: In fact I didn't particularly adjust the difficulty of the test content in this final exam. Because they didn't perform very badly in the mid-term, I felt it would be fine if the difficulty of the test content in the final exam was similar to that in the mid-term one.</p> <p>I: I saw you give students the scripts before listening to the “academic” textbook, so why did you give the scripts to them only after the mid-term but not before it?</p> <p>N: Because they appeared unable to understand the “academic” listening completely before the mid-term and I was afraid that academic listening was too difficult for them, I didn't teach them too much. This time I thought that giving them scripts might help them understand the listening content.</p> <p>I: Did you notice that the students copied the answers from the transcript?</p> <p>N: Well...I did notice their behaviour. I told them that they needed to listen carefully for the first time and then read the transcript. I believed that transcripts could help them understand the lectures better, but they had to bear the consequences after copying answers from the transcripts...I mean if they scored lower in the final exam.</p>
--

Although Dr. N was not very satisfied with the test results in the mid-term exam, she still did not adjust the difficulty level of the test in the final exam (Table 5.33; Question 3). It appeared that she still expected that her students would perform better in the final exam than in the mid-term one. However, she

did modify the way she taught *Mosaic 1* in class after the mid-term exam. Even though the students copied the answers from the transcripts, she still believed that transcripts could help students understand the listening content. The cloze questions in Part II were taken from the textbook to which the students had listened in class. It was thus difficult to establish whether reading transcripts really helped understanding since the text had been listened to before.

Table 5.34 Case 1: Final Survey – Interview with the Teacher (d)

<p>Question 4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p>
<p>N: The types of question were similar to those in the mid-term, because they were common types of question we used in the listening class.</p>

The teacher did not choose the types of comprehension question specifically for the final exam (Table 5.34; Question 4). The types of question in both exams were similar to the exercises they did in class. It was assumed that the students would be very familiar with the types of question involved.

5.9 Summary

In Case 1, a few features of task-based instruction were observed in Dr. N's course. The topics of the teaching materials (i.e. *Impact Listening 3* and *Mosaic 1*) that Dr. N chose were related to everyday life; and she also presented authentic listening input (i.e. the American situation comedy) in class in order to involve the students in relatively authentic target language situations. The talk-show presentations also provided the students with opportunities to practice the target language. However, the interactions between teacher and students tended to restrict the pattern, in the sense that the students only interacted with the teacher by answering questions when asked, and were probably encouraged to do so primarily by the chance to earn extra marks. In the student presentations, the

students, as presenters, tended to enjoy the work, or at least they looked as though they were enjoying entertaining the audience. It seems reasonable to conclude that the students felt more comfortable while interacting with other students in a student presentation than when interacting purely with the teacher. Thus, the answer to Question 1 is that only a few features of task-based instructions were discovered in Dr. N's class: in student presentations, watching the American situation comedy, and focus on the meaning of the language before considering the form.

According to the interview data, the teacher claimed that she tested what she had taught in class in both exams, but she did not specify what she wanted to test. Also, the teacher claimed that the test contents of the two exams were adapted from the teacher's manuals of the two textbooks. Although the majority of the students were familiar with the types of comprehension question used in the two exams (e.g. multiple-choice and cloze test), which they had repeatedly encountered in the two textbooks, test objectives were not clearly established. The answer to Question 2 is that no test objectives were concretely specified for either exam. As for the testing approach, the two exams were paper-based one-way listening questions; speaking tests were not included. Thus, the answer to Question 3 is that communicative testing approaches were not used in both exams.

In both exams, the speed of the spoken texts appeared to be the main problem for nearly half of the students to understand the listening passages. Other linguistic features, however, such as accent, vocabulary, complexity of sentences, conversational type of texts, and colloquial language were not regarded as problems by the majority of the students. The answer to Question 4 is that speech rate was perceived as the major problem for the students while taking the listening exams in Case 1.

The results of the mid-term exam impacted on the teaching to a certain degree, in the sense that Dr. N changed the way she taught academic listening by giving them listening scripts after the mid-term exam, in order to make sure her students understood the listening contents. Nevertheless, this positive intention resulted in negative effects, as some students copied answers from the scripts rather than listening to the passages, and in the final exam where the cloze items were taken from the textbook, which the students had listened to before. In addition, it was also hard to see whether the effect of on-line quizzes had a positive impact on the students' learning and on their performance in the exams, as the students were not tested under the same conditions (i.e. time, place, or environment). It was thus very difficult to establish validity or reliability, or to justify any specific interpretation of the test scores. The quiz scores in particular were likely to be meaningless in relation to the teaching goals. The answer to Question 5 is thus that both positive and negative washback was discovered in Dr. N's group. In addition, since communicative language testing was not used in this case, it was very hard to see if direct tests had positive washback on influencing, changing, or encouraging teaching and learning.

Notes to Chapter Five

¹ The General English Proficiency Test (GEPT) is a language test designed in Taiwan. It categorises test-takers' language ability into five levels: "elementary", "intermediate", "high-intermediate", "advanced", and "superior". This test is intended for all language learners (not a restricted group) and the four language skills are tested at each level. Accordingly, the target group of GEPT includes hundreds of academic institutions and business organisations.

² Because the number of the students was small, it is more useful to aggregate "very satisfied" and "satisfied" as "(very) satisfied". The same logic also applies to the portmanteau expression "(strongly) agreed" in the following analyses.

³ The website is <http://www.cdlponline.org>. The students needed to listen to a topic and then to do the activities on the following web pages. The activities included listening and spelling out the vocabulary, matching words, testing grammar and vocabulary (multiple-choice questions), and asking opinions (short-answer questions).

Chapter Six

Data Analysis – Case 2

6.1 Introduction

In Chapter Five, the analyses of Case 1 were discussed. In this chapter, 33 students in the second group, taking the same English Listening Course, in the same year of study, and from the same department as Case 1 participated in the study. The 33 students were taught by another teacher (Miss T), and the in-class materials used were very different from the materials used by Dr. N in the previous chapter. The background of the course, teacher, and in-class materials are introduced in Section 6.2, and the in-class situation before the mid-term exam, including the teaching approaches and the interaction between the teacher and the students, is summarised in Section 6.3. The results of the mid-term questionnaire are analysed in Section 6.4, and the interview with the teacher regarding the mid-term exam is discussed in Section 6.5. The descriptions of the in-class situation after the mid-term exam are presented in Section 6.6, while Section 6.7 looks at the students' opinions about the final exam. Finally, the interview with the teacher regarding the final exam is discussed.

6.2 Case 2: Background of the Listening Course

There were 33 students and a teacher in the Case 2 listening class, and the course was taught in English. The teacher (Miss T) had obtained a Master's degree in TESOL at a university in Australia. The students in her class were second-year undergraduate students who were majoring in English Language. Miss T said that she was used to teaching English to preschool children or primary school pupils,

after teaching them for more than seven years; this was, however, her first time teaching students at university level. She said that,

children are more eager to express themselves than adult learners, and children often participate actively in class. While facing adult university students, I found it difficult to encourage them to immerse themselves in the communicative teaching environment.

According to Miss T's syllabus, the target of teaching the course was to

focus on intermediate level of listening comprehension in the target language. Students will develop an in-depth understanding of the language they use and of applications of this understanding to classroom discussion. After completing this course, students will improve their listening comprehension, have an understanding of listening models of GEPT, and be able to pass the intermediate level of listening sections of GEPT.

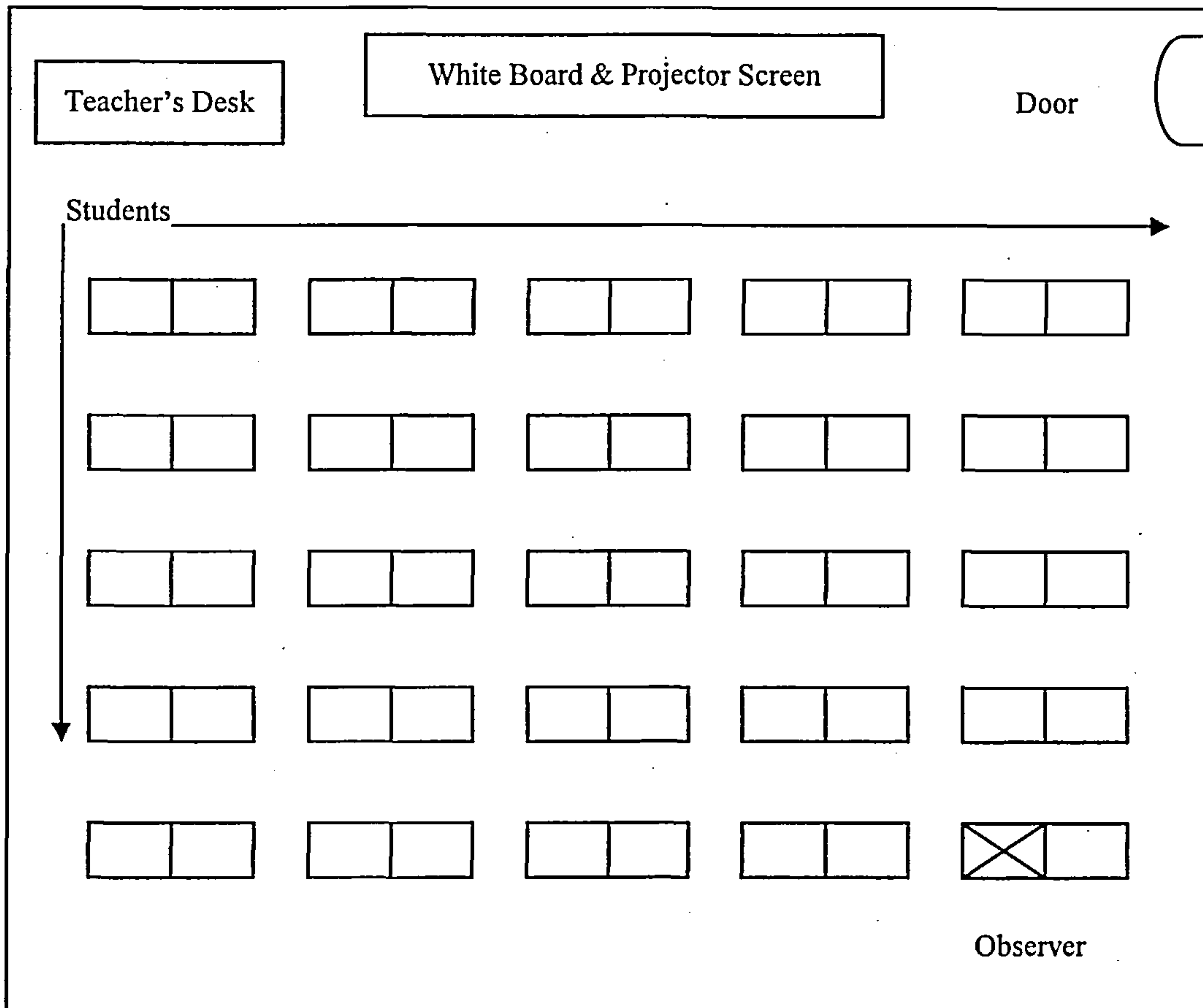
(Extract from Miss T's syllabus note)

It is clear from the above syllabus extract that Miss T's teaching objective focused mainly on training for the listening comprehension part of the General English Proficiency Test (GEPT) in Taiwan. Miss T used two textbooks in class. One textbook was aimed at everyday English (Appendix E.1 and E.2), while the other, which comprised the test battery of the GEPT, was written and published for test purposes (Appendix E.3). In addition to the two textbooks, Miss T excerpted short passages of news reports from an American language learning magazine published in Taiwan, entitled *EZ Talk*, as supplementary in-class listening material. Three observations were carried out before the mid-term exam from 08:10 to 10:00 am on the 16th, 23rd, and 30th of October 2006. It took 12 weeks, from the 16th October 2006 to the 10th January 2007, to complete the research on Case 2. Each lesson lasted 100 minutes with a ten-minute break in the middle.

6.3 Case 2: Classroom Observation before the Mid-term Exam

The students had their lesson in an audio-visual classroom which was specifically designed for English listening classes. The equipment this group had in the classroom was similar to what the students had in Case 1 – personal headsets and microphones (Figure 6.1). Exams also took place in this classroom.

Figure 6.1 Case 2: The Layout of the Classroom



6.3.1 Case 2: First Observation

The first observation was carried out on the 16th of October 2006. At the beginning of the lesson, Miss T discussed the content of the quiz the students had taken the previous week, and pointed out several items on which students made mistakes and gave them correct answers. She also asked the students if they had any questions regarding the quiz, but none of them said they did. Miss T told the students to practice their listening ability after the class, rather than simply rely on

the listening materials they had used from the class. Next, she asked the students to turn to page 7 of Unit 2 in *Listen Up*. Before playing the CD, Miss T asked the students to finish part of the exercises in the textbook while listening to them. After listening to, and finishing, the exercises, Miss T asked the students to answer the questions; however, no one appeared to want to answer them. The teacher asked the questions twice, and only two or three students sitting in the front row answered. Next, the students were divided into pairs, because the teacher wanted them to find the colloquial language used in the listening passages they had just heard. However, while discussing the tasks with the teacher, only one or two pairs of students answered the questions; the other pairs remained quiet. Miss T then moved on to the next listening tasks in the textbook. Again, she asked the students to do the exercises while listening to the CD, and she played the CD again before checking the answers with them.

In the second half of the lesson, the students listened to several excerpts from news reports in *EZ Talk* magazine, and completed the questions by filling in the missing words they had heard (Appendix E.4). The topics of the news reports in the magazine covered politics, economics, society, weather, entertainment, and education. Miss T told the students to listen for main ideas first and then listen for detailed information. She played the CD three times and then discussed the answers with the students. Next, she explained new vocabulary or phrases in the news reports. In the first observation, the majority of the students, apart from a few sitting in the front row, appeared not to interact with the teacher. Although she tried to involve the students in the discussion, most of them tended to work alone – listening to the CD and doing the exercises individually and quietly. Thus, the majority of the students did not speak English in class, even though they were given opportunities to do so. Because the students appeared not to interact with

the teacher, it was very difficult to know if they understood the listening content. The language in the textbook (*Listen Up*) and the news excerpts in the magazine appeared not to be authentic, in the sense that it was too fluent, without a single broken sentence or false start (Appendix E.2), and they were designed for language learning purposes.

6.3.2 Case 2: Second Observation

The second observation was carried out on the 23rd of October 2006. The main activities in this lesson were Unit 3 in *Listen Up* and Unit 2 in GEPT. At the beginning of the class, the students were told to do the exercises while listening to the CD. The teacher checked the answers with the students; however, only a few students sitting in the front row answered the questions; the others did not interact with her. Miss T then moved on to the next exercises. She asked the students to listen for the main ideas, and then played the CD twice. She then divided them into pairs and asked them to find the colloquial language used in the listening passages. When it came to reporting the findings, only two pairs of students sitting in the front row shared their findings with the teacher and other students. Finally, when the teacher checked the students to see if they had understood the listening extracts in the textbook, roughly half of them nodded, but the other half kept silent. The teacher then played the CD again to make sure they understood.

In the second half of the lesson, the students did the exercises in the GEPT textbook in class. There were three types of listening activities in the GEPT test bank for intermediate level – (1) picture description; (2) question or statement response; (3) short conversation (Appendix E.3). She played the CD and the students practiced the test questions, and then she checked the answers with the students. The reason why GEPT was part of the curriculum was that the

department required all students to pass the General English Proficiency Test before they could graduate (see Section 6.5). That is, Miss T taught GEPT primarily for testing purposes. However, when I asked Dr. N from Case 1 why she did *not* teach GEPT in class, her answer was that she had not been told to teach it. I asked the two teachers if they knew each other. Surprisingly, Dr. N and Miss T had not heard of each other and did not know each other's teaching materials or course teaching plans at all. It was thus clear that there was a large discrepancy between the syllabus descriptions and teaching materials for the two groups, although the students were in the same year and technically taking the same course.

6.3.3 Case 2: Third Observation

The last observation before the mid-term exam was carried out on the 30th of October 2006. In this lesson, Unit 4 in *Listen Up* and news reports from *EZ Talk* magazine were the main listening tasks. The procedure for teaching Unit 4 was similar to that used to teach the other units (see Sections 6.3.1 and 6.3.2). After finishing Unit 4 in *Listen Up*, Miss T moved on to the listening tasks of the news reports. The students were also asked to fill in the missing words they heard in the extracts provided by the teacher. Miss T asked the students to listen for the main ideas first, then for detailed information. Next, she explained the new vocabulary and phrases in the news extracts, and checked the answers with the students. The majority of the students, except for three or four sitting in the front row, did not interact with the teacher in class; they completed the listening tasks individually. Again, Miss T tried to encourage the students to answer the questions, but only a few students did so. In this case, only the students sitting in the front row interacted with the teacher, but the remaining students appeared to be isolated

because they worked quietly and individually. It was therefore again hard to know if the students understood the teaching content when half of them kept silent.

Table 6.1 Case 2: Occurrence of Task-based Instruction Features before the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	1 st	✓	Students were assigned a task to find colloquial language used in the conversation.
	2 nd	✓	The same as above.
	3 rd	✓	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st	✓ x	The teacher divided the students into pairs, and asked the pairs to search for the colloquial language used in the listening passages they had just heard. In addition to pair discussions, the students in class were quiet; only a few students sitting in the front row interacted with the teacher.
	2 nd	✓ x	The same as above.
	3 rd	✓ x	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	✓ x	Only one or two pairs of students answered the questions. Other pairs remained quiet.
	2 nd	✓ x	The same as above.
	3 rd	✓ x	The same as above.
4. Authentic texts which reflect a real-life situation are used.	1 st	x	The language in <i>Listen Up</i> and <i>EZ Talk</i> magazine is less authentic than the language in real-life situation (see Section 6.3.1 and Appendix E.3).
	2 nd	x	The teacher felt the conversations in the GEPT were too formulaic (Table 6.13).
	3 rd	x	The same as 1 st observation.
5. The major focus of teaching is on the meaning, and then on the form.	1 st	✓ x	The teacher told the students to listen for main ideas first and then for detailed information, but she did not discuss the grammar point after listening
	2 nd	✓ x	The same as above.
	3 rd	✓ x	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	x	No reflection period was observed in this lesson.
	2 nd	x	The same as above.
	3 rd	x	The same as above.

To summarise, only a few of the features of task-based instruction were observed in Miss T's class. It appeared that Dr. N provided more opportunities with the students to speak English by encouraging them to answer the questions by awarding extra marks and asking them to give presentations in class. Asking students to answer the questions in class helped Dr. N find out if her students understood the listening passages. Miss T, on the other hand, focused more on listening input than on oral production; only a few students were willing to interact with the teacher. Importantly, the teaching objectives of the two teachers were very different, in the sense that Dr. N focused on the understanding of everyday English and academic English lectures, while Miss T put stress on the training for the GEPT test and the understanding of news reports.

6.4 Case 2: Questionnaire Survey on the Mid-term Exam

The students took the exam in the same classroom in which they had the English listening lessons on the 6th of November 2006. The test comprised 35 questions, including multiple-choice, true/false, and short-answer questions (Appendix E.5). Given the use of similar types of question in *Listen Up* and *GEPT*, the students were assumed to be familiar with the multiple-choice and true/false questions at a general level. In Case 2, the test items were not piloted due to the limitation of test preparation time (see Section 6.5). In the mid-term exam, the teacher played the CD twice; however, more than half of the students asked her to play it for a third time. The exam lasted approximately 50 minutes. The questionnaire was group-administered to the students immediately after the exam had finished. Twenty-three of the students were female, and ten were male (Table 6.2); there were no absentees. Miss T was not in the classroom while the students completed

the questionnaire. It took approximately 20 minutes for the group to finish, and the students did not report any problems with filling it in.

Table 6.2 Case 2: Mid-term Survey – Gender of the Respondents

Gender	Frequency	Percent
Male	10	30.3%
Female	23	69.7%
Total	33	100%

6.4.1 Case 2: The Teaching Materials and the Course before the Mid-term Exam

In Question 4, only six students rated them as “frequently” difficult (Table 6.3). However, it appeared that the listening materials for this group were nevertheless more difficult than those rated by Case 1 (cf. Table 5.3), in the sense that 85% in Case 2 found them “sometimes” or “frequently” difficult, compared with just 45.2% in Case 1. For the 32 students who considered the materials difficult to some extent, speech rate and the use of colloquial language were the two main problems cited (Table 6.4), though vocabulary, accents, and text type also led to problems. Eight students reported that they needed to listen to the news broadcasts many times, because speakers often spoke fast in the news broadcasts. In addition, they said that news broadcasts tended to be very professional in terms of terminology and there was new vocabulary which they found hard to understand.

TABLE 6.3 Case 2: Mid-term Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	1	3%
Rarely	4	12.1%
Sometimes	22	66.7%
Frequently	6	18.2%
Always	0	0%
Total	33	100%

TABLE 6.4 Case 2: Mid-term Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	8
Vocabulary	17
Accent	10
Speech rate	29
The use of colloquial language	23
Sound quality	4
Text type	9

In Question 5, all but three students agreed that the course contents were relevant to their listening needs. Nearly half of the students pointed out that teaching for the GEPT helped them prepare for the test before graduation by familiarising them with the question types and test format. The other one third said that the teaching materials were versatile and would let them learn everyday English, news issues, and help them prepare for tests. However, they felt that it was hard to understand most topics in the news broadcasts. In addition, Students 23 and 31 complained that the teacher put too much emphasis on the supplementary materials (i.e. news broadcasts), neglecting the importance of teaching everyday English. Another student (Student 17) said that the teacher should teach listening skills rather than playing the CD again and again, as she could listen to it by herself at home.

6.4.2 Case 2: Students' General Preferences about the English Listening Class

In contrast to the students in Case 1, where 61.3% preferred written tests, more than half of the students (54.4%) in Case 2 preferred oral tasks to written ones for English listening tests (cf. Tables 5.5 and 6.5). One third of the respondents said that it was more interesting to take an oral exam than a written one because the class was very boring. The other third said that it was quicker to answer the questions by "speaking", rather than "writing" things down. Another third felt that answering questions orally allowed them different ways to express answers, but written answers were more fixed. Nevertheless, the remaining fifteen students believed that they performed better in a written exam, because eight of them reckoned their oral ability to be poorer than their written ability. In addition, 25 students (75.8%) regarded "conversations" in the listening texts as easier to

understand than “monologues” (Table 6.6). Indeed, conversational types of listening extract were favoured by the majority of the students in both cases.

TABLE 6.5 Case 2: Mid-term Survey – Preferred Mode of Answering (Q6)

Mode	Frequency	Percent
Writing	15	45.5%
Speaking	18	54.5%
Total	33	100%

TABLE 6.6 Case2: Mid-term Survey – Preferred Type of Speech (Q7)

Types of Speech	Frequency	Percent
Monologue	8	24.2%
Conversations/dialogues	25	75.8%
Total	33	100%

More than three quarters of the students (81.8%) preferred multiple-choice and true/false questions (Table 6.7). Unsurprisingly, they reported that they could select the correct answer according to the pre-determined multiple options. The students also claimed that it was easy to distinguish true/false questions while listening. Short-answer questions were again the type of question nearly two thirds of the students (64.5%) preferred least (Table 6.7); they said that they were liable to spell the vocabulary incorrectly in short-answer questions while writing down the answers and listening to the extracts at the same time. In short, both groups of students in University A preferred and disliked the same types of comprehension questions in the listening tests with multiple-choice and true/false questions preferred most, and short-answer questions preferred least by the majority (Tables 5.7 and 6.7).






Table 6.7 Case 2: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class

Types of Questions	Preferred Question (Q8)		Least Preferred Question (Q9)	
	Frequency	Percent	Frequency	Percent
Short-answer	4	12.1%	21	63.6%
True/false	11	33.3%	0	0%
Multiple-choice	16	48.5%	4	12.1%
Cloze	2	6%	3	9.1%
Dictation	0	0%	4	12.1%
Total	33	100%	33	100%

6.4.3 Case 2: Questions about the Mid-term Exam

The students in Case 2 seemed much less “satisfied” with their mid-term performance than those in Case 1, since only five students (15.2%) said they were “satisfied”. More than a quarter (27.3%), however, said that they were “(strongly) dissatisfied” (Table 6.8; Question 10).

TABLE 6.8 Case 2: Mid-term Survey – Performance in the Mid-term Exam (Q10)

				
1	2	3	4	5

Satisfaction with the Mid-term Exam	Frequency	Percent
Face 1 – Very satisfied	0	0%
Face 2 – Satisfied	5	15.2%
Face 3 – Neither Satisfied nor dissatisfied	19	57.6%
Face 4 – Dissatisfied	6	18.2%
Face 5 – Very dissatisfied	3	9.1%
Total	33	100%

Only two students “agreed” that the test topics were difficult (Table 6.9; Question 11-a). Indeed, the test topics in the two separate mid-term exams were not a problem for the majority of the students in both cases. However, only six students (18.2%) “(strongly) agreed” that the topics of the tasks in the mid-term exam were representative of what had been taught in class (Table 6.9; Question 11-b). This contrasts with the situation in Case 1, where more than half “(strongly) agreed” that the test tasks were considerably representative (cf. Table 5.9). It therefore appeared that, from the students’ perceptions, the topics of test tasks were more representative of the curriculum for Case 1 than Case 2. Less than one third of the students (30.3%) “(strongly) agreed” that the test tasks were harder than the exercises they did in class (Table 6.9; Question 11-c). Compared with Case 1, more students in Case 2 felt that the test tasks were harder than those done in class (cf. Table 5.9 and 6.9).

Table 6.9 Case 2: Mid-term Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q11-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q11-b)		Test Tasks Were Harder than Those in Class (Q11-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	1	3%	1	3%	1	3%
D	7	21.2%	5	15.2%	11	33.3%
Neither A nor D	23	69.7%	21	63.6%	11	33.3%
A	2	6.1%	4	12.1%	8	24.2%
SA	0	0%	2	6.1%	2	6.1%
Total	33	100%	33	100%	33	100%

No more than five students “agreed” that the accent was strong, the vocabulary was too difficult, or the sentences were too complicated to understand (Table 6.10). The majority of Case 1, similarly, felt that accents, vocabulary, and sentences were not a problem (Tables 5.10 and 6.10).

Table 6.10 Case 2: Mid-term Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q11-d)		Vocabulary was too Difficult to Understand (Q11-e)		Sentences were too Complicated to Understand (Q11-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	1	3%	1	3%
D	14	42.4%	11	33.3%	16	48.5%
Neither A nor D	16	48.5%	16	48.5%	12	36.4%
A	3	9.1%	5	15.2%	4	12.1%
SA	0	0%	0	0%	0	0%
Total	33	100%	33	100%	33	100%

Having said that, more than half of the students (57.6%) “(strongly) agreed” that it was hard to understand the listening extracts because the speakers spoke too fast (Table 6.11; Question 11-g). Indeed for both groups of students, the speed of the listening extracts in both exams was considered too fast to understand. The mid-term exam contained both monologue (Part 4, Appendix E.5) and conversational texts (Part 1 to 3, Appendix E.5). However, the monologue appeared to have been more difficult than the conversations, since more than a third of the students (36.3%) “(strongly) agreed” that the it was hard to understand

the monologue. Thus, the students in both groups felt that understanding monologue was more difficult than conversations (Tables 5.11 and 6.11).

Table 6.11 Case 2: Mid-term Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q11-g)		Monologue Speech Hard to Understand (Q11-h)		Conversations Hard to Understand (Q11-i)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	0	0%	1	3%
D	3	9.1%	3	9.1%	9	27.3%
Neither A nor D	11	33.3%	18	54.5%	19	57.6%
A	16	48.5%	11	33.3%	4	12.1%
SA	3	9.1%	1	3%	0	0%
Total	33	100%	33	100%	33	100%

In Question 12, eleven students reported that colloquial language was used, and ten did find it difficult to understand. Colloquial language was found difficult to understand for approximately one third of the students. To summarise, understanding the topics, accents of speakers, vocabulary, and sentences was not perceived as difficult for the majority of the students in either group. The topics of the test tasks in Case 2 mid-term exam were less representative of what had been taught in class than those in Case 1. The fast pace of speech and the use of colloquial language in the two mid-term exams appeared to be the two main obstacles which hindered more than half of the students in two groups from comprehending the listening extracts.

On the question of the test characteristics of the mid-term exam, the results were evenly split. Fourteen students (42.4%) did not have any problems regarding the quality of the recording equipment (Appendix E.8), but a third (33.3%) complained that the audio quality was so poor that they could not hear the texts clearly. Another four pointed out that the volume of part of the recording was too low. The test time appeared to be sufficient for nearly two thirds of the students (63.6%), though five of them said that the testing time was so long that they

listened to the questions too many times. The task/test instructions were considered clear by more than half of the students (54.5%). All except two students reported that the lengths of the listening passages in this exam were similar to what they had listened to in class. The two students complained that the exam texts were longer than the class ones. Multiple-choice questions were regarded as the easiest type of question by nearly two thirds of the students (63.6%) (Table 6.12) while short-answer questions, designed to test students' comprehension of a news report, were considered the most difficult by twenty students (60.6%) (Table 6.12).

Table 6.12 Case 2: Mid-term Survey – Easiest vs. Most Difficult Type of Comprehension

Question in the Mid-term Exam

Types of Questions	Easiest Question (Q14)		Most Difficult Question (Q15)	
	Frequency	Percent	Frequency	Percent
Short-answer	1	3%	20	60.6%
True/false	11	33.3%	6	18.2%
Multiple-choice	21	63.6%	7	21.2%
Cloze	0	0%	0	0%
Dictation	0	0%	0	0%
Total	33	100%	33	100%

6.4.4 Case 2: Students' General Comments on the Mid-term Exam

Twenty-seven students (81.8%) out of the 33 reported that they had learned what they expected to learn thus far; the remaining four students said that the materials were too easy, so that they did not make any progress. In particular, Students 2 and 16 said that it was hard to understand the topics in the exam, because they were irrelevant to what was taught in class. All but two students (93.9%) were satisfied with the way Miss T taught before the mid-term exam, but Students 17 and 33 said that the teacher overestimated their level of English and assessed them with difficult test items. In short, all but three students were satisfied with the assessment methods used; the three concerned said that they did not perform well because the topics were too hard and the speakers spoke too fast.

6.5 Case 2: Mid-term Interview with the Teacher

The face-to-face interview was conducted on the 13th of November 2006, one week after the mid-term exam, when the students' marks were known. The interview took place in a department office in the University and lasted approximately 40 minutes. The interview questions were presented in both English and Chinese; Miss T chose to answer in Chinese. The interview was tape recorded with her permission, transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in Appendix E.9. The translation was as before double-checked by a Chinese speaker who could speak both Chinese and English and a native English speaker.

Table 6.13 Case 2: Mid-term Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Miss T (T): They were from the English magazine and GEPT test battery.</p> <p>Interviewer (I): What kind of magazine?</p> <p>T: It's <i>EZ Talk</i> magazine, because I found the topics in the <i>EZ Talk</i> magazine were practical and related to our everyday life. There were several test questions included in the magazine which tested listeners' comprehension after listening to the content.</p> <p>I: Did the students know that you used the test questions from the magazine?</p> <p>T: They may not have known at the beginning of the course, but they probably did afterwards.</p> <p>I: But this magazine can be purchased; was it possible that the students procured similar test questions?</p> <p>T: I didn't use the test questions in recent issues.</p> <p>I: How about the test questions for GEPT?</p> <p>T: They were from the in-class textbook. Because the chairperson expected the students in our department would be able to pass the GEPT before graduation, I tested those questions.</p> <p>I: Did you design the questions yourself?</p> <p>T: Yes.</p> <p>I: Have the students ever heard the test content?</p> <p>T: I think so. I chose the questions from the last chapter. Because I only taught the first two chapters before the mid-term exam, I chose the questions from the final chapter.</p> <p>I: So it was still possible that students might have heard it before.</p> <p>T: I think the probability was very low, because I found out that the students were passive about learning, unless they really listened to the last chapter....but I changed the questions. I remembered that I tested the students with the same listening content with different questions before, but the students did not perform as well as I expected. You know...there were simply one or two questions in short conversations in GEPT, but in fact there were other questions to ask based on the conversations. So I would change the type of comprehension question, such as true/false or short-answer questions.</p> <p>I: Did the students take notes on their own in class?</p> <p>T: No, so I asked them to find colloquial language in the listening passages, and divided them into groups to discuss the colloquial language for two to three minutes, and asked them to talk about it in class. I told them that they did not have to speak correctly. If they did not know the correct answer, they could still speak out, or guess the answers...I wanted to develop their autonomous learning skills.</p>

I: Did you think that your students' language ability met the level of the listening materials you used in class?

T: I found it OK, because the majority of the students' listening ability was in the middle; only a few students' ability was very good. Some students felt that the two sets of listening materials (*Listen Up* and *GEPT*) I used were too easy, and some felt they were too hard, but the majority said they were OK. For example, they considered that questions in "Picture Description" in the *GEPT* were too easy, but the colloquial language in the short conversations or news items was more difficult for them; it was harder for them to get the meaning in general. It was possible that *GEPT* consisted of more multiple-choice questions, so the probability that they answered correctly was higher; but if I'd changed another type of question, they might not be able to answer correctly. That's why I used the *EZ Talk* magazine; there was a lot of colloquial language in there, but the conversations in *GEPT* were formulaic, because you can understand the contents easily. But I used the questions from *EZ Talk* magazine in the first quiz, the students felt challenged and practical, owing to the colloquial language, but they also told me the quiz was difficult. However, I said to them that they could not always learn something easy; they will not always be able to understand what foreigners say and always follow their speaking speed, so they needed to be familiar with something different.

I: Did you yourself decide to teach *GEPT*?

T: The chairperson asked me to teach that. But for me, I felt that even though you passed the *GEPT*, you simply practiced the listening for testing purposes....well....I taught the *GEPT* since the chairperson asked me to do that.

I: Why did not you teach listening to "lectures"?

T: Because I discovered that the students could not follow the speaking speed while listening to the news....you see, news reports were usually short in length, but they could not cope with those, not to mention the long length of "lectures". Basically I do not like to kill their learning interest in class; they are more likely to feel encouraged when the materials are neither too hard nor too easy.

I: Did you pilot the test items, I mean run a small test, before you test your students in this exam?

T: No, if I had time, I would have tried to pilot those questions.

In this English listening course, Miss T put the emphasis on understanding how to use English in practical ways, in the sense that the listening topics in the in-class materials were related to current English language and socio-economic issues (i.e. news reports), and colloquial language usage (Table 6.13; Question 1). The teacher was aware of the fact that some of the students found colloquial language in the textbooks and news reports challenging and hard to understand. However, she believed that teaching them challenging materials was more beneficial than teaching them easier things. In the mid-term exam, only two students considered the test content hard to understand, but a third of them did regard the colloquial language in the teaching materials and mid-term exam as difficult. It is possible that colloquial English is not very commonly used in academic settings. It is also possible that, from the teacher's point of view, the students did not study autonomously. In addition, Miss T pointed out that the

students' English listening performance was influenced by different types of comprehension question or the way questions were asked. Teaching listening for examination purposes was part of the curriculum, but it appeared that Miss T did not regard it as a positive way of teaching listening.

Table 6.14 Case 2: Mid-term Survey – Interview with the Teacher (b)

<p>Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?</p> <p>T: 60. The mid-term score accounted for 30% of the total score. The final exam accounted for 40%. The quiz accounted for 30%. The students who did not pass the mid-term exam asked me to make the final exam easier, or even test them on the listening extracts they had listened to in class. But I told them that, even if you received full scores by being tested on the listening extracts you had listened to in class, it would not constitute your real language ability. So I never used the questions they had practiced in class. And even though I used different listening contents, those who had scored higher still received higher marks, and those with middle scores were still in the middle. So, if I used the listening passages they had listened to before, then everyone would receive high scores. If they did not want to be failed, they would have to try to gain higher scores in the quizzes, because the quizzes were easier. But I still asked them to work harder, instead of them asking me to give them easier tests.</p>
--

There were five students (15.2%) who did not pass the mid-term exam. Miss T was asked by a few students to decrease the difficulty of the final exam by using the listening extracts that the students *had* listened to before in class; however, she regarded the request as constituting unfair assessment (Table 6.14; Question 2). She thus suggested that the students who did not want to fail could put more effort into the quizzes.

Table 6.15 Case 2: Mid-term Survey – Interview with the Teacher (c)

<p>Question 3. What did you expect the students to have learned from your class?</p> <p>T: Just as I said before, the language in GEPT was formulaic. I would like to teach them the language they would use and hear everyday in the future, such as news reports; they could learn the language and current socio-economical information in news reports. As for language learning, I did not think that there were such learning strategies: you simply listen, practice, and use it. Because even though I told the students what strategies they needed for preparing the examinations, it was still hard to test their real ability. I told the students to listen to different topics, so their listening ability would improve. I also told them testing listening is not like testing vocabulary. It is not the case that if you practice listening the day before the examination, you could perform well. Listening ability needs continuous training. I told them to spend at least half an hour listening to the radio or news reports to strengthen their listening ability.</p>
--

In addition to learning English, Miss T expected that her students would learn knowledge from the listening extracts about recent cultural or socio-

economic issues (Table 6.15; Question 3). Interestingly, she did not believe that learning strategies improved students' listening ability; rather she suggested that listening, practicing, and using the English persistently were more useful to strengthen listening ability.

Table 6.16 Case 2: Mid-term Survey – Interview with the Teacher (d)

<p>Question 4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?</p>
<p>T: I did not want them to practice their listening ability for examination purposes. I hope that my students know how to use the language. I felt that they were improving after teaching half a semester. At least they improved a bit, and met a small part of my expectation. However, because each student's learning situations, preferences, and learning speed varied, I found it difficult to completely achieve my goals, but on average, they are improving.</p>
<p>I: On which part of the test did the students perform worst?</p>
<p>T: There were no obvious differences, but there were still several students who did not get the meaning of the short-answer questions.</p>
<p>I: I notice that you did not test students' speaking ability in the mid-term exam. Were there any reasons?</p>
<p>T: Well...based on the course aim in our department, I think this course was designed to train listening rather than speaking. I just followed the departmental curriculum.</p>
<p>I: Will you use any speaking test for the final exam?</p>
<p>T: uh...I don't think so.</p>

The teacher wanted to know if the students understood how to use the language they had learned in class (Table 6.16; Question 4). However, Miss T found it hard to completely achieve her goals owing to the individual students' learning situation, English level, and preferences. Thus, she felt that the improvement in her students' learning was limited. In addition, students' speaking ability was not tested in the mid-term exam and would not be tested in the final exam owing to the curriculum design. In this group, the mean of the mid-term scores for Case 2 was 69.8.

Table 6.17 Case 2: Mid-term Survey – Interview with the Teacher (e)

<p>Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?</p>
<p>T: I decided on the difficulty of the mid-term exam from the results of the quizzes, their in-class performance, and supplementary materials. If they were not interested in certain topics or they did not perform well, I adjusted the difficulty in the light of their reactions and performance in class.</p>

Although the teacher said that she did not like her students to listen to easy English materials, she still took the students' reactions to the listening materials and performance in class into consideration when deciding the content in the mid-term exam (Table 6.17; Question 5). Even so, nearly one third of the students (30.3%) felt that the test tasks in the mid-term test were harder than the exercises used in class (Table 6.9).

Table 6.18 Case 2: Mid-term Survey – Interview with the Teacher (f)

<p>Question 6. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p>
<p>T: The reason I tested them using short-answer questions was to prevent the students from guessing the correct answers in the multiple-choice questions, so I tested them with different types of comprehension question. I thought that they were busy studying, so I did not need to test them with too difficult questions. If I used short-answer questions, I would know whether they really understood and improved their listening. Well...of course students preferred the multiple-choice questions, but I thought that if I tested them with multiple-choice questions, they might guess the answers, so then I would not be able to know if they really understood the contents. So I preferred to use different types of question to assess their real ability. I did not want to fail them, so if they wanted to pass the exam, it would need to depend on their true ability. They also guessed the answers in the true/false questions; they chose either "true" or "false" for all questions, so I preferred to use short-answer questions to prevent them from guessing the answers.</p>

In Question 6, Miss T was aware of the situation that students preferred multiple-choice and true/false questions, but she knew that it was hard to prevent the students from guessing answers in the two types of question (Table 6.18). The teacher employed short-answer questions in order to decrease the probability of guessing answers.

Table 6.19 Case 2: Mid-term Survey – Interview with the Teacher (g)

<p>Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)</p>
<p>T: I would like to increase the difficulty after seeing their improvement in the mid-term exam, so I will increase the "news report" in class materials. I want to give music and films to the students in English in class, but I have not tried it before the mid-term exam. I was unsure if I wanted to give or not to give them the music in class before I knew their level of English in the mid-term exam. However, because they did improve in this exam, I think I will play music and films in class after several weeks. I could never stay in the same place, otherwise they will not improve. Although moving to the next level might increase the burden on those whose level of English was lower, I still have to consider the students with a higher level of English. Choosing appropriate materials for the students is a challenge for me.</p>

I: Will you still use the two textbooks after the mid-term exam?

T: Yes, I will, but I will give them extra advanced listening materials. Because their textbooks were fairly fundamental, I will give some extra materials.

I: Will you increase the difficulty of test contents in the final exam?

T: I want to increase the difficulty, but many students will fail. But it is not possible for me to give an easier test, so I think it will be a bit harder than the mid-term exam. I hope that those whose English abilities are lower will be able to perform well in the quizzes. The score for the final exam accounted a large percentage of the total score; if they cannot perform well in quizzes, I am not sure what I will do.

The teacher thought that her students' listening ability had improved in the mid-term exam, even though their performance had not met her expectation (Table 6.19; Question 7). Miss T understood that increasing the difficulty of the final exam could result in more students failing, but she was eager to improve their listening ability by using advanced listening materials such as news reports, music, and films in English. She was concerned about a situation in which the students with lower English listening ability might not be able to pass the final exam because it accounted for a large percentage of the total score. Again, Miss T strongly recommended them to try and score higher in the quizzes.

6.6 Case 2: Classroom Observation after the Mid-term Exam

Listen Up and *GEPT* continued to be used in class after the mid-term exam.

Another three observations were carried out from 08: 10 am to 10: 00 pm on 27 November, 11 and 25 December 2006.

6.6.1 Fourth Observation

The first observation after the mid-term exam was carried out on the 27th of November 2006. The activities in this lesson focused on listening to Unit 6 in *Listen Up* and *EZ Talk* magazine. At the beginning of the lesson the students were asked to do the first two parts of the exercises in *Listen Up* while listening. The teacher played the CD twice and then asked the students if they understood. Only six or seven students sitting in the front row nodded; the remaining students kept

silent. Next the teacher discussed the answers to the exercises with the class; however, just three or four students sitting at the front row interacted with her. Miss T then moved on to next exercises; she again asked the students to do the exercises while listening to the CD. After listening twice, she checked the answers with the students. Again, only three or four students sitting in the front interacted with her.

In the second half of the lesson, Miss T asked the students to listen to two conversations and ten short news reports taken from *EZ Talk* magazine. After listening to two conversations, the teacher divided the students into pairs for discussion and asked them to find the colloquial language in the two conversations. The pair discussion lasted ten minutes. Three to four pairs of students got involved in the interaction when the teacher discussed the colloquial language they had found. As for listening to news reports, the students were, again, given ten questions covering the excerpts they listened to. They needed to fill in the missing words they heard in the news; then Miss T checked the answers with them.

6.6.2 Case 2: Fifth Observation

The lesson was observed on the 11th of December 2006. At the beginning of the lesson, Miss T discussed the quiz the students took in the previous week. The teacher explained three or four questions which students were confused by. She told the students to study harder because they had not performed well generally. She then asked the students if the quiz was too hard, and the students sitting in the front two rows said it was not, but the remaining students did not say anything. The teacher did not ask those who did not express an opinion but went on teaching. Next, she taught Unit 3 in *GEPT*. She told the students to do the tasks

in a “Picture Description” while listening to the CD. While discussing the answers, none of the students had any questions. The teacher then moved on to the second part “Question and Statement Response” in this unit. She played the listening contents twice and discussed the answers with the students. The class then took a break.

In the second half of the class, the teacher played the listening passages the students had just listened to in “Question and Statement Response” and asked them to find the colloquial language. Students were again divided into pairs. After ten minutes, the teacher discussed the answers with the students. Again, only three to four pairs sitting in the front row interacted with her. Next, she moved on to “Short Conversations”. She played the conversations twice, but approximately a third of the students asked her to play the tape one more time. The colloquial language in these “Short Conversations” was also pair-discussed for ten minutes; however, the situation of discussing the answers with the teacher was the same as before. Next, the teacher played a pop song in English; and the students were given a piece of paper to fill in the missing words they heard.

6.6.3 Case 2: Final Observation

The final observation took place on the 25th of December 2006. In the lesson, students did a quiz at the beginning of the class. The quiz consisted of filling in the missing words of a song, a news report, and a conversation (Appendix E.7). The quiz lasted thirty minutes and Miss T played the passages three times, but five or six students asked her to play the song and news for a fourth time. After the quiz, the teacher taught Unit 8 in *Listen Up*. The procedure for teaching *Listen Up* was similar to what had been observed before (see Sections 6.3.1, 6.3.2, 6.3.3, 6.6.1). In the second half of the lesson, the teacher finished Unit 8 and moved on

to the conversations in *EZ Talk* magazine. The students were again divided into pairs to discuss the colloquial language. Unsurprisingly, only three to four pairs interacted with the teacher while discussing the answers.

In the three observations after the mid-term exam, the majority of the students appeared to be uninterested in interacting with the teacher, since less than ten students did so. Although the students were all given opportunities to speak English via pair discussions, only ten minutes were given for them, which was inadequate.

Table 6.20 Case 2: Occurrence of Task-based Instruction Features after the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	4 th	✓	Students were assigned a task to find out colloquial language used in the conversation.
	5 th	✓	The same as above.
	6 th	✓	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	4 th	✓ x	The teacher divided the students into pairs, and asked the pairs to search for the colloquial language used in the listening passage. In addition to pair discussions, the students in class were quiet; only a few students sitting in the front row interacted with the teacher.
	5 th	✓ x	The same as above. Students did the listening tasks individually.
	6 th	✓ x	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	4 th	✓ x	Only three to four pairs of students answered the questions. The other pairs remained quiet.
	5 th	✓ x	The listening extracts in GEPT were also inauthentic since they were too formulaic.
	6 th	✓ x	The same as above.
4. Authentic texts which reflect a real-life situation are used.	4 th	x	The listening extracts in <i>Listen Up</i> were inauthentic because the texts were not based on language in real-life situations (see Section 6.2). However, the teacher felt that the language in the news reports was more authentic.
	5 th	x	The same as above.
	6 th	x	The same as above.
5. The major focus of teaching is on the meaning, and then on the form.	4 th	✓ x	The teacher told the students to listen for main ideas, but she did not teach grammar.

	5 th	✓ x	The same as above.
	6 th	✓ x	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	4 th	x	No reflection period was observed in this lesson.
	5 th	x	The same as above.
	6 th	x	The same as above.

6.7 Case 2: Questionnaire Survey on the Final Exam

The students took the final exam on the 15th of January 2007. It consisted of 30 questions, involving true/false, multiple-choice, and short-answer questions (Appendix E.6). In addition to the 30 questions, a song with 77 cloze blanks was also included in the test for bonus marks. Using short-answer questions in tests reduces the possibility of guessing in true/false or multiple-choice items, however, it is recommended that answers should be kept short so that students would not overhear the next question (Hughes, 2002: 145). Miss T played the whole passage three times, but more than half of the students asked her to play it a fourth time. However, the test items in the final exam were again not piloted, Miss T said, owing to insufficient preparation time. This raised the question of how far course instructors can realistically pilot their test items before administering formal classroom examinations. The exam lasted approximately one hour. After the exam and Miss T had left the class, the final questionnaire was group-administered to the students. It took them 20 minutes to finish.

6.7.1 Case 2: The Teaching Materials and the Course after the Mid-term Exam

Compared with the students' feelings about the difficulty of the in-class materials before the mid-term, the proportion of the students who felt the materials were "frequently" difficult after the mid-term exam increased from 18.2% to 33.3% (Table 6.21). The data collected from the classroom observations showed the teacher incorporated "listening to songs in English" into the syllabus, and it was

possible that students found it hard to follow the speed. Indeed, Table 6.22 shows that speech rate was still the major problem for 25 students (75.8%) in the second half of the course. Colloquial language in the listening materials was again a problem for almost half of the students (45.5%).

TABLE 6.21 Case 2: Final Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	0	0%
Rarely	4	12.1%
Sometimes	18	54.5%
Frequently	11	33.3%
Always	0	0%
Total	33	100%

TABLE 6.22 Case 2: Final Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	8
Vocabulary	9
Accent	4
Speech rate	25
The use of colloquial language	15
Sound quality	3
Text type	8

6.7.2 Case 2: Students' General Preferences about English Listening Class

After the mid-term exam, more than half of the students (57.6%) still preferred to be tested in an oral form than in a written form (Table 6.23). The preferences for mode of answering remained almost the same before and after the mid-term exam, as did the preferences for the type of speech (Table 6.24).

TABLE 6.23 Case 2: Final Survey – Preferred Mode of Answering (Q5)

Mode	Frequency	Percent
Writing	14	42.4%
Speaking	19	57.6%
Total	33	100%

TABLE 6.24 Case 2: Final Survey – Preferred Type of Speech (Q6)

Types of Speech	Frequency	Percent
Monologue	8	24.2%
Conversations/dialogues	25	75.8%
Total	33	100%

Multiple-choice and true/false questions were again preferred by almost everybody (Table 6.25), though the proportion of students who disliked the short-answer questions increased slightly from 63.6% to 75.8% (Tables 6.7 and 6.25). It appeared that more students found it demanding to answer listening questions in an open-ended format after the mid-term exam.

Table 6.25 Case 2: Final Survey – Preferred vs. Least Preferred Type of Comprehension

Types of Questions	Preferred Question (Q7)		Least Preferred Question (Q8)	
	Frequency	Percent	Frequency	Percent
Short-answer	2	6%	25	75.8%
True/false	13	39.4%	0	0%
Multiple-choice	16	48.5%	4	12.1%
Cloze	2	6%	3	9.1%
Dictation	0	0%	0	0%
Total	33	100%	33	100%

6.7.3 Case 2: Questions about the Final Exam

More than half of the students (54.5%) were “(very) dissatisfied” with their performance in the final exam (Table 6.26); almost twice the number dissatisfied in the mid-term exam. The situation in the Case 2 also contrasted with that in Case 1, where more students were satisfied with their test results in the final exam than in the mid-term one.

TABLE 6.26 Case 2: Final Survey – Performance on the Final Exam (Q9)



Satisfaction with the Final Exam	Frequency	Percent
Face 1 – Very satisfied	0	0%
Face 2 – Satisfied	2	6.1%
Face 3 – Neither Satisfied nor dissatisfied	13	39.4%
Face 4 – Dissatisfied	14	42.4%
Face 5 – Very dissatisfied	4	12.1%
Total	33	100%

More than half of the students (51.5%) “(strongly) agreed” that the test topics were difficult, and almost three quarters (72.7%) reported that the test tasks were harder than tasks done in class (Table 6.27; Q11-a and Q11-c). Indeed, compared with the test topics in the mid-term exam, the proportion of students who felt the topics were hard increased drastically from 6.1% to 51.5% (cf. Tables 6.9 and 6.27). The proportion of students who felt the test tasks were harder than those done in class also increased markedly, from 30.3% to 72.7%. In addition, almost a

third of the students (30.3%) “(strongly) agreed” that the topics were representative of what they had been taught in class (Table 6.27; Q11-b). So, although the test tasks in the final exam were, to a certain degree, representative of the in-class materials, the majority of the students regarded the tasks as hard to understand.

Table 6.27 Case 2: Final Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q10-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q10-b)		Test Tasks Were Harder than Those in Class (Q10-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	0	0%	0	0%
D	4	12.1%	2	6.1%	0	0%
Neither A nor D	12	36.4%	21	63.6%	9	27.3%
A	16	48.5%	8	24.2%	23	69.7%
SA	1	3%	2	6.1%	1	3%
Total	33	100%	33	100%	33	100%

Only five students (15.2%) “agreed” that the accent of speakers was too strong to understand (Table 6.28; Q11-d). However, understanding the vocabulary and sentences in the listening passages were problems for a third (33.3%) and for nearly a quarter (24.2%) of the students respectively – double the number of students in the mid-term exam (cf. Tables 6.10 and 6.28).

Table 6.28 Case 2: Final Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q10-d)		Vocabulary was too Difficult to Understand (Q10-e)		Sentences were too Complicated to Understand (Q10-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	3	9.1%	0	0%	2	6.1%
D	3	9.1%	4	12.1%	4	12.1%
Neither A nor D	22	66.7%	18	54.5%	19	57.6%
A	5	15.2%	10	30.3%	7	21.2%
SA	0	0%	1	3%	1	3%
Total	33	100%	33	100%	33	100%

Nearly three quarters of the students (75.8%) “(strongly) agreed” that it was hard to understand the listening contents because the speakers spoke fast (Table 6.29; Q11-g). It appeared that the final test was perceived as harder than

the mid-term one, since the number of students who found that the speakers spoke too fast increased from 57.6% (mid-term exam) to 75.8%. In the final exam, both conversations (Part 1 and 2; Appendix E.6) and monologue speech (Part 3 to 5) were included. Almost two thirds (60.6%) “(strongly) agreed” that it was hard to understand the monologue speeches (i.e. a short talk, a news report, and a lecture). In the mid-term exam, the teacher only tested her students with a single monologue (i.e. a news report), but she increased this in the final exam to three different topics. Listening to many monologues proved difficult for students who were not used to listening to long or fast discourses.

Table 6.29 Case 2: Final Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q10-g)		Monologue Speech Hard to Understand (Q10-h)		Conversations Hard to Understand (Q10-i)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	2	6.1%	0	0%
D	1	3%	1	3%	7	21.2%
Neither A nor D	7	21.2%	10	30.3%	15	45.5%
A	20	60.6%	17	51.5%	10	30.3%
SA	5	15.2%	3	9.1%	1	3%
Total	33	100%	33	100%	33	100%

In Question 11, more than a third of the students (39.4%) said that colloquial language was used, but this time all of them found it hard to understand. Although the students did many exercises and had discussions about colloquial language in class, approximately a third again found it hard to recognise and understand it in the listening passages. Based on the students’ points of view in the two exams, the final exam was perceived as the harder of the two in terms of the topics, vocabulary, sentences, speech rate, monologue texts, and colloquial language. In Case 1, on the other hand, the final exam was perceived as easier than the mid-term one.

In Question 12 (Appendix E.8), fourteen students (42.4%) complained that the volume was too low. Testing time in both exams was sufficient for 63.6% of students in the mid-term exam, and nearly three quarters of the students (72.7%) in the final exam. Test and task instructions in the final exam were improved in the sense that the number of the students who said the instruction was clear increased from 54.5% (mid-term exam) to 97% in the final exam. The proportions of the students who reported that the lengths of the texts in the test were similar to those used in class in both exams were similar – 94% in the mid-term exam and 87.9% in the final one. In the final exam, more than half of the students reported that multiple-choice questions were the easiest type of question (Table 6.30). Short-answer questions (i.e. for the monologue in the news report, the lecture, and the short talk) and cloze questions (i.e. for the song) were considered the most difficult types of question by more than three quarters of the students (Table 6.30).

Table 6.30 Case 2: Final Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam

Types of Questions	Easiest Question (Q13)		Most Difficult Question (Q14)	
	Frequency	Percent	Frequency	Percent
Short-answer	3	9%	15	45.5%
True/false	9	27.3%	4	12.1%
Multiple-choice	19	57.6%	3	9%
Cloze	2	6%	11	33.3%
Dictation	0	0%	0	0%
Total	33	100%	33	100%

6.7.4 Case 2: Students' General Comments on the Final Exam

All except three students stated that they had learned what they expected to learn after the mid-term exam (Question 15-a); Students 8 and 15 said that they did not think the teacher had an appropriate teaching plan, because what she had taught was so varied that it was hard for them to focus on particular topics. In addition, there was general satisfaction with the way the teacher taught after the mid-term exam (Question 15-b), though Students 8 and 15 again reported that the teacher

did not take the students' level of English or their feelings towards the in-class materials into consideration. After finishing this course, all but two students felt that their English listening skills had improved. In general, the majority of the students in both groups were satisfied with the teaching materials, teaching and assessment methods and felt positive about their improvement in English listening ability (Question 15-c).

6.8 Case 2: Final Interview with the Teacher

The final interview was conducted on the 16th of January 2007 after the final exam. The face-to-face interview again took place in the department office in the University and lasted approximately twenty minutes. In this interview, Miss T again chose to answer in Chinese. The interview was also tape recorded with permission, transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in the Appendix E.10. The translation was double-checked as before by a Chinese speaker who could speak both Chinese and English and by a native English speaker.

Table 6.31 Case 2: Final Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Miss T (T): It was also taken from the test battery of GEPT in the textbook and the extracts from <i>EZ Talk</i> magazines. In addition to the GEPT questions in the form of multiple-choice, I tested more short-answer questions this time – one part was a short talk, another was a news report, and the other was a lecture.</p> <p>Interviewer (I): Had your students listened to “lectures” during the term?</p> <p>T: No....but they had listened to my lecture, it was a way of listening to “lectures.”</p> <p>I: In the mid-term interview, you said that you would not ask them to listen to “lectures” because their English ability was not good enough to cope with the long length of “lectures”. But why did you test them with “lectures”?</p> <p>T: In fact this lecture was not very long, because I simply selected the first three paragraphs to test them, and I only designed two questions regarding the lecture which I thought were easy.</p> <p>I: What did you think of the students' performance this time?</p> <p>T: They did not perform well on the short-answer questions. Because there were too many short-answer questions, the students with lower English ability failed.</p>

In the final exam, the teacher used a different type of text – a lecture – that she had never tested in the mid-term exam or in the in-class quizzes (Table 6.31;

Question 1). Although the teacher claimed that the text of the lecture was not long and the questions were easy, she found that her students did not perform well in this part.

Table 6.32 Case 2: Final Survey – Interview with the Teacher (b)

<p>Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?</p>
<p>T: I hoped that they would perform better than they did in the mid-term exam, and they would be able to comprehend advanced listening content. That is, I hoped that their English ability was higher than it was in the mid-term exam. But they did not perform well in the final exam. It was possible that there were too many short-answer questions which accounted for a large proportion of the total score, so they did not score highly. I hoped that they would perform well this time because they made progress every week, though slowly....I think I may have used too difficult questions this time, so they failed to perform well.</p>
<p>I: How about the song for extra marks? How was the students' performance?</p>
<p>T: They did not perform very well on that either.</p>

The teacher expected her students would understand advanced English listening content after the mid-term exam, because she felt that they were improving, but they did not meet her expectation. The mean of the final scores was 61.1 which was slightly lower than the 69.8 in the mid-term exam. The teacher believed that the reason why the students did not perform well in the final exam was because listening extracts were too hard for them (Table 6.32; Question 2).

Table 6.33 Case 2: Final Survey – Interview with the Teacher (c)

<p>Question 3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria?</p>
<p>T: I defined the difficulty of the final exam based on the students' in-class performance this term, but it was possible that I used too difficult questions, so they could not understand the final exam.</p>

In the mid-term interview, the teacher said that she would increase the difficulty of the final exam, and she did use harder listening extracts in the final exam, but she found that her students could not always cope with them (Table 6.33; Question 3). However, it appeared that this did not result in better perceived performance on average, since more than half of the students were not satisfied with their own performance (Table 6.26).

Table 6.34 Case 2: Final Survey – Interview with the Teacher (d)

<p>Question 4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p>
<p>T: I increased the number of short-answer questions, because I felt that it would result in students guessing the answers if I used too many multiple-choice questions. Short-answer questions really tested whether students understood the questions or not. In fact most of them could not understand the passages, or else they answered the first one or two questions and then got lost in the following questions.</p>
<p>I: Was it possible that the listening extracts were too long?</p>
<p>T: Well, I found the length was OK. I felt that it was similar to what they had listened to in class.</p>
<p>I: Did you pilot the test items before the final exam?</p>
<p>T: No, I was busy at the end of the term, so I did not have time to pilot.</p>

In the final exam, Miss T believed that using short-answer questions really tested the students' understanding of test items and reduced the probability of guessing. Nevertheless, answering short-answer questions was not preferred by nearly two thirds of the students (Table 6.30). The comprehension questions that the students usually answered in the textbooks had multiple-choice and cloze formats; short-answer questions were only tested in the quizzes and the two exams. It was thus quite possible that the students could cope with the exercises in the textbooks without difficulty, but felt challenged while answering short-answer questions in both exams.

6.9 Summary

In Case 2, the features of task-based instruction in this class were not very noticeable. Though the teaching was implemented in terms of different activities in the textbooks and based on the news reports, the opportunities for students to interact with the teacher and the other classmates were limited to pair discussions between students. The only problem-solving tasks involved finding, in pairs, colloquial language in the listening extracts in the textbooks and the news reports. However, the students in class did not respond actively to the teacher's questions or attempts at discussion, it was very hard to know whether they understood the teaching or not, which only served to increase the difficulty of evaluating the

teaching effectiveness. The teacher's teaching experiences, moreover, may have influenced how she managed the classroom, encouraged her students, and interacted with them in the sense that she said that her lack of experience in teaching adults was a problem. The language in the listening materials was less authentic than language in real-life situations (see 6.3.1 and Appendix E.2 and E.3). Reflection periods at the end of the lessons were not employed. Thus, the answer to Question 1 for Case 2 is that a few features of task-based instruction were implemented in the situation of problem-solving tasks, pair discussions, and a focus on the meaning followed only subsequently by on the form.

Turning to test objectives, the teacher did not clearly specify what she wanted her students to have achieved in the two exams; the answer to Question 2 is that the test objectives were specified neither in the mid-term exam nor in the final exam. As with the test method in Case 1, only paper-based assessment was used in the two exams. The answer to Question 3 is accordingly that communicative language testing approaches were not used in either listening exam. In the mid-term exam, speech rate, monologue speech, and colloquial language were the three main problems for approximately half of the students. Topics became another problem for the students in the final exam, because the teacher said that she increased the number of test items for the monologue extracts. The answer to Question 4 in Case 2 is that speech rate, monologue-related topics, and colloquial language were the main problems for the students to understand.

The results of the mid-term test impacted on the teaching after the mid-term exam, in the sense that Miss T started to teach English songs to the students. The results of the mid-term exam also impacted on the difficulty of the final test content, because the teacher tested the students with more monologue-related topics. However, due to a large discrepancy in the item difficulty levels between

the two exams, it was very hard to compare students' progress based on their test scores. The answer to Question 5 is thus that although there was washback on the teaching and testing based on the mid-term results, it was hard to know if the harder test items in the final exam reflected positive washback on the teaching after the mid-term exam, since the differing difficulty level of the two exams made it hard to compare students' progress.

In the two cases, the two teachers – Dr. N and Miss T – used different teaching materials and assessment methods to teach their students. The impact of the mid-term exam results on the teaching of listening and the final exam was also different between the two groups. Though the two groups of students were in the same year of the study and registered for the same course title, they received very different courses in terms of teaching approaches, in-class listening materials, listening tasks and assessment methods. The results of the mid-term exams also had different impacts on teaching after the mid-term and the design of the final exam.

Teaching Methods and Classroom Activities

In Case 1, Dr. N's listening course focused on the understanding of everyday English and academic lectures, while Miss T in Case 2 put more emphasis on teaching listening for examination purposes (GEPT) and current topics in political or socio-economic news reports. While Dr. N encouraged her students to answer questions individually by giving extra marks, Miss T tended to use pair discussions in class. The classroom observations showed that, the students in Case 1 were more willing to interact with the teacher than those in Case 2. The listening activities in Case 1 appeared to be more diverse, in the sense that group presentation and drama appreciation increased the opportunities for the students to

practice and be familiar with authentic English, rather than simply listening to textbooks. The listening activities in Case 2 only involved doing exercises from textbooks or based on news reports. The observations also showed that students in neither group were given any opportunities to reflect on what they had learned and how well they had done on the course. It was thus hard to know if the students really understood what they had learned in class or not. This might be the reason why the students did not perform as well as their teachers expected, since they might simply not have learned what had been taught.

Classroom Assessment

When it came to assessment, Dr. N tended to use multiple-choice, true/false, or cloze tests to assess students' listening abilities. Miss T, however, preferred to use short-answer questions. Speech rate was a serious problem for both groups of students, though, topics, accents, vocabulary, and sentences in the mid-term exams were considered comprehensible by approximately three quarters of the students in both groups. Test characteristics such as the test instructions, the testing time, and the length of the listening passages were considered acceptable for the majority of students. However, the quality of the testing equipment needed to be improved. The students appeared to be sensitive to the types of comprehension question used and the way in which these questions were asked. The two teachers claimed that their students had difficulties in answering items modified by the teachers. The mean scores in both exams in both groups showed that Case 1 students scored higher than Case 2 ones, but it was very difficult to judge their comparative language abilities, since they did not use the same in-class materials or did not take the same tests.

Impact of the Test on Teaching and Testing

In both groups, the mid-term exam impacted on the teaching of listening in the second half of the term. Dr. N (Case 1) checked whether her students understood the situation comedy by asking them several questions regarding the contents. She also gave students scripts of academic lectures in order to help them understand the academic listening texts. Similarly, Miss T incorporated a new task after the mid-term exam – listening to English songs. However, neither Dr. N nor Miss T used any direct test, i.e. oral test, to test students' listening skills; that is, there were only one-way inputs of the test method rather than two-way interactions. Therefore, in the two cases, it was impossible to know if using direct tests would have resulted in more positive washback than indirect tests.

Chapter Seven

Data Analysis – Case 3

7.1 Introduction

Chapters Seven and Eight cover the analyses of English listening courses from two groups of students from University B. University B is a private university in southern Taiwan which includes four schools – Nursing, Medical and Health Sciences, Environmental and Life Sciences, and Humanities and Management. It comprises three different systems with regard to the length of study: (1) the “four-year system” requires four years of study for students who have finished their secondary education; (2) the “two-year system” requires a two-year programme for those who have graduated from five-year junior colleges, and (3) the “five-year junior college system” requires five years of study for students who have finished junior high school education¹. In University B, there were 48 students in total from the Department of Foreign Languages taking the “*English Listening and Speaking Practice*” course. The 48 students were divided into two separate classes and were taught by two different teachers. The department provided two teachers for the course, and the students, who were in the first year of the “two-year system” study, could choose the class led by either teacher.

Case 3 from University B is analysed in this chapter while Case 4 is discussed in Chapter Eight. Firstly, the descriptions of the course, the teacher and the in-class materials are introduced in Section 7.2. The teaching situations before the mid-term exams are illustrated in Section 7.3, and students’ opinions about the teaching materials and the mid-term exam are reported in Section 7.4. The mid-term interview with the teacher is presented in Section 7.5, and Section 7.6 looks at the in-class situation after the mid-term exam. The results of the final

questionnaire are analysed in Section 7.7. Finally, the final interview with the teacher is discussed in Section 7.8.

7.2 Case 3: Background of the Listening Course

There were 23 students and a teacher in Case 3. The teacher (Dr. C) had obtained a PhD degree in Research in Education at a university in the UK and had had ten year's experience of teaching undergraduate and college students. The students in her class were first-year undergraduate students of the "two-year" university programme who specialised in Foreign Languages. According to Dr. C's syllabus, the course was aimed at

Advanced students who wish to improve their listening and speaking skills. The course comprises integrated coverage and development of oral and aural skills. Reproduction of the text contents into role play format is one of the routine tasks of the course. It is hoped that by the end of the semester, students are able to listen to live broadcasting and express their opinions afterward.

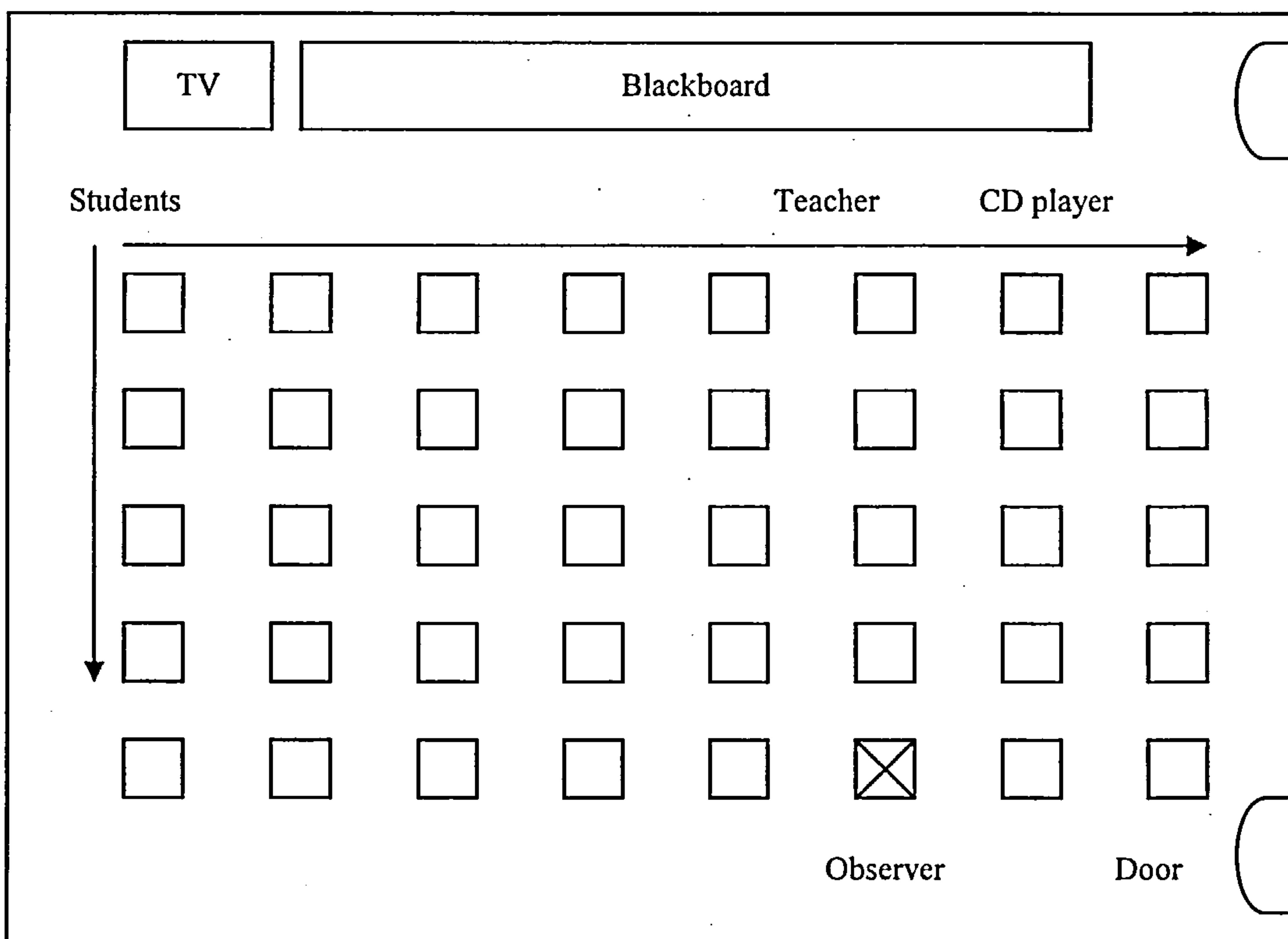
(Extract from Dr. C's syllabus notes)

From Dr. C's syllabus, it was clear that speaking skills were emphasised, and these were developed via oral reproductions of the listening texts and role play formats. The teacher only used one textbook – *Way Ahead* – in class (Appendix F.1). According to the Preface in the textbook (Sampson, 1999), the listening dialogues inside the book involved "everyday language in a wide variety of real-life settings and situations", and provided "guided practice in using many conversational functions and strategies" based on the listening dialogues. Three observations were carried out before the mid-term exam, from 13:10 to 15:00 on the 26th of October, the 2nd and 9th of November 2006. Researching the first group lasted 12 weeks from the 26th of October 2006 to the 19th of January 2007. Each lesson lasted 100 minutes with a ten-minute break in the middle.

7.3 Case 3: Classroom Observation before the Mid-term Exam

The students had their lessons in an ordinary classroom which was not specifically designed for English listening classes; noises outside the classroom could be heard and it was particularly noisy outside at the beginning and the end of each lesson. The students were situated in the classroom without personal headsets or microphones; there was instead a tape recorder, which was controlled by the teacher, and a TV. Exams also took place in this classroom.

Figure 7.1 Case 3: The Layout of the Classroom



7.3.1 Case 3: First Observation

The first observation was carried out on the 26th of October 2006. Dr. C told the students that she would finish Chapter 3 in *Way Ahead* that day. At the beginning of the class, it was noisy outside the classroom because students from other courses were talking in the corridor. Dr. C ignored the noise and told the students to answer the questions in the textbook and to discuss the sight worth seeing in

their home town with each other. After ten-minutes of discussion, the teacher nominated four students to share their experiences. The teacher then moved on to the listening part; she played the CD twice and the students listened to the dialogues. The listening dialogues were included in the textbook (Appendix F.1), so the students could listen while reading the passages. While listening to the dialogues, the teacher did not tell the students to focus either on the meaning or on the form. Next, she divided the group into two halves, and asked them to read and reproduce the dialogues in the listening passages, in the sense that half of the students pretended to be Speaker A, and the other half pretended to be Speaker B, then did a swap and repeated it again. After reproducing the dialogue, the teacher went on teaching the sentence patterns the students heard in the dialogues.

In the second half of the lesson, Dr. C moved on to the next listening dialogue. Students listened to it twice, then read and reproduced it in the textbook. Again, the students were divided into pairs to practice the sentence patterns. Next, the teacher told the students to do the exercises in the textbook while listening. She then nominated students to answer the questions and explained new vocabulary and phrases in the passages. Before the end of the lesson, the teacher nominated a pair of students to present the dialogue in Chapter Two which had been taught in the previous lesson. The two students read the dialogue in the textbook in front of the students. In this lesson, the students were given opportunities to discuss in pairs. They did not report their findings in groups, but answered individually when nominated by the teacher. In addition, the students were asked to reproduce the listening texts by reading them aloud. With respect to authenticity, the listening dialogues appeared to be fairly inauthentic because the conversations between the two interlocutors were very fluent, without any

interruptions or broken sentences (Appendix F.1). Finally, the teacher did not give students time to reflect on what they had learned at the end of the lesson.

7.3.2 Case 3: Second Observation

The second observation was carried out on the 2nd of November 2006. The two main components comprising this lesson were the teaching of Chapter 4 in *Way Ahead* and student presentations. At the beginning of the lesson, again, there was noise outside the classroom. Dr. C ignored the noise and told the students to discuss their own experiences of “Shopping” in pairs, and nominated three students to share their stories. The teacher then moved on to listen to the dialogues in the textbook. After listening twice, she explained the new vocabulary and phrases in the listening passages, and then divided the students into two halves to reproduce the dialogues. Next, the teacher asked the students to practice the sentence patterns in pairs. In the second half of the lesson, the teacher moved on to the next listening dialogues. Again, the students practiced the listening passages by reproducing the sentences, and then they were told to do the exercises while listening to the passages. After finishing Chapter 4, two pairs of students shared two English pop songs with other students. The audience was given the lyrics to help them understand. While introducing and explaining the contents of the song, both of the pairs spoke English first, but found it hard to continue, so they changed to Chinese. The teacher did not say anything when they spoke Chinese. Before the end of the lesson, the teacher again asked another pair of students to practice the dialogues they had learned the previous week. The pair reproduced the dialogue by reading it from the textbook.

In this observation, opportunities for the students to speak English were mainly restricted to the reproduction of listening dialogues and sentence patterns.

Although a pair of students was asked to present a song in front of other students, Chinese was allowed to be used during presentations, which did nothing to help improve students' English speaking abilities. The classroom was quiet; the students, in this class, were not particularly active in interacting with the teacher; they only interacted when told to do so. However, at the same time the teacher did not give students much of a chance to interact with her, because she controlled the initiation of communication in the class. The environment of the classroom was not suitable for an English listening class because the noises outside the classroom frequently drowned out the volume of the tape recording. The students read and repeated the sentences without any difficulty or questions, and the in-class teaching material appeared to be comprehensible for them.

7.3.3 Case 3: Third Observation

The third observation was carried out on the 9th of November 2006. The main activities in this lesson were Chapter 5 in *Way Ahead* and film appreciation. The procedure and the teaching situations were similar to what was found in the first observation (see Section 7.3.1 and 7.3.2). The teacher did not finish Chapter 5 in the first half of the lesson. In the second half, the students watched the film "*Harry Potter and the Prisoner of Azkaban*" without any subtitles. She briefly introduced the background of the film and then sat at the back of the classroom watching it with the students. The students appeared to concentrate on the film, but they only watched part of it because it was longer than the allocated time. She told them that they would finish the film after the mid-term exam.

At the end of the film, no questions for discussion or as tests were raised; the students simply enjoyed the film. The language in the film appeared to be more authentic than that in *Way Ahead*, in the sense that the students could not

only listen at native English speaker speed but also see the body language used by the actors. No problem-solving tasks were used and no reflection period was observed in any of the three observations.

Table 7.1 Case 3: Occurrence of Task-based Instruction Features before the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	1 st	✗	There was no problem-solving task in the class.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st	✓✗	The students were told to reproduce the listening dialogues in the textbook. The teacher divided the students into pairs to practice the sentence patterns in the textbook. Students answered the questions when nominated. Although practicing the dialogue taught in the previous lesson was an opportunity for students to speak English, only one pair of students was selected to present it.
	2 nd	✓✗	The same as above.
	3 rd	✓✗	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	✗	Students answered the questions individually.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
4. Authentic texts which reflect a real-life situation are used.	1 st	✗	The listening extracts in <i>Way Ahead</i> were considered less authentic than the language in real-life situations.
	2 nd	✓✗	The language in student presentation was more authentic than the dialogues in <i>Way Ahead</i> , but the students tended to use Chinese when they could not express their ideas properly.
	3 rd	✓✗	The language in the " <i>Harry Potter</i> " film appeared to be more authentic than that in <i>Way Ahead</i> .
5. The major focus of teaching is on the meaning, and then on the form.	1 st	✗	The teacher did not specify what students should focus on listening.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	✗	No reflection period was observed.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.

7.4 Case 3: Questionnaire Survey on the Mid-term Exam

The students took the mid-term exam in the same classroom in which they had the English listening lessons on the 16th of November 2006. The test comprised 86 questions and an extra bonus question; however, only fourteen multiple-choice questions and one bonus short-answer question out of 86 related to listening comprehension questions (see Parts IX and X, Appendix F.2), the remaining 71 were reading comprehension questions, including matching, cloze, and multiple-choice items (see Parts I to VIII, Appendix F.2). The test items in this exam were not piloted; the teacher said that she did not have time to pilot the test items because she needed to prepare exams for other courses she taught. The students answered the reading comprehension questions for 40 minutes and then moved on to the listening questions (see Section 7.5). The teacher only played the tape once, which lasted fifteen minutes; replays were not allowed. The students were allowed to finish the reading part after finishing the listening part, and the exam lasted 90 minutes in total. Though the course was entitled *English Listening and Speaking Practice*, the students simply answered fifteen listening questions for fifteen minutes, but spent more than one hour completing the reading comprehension questions, which had nothing to do with testing their listening ability. After the exam, the questionnaire was group-administered to the students. Twenty students were female, and three were male; there were no missing responses (Table 7.2). Dr. C was not in the classroom while the students completed the questionnaire. It took approximately 20 minutes for the group to finish as anticipated.

Table 7.2 Case 3: Mid-term Survey – Gender of the Respondents

Gender	Frequency	Percent
Male	3	13%
Female	20	87%
Total	23	100%

7.4.1 Case 3: The Teaching Materials and the Course before the Mid-term Exam

In Question 4, none of the students found the in-class listening materials “always” difficult, and only two (8.7%) regarded the textbook as “frequently” difficult to understand (Table 7.3). For the eighteen students who felt the textbook was hard at points, speech rate and accent were the two main problems for approximately half (Table 7.4). Vocabulary and colloquial language were also listed as minor problems.

TABLE 7.3 Case 3: Mid-term Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	5	21.7%
Rarely	7	30.4%
Sometimes	9	39.1%
Frequently	2	8.7%
Always	0	0%
Total	23	100%

TABLE 7.4 Case 3: Mid-term Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	0
Vocabulary	7
Accent	11
Speech rate	14
The use of colloquial language	6
Sound quality	1
Text type	2

In Question 5, the students all said that the course contents so far were relevant to their listening needs. Nearly a third also said that the topics of the textbook were related to their everyday life, which was very useful. The other five students said that they did not have particular learning needs; they were taking the course because it was compulsory, but they were satisfied with the teaching contents because they were easy to understand.

7.4.2 Case 3: Students’ General Preferences about the English Listening Class

While doing listening activities, more than half of the students (56.5%) preferred answering in speaking mode to answering in writing mode (Table 7.5; Question 6). Six out of thirteen said that they could train their pronunciation, listening and speaking abilities at the same time by speaking answers. Another four claimed that it was faster to answer questions by speaking than by writing. In addition, all

but two students (91.3%) preferred listening to conversations than monologues (Table 7.6; Question 7), and seven stated that conversational type speech was more interactive and interesting than monologues. Another five said that conversations were useful and practical, in the sense that they could learn how to communicate with foreigners.

TABLE 7.5 Case 3: Mid-term Survey – Preferred Mode of Answering (Q6)

Mode	Frequency	Percent
Writing	10	43.5%
Speaking	13	56.5%
Total	23	100%

TABLE 7.6 Case 3: Mid-term Survey – Preferred Type of Speech (Q7)

Types of Speech	Frequency	Percent
Monologue	2	8.7%
Conversations/dialogues	21	91.3%
Total	23	100%

More than half of the students (52.2%) preferred to be tested with multiple-choice questions (Table 7.7; Question 8). Seven of them, unsurprisingly, reported that pre-determined answers in multiple-choice items helped them decide the correct answer. Another three said that multiple-choice questions were very common and typical in language tests which they were familiar with. Short-answer questions, again, remained the least preferred type of question by more than half of the students (Question 9). Six of them reported that even though they understood the listening passages, it was sometimes hard for them to figure out the answers in their own words.

Table 7.7 Case 3: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class

Types of Questions	Preferred Question (Q8)		Least Preferred Question (Q9)	
	Frequency	Percent	Frequency	Percent
Short-answer	3	13%	13	56.5%
True/false	5	21.7%	0	0%
Multiple-choice	12	52.2%	1	4.3%
Cloze	3	13%	0	0%
Dictation	0	0%	9	39.1%
Total	23	100%	23	100%

7.4.3 Case 3: Questions about the Mid-term Exam

Nearly half of the students (47.8%) were “satisfied” with their performance in the mid-term exam (Table 7.8; Question 10); on average, it appeared that they felt that they performed well in the exam. However, for those who were “(very) dissatisfied” with their mid-term performance, it was hard to know whether they did not perform well in the listening part or the reading part, or both. It was also possible that students’ performance on one part could influence their performance on the other.

TABLE 7.8 Case 3: Mid-term Survey – Performance in the Mid-term Exam (Q10)



Satisfaction with the Mid-term Exam	Frequency	Percent
Face 1 – Very satisfied	0	0%
Face 2 – Satisfied	11	47.8%
Face 3 – Neither Satisfied nor dissatisfied	6	26.1%
Face 4 – Dissatisfied	5	21.7%
Face 5 – Very dissatisfied	1	4.3%
Total	23	100%

Only five students (21.7%) “agreed” that the test topics in the mid-term exam were difficult (Table 7.9; Question 11-a). More than a third “(strongly) agreed” that the test tasks in the exam were harder than those used in class (Table 7.9; Question 11-c). However, nearly half of the students (43.5%) “(strongly) agreed” that the topics of the test were representative of what had been taught in class.

Table 7.9 Case 3: Mid-term Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q11-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q11-b)		Test Tasks Were Harder than Those in Class (Q11-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	0	0%	0	0%
D	7	30.4%	1	4.3%	3	13%
Neither A nor D	11	47.8%	12	52.2%	12	52.2%
A	5	21.7%	8	34.8%	7	30.4%
SA	0	0%	2	8.7%	1	4.3%
Total	23	100%	23	100%	23	100%

Less than four students “agreed” that the accent was too strong, the vocabulary was too hard, or the sentences were too complicated to understand (Table 7.10). It thus seemed that, in the mid-term exam, accents, vocabulary, and sentence structures were comprehensible for the majority of the students.

Table 7.10 Case 3: Mid-term Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q11-d)		Vocabulary was too Difficult to Understand (Q11-e)		Sentences were too Complicated to Understand (Q11-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	2	8.7%	1	4.3%	1	4.3%
D	7	30.4%	11	47.8%	11	47.8%
Neither A nor D	12	52.2%	10	43.5%	8	34.8%
A	2	8.7%	1	4.3%	3	13%
SA	0	0%	0	0%	0	0%
Total	23	100%	23	100%	23	100%

Nevertheless, more than half of the students (52.1%) “(strongly) agreed” that it was hard to understand the listening passages because the speakers spoke too fast (Table 7.11; Question 11-g). In the listening part, a monologue and fourteen conversations were included. More than half of the students (52.2%) “(strongly) agreed” that it was hard to understand the monologue (Question 11-h), which was a news report. Conversations, on the other hand, seemed to be easier than the monologue, since only four students found it hard to understand them (Question 11-i).

Table 7.11 Case 3: Mid-term Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q11-g)		Monologue Speech Hard to Understand (Q11-h)		Conversations Hard to Understand (Q11-i)	
	Frequency	%	Frequency	%	Frequency	%
SD	1	4.3%	0	0%	1	4.3%
D	4	17.4%	3	13%	9	39.1%
Neither A nor D	6	26.1%	8	34.8%	9	39.1%
A	9	39.1%	10	43.5%	3	13%
SA	3	13%	2	8.7%	1	4.3%
Total	23	100%	23	100%	23	100%

In Question 12, only one student said that colloquial language was used, and he found this hard to understand. To summarise, speech rate in the mid-term exam was again a problem for this group of students, as well as in Cases One and Two. Linguistic features such as accent, vocabulary, and sentence structures were, on the other hand, comprehensible.

When it came to the test characteristics, more than half of the students complained that the background noise outside the testing environment was too loud (Appendix F.4). (I noted in Section 7.3 that the environment outside the classroom was very noisy at the beginning and the end of the lessons.) In addition, five students pointed out that the volume from the tape recorder was too low and the sound quality was very poor, so the noisy environment and the poor sound quality might well have contributed to poor performance. The testing time was sufficient for half of the students (52.2%), and the test instructions were clear for nearly three quarters of them (73.9%). Almost three quarters also reported that the lengths of the texts in the test were similar to those they had listened to in class. Since the conversation part was taken from the textbook exercises, it was reasonable to assume that the students were used to the length of the conversations in the mid-term exam. All but two students found that the multiple-choice questions were the easiest type of test item (Table 7.12; Question 15); however, they found it hard to answer the short-answer questions (Table 7.12; Question 16). Although the title of this course was *English Listening and Speaking Practice*, no speaking tasks were included in the mid-term exam; this ran counter to the claim that oral skills were included in the teaching syllabus (see Section 7.2).

Table 7.12 Case 3: Mid-term Survey – Easiest vs. Most Difficult Type of Comprehension

Question in the Mid-term Exam

Types of Questions	Easiest Question (Q14)		Most Difficult Question (Q15)	
	Frequency	Percent	Frequency	Percent
Short-answer	2	8.7%	21	91.3%
True/false	0	0%	0	0%
Multiple-choice	21	91.3%	2	8.7%
Cloze	0	0%	0	0%
Dictation	0	0%	0	0%
Total	23	100%	23	100%

7.4.4 Case 3: Students' General Comments on the Mid-term Exam

All but one student stated that they had learned what they expected to learn before the mid-term exam. Student 5 pointed out that the textbook was too easy for her. All students were satisfied with the way the teacher taught in class. All but two students were satisfied with the assessment method the teacher used in the exam; these two students were concerned that the teacher used too many test items in the reading part, which made it hard for them to answer all questions properly.

7.5 Case 3: Mid-term Interview with the Teacher

The face-to-face interview was conducted on the 23rd of November 2006, one week after the mid-term exam. The interview took place in a classroom in the University and lasted approximately twenty minutes. The interview questions were presented in both English and Chinese; Dr. C chose to answer in Chinese. The interview was tape recorded with her permission, transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in Appendix F.5. The translation was, as before, double-checked by a Chinese speaker who could speak both Chinese and English and a native English speaker.

Table 7.13 Case 3: Mid-term Survey – Interview with the Teacher (a)

Question 1. Where did the content of the test come from?
Teacher (C): It was taken from the teacher's manual.
Interviewer (I): In addition to the listening questions, you also tested students with reading comprehension questions. Were the reading questions also from the teacher's manual?

C: Yes, the reading part was taken from the teacher's manual, it was related to the teaching contents in the textbook, but it was in the "Practice" part. I was going to use the reading questions as in-class exercise activities for students, but the class schedule was so tight that I did not have time to use it, that's why I tested them with those questions in the mid-term exam. The listening part was from the textbook.

I: Was it possible that they got access to the listening questions before the mid-term exam?

C: Yes, they could have done, because the test items were in the textbook, but they could only see the test items without listening to the questions.

I: Did they know that you used the questions from the textbook?

C: No, they did not, so the probability that they noticed the fact was very low.

I: Where did the extra bonus (see Section 7.4) news report come from?

C: It was a news excerpt from ICRT (International Community Radio Taipei) radio.

I: I found that you did not use any speaking test in the mid-term exam. Could you tell me why you did not test the students' speaking ability?

C: uh...I had already given them too many reading questions, which taken up too much testing time, so there was no more time for a speaking test.

I: Did you pilot the test items, I mean run a small test, before you tested your students in this exam?

C: No, I had to prepare exams for other courses I taught. I did have time to run a pilot.

In Table 7.13, Dr. C explained the reason why she tested the students with the reading comprehension questions in the mid-term exam – she tested the reading questions which she had planned to use as in-class exercises. However, using reading tests to assess the students' listening ability threatened the content and construct validity of the test, since the reading tests could not test listening ability. Although the teacher claimed that the reading test was related to the teaching contents based on the textbook and the teacher's manual, assessing listening ability with inappropriate test items placed a large burden on the students, as they had to spend disproportionate time concentrating on the items (see Section 7.4.4). It would have been preferable to have direct tests to test students' speaking skills, since this would also have tested students' listening ability.

Table 7.14 Case 3: Mid-term Survey – Interview with the Teacher (b)

Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score for the course?

C: 60. The mid-term and the final scores all accounted for 30% of the total score.

I: Were there any quizzes?

C: No, the students needed to do a presentation, which accounted for their in-class participation. I arranged three pairs of students to practice the dialogues in the textbook before the end of every lesson. And I chose a pair randomly to present the dialogue they practiced.

I: Could they read the book while presenting?

C: Yes they could.

Quizzes were not used in Dr. C's class; presentations of dialogues in the textbook accounted for students' in-class participation mark (Table 7.14; Question 2). Although reading the dialogues in the textbook was taken to constitute giving a presentation in class, it was more like a presentation of pronunciation than a presentation of ideas using the students' own language.

Table 7.15 Case 3: Mid-term Survey – Interview with the Teacher (c)

<p>Question 3. What did you expect the students to have learned from your class?</p> <p>C: I hope that my students learned speaking skills while learning listening, because listening and speaking are connected. In addition to understanding the listening contents in the textbook, I also expect that they can understand the English on ICRT radio.</p> <p>I: Did you give any ICRT listening materials to the students in class?</p> <p>C: No, I focused on the textbook in class, but I asked them to listen to ICRT after the class.</p> <p>I: So, you would include ICRT listening passages in the mid-term or final exams.</p> <p>C: Yes, but only one or two questions.</p>

The students were asked to listen to the news reports from ICRT radio spontaneously after class, but listening to the radio was not included as a part of the in-class teaching materials (Table 7.15, Question 3). However, it was hard to know whether the students listened to the radio or not, or how far they understood the language on the radio, as the teacher did not discuss this with them regularly. The teacher concluded that the outcome of asking the students to listen to ICRT radio after class was not satisfactory, as only some of them were able to understand the news report in the mid-term exam (Question 5 below).

Table 7.16 Case 3: Mid-term Survey – Interview with the Teacher (d)

<p>Question 4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?</p> <p>C: 80% of the test objective was to know how much they understand in terms of the in-class teaching material, the other 20% was to establish their level of English proficiency.</p> <p>I: Do you mean that your mid-term exam had two different purposes?</p> <p>C: Yes.</p> <p>I: How did you decide that 80% was to assess the understanding of the textbook and the other 20% was for their level of English proficiency in terms of the test questions?</p> <p>C:(silence).....well...I can tell from their marks. If they scored higher, it meant they understood more and their English ability was better, and vice versa.</p> <p>I: So, you believe that the students' marks told you everything about their understanding of the teaching material and their English level.</p> <p>C: Yes, but they simply understood 50% of ICRT news report, so I don't think I achieved the test objective completely.</p>
--

In Question 4, the teacher claimed that the mid-term exam had two different functions and purposes – achievement and proficiency – with different proportions (Table 7.16). However, she did not justify how she used the test items to interpret the students’ achievement on the mid-term exam specifically and their English proficiency in general. In fact, the purposes of achievement, or proficiency tests are very different, but the test items simply covered the teaching contents before the mid-term exam. At this point, it is hard to generalise about the students’ English proficiency by using a test which only covered the topics from the five chapters in the textbook.

Table 7.17 Case 3: Mid-term Survey – Interview with the Teacher (e)

<p>Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?</p> <p>C: The test questions were from the textbook. Because they understood the teaching material in class, they should have understood the test questions because they were from the textbook.</p> <p>I: Did you know the students’ level of English before you chose the teaching material?</p> <p>C: At the beginning of the course, I asked the students to introduce themselves in English in order to find out their level of English and then I chose the teaching material. The textbook I am using right now was easy for them, because if you want to teach listening, you need to choose materials they can understand.</p> <p>I: What did you think about the students’ performance in the mid-term exam?</p> <p>C: Their marks were lower than I expected; although the students understood the listening contents in class, they did not perform well if I changed to a different type of question, I mean...the reading comprehension questions. They understood approximately 70% of the test content. They could only understand the ICRT news report partially – approximately 50%. Though I asked them to listen, the outcome was not satisfactory for me.</p>
--

In Question 5, the teacher believed that the difficulty level of the teaching material should be lower than the students’ level of English so that they could handle it (Table 7.17). In this group, the mean of the mid-term scores was 63.7. Dr. C was not satisfied with the results of the mid-term exam, even though they could understand the materials used in class. The teacher claimed that the reason why they did not perform well might be that she used different types of test question (i.e. reading test items). Thus, using reading test items to test listening comprehension did appear to threaten the validity of the listening test as the test scores could not really be used to measure the students’ listening ability.

Table 7.18 Case 3: Mid-term Survey – Interview with the Teacher (f)

Question 6. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?
C: I did not choose the questions in particular. I simply used the test items in the Teacher's Manual and the textbook. For the listening part, I used multiple-choice items. They were allowed to listen to the conversations once, and they had to choose the best answer immediately after listening to the questions.
I: Why did not you use other types of question?
C: Because there was not enough testing time.

Dr. C, like Dr. N in Case 1, depended on the test questions that had been designed by the textbook publisher (Table 7.18; Question 6). Multiple-choice questions and a short-answer question were included in the listening part, but due to the limitation of testing time, the teacher was unable to use other types of question.

Table 7.19 Case 3: Mid-term Survey – Interview with the Teacher (g)

Question 7. Will the result of the mid-term examination impact on the teaching of the following half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)
C: I think I will increase the difficulty of the teaching content. Because they scored lower this time, they needed to be trained.
I: How about the final exam?
C: Perhaps it will be harder; the number of ICRT news questions will be increased.
I: Aren't you worried that the students might not be able to pass the course due to the harder test items?
C: I am not worried about it. If they can't pass, I am afraid they will just have to re-take the course.

The teacher claimed that she would increase the difficulty of the teaching content after the mid-term exam, and probably increase the difficulty of the final exam (Table 7.19; Question 7).

7.6 Case 3: Classroom Observation after the Mid-term Exam

Way Ahead continued to be used in class after the mid-term exam. Another three observations were carried out from 13:10 to 15:00 on the 7th and 21st of December 2006, and the 4th of January 2007.

7.6.1 Case 3: Fourth Observation

The first observation after the mid-term exam was carried out on the 7th of December 2006. The lesson involved teaching Chapter 7 in *Way Ahead* and watching a movie. At the beginning of the class, Dr. C told the students to discuss their experiences of doing part-time jobs with each other, and then nominated four students to share their working experiences. After the discussion, the students listened to the dialogue in the textbook twice, and then they were divided into halves to read it. The teacher played the dialogues again and told the students to do the exercises in the textbook. While discussing the answers, she again nominated the students to answer the questions. Next, the students were told to practice the sentence patterns in the textbook in pairs.

In the second half of the lesson, the teacher played the second half of the movie (without any subtitles) that the students had not finished before the mid-term exam – “*Harry Potter and the Prisoner of Azkaban*”. She sat in the back row watching the film with the students and they appeared to concentrate on it. The teacher did not give the students any tasks regarding the film. Compared with the language in the film, the dialogues in the textbook were, as argued in Section 7.4.1, less authentic.

7.6.2 Case 3: Fifth Observation

The fifth observation was carried out on the 21st of December 2006. The lesson involved Chapter Eight in *Way Ahead* and student presentations. The procedure for teaching Chapter Eight was similar to that which was observed in Sections 7.6.1, 7.3.1, and 7.3.2. After teaching Chapter Eight, a pair of the students was selected randomly by the teacher to practice the dialogue in Chapter Seven. The pair was again allowed to read the dialogue in the textbook in front of all students.

Next, two pairs of students presented and played two English pop songs to the rest of the class. The audience was given the lyrics in order to help them understand. The students tried to speak English but changed to Chinese when they did not know how to express their ideas.

7.6.3 Case 3: Final Observation

The final observation was undertaken on the 4th of January 2007. The activities in this lesson were to finish Chapter Ten and to have student presentations. Again, the procedure for teaching Chapter Ten was similar to the teaching observed in Sections 7.6.1, 7.3.1, and 7.3.2, where the students practiced their oral English by reproducing the dialogues and the sentence patterns in the textbook. The interactions between the teacher and the students occurred when they were nominated to answer the questions regarding the exercises in the textbook. Pair discussions and practice between students took up approximately twenty minutes of the lesson. Next, the teacher moved on to the student presentations, where a pair of students practiced the dialogues in Chapter Nine and the other two pairs presented two English pop songs in front of all students. Similarly, the presentations in this observation followed the same pattern in Section 7.6.2. In none of the six observations in this academic term, were any problem-solving tasks used.

Table 7.20 Case 3: Occurrence of Task-based Instruction Features after the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	4 th	✗	Problem-solving tasks were not used in class.
	5 th	✗	The same as above.
	6 th	✗	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve	4 th	✓✗	Reproductions of the listening dialogues and practice of the sentence patterns in the textbook were again observed after the mid-term exam. Students answered the questions when nominated.

problems/tasks.			Practicing the dialogues and the sentence patterns were the only opportunities in which all students could speak English in class. Presentations, on the other hand, only allowed a few pairs of students to speak English.
	5 th	✓ x	The same as above.
	6 th	✓ x	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	4 th	x	Students discussed tasks and practiced English in pairs, but answered questions individually.
	5 th	x	The same as above.
	6 th	x	The same as above.
4. Authentic texts which reflect a real-life situation are used.	4 th	✓ x	The language in the "Harry Potter" film appeared to be more authentic than the dialogues in <i>Way Ahead</i> .
	5 th	✓ x	The language in student presentation was more authentic than the dialogues in <i>Way Ahead</i> , but the students tended to use Chinese when they could not express their ideas properly.
	6 th	✓ x	The same as above.
5. The major focus of teaching is on the meaning, and then on the form.	4 th	x	The teacher did not specify what to focus on when listening.
	5 th	x	The same as above.
	6 th	x	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. a reflection period).	4 th	x	No reflection period was observed in this lesson.
	5 th	x	The same as above.
	6 th	x	The same as above.

To summarise, the teacher did not adjust the difficulty of the teaching after the mid-term exam, although she had claimed she would do so in the interview. The procedures for teaching and student presentations remained very similar to what was observed before the mid-term exam.

7.7 Case 3: Questionnaire Survey on the Final Exam

The students took the final exam on 19 January 2007. It consisted of 22 listening questions, involving dictation (sentence-practice based on the textbook and an English pop song), multiple-choice and short-answer questions (five ICRT news reports) (see Part A to C, Appendix F.3). The listening dictations accounted for

50% of the total score. However, Brindley (1998: 179) and Buck (2001: 77) consider that dictation makes too many demands on students' grammatical, spelling, and lexical abilities, and it was rare for tests to infer meaning in communicative contexts. Weir (1993) also argues that dictations are significantly different from oral conversations in real-life situations. In the final exam, the test items were again unpiloted, and the reason was the same as the one given for the mid-term exam (see Sections 7.8 or 7.5). However, a speaking test was included, where the students read a dialogue in the textbook individually to the teacher, and she marked the result in terms of the students' pronunciation, fluency, and intonation (see Part D, Appendix F.3). Dr. C played all the passages only once, and no replays were allowed. The exam lasted approximately 45 minutes. After the exam and Dr. C had left the class, the final questionnaire was group-administered to the students. It took them 20 minutes to finish, which was well within the anticipated time.

7.7.1 Case 3: The Teaching Materials and the Course after the Mid-term Exam

Only three students (13%) felt that the teaching material after the mid-term exam was "frequently" or "always" difficult (Table 7.21). My impression that the teacher did not adjust the difficulty of the teaching contents after the mid-term exam was supported, since the number of the students who regarded the textbook as difficult before and after the mid-term exam were very similar (Tables 7.3 and 7.21). Speech rate, vocabulary, accent, and colloquial language were again the main problems with the listening material for nineteen students after the mid-term exam (Tables 7.22 and 7.4).

TABLE 7.21 Case 3: Final Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	4	17.4%
Rarely	8	34.8%
Sometimes	8	34.8%
Frequently	2	8.7%
Always	1	4.3%
Total	23	100%

TABLE 7.22 Case 3: Final Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	1
Vocabulary	10
Accent	9
Speech rate	11
The use of colloquial language	9
Sound quality	3
Text type	4

7.7.2 Case 3: Students' General Preferences about the English Listening Class

Nearly two thirds of the students preferred to answer listening tasks in a speaking mode (Table 7.23); this proportion was slightly higher than it was in the mid-term exam (Table 7.5). However, there was still a high percentage of students (93.1%) who felt they could understand conversations better than monologues in listening passages (Table 7.24). The proportions for the preferred type of speech stayed the same before and after the mid-term exam (Table 7.6).

TABLE 7.23 Case 3: Final Survey – Preferred Mode of Answering (Q5)

Mode	Frequency	Percent
Writing	8	34.8%
Speaking	15	65.2%
Total	23	100%

TABLE 7.24 Case 3: Final Survey – Preferred Type of Speech (Q6)

Types of Speech	Frequency	Percent
Monologue	2	8.7%
Conversations/dialogues	21	91.3%
Total	23	100%

For more than half of the students (61%), multiple-choice items remained the most preferred type of comprehension question in the listening tasks after the mid-term exam (Table 7.25; Question 8). Short-answer items, on the other hand, were considered the least preferred type of question for more than two thirds of the students (69.9%) after the mid-term exam (Question 9).

Table 7.25 Case 3: Final Survey – Preferred vs. Least Preferred Type of Comprehension

Question in Class

Types of Questions	Preferred Question (Q7)		Least Preferred Question (Q8)	
	Frequency	Percent	Frequency	Percent
Short-answer	1	4.3%	16	69.6%
True/false	5	21.7%	0	0%
Multiple-choice	14	61%	2	8.7%
Cloze	3	13%	5	21.7%
Dictation	0	0%	0	0%
Total	23	100%	23	100%

7.7.3 Case 3: Opinions about the Final Exam

The number of students who were “satisfied” with their performance in the final exam went down to three (13%), compared with eleven (47.8%) in the mid-term exam (Table 7.26 and 7.8). Conversely, the proportion of students who were “(strongly) dissatisfied” with their performance in the final exam increased from 26.1% (mid-term exam) to 43.4% (final exam). In other words, fewer students were satisfied with their test results in the final exam, which supported the fact that the final exam was harder than the mid-term one.

TABLE 7.26 Case 3: Final Survey – Performance in the Final Exam (Q9)



Satisfaction with the Final Exam	Frequency	Percent
Face 1 – Very satisfied	0	0%
Face 2 – Satisfied	3	13%
Face 3 – Neither Satisfied nor dissatisfied	10	43.5%
Face 4 – Dissatisfied	7	30.4%
Face 5 – Very dissatisfied	3	13%
Total	23	100%

Compared with the 27.1% of the students who “(strongly) agreed” that the test topics in the mid-term exam were difficult, the proportion of the students who “(strongly) agreed” that the test topics in the final exam were difficult reached 60.8% (Tables 7.27 and 7.9). Nearly two thirds of the students (65.2%) “(strongly) agreed” that the test tasks were harder than those they did in class,

which almost doubled the number of the students in the mid-term exam (Tables 7.27 and 7.9). Although the final exam was harder than the mid-term one, the test tasks remained representative for ten students (43.4%) in both two exams (Tables 7.20 and 7.9).

Table 7.27 Case 3: Final Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q10-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q10-b)		Test Tasks Were Harder than Those in Class (Q10-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	0	0%	1	4.3%
D	2	8.7%	4	17.4%	1	4.3%
Neither A nor D	7	30.4%	9	39.1%	6	26.1%
A	9	39.1%	7	30.4%	9	39.1%
SA	5	21.7%	3	13%	6	26.1%
Total	23	100%	23	100%	23	100%

Just as the test tasks in the final exam were perceived as harder than the tasks used in the mid-term exam, so the accent, vocabulary, and sentence structures were also perceived harder. The number of the students who “(strongly) agreed” that the accent and the sentence structures in the final exam were hard to understand was treble that in the mid-term exam (Tables 7.28 and 7.10). In addition, the vocabulary in the listening passages appeared to be a big problem for almost half of the students because the number increased dramatically from 4.3% (mid-term) to 47.8% in the final exam (Tables 7.28 and 7.10).

Table 7.28 Case 3: Final Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q10-d)		Vocabulary was too Difficult to Understand (Q10-e)		Sentences were too Complicated to Understand (Q10-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	0	0%	0	0%	1	4.3%
D	2	8.7%	2	8.7%	4	17.4%
Neither A nor D	14	60.9%	10	43.5%	8	34.8%
A	7	30.4%	9	39.1%	6	26.1%
SA	0	0%	2	8.7%	4	17.4%
Total	23	100%	23	100%	23	100%

In addition to the vocabulary, the fast speech rate in the listening passages was again perceived as a serious problem by almost two thirds of the students (Table 7.29). In the final exam, only monologues were tested, which included dictations of sentences, an English song, and five news reports; no conversation was included. However, compared with the monologue in the mid-term exam, less than half of the students (43.5%) “(strongly) agreed” that the monologues in the final exam were hard to understand (Table 7.29).

Table 7.29 Case 3: Final Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q10-g)		Monologue Speech Hard to Understand (Q10-h)	
	Frequency	%	Frequency	%
SD	0	0%	1	4.3%
D	3	13%	1	4.3%
Neither A nor D	5	21.7%	11	47.8%
A	11	47.8%	6	26.1%
SA	4	17.4%	4	17.4%
Total	23	100%	23	100%

In Question 11, only six students reported that colloquial language was used, but this time none of them regarded it as a problem. As for the test characteristics in Question 12, nearly half of the students (47.8%) – double the number in the mid-term exam – complained that the sound quality was so poor that they couldn’t hear the text clearly (Appendix F.4). Testing time in the final exam was less problematic than it was in the mid-term exam, since all but four students now claimed that it was adequate. The test instructions were clear for almost all students; however, it appeared that the listening texts in the final exam were longer than those used in class, since the number of the students who found the lengths of the listening extracts the longer increased from 13% (mid-term exam) to 26.1% (final exam). In the final exam, multiple-choice questions were the easiest type of question for nearly three quarters of the students (74%) (Table

7.30; Question 13), while dictations and short-answer questions were found the most difficult for more than a third (Table 7.30; Question 14).

Table 7.30 Case 3: Final Survey – Easiest vs. Most Difficult Type of Comprehension Question in the Mid-term Exam

Types of Questions	Easiest Question (Q13)		Most Difficult Question (Q14)	
	Frequency	Percent	Frequency	Percent
Short-answer	3	13%	9	39.1%
True/false	0	0%	0	0%
Multiple-choice	17	74%	3	13%
Cloze	0	0%	0	0%
Dictation	3	13%	11	47.8%
Total	23	100%	23	100%

7.7.4 Case 3: Students' General Comments on the Final Exam

All but one student said that they had learned what they expected to learn after the mid-term exam (Question 15-a). Student 13 said that his English listening ability improved very slowly because his English ability was poor. All students said that they were satisfied with the way the teacher taught after the mid-term exam (Question 15-b). More than three quarters (78.3%) claimed that they were satisfied with the assessment method the teacher used in the final exam (Question 15-c); three of them pointed out that testing English listening with a song was very interesting for them. However, the remaining five (21.7%) complained that the final exam was so difficult that they felt frustrated. After taking the English listening course, all but one student reported that their English listening ability had improved over the term (Question 15-d). Student 22 suggested that the reproduction activities in class were very boring, and the presentations of the dialogues in the textbook were too rigid.

7.8 Case 3: Final Interview with the Teacher

The final interview was conducted on 19 January 2007 after the final exam. The interview was face-to-face and again took place in a classroom in the University; it

lasted approximately fifteen minutes. Dr. C again chose to answer in Chinese. The interview was also tape recorded with permission, transcribed into Chinese and then translated into English. Both Chinese and English transcriptions are included in the Appendix E.10. The translation was double-checked as before by a Chinese speaker who could speak both Chinese and English and by a native English speaker.

Table 7.31 Case 3: Final Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Dr. C (C): The “Dictation” was taken from the Teacher’s Manual, the English song was taken from an English teaching textbook, and the news reports were recorded from the ICRT radio by me.</p> <p>Interviewer (I): How about the oral test?</p> <p>C: I asked them to read a dialogue from the textbook which they had practiced in class.</p> <p>I: Did every student read the same dialogue?</p> <p>C: They read the dialogue I chose randomly from Chapter Six to Ten in the textbook.</p> <p>I: What dimensions did you look at while assessing their speaking ability?</p> <p>C: I put it in the final exam answer sheet (see Part D, Appendix F.3). I looked at three dimensions – pronunciation, fluency, and intonation.</p> <p>I: Did you use any rating scales with specific descriptions to mark the students’ proficiency in the three dimensions? For example, point 5 meant that their pronunciation was correct, point 4 assumed that they made a few mistakes on pronunciation but they were reasonably correct in general.</p> <p>C: I did not use a rating scale with specific descriptions....because it would have taken a lot of time to score it.</p> <p>I: How did you score their speaking ability?</p> <p>C: The more accurate their pronunciation was and the more fluently they spoke, the higher the scores they could get.</p> <p>I: Did you pilot the test items before the final exam?</p> <p>C: No. As I have told you for in the mid-term exam, I needed to prepare exams for other courses I taught.</p>
--

While marking the students’ English speaking abilities, the teacher knew what she would look at in terms of specific dimensions in the speaking test (Table 7.31; Question 1). However, she used a scoring system without any detailed descriptions designed by herself to mark their English proficiency (see Part D, Appendix F.3). This meant that marking the speaking test in the final exam was inevitably subjective and intuitive, in the sense that the teacher did not have any descriptions of speaking levels or bands by which to judge the students’ oral performance.

Table 7.32 Case 3: Final Survey – Interview with the Teacher (b)

<p>Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?</p>
<p>C: The final exam was an integrated test which tested what they had learned this term.</p>
<p>I: Were the teaching contents before the mid-term exam included in the final exam?</p>
<p>C: No. Take the English song in the final exam as an example, the students had presented different songs in class for the whole term, so they should have been familiar with listening to English songs. Similarly, I asked them to listen to the ICRT radio after class, and they should have listened for a term.</p>
<p>I: What did you think about their performance in the final exam?</p>
<p>C: I found it OK. I thought they would perform poorly in the ICRT news part, but many students answered the questions correctly – maybe I did not design too difficult items. I think they performed well in the speaking part. After all, they had practiced so many times in class and in presentations; many students scored highly in this part.</p>
<p>I: So you mean you are satisfied with their oral test scores?</p>
<p>C: Yes, those dialogues were not difficult for them to understand, and they had practiced them so many times, there should not be any reasons why they would not perform well!</p>

In the final exam, the mean of the final scores was 72, which was higher than the 63.7 in the mid-term exam. The teacher felt that her students performed better in the final exam (Table 7.32; Question 2). In general, the teacher appeared to be satisfied with the students' performance on the final exam, particularly in the speaking part. The majority of the students should have been able to cope with the listening dialogues in the textbook without difficulty, based on the teacher's descriptions of their oral performance.

Table 7.33 Case 3: Final Survey – Interview with the Teacher (c)

<p>Question 3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria?</p>
<p>C: Since the "Dictation" part was from the teacher's manual, the difficulty of the questions was similar to that in the textbook, and the majority of the students could understand the content; I think they could answer this part correctly. But this time I tested them with more ICRT news questions, I knew that it was challenging for them, so I designed easier test items so that they would not feel frustrated.</p>

The teacher decided on the difficulty of the test items based on the difficulty of the textbook used in class because she knew that the majority of the students understood the teaching contents (Table 7.33; Question 3). In addition, she increased the difficulty of the final exam by using more news reports, as she had said she would in the mid-term interview. However, she tried not to use too difficult test items while taking the students' level of English into account.

Table 7.34 Case 3: Final Survey – Interview with the Teacher (d)

<p>Question 4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p>
<p>C: The reason I tested them with dictation was that I have never used this type of test item, and I wanted to change to another type so that they would not be tested by the same types of questions all the time.</p>
<p>I: Weren't you worried that they might lose marks because they might not be familiar with dictation?</p>
<p>C: Dictation is the type of question where they write down what they listen to. This is a very easy type of question which I thought they would be familiar with.</p>
<p>I: How about the test question for the ICRT news reports?</p>
<p>C: Well....as I said before, it could be challenging to test their listening ability with news reports. If I designed too difficult test items, they might perform poorly. That's why I used multiple-choice questions, so the probability that they answered correctly would be higher.</p>

As for the test items, the teacher believed that her students would not have problems in answering dictations (Table 7.34; Question 4). She also believed that her students could score higher on multiple-choice items because they were easier to answer.

7.9 Summary

In the six lessons observed, a few features of task-based instructions were implemented. The dialogues in the textbook, however, were inauthentic with respect to real-life situations in that they were too fluent without any hesitations, false starts, or interruptions, so reproducing them cannot really be regarded as using authentic language in authentic target language situations. When the students presented English songs in class, speaking English was not compulsory. In addition, the opportunities for the students to speak English were limited to reproductions of listening passages in the textbook. The students appeared to understand the listening contents and exercises in the textbook, since they answered the questions and reproduced the dialogues without difficulty. Thus, the answer to Question 1 was that only a few features of task-based instruction were discovered in Case 3 in terms of watching English films and student presentations of English songs.

As for test objectives, there were no specific test objectives for the written tests in both the mid-term and final exams, since the teacher simply used the test bank in the teacher's manual in the mid-term. Besides, the testing contents in both exams were very different, in the sense that reading comprehension was included in the mid-term exam, and this did aimed neither to test the students' listening nor speaking abilities. This inclusion has been seen as seriously affecting the validity of the scores as indicators of listening ability. In addition to the written exams, the teacher had a specific "educational" objective for final oral exams, where she looked at students' pronunciation, fluency, and intonation. Thus, the answer to Question 2 is that there were no specific test objectives for the written exams, but there was an "educational" objective for the oral exam. However, although students' speaking skills were tested in the final exam, the oral test was much more a reading and pronunciation test than a communicative test that involved meaningful exchanges. Thus, the answer to Question 3 is that communicative language testing approaches were neither used in the mid-term nor in the final exams.

The speech rate and the monologues constituted major problems for more than half of the students in both exams. In the final exam, however, the number of students considered that the difficulty of vocabulary and sentences increased; it is very possible that Dr. C increased the number of monologue-related topics so that the students found them harder than those in the mid-term exam. The answer to Question 4 is that speech rate and monologue-related items were the two main problems for students in Case 3.

When it comes to ascertaining whether test results impacted on the teaching and testing or not, the results of the mid-term exam did impact on the difficulty of the final exam but did not impact on the teaching after the mid-term,

although the teacher claimed she had increased the difficulty level of the teaching content. Interestingly, although the final exam included more listening passages regarding the news reports than the mid-term, and the students felt that the final exam was harder than the mid-term one, they performed better on the final exam than the mid-term one. The teacher said that her students received higher scores in the speaking section, because the dialogues they spoke in the oral exam were practiced very frequently in class. Thus, the answer to Question 5 is that the results of the mid-term exam did not washback on the teaching after the exam, but it did influence how the final exam was designed. Although the students' speaking ability was tested in the final exam, it was more like a reading and pronunciation exercise that they usually practiced in class, rather than a direct test. Thus, it is doubtful whether the pronunciation tasks students did in class and in the final exam would be accurate prediction of their future oral communicative ability.

Notes to Chapter Seven:

¹ A junior high school is for pupils who are 13 to 15 years old.

Chapter Eight

Data Analysis – Case 4

8.1 Introduction

In Chapter Seven, the analyses of Case 3 were discussed. In this chapter, twenty-five students in the second group, taking the same English Listening Course – *English Listening and Speaking Practice* – in the same year of study, and from the same department as Case 3 participated in the study. The 25 students were taught by another teacher (Dr. D), and the in-class materials used were completely different from the materials used by Dr. C in Chapter Seven. Firstly, the background of the course, syllabus, and in-class materials are introduced in Section 8.2, and the teaching approaches and the interaction between the teacher and the students in class before the mid-term exam are summarised in Section 8.3. The results of the mid-term questionnaire are analysed in Section 8.4; Section 8.5 discusses the mid-term interview with the teacher. Section 8.6 presents the descriptions of the in-class situation after the mid-term exam; Section 8.7 looks at the students' opinions about the final exam. Finally, the interview with the teacher regarding the final exam is discussed in Section 8.8.

8.2 Case 4: Background to the Listening Course

There were 25 students and a teacher in Case 4. The teacher (Dr. D) was an American, obtained his PhD degree in Psychology at a university in the USA and had had six year's experience of teaching Taiwanese undergraduate and college students in Taiwan. The students in his class were also first-year undergraduate students of the "two-year" university programme who specialised in English Language. According to Dr. D's syllabus, the purpose of this course was to

bring intermediate and advanced English to bear in oral discourse. Speaking, vocabulary, and listening skills will be developed and reinforced through the production of original dialogues, information exchange, audio reviews, and by practicing model conversations in the text using the look-up-and-say¹ method. The mid-term and final exams are based on both written and oral measures.

(Extract from Dr. D's syllabus notes)

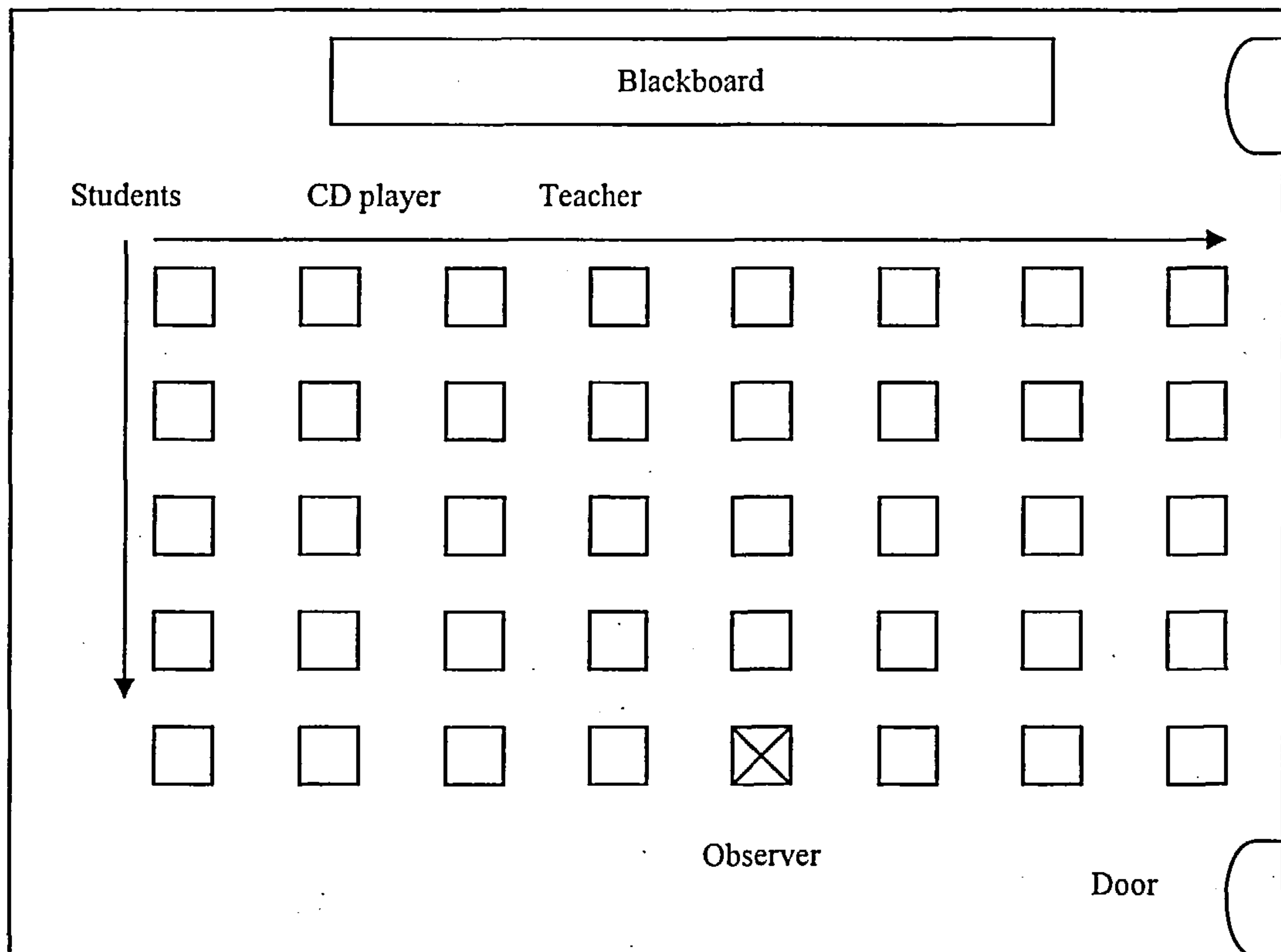
It was clear from above extract that Dr. D expected that his students would develop speaking, vocabulary and listening skills in this class. Reproductions of dialogues and model conversations in the textbook were also considered the main tasks in class. In this course, the teacher only used one textbook – *New Interchange* (Appendix G.1). According to the book's Introduction (Richards, 2001: iii), the listening activities included "listening to narratives, commercials, discussions, and activities" while the reading exercises "derive from authentic sources, and often reflect cross-cultural themes, exploring life-styles and values in different countries." The listening exercises in the textbook appeared to be based on everyday life topics. Three observations were carried out before the mid-term from 13:10 to 15:00 on the 27th of October, and the 3rd and 10th of November 2006. Researching the second group in Case 2 lasted 12 weeks, from the 27th of October 2006 to the 18th of January 2007. As with Case 3, each lesson lasted 100 minutes with a ten-minute break in the middle.

8.3 Case 4: Classroom Observation before the Mid-term Exam

The students took their lessons in an ordinary classroom which was not specifically designed for English listening classes; noises outside the classroom could be heard and it was very noisy outside at the beginning and the end of each lesson. The students were in the classroom without personal headsets,

microphones, or other audio-visual equipment; there was only a tape recorder controlled by the teacher. The exams also took place in this classroom.

Figure 8.1 Case 4: The Layout of the Classroom



8.3.1 Case 4: First Observation

The first observation took place on the 27th of October 2006. The main activity was to teach Unit 3 in *New Interchange*. At the beginning of the lesson, there was noise outside the classroom but the teacher ignored it. He shared his experiences of asking a favour from his friends and of doing a favour to help his friends. He then asked the students to share their experiences as well, but none of them raised their hands and shared anything. The teacher kept on asking the students three or four times, and finally one student did respond in English. He told all students that they needed to speak English so that they could practice it. After discussion, the teacher read the conversation in the textbook to the students and explained the gist, and then he played the CD twice and explained the vocabulary and phrases in

the conversation. Next, he moved on to the grammar and pronunciation activities in the textbook. He taught the grammar in the example and told them to do the exercises; the grammar focus in Unit 3 was on “if” clauses. While discussing the answers, only three or four students answered the questions after being asked by the teacher three or four times. When practicing the pronunciation, the students listened to the teacher pronouncing the practice sentences in the textbook, then they read them aloud together.

In the second half of the lesson, Dr. D played the conversation again, divided the students into pairs and told them to reproduce the conversation themselves. After ten minutes, the students were told to do the exercises in the textbook while listening to the CD. While being asked to answer the questions by the teacher, the students again did not answer them. He kept on asking the students to answer, and finally, two or three students answered. The teacher then moved on to the vocabulary exercises; he explained the vocabulary first and told the students to complete it.

In this observation, the students did not actively respond to the teacher’s questions. Though the teacher tried to involve his students in interactions, they remained quiet until they were told to read the conversations themselves or to answer the questions. While doing the exercises in the textbook, no group or pair discussions occurred and the students reported answers individually. There were no problem-solving tasks involved in the class. The teacher focused both on the meaning and on the grammar and pronunciation practice.

8.3.2 Case 4: Second Observation

The second observation was carried out on the 3rd of November 2006. The main activity in this lesson was to finish Unit 3, as it had not been finished the previous

week. At the beginning of the lesson, the teacher chatted with the students for around ten minutes and then started the lesson. He read another conversation, which was different from the one he had taught the previous week, but the topics of the two conversations were similar. After reading the conversation, he played the CD twice and told the students to reproduce the conversation themselves in pairs. After ten minutes, the teacher taught the grammar in the conversations and model sentences, and then asked the students to do the exercises in the textbook. While discussing the answers, none of the students raised their hands or responded actively to the teacher. After he had asked the question three or four times, two or three students gave answers. Next, the teacher told the students to read a short story regarding different cultural experiences and to complete the reading exercises in the textbook. However, this time he did not ask the students to answer the questions, but told them the answers.

In the second half of the lesson, the teacher started Unit 4 in *New Interchange*. He began by talking about which kind of news he liked to read every day and which news he was not very interested in. Then he asked the students about their favourite type of news and why they liked it. Again, the students kept quiet, but the teacher continued to try and involve them in the talk; after being asked three or four times, two students sitting in the front row shared their favourite type of news. Next, Dr. D read the conversation in the textbook and played the CD twice. He explained the gist of the conversation and moved on to the pronunciation activities in the textbook. He read the model sentences with clear pronunciation and told the students to read aloud together. The class did not finish Unit 4.

In short, the interaction between the teacher and the students was fairly minimal and the teacher had to try several times to elicit the answers from just two

or three students. The students again completed listening or reading exercises individually, and no problem-solving tasks were employed. Although the intonation and pronunciation in the model conversations were highlighted clearly, the language was again very fluent, without any interruptions or false starts, and thus not as authentic as the language used in real-life situations.

8.3.3 Case 4: Third Observation

The third observation was carried out on the 10th of November 2006. The main activity in this lesson was to finish Unit 4, as it had not been finished the previous week. At the beginning of the lesson, the teacher chatted with the students for ten minutes, and played the CD to remind them of the conversation in Unit 4 they had listened previously. Next, he explained the grammar in the conversation and told them to do the grammar exercises in the textbook; the grammar focus in Unit 4 was the “past continuous” and the “simple past”. After ten minutes, the teacher did not ask for answers from the students, but told them the answers. Next, he moved on to the second conversation in Unit 4. The procedure for teaching this conversation was similar to what was observed in Sections 8.3.1 and 8.3.2.

In the second half of the lesson, the teacher taught the grammar in the conversation and asked the students to do the exercises in the textbook. When it came to the discussion, the students again did not respond actively to their teacher; just two or three answered the questions after being asked three or four times. Next, the students were told to read four short news reports regarding tabloid stories and to complete the questions in the textbook. Once again, the teacher did not ask the students to answer the questions, but told them the answers. Although no tasks involving groups or pairs were used, reproductions of conversations in pairs did occur.

Table 8.1 Case 4: Occurrence of Task-based Instruction Features before the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	1 st	✗	No problem-solving tasks were used in class.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st	✓ ✗	The students were told to reproduce the conversations in the textbook in pairs. Students did not answer the questions actively; they answered after being asked three or four times.
	2 nd	✓ ✗	The same as above.
	3 rd	✓ ✗	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	✗	Students answered the questions individually. No discussion observed.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
4. Authentic texts which reflect a real-life situation are used.	1 st	✗	The listening extracts in <i>New Interchange</i> were considered less authentic than the language in real-life situations.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.
5. The major focus of teaching is on the meaning, and then on the form.	1 st	✓	The teacher focused on the meaning, and then on the grammar.
	2 nd	✓	The same as above.
	3 rd	✓	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	✗	No reflection period was observed.
	2 nd	✗	The same as above.
	3 rd	✗	The same as above.

To summarise, only a textbook was used in this class; there were no supplementary materials, such as English films or songs, as there were in Case 3. Similar to the other three cases reported in Chapters Five to Seven, the students were again given no opportunities to reflect on what they had learned or on how well they were doing, at the end of the lesson.

8.4 Case 4: Questionnaire Survey on the Mid-term Exam

The students took the mid-term exam on the 17th of November 2006 in the same classroom in which they had the English listening lessons. The test comprised 35

reading comprehension questions, including multiple-choice, sentence completion and rewriting, vocabulary correction, cloze, and grammar items (Appendix G.2), and a speaking test. No listening comprehension items were included; items in reading test were not piloted (see Section 8.5). It took the students 40 minutes to finish the reading items, and they then moved on to the speaking test. In the speaking test, the students reproduced the conversation in the textbook individually with the teacher; the conversations ranged from Units 1 to 4, which had been taught before the mid-term exam. In other words, the students and the teacher read a conversation in the textbook, and the teacher told me that the conversation was chosen randomly by him. The students were allowed to take the textbook with them, but it was suggested they should not “read” the sentences, but “say” them. Though what the teacher said in the conversations could be regarded as listening input for the students, it was hard to know if the students really paid attention to what the teacher had said and responded accordingly, since they could also read the same thing in the textbook. In this case, oral responses to what the teacher said became invalid if the students could read without listening. This group is the only group of the four in the present study who did not use any listening items in the mid-term exam; therefore, the analyses in Section 8.4.3 were based on what English the students had listened to (from their teacher) in the speaking test. After the exam, the questionnaire was group-administered to the students. Twenty-two students were female, and three were male (Table 8.2); there were, again, no missing responses. Dr. D was not in the classroom while the students completed the questionnaire. It took approximately 20 minutes for the group to finish, and they did not report any problems with filling it in.

Table 8.2 Case 4: Mid-term Survey – Gender of the Respondents

Gender	Frequency	Percent
Male	3	12%
Female	22	88%
Total	25	100%

8.4.1 Case 4: The Teaching Materials and the Course before the Mid-term Exam

In Question 4, none of the students found the in-class listening materials “always” or “frequently” difficult; around a quarter (20%) considered the listening contents “sometimes” difficult to understand (Table 8.3). For seventeen students who felt the textbook was hard at different points, less than eleven said that the speech rate and vocabulary were the two main problems (Table 8.4). The rapid speech rate made it difficult for approximately half of the students to comprehend the listening contents in both cases in University B.

TABLE 8.3 Case 4: Mid-term Survey – Perceived Difficulty of In-class Listening Materials (Q4)

	Frequency	Percent
Never	8	32%
Rarely	12	48%
Sometimes	5	20%
Frequently	0	0%
Always	0	0%
Total	25	100%

TABLE 8.4 Case 4: Mid-term Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	2
Vocabulary	10
Accent	6
Speech rate	11
The use of colloquial language	0
Sound quality	0
Text type	0

In Question 5, all except two students said that the course contents were relevant to their learning needs, and nearly a third claimed that they would need the listening skill after graduation. The other half said that the topics in each unit covered different types of lifestyle around the world, which they found highly interesting. The remaining students reported that they did not have specific needs for learning listening; they were simply taking the course because it was compulsory. In all four cases, only 16 (14%) of the 112 said they were following the course purely because it was compulsory. However, two students who did not

think the course contents relevant to their needs said that they hoped that the teacher would focus more on listening than speaking, because they wanted to improve their listening ability before moving onto speaking.

8.4.2 Case 4: Students' General Preferences about the English Listening Class

The proportion of the students in Case 4 (72%) who preferred to answer listening questions by speaking was higher than in Case 3 (56.5%) (Tables 8.5 and 7.5). A third of Case 4 students said that they did not have many opportunities to practice their oral English in other English classes; it appeared that they hoped that they could speak English in class. However, according to the classroom observation, the teacher had tried to involve the students in discussion, but the students themselves did not interact with the teacher actively. Although the MOE requirement implies that a communicative language learning environment needs to be established to provide students with opportunities to speak English and the teacher in Case 4 tried to establish a communicative environment, the implementation was not achieved. The other six students claimed that speaking out the answers prevented them from misspelling vocabulary. Conversational types of speech were regarded as the preferred type of listening texts for almost all except two to three students in both groups at University B (Tables 8.6 and 7.6). More than half of the students in Case 4 stated that it was easier to understand the content by guessing the conversations between interlocutors.

TABLE 8.5 Case 4: Mid-term Survey –

Preferred Mode of Answering (Q6)

Mode	Frequency	Percent
Writing	7	28%
Speaking	18	72%
Total	25	100%

TABLE 8.6 Case 4: Mid-term Survey –

Preferred Type of Speech (Q7)

Types of Speech	Frequency	Percent
Monologue	3	12%
Conversations/dialogues	22	88%
Total	25	100%

In both cases in University B, multiple-choice items were the favourite type of the listening comprehension question for more than half of the students (56%) (Table 8.7 and 7.7). Unsurprisingly, short-answer items remained the type of comprehension question more than half of the students preferred least in both groups (60%).

Table 8.7 Case 4: Mid-term Survey – Preferred vs. Least Preferred Type of Comprehension Question in Class

Types of Questions	Preferred Question (Q8)		Least Preferred Question (Q9)	
	Frequency	Percent	Frequency	Percent
Short-answer	4	16%	15	60%
True/false	7	28%	0	0%
Multiple-choice	14	56%	2	8%
Cloze	0	0%	3	12%
Dictation	0	0%	5	20%
Total	25	100%	25	100%

8.4.3 Case 4: Questions about the Mid-term Exam

In Case 4, 40% of the students were “(very) satisfied” with their performance in the mid-term exam (Table 8.8), which was close to the number of the students in Case 3 (47.8%) (Table 7.8). The proportion of students (20%) who were “(very) dissatisfied” with their performance in Case 4 was also close to the proportion in Case 3 (26%). Similar to the situation in Case 3, it was hard to know on which skill (reading, speaking, or both) the students performed well or poorly. It was also very possible that the students’ performance on the reading test affected their performance on speaking test.

TABLE 8.8 Case 4: Mid-term Survey – Performance in the Mid-term Exam (Q10)



Satisfaction with the Mid-term Exam	Frequency	Percent
Face 1 – Very satisfied	4	16%
Face 2 – Satisfied	6	24%
Face 3 – Neither Satisfied nor dissatisfied	10	40%
Face 4 – Dissatisfied	5	20%
Face 5 – Very dissatisfied	0	0%
Total	25	100%

Because no listening items were included in the mid-term exam, I told the students to answer Question 11 based on their opinions of the speaking test, before giving them the questionnaire. That is, the students' answers in Question 11 were analysed on the basis of their perceptions regarding their speaking test, as the speaking test still required some understanding of listening input. In this group, the listening input was from the teacher speaking. Thus, in Question 11, none of the students "agreed" that the test topics were difficult (Table 8.9; Question 11-a). This is unsurprisingly; since the topics of the speaking test were the conversations taught in the textbook, the students might be expected to be highly familiar with the topics. The topics of the test tasks in the Case 4 exam appeared to be more representative of what had been taught in class than those in the Case 3 mid-term exam, because more than two thirds (68%) "(strongly) agreed" that the topics in the speaking test were representative (Table 8.9; Question 11-b), compared with 43.5% of the students in Case 3 (Table 7.9; Question 11-b). The number of the students (12%) who "agreed" that the tasks were harder than those used in class in Case 4 (Table 8.9; Question 11-c) was lower than for Case 3 (34.7%) (Table 7.9; Question 11-c).

Table 8.9 Case 4: Mid-term Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q11-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q11-b)		Test Tasks Were Harder than Those in Class (Q11-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	3	12%	0	0%	2	8%
D	16	64%	2	8%	13	52%
Neither A nor D	6	24%	6	24%	7	28%
A	0	0%	12	48%	3	12%
SA	0	0%	5	20%	0	0%
Total	25	100%	25	100%	25	100%

Just three students in both two groups at University B "agreed" that accents, vocabulary, or sentences were hard to understand (Tables 8.10 and 7.10).

Table 8.10 Case 4: Mid-term Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q11-d)		Vocabulary was too Difficult to Understand (Q11-e)		Sentences were too Complicated to Understand (Q11-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	4	16%	5	20%	3	12%
D	15	60%	17	68%	15	60%
Neither A nor D	3	12%	2	8%	5	20%
A	3	12%	1	4%	2	8%
SA	0	0%	0	0%	0	0%
Total	25	100%	25	100%	25	100%

In Case 4, only three students (12%) “(strongly) agreed” that it was hard to understand the conversations because the teacher spoke fast (Table 8.11; Question 11-g), which was much fewer than the 52.1% of the students in Case 3 who “(strongly) agreed” that the speakers on the listening extracts spoke too fast (Table 7.11; Question 11-g). It was possible that Dr. D adjusted his speed of speaking while talking to individual students, but the speakers in the pre-recorded tape (for Case 3) clearly could not do so. That might explain why the majority of the students in Case 4 did not find Dr. D’s speech rate to be a problem.

Table 8.11 Case 4: Mid-term Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q11-g)	
	Frequency	%
SD	4	16%
D	13	52%
Neither A nor D	5	20%
A	3	12%
SA	0	0%
Total	25	100%

In Question 12, no students reported that colloquial language was used in the speaking test. So, overall in University B, colloquial language was not problematic. With regard to the test characteristics, since no tape recording or visual equipment was used in the mid-term exam, Question 13-1 was ignored. In Question 13-2, all students said that the testing time was sufficient for them to

answer all the questions properly, and the test instructions were clear (Appendix G.5). However, only half of the students in Case 3 said that the testing time was sufficient in the mid-term exam (Appendix F.4). Unsurprisingly, all students reported that the lengths of the texts in the test were similar to those they had listened to in class, since the texts were taken directly from the conversations in the textbook. In addition, since there were no listening comprehension questions in the mid-term exam for Case 4, Questions 15 and 16, which asked which type of the listening question the students found easiest and hardest, could not be included in the data analysis.

8.4.4 Case 4: Students' General Comments on the Mid-term Exam

In Question 17, students all said that they had learned what they expected to learn thus far, and they were satisfied with the way the teacher taught. Exceptionally, one student (Student 8) said that she had used the textbook previously in another class, but Dr. D taught it in a different way. She said that even though she was very familiar with the content, she was satisfied with the way Dr. D taught. All students were satisfied with the assessment method the teacher used in the mid-term exam.

8.5 Case 4: Mid-term Interview with the Teacher

The face-to-face interview was conducted on the 24th of November 2006, one week after the mid-term exam. The interview took place in the teacher's office in the University and lasted approximately twenty minutes. As Dr. D is a native speaker of English and cannot speak Chinese, the interview questions were presented in English. The interview was tape recorded with his permission.

Table 8.12 Case 4: Mid-term Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Dr. D: You had the textbook last week, the written exam was from the publisher of the book, so that was a test bank from the textbook.</p> <p>I: Do you use other textbooks?</p> <p>D: No, only the textbook.</p> <p>I: Was it possible for the students to get access to the test questions and answers before the test?</p> <p>D: No, it is the teacher edition.</p> <p>I: How about the speaking test?</p> <p>D: It was a reproduction test of the conversations in the textbook.</p>

Similar to the situation in Case 3, Dr. D also used the reading test from the Teacher’s edition, which was based on the textbook contents (Table 8.12; Question 1). The speaking test was based on reproducing the conversations in the textbook; the students should have been very familiar with the content, which might account for the higher average score.

Table 8.13 Case 4: Mid-term Survey – Interview with the Teacher (b)

<p>Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?</p> <p>D: 60 was the lowest passing score.</p> <p>I: How did you assess the students' oral ability in the exam?</p> <p>D: For the oral exam, students were rated from 0 to 5 in three categories. The first one was “Speaking”; this referred to pronunciation and intonation. Then, it was “Voice”; this referred to voice volume, louder was better, and enthusiasm. The last one was “Look up and say”; this is the practicing technique we used in class. The goal was to look up and speak <i>after</i> reading some new words, not during reading. Students who just read the conversations got low scores here.</p> <p>I: Did you use any rating scale with specific descriptions to specify their ability?</p> <p>D: No, I didn’t. The three categories I used were enough for a small classroom exam.</p>
--

In Question 2, the assessment Dr. D used to mark his students’ oral ability was similar to the assessment Dr. C used in Case 3 (Table 8.13). In Case 4, Dr. D had specific characteristics of oral abilities he wanted to look at in terms of pronunciation, intonation, voice, and “look-up-and-say”. However, he rated the students’ ability intuitively rather than via rating scales with specific descriptions of skill dimensions. Rating speaking ability was thus fast but could be subjective.

Table 8.14 Case 4: Mid-term Survey – Interview with the Teacher (c)

<p>Question 3. What did you expect the students to have learned from your class?</p> <p>D: Well...As for the skills, one of the skills was to get them feel free about speaking. It was a problem in all the conversation courses in our university. We have students, say, junior college, four-year university or two-year university there. They were very afraid to speak, so one of the goals was to get down to speaking regardless of their level. So I got students at low level and high level and they were all afraid to speak. So one of the goals is to get them to speak...</p>

I: How about listening?

D: I didn't focus on listening because every class, they had to listen to a foreigner. Because I teach the class all in English, so in class they always listen to a foreigner.

I: Do you think they can understand your speaking in class?

D: No, they couldn't. They couldn't understand my speaking at the beginning of the class. But as the weeks went on, they got used to my voice...they got used to the speed of my voice. So at the end of the semester, they were hearing more words. I knew this from my experience.

I: So they are getting better now.

D: They are getting better because they have a foreign teacher.

In Question 3, the teacher expected that his students could learn to speak actively in his course, but he did not specifically focus on improving their listening ability (Table 8.14). He found that the students could not understand what he spoke in class and they were also afraid of speaking. However, he was aware that his students were making progress every week.

Table 8.15 Case 4: Mid-term Survey – Interview with the Teacher (d)

Question 4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?

D: By the time we got the mid-term done. One of the goals, I usually achieved this goal, was that students are starting to get relaxed about speaking. So this goal – feel free about speaking – was achieved by the mid-term. I focused a lot on accurate pronunciation so my students were also being pressured to pay attention to “s” and “ed”. So my students would come to drop “s” and drop “ed”, by the time of the mid-term they started to remember to pronounce their “s” and “ed” so you will hear them pronounce “s” and “ed”.

I: What do you think about their performance this mid-term?

D: The speaking performance was good, I would give it about 80. The written performance in the test, erm, I felt OK but it was a bit lower.

I: Why did you think they performed better on speaking instead of writing?

D: Because they spent a lot of time practicing speaking in class.

In this case, Dr. D wanted to see if his students had become relaxed about speaking, and he said that this goal has been achieved in the mid-term exam (Table 8.15; Question 4). In addition, he had a specific objective which he expected his students to achieve – pronouncing “s” and “ed” while speaking. However, the speaking test was hardly communicative, as the students were allowed to “read” the conversations in the textbook to the teacher (see Section 8.4). For this group, the mean of the reading scores was 74.6, and the mean of the speaking scores was 80.4. On average, then, the students performed slightly better in the oral exam than in the reading one. Dr. D was more satisfied with the speaking test than with the written test.

Table 8.16 Case 4: Mid-term Survey – Interview with the Teacher (e)

<p>Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?</p>
<p>D: I didn't decide on the items because once I chose the test, the test has been written and developed by the publisher of the textbook. So the publisher of the textbook designed the test and I trust their test validity.</p>
<p>I: Do you think your students' level of English can meet the difficulty of the test from the publisher?</p>
<p>D: Yes, actually I would say this book is a little bit easy. I would choose a more difficult book...so I don't think they would have problems.</p>
<p>Question 6. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?</p>
<p>D: I didn't choose the type of question particularly. I thought this test has been, erm, was designed very well, because students have troubles with grammar, it is hard for them to cheat. If they didn't understand, they just didn't know the answers.</p>
<p>I: Did you pilot the test items, I mean run a small test, before you tested your students in this exam?</p>
<p>D: No, I used the book for several years and I found it's a good book, and I found out the exam synchronised with the book very well. It's reliable. I could always get the same results. From my experience I knew this test was good.</p>

The teacher did not decide the difficulty of the test items himself, but used ready made items (Table 8.16; Question 5). He believed the test items had been validated by the publisher, but, similar to Cases 1 and 3, whether the test items were piloted or not was not clarified by the textbook publisher (Question 6). The teacher felt that the textbook was fairly easy for his students and they could understand the written test without any real problems. However, the students did not perform any better on the written test than on the oral one. It was thus possible that the written test questions were not properly designed to assess the students' understanding of the teaching contents. Also, there might have been a gap between what had been taught, what had been learned, and what the students could actually perform.

Table 8.17 Case 4: Mid-term Survey – Interview with the Teacher (f)

<p>Question 7. Will the result of the mid-term examination impact on the teaching of the following half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)</p>
<p>D: Not in this class, because the scores are never so low that I feel like there is something bad going to happen. Their scores were usually very good on average, so I don't think I have to change it.</p>
<p>I: Do you think you will increase the difficulty of the items for the final exam, because you said they did well in average in the mid-term?</p>

D: No, I won't.

I: Why not?

D: Because I tested them on units 1 to 4 for the mid-term and I am teaching them 5 to 8 now, so I will test them on units 5 to 8 in the final. The contents of units 5 to 8 are different from 1 to 4 and they haven't learned 5 to 8, so I don't need to change the difficulty. They are different exams.

The teacher would not adjust the difficulty of the test items in the final exam because he was satisfied with the students' performance so far. In addition, he claimed that the test contents in the final exam would be different from the mid-term test content, so it was unnecessary to change the difficulty of any final test items.

8.6 Case 4: Classroom Observation after the Mid-term Exam

New Interchange continued to be used in class after the mid-term exam. Another three observations were carried out from 13:10 to 15:00pm on the 8th and 22nd of December 2006, and the 5th of January 2007.

8.6.1 Case 4: Fourth Observation

The first observation after the mid-term exam was carried out on the 8th of December, 2006. In this lesson, teaching Unit 6 was the main activity. At the beginning of the lesson, the teacher chatted with the students for ten minutes, and talked about the topic in the textbook regarding complaints he had made after receiving poor service. Then he told the students to share their experiences about which kind of service they had complained about, and why they complained about it. However, yet again none of the students responded. After asking the students for the third or fourth time, two students finally shared their experiences. Next, the teacher read the conversation in the textbook and explained the vocabulary and phrases in it. He played CD twice and told the students to reproduce the conversation in pairs.

In the second half of the lesson, the teacher moved on to the grammar exercises in the textbook; the grammar focus in Unit 6 was “past participles”. The teacher firstly taught the students the grammar rules in the model sentences and asked them to do the exercises. While discussing the answers, the students did not interact with the teacher actively, but answered the questions after being asked three or four times. It was hard to know if all the students understood the grammar exercises they did, because only six students answered the questions; the remaining students kept silent. Next, the teacher played the conversation again and told the students to do the listening exercises in the textbook.

8.6.2 Case 4: Fifth Observation

The fifth lesson was observed on the 22nd of December, 2006. Unit 7 in *New Interchange* was the main activity for this lesson. At the beginning of the lesson, the teacher again chatted with the students for ten minutes, talked about the environment problems in the textbook and expressed his own opinions. Then he told the students to express their opinions about which environment problem they were worried about. In short, the procedure of classroom discussion and the teaching of Unit 7 were similar to what was observed in Sections 8.4.1, 8.4.3 and 8.6.1. Next, the teacher moved on to the pronunciation exercises to practice the “reduction of auxiliary verbs” (i.e. “has”, “have”, “is”, and “are”). Again, he read the sentence models and then asked the students to reproduce them in pairs.

8.6.3 Case 4: Final Observation

The final observation took place on the 5th of January, 2007. The teacher told the students that they would finish the last unit – Unit 8. He chatted with the students for ten minutes and moved on to Unit 8; he talked about what continuing

education was and told the students to talk about which kind of class they wanted to take after formal education. The procedure for the classroom discussion and the teaching of Unit 8 was similar to what was observed in Section 8.4.1, 8.4.3 and 8.6.1. The teacher also focused on pronunciation (i.e. intonation in questions) and on the grammar rules (i.e. “would rather” and “would prefer”). In none of the six observations were the students given any opportunities to reflect on what they had learned in class.

Table 8.18 Case 4: Occurrence of Task-based Instruction Features after the Mid-term Exam

TBI Characteristics	No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	4 th	✗	No problem-solving tasks in class.
	5 th	✗	The same as above.
	6 th	✗	The same as above.
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	4 th	✓ ✗	Students were told to reproduce the conversations in the textbook in pairs. When it came to the discussions, the students did not respond actively to the teacher’s questions. Less than five students answered, despite being asked three or four times.
	5 th	✓ ✗	The same as above.
	6 th	✓ ✗	The same as above.
3. Students report findings of a task to class, in groups or pairs, after problem solving.	4 th	✗	Students completed the tasks and answered the questions individually.
	5 th	✗	The same as above.
	6 th	✗	The same as above.
4. Authentic texts which reflect a real-life situation are used.	4 th	✗	The conversations in the textbook were not authentic compared with the language in real-life situations.
	5 th	✗	The same as above.
	6 th	✗	The same as above.
5. The major focus of teaching is on the meaning, and then on the form.	4 th	✓	The major focus was on the meaning; grammar was taught afterwards.
	5 th	✓	The same as above.
	6 th	✓	The same as above.
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	4 th	✗	No reflection period was observed.
	5 th	✗	The same as above.
	6 th	✗	The same as above.

In this case, the teaching of English listening and speaking did not differ much before and after the mid-term exam. The teacher mainly employed a lecture format in class, though discussions of exercises and reproductions of conversations also took place. However, the interactions between teacher and the students were not very productive, because the students rarely responded to or asked questions. When compared with the situation in Case 3, it is apparent that nominating the students to answer the questions was a more effective way to bring them into discussions. Dr. D again did not use any audio-visual materials (e.g. English songs or films) in class after the mid-term exam.

8.7 Case 4: Questionnaire Survey on the Final Exam

The students took the final exam on the 18th of January, 2007. It consisted of only four listening multiple-choice questions (Part A, Appendix G.3), and 38 reading and writing items, including matching, sentence completion, rewriting, and true/false questions (Parts B to I, Appendix G.3). A speaking test was also included in the final exam, where the students read a conversation created by the teacher in groups, and were marked in terms of pronunciation, intonation, volume, and “look-up-and-say” technique. The teacher created six conversations based on the topics taught after the mid-term exam (Appendix G.4), and the students were grouped randomly into twos or threes to read them. In the listening test, Dr. D played the four conversations twice; the whole exam lasted approximately one and a half hours. After the exam and Dr. D had left the class, the final questionnaire was group-administered to the students. It took them 20 minutes to finish.

8.7.1 Case 4: The Teaching Materials and the Course after the Mid-term Exam

After the mid-term exam, no students felt that the textbook was “frequently” or “always” difficult (Table 8.19; Question 4). Speech rate and vocabulary were again the two main problems which impeded their understanding of the listening extracts (Table 8.20).

TABLE 8.19 Case 4: Final Survey – Perceived Difficulty of In-class Listening

Materials (Q4)		
	Frequency	Percent
Never	8	32%
Rarely	11	44%
Sometimes	6	24%
Frequently	0	0%
Always	0	0%
Total	25	100%

TABLE 8.20 Case 4: Final Survey – The Frequency of Reasons for Difficulty (Q4)

Factors	Frequency
Topical content	1
Vocabulary	9
Accent	6
Speech rate	10
The use of colloquial language	3
Sound quality	0
Text type	0

8.7.2 Case 4: Students’ General Preferences about the English Listening Class

Students’ preferences for the mode of answering remained relatively constant after the mid-term exam where “answering by speaking” had been preferred by more than three quarters of the students (Table 8.21). The proportion of the students who preferred to listen to conversational speech also remained the same (Table 8.22).

TABLE 8.21 Case 4: Final Survey – Preferred Mode of Answering (Q5)

Mode	Frequency	Percent
Writing	6	24%
Speaking	19	76%
Total	25	100%

TABLE 8.22 .Case 4: Final Survey – Preferred Type of Speech (Q6)

Types of Speech	Frequency	Percent
Monologue	3	12%
Conversations/dialogues	22	88%
Total	25	100%

Multiple-choice questions remained the most preferred type of comprehension question for 60% of the students, while short-answer questions also remained the least preferred type of question for more than half of them (Table 8.23).

Table 8.23 Case 4: Final Survey – Preferred vs. Least Preferred Type of Comprehension

Question in Class

Types of Questions	Preferred Question (Q7)		Least Preferred Question (Q8)	
	Frequency	Percent	Frequency	Percent
Short-answer	4	16%	14	56%
True/false	6	24%	0	0%
Multiple-choice	15	60%	2	8%
Cloze	0	0%	4	16%
Dictation	0	0%	5	20%
Total	25	100%	25	100%

8.7.3 Case 4: Opinions about the Final Exam

More than a quarter of the students (28%) were “(very) satisfied” with their performance in the final exam (Table 8.24); this represented a marked drop from the 40% satisfaction rate in the mid-term exam. The number of the students who were “(very) dissatisfied” with their test results in the final exam was the same as for the mid-term exam (20%).

TABLE 8.24 Case 4: Final Survey – Performance in the Final Exam (Q9)



Satisfaction with the Final Exam	Frequency	Percent
Face 1 – Very satisfied	1	4%
Face 2 – Satisfied	6	24%
Face 3 – Neither Satisfied nor dissatisfied	13	52%
Face 4 – Dissatisfied	4	16%
Face 5 – Very dissatisfied	1	4%
Total	25	100%

In the mid-term exam, because there were no listening items, the students’ answers to Questions 11 were based on their opinions of listening to Dr. D’s speech in the speaking test. However, in the final exam, Dr. D was no longer the only listening input in the speaking test, as the groups of two or three students listened and spoke to each other. In such a situation, the listening input and linguistic features of every student’s output could be diverse; in the sense that each student only listened to their partners’ speaking, listening input was not the same

for everyone in speaking test. Thus, Question 10 in the final questionnaire was analysed in the light of the four listening items used in the final exam (see Part A, Appendix G.3).

In the final exam, only three students (12%) “(strongly) agreed” that the test topics were difficult (Table 8.25) which was much less than the 60.8% for the students in Case 3 (Table 7.27). Only five students (20%) in Dr. D’s group “(strongly) agreed” that the test tasks were harder than those they did in class, but nearly two thirds of the students (65.2%) in Case 3 felt this to be true. The listening questions in Case 4 were likely to be easier than those used in Case 3, because the Case 4 students were tested with only four conversations, while the Case 3 students were tested with monologues. The topics of the test tasks in the Case 4 final exam appeared to be more representative than those used in Case 3, since 60% of the students in Case 4 “(strongly) agreed” that the topics were representative of what had been taught in class, compared to just 43.4% in Case 3.

Table 8.25 Case 4: Final Survey – Topics of Test tasks

Test Rating scale	Test topics were Difficult (Q10-a)		Topics of Test Tasks were Representative of What Had Been Taught (Q10-b)		Test Tasks Were Harder than Those in Class (Q10-c)	
	Frequency	%	Frequency	%	Frequency	%
SD	4	16%	1	4%	7	28%
D	11	44%	1	4%	4	16%
Neither A nor D	7	28%	8	32%	9	36%
A	3	12%	9	36%	3	12%
SA	0	0%	6	24%	2	8%
Total	25	100%	25	100%	25	100%

Only two students (8%) “(strongly) agreed” that the accent was too strong, vocabulary was too hard, or the sentences were too complicated to understand in the four listening extracts (Table 8.26). This contrasts starkly with the 40% of the students in Case 3 who considered the accent, vocabulary, or sentence structures to be problematic (Table 7.28).

Table 8.26 Case 4: Final Survey – Linguistic Features of the Listening Extracts (1)

Linguistic Features Rating scale	Accent was too Strong to Understand (Q10-d)		Vocabulary was too Difficult to Understand (Q10-e)		Sentences were too Complicated to Understand (Q10-f)	
	Frequency	%	Frequency	%	Frequency	%
SD	6	24%	8	32%	7	28%
D	8	32%	9	36%	10	40%
Neither A nor D	10	40%	6	24%	6	24%
A	1	4%	2	8%	2	8%
SA	0	0%	0	0%	0	0%
Total	25	100%	25	100%	25	100%

The proportion of the students who “(strongly) agreed” that it was hard to understand the test contents because the speakers spoke fast in Case 4 (24%) was much lower than that in Case 3 (65.2%) (Tables 8.27 and 7.29). Only two students “agreed” that the conversations were hard to understand.

Table 8.27 Case 4: Final Survey – Linguistic Features of the Listening Extracts (2)

Linguistic Features Rating scale	Hard To Understand because The Speaker(s) Spoke Fast (Q10-g)		Conversations Hard to Understand (Q10-i)	
	Frequency	%	Frequency	%
SD	5	20%	4	16%
D	2	8%	10	40%
Neither A nor D	12	48%	9	36%
A	6	24%	2	8%
SA	0	0%	0	0%
Total	25	100%	25	100%

In Question 11, only three students reported that colloquial language was used, but none of them found it to be a problem. As for the test characteristics (Question 12), 80% of the students stated that the quality of the recording was good, and they could hear the texts clearly (Appendix G.5). Testing time in the final exam was again sufficient for all students. Moreover, all except three students said that the test/task instructions were clear, and all but two reported that the lengths of the texts in the test were similar to those they listened to in class. In short, the test characteristics in the mid-term and the final exam did not differ greatly. That is, at least 80% of the students were satisfied with the quality of the

recording, testing time, test/task instructions, and the length of the listening extracts in both exams. In the final exam, Questions 13 and 14 were ignored, as it is pointless to compare which type of comprehension question the students found easiest or hardest, because only multiple-choice items were used in the listening test, and all except two “(strongly) agreed” that it was hard to understand to understand the conversations (Table 8.27; Question 10-i).

8.7.4 Case 4: Students’ General Comments on the Final Exam

All except one student claimed that they had learned what they expected to learn after the mid-term exam (Question 15-a). Student 4 pointed out that he did not like the listening course; he took it purely because it was compulsory. Again, all students were satisfied with the way the teacher taught after the mid-term exam (Question 15-b). All but two students were satisfied with the assessment method the teacher used in the final exam (Question 15-c). The two students concerned complained that the teacher tested them with too many reading and writing questions which took too much time to finish. After finishing the course, all except two students reported that their English listening skills had improved (Question 15-d). Student 10 said that the teacher put too much emphasis on oral training, neglecting to listen to different types of listening discourses, such as English films. The other student, Student 25, felt that her listening had not improved, but her speaking had become a little better.

8.8 Case 4: Final Interview with the Teacher

The final interview was conducted on the 19th of January 2007 after the final exam. The face-to-face interview was again conducted in the teacher’s office in

the University and lasted approximately fifteen minutes. The interview was also tape recorded with permission, and then transcribed.

Table 8.28 Case 4: Final Survey – Interview with the Teacher (a)

<p>Question 1. Where did the content of the test come from?</p> <p>Dr. D (D): The written exam was from the test bank in the teacher’s edition, but different units.</p> <p>Interviewer (I): How about the oral test content?</p> <p>D: I designed the conversations based on the four units – Unit 5 to 8 in the textbook.</p> <p>I: Why did you design the conversations which were longer than those in the textbook? Did you think that your students could digest or memorise the content in a few minutes before the oral exam?</p> <p>D: Erm...The conversations were a little bit longer, but they were easy, because they were based on the units I taught in class, and they had scripts to read. Actually, they performed quite well in the oral exam. So...I don’t think they would have problems.</p>
--

In the final exam, the written test was again taken from the teacher’s manual, but the oral conversations were designed by the teacher (Table 8.28; Question 1). Although the designed conversations were a little longer than the conversations in the textbook, the teacher felt that the students performed very well without difficulty in the final oral exam. The mean of the final written exam was 89.5, which was higher than the 74.6 for the mid-term exam. For the oral exam, the mean of the final oral exam was 84.5, which was again slightly higher than the 80.4 recorded for the mid-term exam. Dr. D was again satisfied with the result of his students’ performance in both exams.

Table 8.29 Case 4: Final Survey – Interview with the Teacher (b)

<p>Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and the students’ learning outcomes? Do you think you have achieved them?</p> <p>D: At the end of the semester, one of the goals is that students are starting to feel more confident about speaking. So this goal – feel more confident about speaking – is achieved at the end of the semester. I focused a lot on intonation and stress so my students paid more attention to key words and intonation in questions. So my students would come to rise or fall intonation in sentences, and pronounced words with correct stress.</p> <p>I: Did you use any rating scale to rate their oral ability in the final exam?</p> <p>D: No, I simply used the same three categories that I used in the mid-term oral exam.</p>

The teacher did have clear test objectives which he wanted to achieve in relation to his teaching plan and the students’ learning outcomes (Question 8.29; Question 2). Before the mid-term exam, Dr. D focused on including “s” and “ed” in sentences, while after the mid-term exam, he put emphasis on intonation and

stress in pronunciation. He also expected that the students would feel more confident about speaking throughout this semester, and he found that his students did improve their speaking.

Table 8.30 Case 4: Final Survey – Interview with the Teacher (c)

Question 3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria?
D: Well...as I said before, I don't decide on the items because the test has been written and developed by the publisher of the textbook. So I trust the test the publisher of the textbook designed.
I: Did you pilot the test items for this exam?
D: No, because the exam in the textbook is very reliable.
Question 4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions?
D: No, I don't. As I said in the mid-term exam, I think this test was designed very well. And I trust the reliability of the test.

The teacher again identified and decided on the difficulty of the test items based on the test bank in the teacher manual, and he appeared to rely very much on the test from the textbook publisher (Table 8.30; Questions 3 and 4).

8.9 Summary

Only one feature of TBI was observed in Dr. D's class. Opportunities for the students to speak English in class were limited to the reproductions of the conversations in each unit; although the teacher was eager to involve them into discussions, the students did not appear to interact willingly with him. While teaching the listening extracts, the teacher focused more on understanding the meaning than on the form. He did not, however, give the students opportunities to reflect on what they had learned at the end of the lesson. The teaching materials appeared to match the majority of the students' learning needs even though a number of the students did not have particular expectations for this course. In addition, the conversations in the textbooks were not as authentic as those in real-life situations (see Section 8.3.2). The answer to Question 1 is thus that only one

feature of task-based instruction was employed: Dr. D focused on the meaning rather than on the form of the discourse.

Dr. D had specific test objectives in both mid-term and final oral exams; he focused on his students' accuracy of pronunciation, intonation, vocal volume and "look-up-and-say" skills. The test contents in both mid-term and the final exams were very similar, except that listening test was not included in the mid-term exam. The assessment methods were also similar in the two exams, where teacher-made categories were used to rate the students' oral ability. The answer to Question 2 is that test objectives in oral exams were specific at the "educational" level; no specific test objectives were discovered in the two written exams, because the test items were taken from the teacher's manual. Although speaking skills were assessed in both exams, the teacher focused too much on the pronunciation rather than communicative skills. Thus, the answer to Question 3 is that communicative language testing approaches were not used in either exam.

The answer to Question 4 is that the students did not find particular problems while taking the two exams. When it comes to how far the test results impacted on the teaching and testing after the mid-term exam, it is apparent that they neither impacted on the teaching nor influenced the difficulty of the final exam. Thus, the answer to Question 5 is that no washback effect was discovered with respect to teaching or the difficulty of the final exam.

Teaching Methods and Classroom Activities

In Cases 3 and 4, reproducing the listening extracts in the textbooks was the main activity in class, but other in-class activities differed between the two groups. Dr. D in Case 4 emphasised pronunciation, grammar, and reading tasks, while Dr. C in Case 3 focused on the oral practices of sentence patterns. In addition, student

presentations were frequently used in Case 3, which provided the students with more opportunities to practice English than was the case in Case 4. The listening materials in Case 3 appeared to be more varied, in the sense that not only the listening extracts in the textbook were taught, but English music and films were also used as supplements. However, the conversations in the textbook in Cases 3 and 4 were not authentic, because they were too fluent in both cases, without broken sentences, false starts, or interruptions. The interactions between the teacher and the students in neither case were active nor spontaneous, and reflection periods were not used in either group. It was thus hard to know if the students understood the teaching contents since they neither were given opportunities to reflect on what had been taught nor did they express their ideas or opinions in class.

Classroom Assessment

In the mid-term exam, the assessment methods in both groups were very similar, in the sense that the reading exam accounted for a large proportion of the mid-term listening and speaking scores. As discussed before, assessing listening and speaking ability with so many reading items could threaten the validity of the test. In the oral exam, the two teachers did have test objectives designed to incorporate different dimensions of speaking ability (i.e. pronunciation, fluency, and intonation, voice, and look-up-and-say). However, their rating procedures were subjective as they did not mark students by using rating scales but instead through intuitive judgement.

Impact of the Test on Teaching and Testing

In both cases, the results of the mid-term exam neither impacted on the teaching procedure nor influenced any change in the listening materials used in class. The test consequences did not impact on the difficulty of the final exam in Case 4 because the teacher was satisfied with the students' performance, and he felt that there was no need to adjust the difficulty of the final test. In Case 3, however, the students' test results in the mid-term exam did impact on the difficulty of the final exam in the sense that the teacher used more news excerpts from the ICRT.

Notes to Chapter Eight

¹ Look-up-and-say method means that when the students read the listening extracts in the textbook, they need to “speak” or “say” the conversations rather than simply read them.

Chapter Nine

Discussion

9.1 Introduction

This chapter pulls together the findings from Chapters Five to Eight, and attempts to answer the various research questions with which this thesis is concerned. The overall research question was “*Are the general listening test practices within and between Taiwanese universities similar or markedly different?*” and this was investigated at two universities, A and B. The classes and instructors are summarised for ease of reference in Table 9.1. There were five specific research questions:

- (1) *How far is task-based instruction implemented in English listening classes?*
- (2) *How far are the test objectives clearly specified in relation to the curriculum and teaching contents?*
- (3) *How far are communicative language testing approaches applied in the two listening exams?*
- (4) *What kind of problems will influence students’ listening comprehension in the two listening exams?*
- (5) *How far do the results of the mid-term exam have washback effects on teaching?*

Question 1 is discussed in Section 9.2; followed by Question 2 in Section 9.3. Question 3 is analysed in Section 9.4, Question 4 in Section 9.5, and Section 9.6 looks at Question 5.

Table 9.1 Course Instructors in the Main Study

University	University A		University B	
Cases	Case 1	Case 2	Case 3	Case 4
Course Instructor	Dr. N	Miss T	Dr. C	Dr. D

9.2 The Implementation of Task-based Instruction

Question 1: How far is task-based instruction implemented in the English listening classes?

Field (2002; see Chapter 1.8.2) suggests that there are three stages involved in teaching listening: (1) finding a topic to motivate the students, (2) giving listening tasks/exercises, and (3) teaching grammar and vocabulary. The data from the four classes show that, Dr. N and Miss T at University A skipped the first stage, but Dr. C and Dr. D at University B followed all three stages.

With regard to the implementation of task-based instruction (TBI) in teaching listening, six features were selected as characterising TBI: problem-solving tasks, oral practice of English, report finding in pairs or in groups, a focus on meaning and then on form, and the authenticity of listening materials. The four groups of students were all given opportunities to speak English in class, particularly when they were asked to answer the exercise questions in the textbooks, but answers were limited to vocabulary, phrases, and a few sentences (Table 9.2). In addition to the textbook exercises, opportunities for the students to speak English occurred in student presentations, for example, the radio talk show in Dr. N's class (Chapters 5.3.2 and 5.6.1), and the presentations of English songs in Dr. C's class (Chapters 7.3.2, 7.6.2, and 7.6.3). However, only one or two groups/pairs were allowed to present every week; that is, the opportunity for the whole body of students to practice English in every lesson was very limited.

Reproductions and role-plays of the listening extracts in the textbook were frequently used in both cases at University B (Chapters 7.3.1, 7.3.2, and 8.3.1). However, from Willis's (1996), Ellis's (2003), and Nunan's (2004) viewpoints of task-based instruction, activities that involved role-plays and sentence practice modules do not count as "tasks", since there is no manipulation or production of

the target language, or meaningful exchanges of ideas in the language (see Chapter 1.8.3). Thus, the oral reproduction of conversations at University B could only be counted as oral practice of language skills in English (i.e. pronunciation, intonation, and sentence patterns), and not as a valid communication “task” which required spontaneous speaking of English in a meaningful way. Interaction between the teachers and the students in the four cases was largely confined to a situation where the students talked when asked to do so by the teacher. However, Malamah-Thomas (1987; see Chapter 1.8.3) does not regard students’ responses to teachers’ prompts as real interaction, which requires reciprocal interaction. Also, Savignon (2005; see Chapter 1.8.3) suggests that communicative interactions require the speakers’ ability to make meaningful interactions with others, rather than reciting dialogues or performing in “discrete-point tests of grammatical knowledge”. When interacting with the students, Dr. C tended to nominate her students to answer questions in class, but the other three teachers left the available opportunities for the students to take up themselves, albeit reluctantly. Even though the students were encouraged to answer questions in all the classes, only a few students in Miss T’s and Dr. D’s classes did so; the students did not ask any questions in class. It was thus very hard to know if the students really understood the listening contents or the classroom exercises. Dr. N, conversely, motivated her students to answer questions by giving extra marks, which did encourage more students to raise their hands. Discussion or problem solving in pairs or in groups was only found in Miss T’s class, where the students were told to search for colloquial language and reported it in pairs.

Table 9.2 Comparison of TBI Characteristics across the Four Teachers

TBI Characteristics	Ob. No.	University A		University B	
		Case1 Dr. N	Case2 Ms T	Case 3 Dr. C	Case 4 Dr. D
1. There is at least one problem-solving task for students to do in class.	1 st	x	✓	x	x
	2 nd	x	✓	x	x
	3 rd	x	✓	x	x
	4 th	x	✓	x	x
	5 th	x	✓	x	x
	6 th	x	✓	x	x
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st	✓ x1	✓ x1	✓ x1	✓ x1
	2 nd	✓ x1	✓ x1	✓ x1	✓ x1
	3 rd	✓ x1	✓ x1	✓ x1	✓ x1
	4 th	✓ x1	✓ x1	✓ x1	✓ x1
	5 th	✓ x1	✓ x1	✓ x1	✓ x1
	6 th	✓ x1	✓ x1	✓ x1	✓ x1
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st	x	✓ x	x	x
	2 nd	x	✓ x	x	x
	3 rd	x	✓ x	x	x
	4 th	x	✓ x	x	x
	5 th	x	✓ x	x	x
	6 th	x	✓ x	x	x
4. Authentic texts are used, which reflect a real-life situation.	1 st	x2	x2	x2	x2
	2 nd	✓ x2	x2	✓ x2	x2
	3 rd	✓ x2	x2	✓ x2	x2
	4 th	✓ x2	x2	✓ x2	x2
	5 th	✓ x2	x2	✓ x2	x2
	6 th	x2	x2	✓ x2	x2
5. The major focus of teaching is on the meaning and then on the form.	1 st	✓	✓ x	x	✓
	2 nd	✓	✓ x	x	✓
	3 rd	✓	✓ x	x	✓
	4 th	✓	✓ x	x	✓
	5 th	✓	✓ x	x	✓
	6 th	✓	✓ x	x	✓
6. Students are given opportunities to reflect on what they have learned and how well they are doing (i.e. reflection period).	1 st	x	x	x	x
	2 nd	x	x	x	x
	3 rd	x	x	x	x
	4 th	x	x	x	x
	5 th	x	x	x	x
	6 th	x	x	x	x

¹ Answers to the exercise questions in the textbooks which did not involve meaning exchange in oral interactions (i.e. answers to multiple-choice, true/false, cloze, sentence completion questions) did not count as real oral information communication in TBI.

² As discussed in Chapters Five to Eight, the language in the English listening textbooks in the four classes was not authentic with respect to real-life situations (see Chapter 5.3.1 and Appendix D.3; Chapter 6.3.1 and Appendix E.2; Chapter 7.3.1 and Appendix F.1; and Chapter 8.3.2 and Appendix G.1). On the other hand, English films and the situation comedy, and the student presentations in English appeared to be more authentic compared with the English in the textbooks.

While teaching the listening contents, Dr. N and Miss T at University A and Dr. D at University B told their students to focus on the meaning and main ideas of the listening passages, and then the three teachers explained the new vocabulary, grammar, or phrases. Based on the classroom observation data, Dr. C at University B did not specifically tell the students to focus on the main ideas before going to the detailed grammar teaching; they simply followed the structures of each unit in the textbook. As for the in-class listening materials, the four teachers all used textbooks in class. According to the prefaces of the three textbooks used in Dr. N's class (*Impact Listening 3*), Dr. C's class (*Way Ahead*), and Dr. D's class (*New Interchange*), the listening extracts or activities were "derived" or "based on" everyday English in real-life settings (Chapters 5.2, 7.2, and 8.2). However, the language and conversations in the three textbooks for the four English classes were relatively inauthentic compared with the language in real-life situations, since the language in the listening passages or conversations was highly fluent, without false starts, interruptions, or broken sentences (Appendix D.3, E.2, F.1, and G.1). The preface of *Listen Up* in Miss T's class, on the other hand, did not claim that authentic texts were used (see Chapter 6.2). Although the listening extracts in the textbooks were not authentic discourse, meaningful connections between the teachers, the students, and the teaching contents, such as creating a task that required the students' language ability to

achieve it, could be established based on the teaching content. Nevertheless, problem-solving tasks were only found in Miss T's class, as discussed before. In the other three listening classrooms, neither the language nor the activities was authentic, because the listening activities were based on textbook exercise questions and oral practice of skills, rather than on the spontaneous use of English in real-life or simulated situations. In addition to the textbook contents, supplementary listening materials, such as a situation comedy, films, and songs in English were also used. The language when there was an audiovisual context was closer to a real-life setting, since the students could not only hear the language, but also see the facial expressions and the body language of actors, as well as the background settings where the conversations took place. This gave the students an opportunity to visualise what they might encounter in the target language situations. In addition, in none of the four cases, were any the opportunities given to the students to reflect on what they have learned and how well they were doing at the end of the lessons. The answer to Question 1 is therefore that only some of the six aspects of task-based instruction were discovered in the four English listening classrooms: in student presentations, watching English films, the problem-solving task of finding out colloquial language, and putting the focus on the meaning before the form.

9.3 The Test Objectives

Question 2: How far are the test objectives clearly specified in relation to the curriculum and teaching contents?

Question 2 investigates how far listening skills or ability were clearly indicated as the main test objective. In both mid-term and final exams, the four teachers (Chapter 5.5, 6.5, 7.5, and 8.5) all claimed in the interviews that they tested what had been taught in class, and to this end, three teachers – Dr. N, Dr. C, and Dr. D –

took existing listening questions directly from the relevant teacher's manual(s) as the exam questions. However, Hughes (2003), and Anastasi and Urbina (1997) suggest that classroom tests should cover the instructional objectives rather than simply test detailed teaching and textbook contents, as these provide a more precise indication of what has actually been achieved (see Chapter 2.2.1). Miss T, on the other hand, designed the test questions herself. Nevertheless, none of the test questions used in the tests observed was piloted by the teachers so as to fit the students' level of English before being administered. The teacher's manuals do not clarify whether their questions were designed and piloted on participants similar to those in present study; using tests without clearly defined instructional objectives and without taking students' level of English into account must inevitably have led to a degree of inaccuracy in the assessment. In partial mitigation, it was discovered that, from the interview data, it was logistically hard for the teachers to pilot the test items before the exams.

Brown and Hudson (2002; see Chapter 2.2.1) define test objectives as being of three types – *educational*, *instructional*, and *performance*. For the written exams, none of the four teachers had specified a clear test objective which related to their syllabus. In the speaking exams, however, the two teachers at University B did have specific “educational” objectives; Dr. D set goals to examine the accuracy of pronunciation (i.e. pronouncing “s” and “ed”), intonation, vocal volume, and “look-up-and-say” techniques in the speaking tests (Chapter 8.6 and 8.8), while Dr. C aimed to test the students' pronunciation, fluency, and intonation (Chapter 7.8). In addition to their oral test objectives, their ratings were fairly subjective, as neither employed rating scales with detailed descriptions of expected oral performance. Moreover, in Dr. C's and Dr. D's mid-term exams, the use of non-listening-based test items (i.e. reading and grammar questions)

served to reduce content validity, because the reading tasks could not represent the listening tasks that the students had to master after taking the listening courses (Weir, 1993; see Chapter 2.4). In addition, the inappropriate test contents and the lack of specific test objectives at University B must have biased the assessment and the interpretation of students' achievement effectively reducing the construct validity (Messick, 1993; see Chapter 2.4). The answer to Question 2 is therefore that test objectives were specific at the "educational" level in the oral exams, because the teachers claimed to test what they had taught in class, which was based on the textbook contents; no specific or appropriate test objectives were given for the written exams.

In addition to the absence of test objectives in the written exams, the foci of the teaching objectives between the two teachers in University A were diverse. Table 9.3 lists a comparison of the teaching objectives of the four teachers. While Dr. N emphasised the students' understanding of everyday and academic English, Miss T stressed training for GEPT examinations. At University B, Dr. C focused on the students' ability to understand live news broadcasts and to express ideas orally. Although Dr. D in Case 4 did not have specific teaching objectives, his teaching contents were based on the theme of cross-cultural life-styles (Chapter 8.2), which was different from the emphasis on listening to everyday English and live news reports in Dr. C's class. This sort of marked diversity between the four groups led to a problem with interpreting the students' listening ability and achievement both within and across the two universities.

Table 9.3 Comparison of the Teaching Objectives of the Four Teachers

Comparison of the Teaching Objectives of the Four Teachers	
Case 1	<i>The course was designed to improve undergraduate students' listening ability at intermediate level to understand English in "general and academic" situations. (Extract from Dr. N's syllabus notes, Chapter 6.2)</i>
Case 2	<i>Students will develop an in-depth understanding of the language they use and of applications of this understanding to classroom discussion. After completing this course, students will improve their listening comprehension, have an understanding</i>

	<i>of listening models of GEPT, and be able to pass the intermediate level of listening sections of GEPT.</i> (Extract from Miss T's syllabus note, Chapter 7.2)
Case 3	<i>It is hoped that by the end of the semester, students are able to listen to live broadcasting and express their opinions afterward.</i> (Extract from Dr. C's syllabus notes, Chapter 8.2)
Case 4	No specific teaching objectives.

9.4 The Communicative Language Testing Approaches

Question 3: How far are communicative language testing approaches applied in the two listening exams?

The listening exams in the four cases tended to be one-way listening tests which did not require two-way oral interactions between students. Although Dr. C and Dr. D in University B used oral tests in their final exams, the oral assessments in the two teachers' classes were essentially skill-based, because the speaking test contents were based on reading the conversations in the textbook, or those created by the teacher. This runs contrary to the recommendation by Hughes (2003; see Chapter 2.6.4) that it is better not to use "prepared" texts or "read aloud" techniques in a speaking test, since this does not test students' real oral interaction ability, but rather their reading ability. Thus, the answer to Question 3 is that neither communicative nor direct tests, which involved spontaneous use of English and meaning exchanges were used to elicit students' speaking ability in either the mid-term or the final exams.

9.5 The Difficulty the Students Encountered in the Two Exams

Question 4: What kind of problems will influence students' listening comprehension in the two listening exams?

Research Question 4 asks about the main problems that influence Taiwanese students' listening comprehension in the two exams. Table 9.4 lists the problems that the students appear to have encountered in the listening extracts while taking the two exams; the data were collected from the two questionnaires the students answered after the exams. The percentages are based how many of the students

“(strongly) agreed” that particular features were difficult to understand in the two exams. In both exams, the main difficulty for more than half of the students from Dr. N, Miss T, and Dr. C’s classes was the fast speech rate of speakers. In the same three listening classes, the speech rate problem recurred with respect to understanding the listening passages in the textbooks, where a high percentage of the students (approx. 75%) regarded it as a major problem (Tables 5.4, 6.4, 7.4).

Table 9.4 Difficulties in Comprehending the Listening Contents in the Two Exams

University	University A				University B			
	Case 1 (Dr. N)		Case 2 (Ms T)		Case 3 (Dr. C)		Case 4 (Dr. D)	
Group	Mid	Final	Mid	Final	Mid	Final	Mid	Final
Test Time	6.5%	6.5%	6.1%	51.5%	21.7%	60.8%	0%	3%
Topic	6.5%	6.5%	6.1%	51.5%	21.7%	60.8%	0%	3%
Accent	16.2%	22.6%	9.1%	15.2%	8.7%	30.4%	12%	4%
Vocabulary	6.5%	9.7%	15.2%	33.3%	4.3%	47.8%	4%	8%
Sentence	3.2%	6.5%	12.1%	24.2%	13%	43.5%	8%	8%
Speed	51.6%	42%	57.6%	75.2%	52.1%	65.2%	12%	24%
Monologue	25.8%	22.6%	36.3%	60.6%	52.2%	43.5%	 	
Conversation	6.5%	12.9%	12.1%	33.3%	17.3%	 	8%	8%

The type of listening text also influenced nearly half of the students’ comprehension of listening passages in Miss T’s and Dr. C’s classes, particularly when too many monologue-related items were tested. In Miss T’s and Dr. C’s classes, again, the number of the students who were worried about the topics tested in the mid-term exam increased markedly in the final exam. It was discovered that Miss T added two different topics and more monologue-related questions into the final exam (see Appendix E.5 and E.6), while Dr. C tested the students with both a completely different type of test item and different listening texts (i.e. using a reading test and listening conversations in the mid-term exam, but listening dictations, an English song, and news reports in the final exam) (see Appendix F.2 and F.3). Compared with lectures and dialogues, news broadcasts proved to be the most difficult type of text, due in all probability to less redundant and denser sentences, fewer pauses, and repetition (Rubin, 1994; Shohamy and Inbar, 1991; see Chapter 2.5). This would explain why longer and faster

monologue texts were claimed as a problem by the two groups of students concerned. In Miss T's and Dr. C's groups, the longer and the faster the listening texts were, the more likely it was that the students would find difficulties in understanding the topics, vocabulary, and sentences, since they had to digest the fast listening input, the language, the meaning, and to decide the correct answer in a limited response time (see Miss T's and Dr. C's final exams in Table 9.4). This would support Ur's (1984; see Chapter 2.3) point that listeners may fail to recognise the vocabulary they have learned in class because they are not very familiar with it when it occurs in a stream of spoken discourse. However, when it comes to the assessment, comparing and judging student progress from the two exams becomes a serious problem both *within* and *across* universities. Firstly, the difficulty of the two exams in Cases 1 and 2 varied a lot, in the sense that the difficulty between Dr. N's two exams was similar, but there was a large jump in difficulty in Miss T's two exams. Secondly, test methods and formats were diverse in Cases 3 and 4, where there was a big disparity between the listening extracts and test items in Dr. C's two exams, while the test formats and methods in Dr. D's two exams remained very similar. Such marked diversities in both the mid-term and final exams within and across universities reduce the fairness of marking system, and argue strongly for quality control measures of teaching, testing, and score interpretation.

In contrast to the other three groups, only a few students in Dr. D's class said they had problems while listening to the two exams. As discussed in Chapter 8.4.3, the students were familiar with the conversations in their oral exam, since they had listened to them in class and they were also allowed to read the textbook in the exam. In addition, during the oral exam, Dr. D might have adjusted his speed of speaking while talking to individual students so that each student could

hear clearly. The familiarity of test topics and the modified speed of speaking from the teacher may well have served to reduce the difficulty of listening to the conversations in the mid-term exam.

As for the test task characteristics, the quality of the recording was a problem in both exams in Miss T and Dr. C's classes, in the sense that the quality was too poor or the volume was too low for many students to hear the text clearly (Appendix E.8 and F.4). In particular, the background noise outside the test room was loud in Dr. C's mid-term exam as well as in the final one. The quality of the recording and the noise outside the classroom had in both cases, the students claimed, affected their performance. Other characteristics, such as the test time, test instructions, and the length of the listening texts were not perceived as problematic for the majority of the students in any of the four groups. In short, the answer to Question 4 is that the speech rate, monologue-related texts, and the quality of tape recording were perceived as the three major problems by the students while taking the listening exams.

9.6 The Washback of Test Results on the Teaching

Question 5: How far do the results of the mid-term exam have washback effects on teaching?

“Washback” refers to the influence from the tests on teaching and learning (see Chapter 2.7). The two teachers – Dr. N and Miss T – at University A slightly amended their teaching after the mid-term exam. Dr. N asked the students to answer four questions on a piece of paper, which she did not do before the mid-term exam, to see how much they understood the situation comedy they saw in class (see Chapter 5.6.2). Similarly, Miss T also adjusted her teaching after the mid-term exam by adding a different activity (i.e. listening to English songs) in class (see Chapter 6.6.2); because, she said, that her students had improved their

listening by the mid-term point, she wanted to give them different listening activities after it, and her students were also given a sheet of paper to fill in the missing lyrics (see Chapter 6.5). After establishing the students' performance in the mid-term exam, Miss T claimed that she would give harder test items in the final exam, since she wanted her students to improve their listening ability (see Chapter 6.5). However, based on the students' marks in the final exam, it was hard to know if the harder test items in the final exam reflected beneficial consequences of the teaching after the mid-term exam, since her students scored lower on average in the final exam than in the mid-term one (see Chapter 6.8). In addition, in order to ensure that her students understood the academic listening passages in *Mosaic 1*, Dr. N gave her students the transcripts, so that they could check the extracts they had listened to. This resulted in a problem when a positive intention from the teacher turned out to have negative washback in learning, as some of the students answered the exercise questions and the teacher's questions in the textbook by copying (see Chapter 5.6.3). It was thus hard to know if the students really "listened to" and understood the passages, or if they simply understood the contents by "reading" them. Messick (1996; see Chapter 3.7) notes that a test can influence *what* is taught much more than *how* it is taught in class. In University A, the mid-term test results did seem to influence *what* was taught in class more than *how* it was taught. The answer to Question 5 is thus that both positive and negative washback effects were discovered in the teaching at University A. Positive washback was detected when the teaching provided the teacher with the information about the extent to which her students understood the episode or the English songs, and about the effectiveness of using the videos or songs in class, while negative washback effects resulted from the students' negative learning. On the other hand, no washback effects from test to teaching

were found in the University B classes. In addition to the teaching after the mid-term exam, the mid-term results influenced the design and the difficulty of the final exams in Cases 2 and 3 (see Section 9.5). However, since none of the four cases applied communicative language testing approaches to test students' speaking skills, it was therefore impossible to know if direct tests would result in positive washback effects on teaching and learning.

9.7 Summary

From the above discussion, it is clear that the teaching objectives were specific at the curriculum and syllabus level, but none of the four teachers determined what kind of objective or goal they would like to achieve in each lesson. Test objectives were clear for the oral exams; none of the four teachers, however, clarified their objectives in the written exams, where they used existing test questions from the Teacher's Manuals. The absence of clear test objectives for the both mid-term and final written exams led to problems of: (1) matching the students' scores to their listening ability, as test items were often intended to test, for example, problems of reading and writing rather than listening; (2) matching the students' scores to the skill levels which they were expected to achieve in their year of study. In addition, there was a large discrepancy in the difficulty level and the test contents between the mid-term and the final exam in Miss T's and Dr. C's group which made it impossible to compare: (1) the students' progress between the two tests within each case (see Section 9.4), and (2) the students' scores between groups (e.g. Cases 1 and 2). At University A, it was hard to compare the students' mid-term and final scores and the progress made between Dr. N's and Miss T's groups because the difficulty of the two exams in Dr. N's class was similar, but varied greatly in Miss T's class. At University B, similarly, the

formats of Dr. D's two exams were similar, while those of Dr. C's two exams were very different. Only a few aspects of task-based instruction were observed in the four groups; the teaching was generally far more teacher-centred than focused on students' discussions and task completion. Although the test results did have washback effects on the teaching (Cases 1 and 2) and did influence the final exam to a certain extent (Cases 2 and 3), the impact did not always take the form of beneficial teaching and testing situations. In short, it was discovered that the general listening test practices within and between the Taiwanese universities were markedly different. The next chapter will consider the implications of the findings in the light of current educational policy in Taiwan.

Chapter Ten

Conclusions and Suggestions

10.1 General Aims of the Study

The primary focus of the research was to evaluate the teaching and assessment of university English listening courses in a situation where government initiatives to develop an English teaching environment, and to ensure the quality of teaching is being heavily promoted. I discussed in Chapter 1.2 how the “liberalisation” of Martial Law allowed greater freedom for individual universities to establish and operate their institutions, and how the join of the WTO simulated the quantity of universities. With an urgent need to ensure the quality of teaching and learning in higher education, a national teaching evaluation system (i.e. HEEACT) across universities was established, and individual self-evaluation within universities were developed. While the HEEACT focuses on overall teaching evaluation of a department and a university, individual self-evaluation specifically serves as a basis for rewarding teachers’ behaviour and teaching performance in class. Both evaluation systems were established to ensure the quality of teaching in universities. As one of the goals of the Ministry of Education (MOE) in 2002 was to promote English teaching, the MOE decided that English teaching should be included in the national evaluation project. The aim of this study was to explore one part of English teaching, namely listening, to examine how it was taught and tested, and to evaluate how far approaches appropriate to implementing the MOE requirements (namely a communicative or a TBI approach) were employed at a sample of Taiwanese universities. Evaluating the programmes included examining the teaching approaches, teaching materials, classroom interactions, test methods, testing contents, and test washback effects on teaching.

The thesis does not focus on how evaluation procedures should be established or operationalised by the Taiwanese government, but rather on evaluating how far the English listening courses meet the spirit of the government legislation and help undergraduate students develop their communicative skills in class or later in work contexts. The study was classroom-based; it used a case study approach and involved four cases constituted by 112 undergraduate students and four course instructors from two private universities, with a different group of 41 undergraduates and a teacher from another university taking part in the pilot study. Both qualitative (observation, interview, and document survey) and quantitative (questionnaire and test score) research approaches were used.

The implications of the results of the study are based on the two interrelated dimensions of the purposes of evaluation outlined by Rea-Dickins (1990; 1994; see Chapter 1.5) – *accountability* (i.e. purpose of evaluation) and *development* (i.e. course and programme planning). I noted in Chapter One that the common purpose of the national HEEACT and self-evaluation systems was to ensure the quality of teachers and teaching within universities. However, due to a lack of unified evaluation standards for self-evaluation systems across the university sector, the “accountability” of self-evaluation may vary within different universities. The “development” of evaluation in this chapter is discussed in the context of instruction, curriculum, in-class listening materials, assessment methods and test items. There is a third dimension – *awareness raising* (i.e. teacher training or teacher development) – suggested by Rea-Dickins; however, because neither teacher training nor development is included in the Taiwanese higher education evaluation system (see Chapter 1.5), policy and practice will be suggested, rather than teacher training and development. The details of the main findings were given in Chapters Five to Eight and pulled together in Chapter Nine.

For this final chapter, I want to make a bridge between the important issues arising from the case studies and the current policy of higher education in Taiwan.

10.2 Implications for Instruction and Curriculum Development

In Chapter 1.8.4, I pointed out that research suggests that task-based instruction has been found to have a positive effect on Taiwanese primary and secondary pupils' development of the four English language skills, particularly as regards their speaking ability, and on their improvement generally of communicative skills (Lun, 2004; Tseng, 2006). In the present study, it was discovered that only a few features of TBI instruction were implemented in the English listening and speaking classes at university level (see Chapter 9.3). The students did not ask questions in class, even though they were given opportunities to speak; they tended to speak while answering the questions in the textbooks, rather than having meaningful discussions with the teachers in class. Reproductions of conversations and sentence patterns in the textbooks by the students were found in every lesson observed at University B. Pair discussions were not used frequently in any of the lessons I observed; even where such discussions *were* used, they were limited to a mere ten minutes. The lessons were all found to be very teacher-centred, in the sense that the teachers talked and dominated the conversations and interactions. Based on the students' responses in the questionnaires, the reasons why the students preferred writing to speaking in listening classes were that they reported that they performed better and felt more relaxed in writing than speaking, although they still said that they needed more opportunities to practice speaking English in class. In other words, although the students expected to improve their oral skills, they were inclined to avoid taking risks by making mistakes in speaking. In short, the English listening and speaking classes in the present study remained

more audiolingually-oriented than communicatively-based, and this is unlikely to further the students' ability to communicate in English as part of their degree or in a career context after taking the courses. The results of the study suggest that, firstly, students could be encouraged to speak English via student presentations of topics related to the lessons before they were asked to raise their hands to speak in class. Secondly, more time for group discussions needs to be given to the students, so that it is not always the teachers who dominate the class. At the same time, exercise questions in class should not always focus on close-ended items (i.e. true/false, multiple-choice, or cloze), which imply that only one answer is correct; open-ended questions which ask for students' opinions concerning listening topics could usefully be employed to help students develop confidence in speaking English.

In addition, teaching may be influenced by test consequences, both intended and unintended (Stobart, 2003). In the present study, it was hypothesised that it was more likely for the students to improve their listening ability if the teachers used communicative or direct tests than indirect ones. However, in the two exams, communicative approaches were neither used in Cases 1 and 2 nor in Cases 3 and 4, plus there were very limited opportunities for students to speak English in class, so it was hard to discover whether or not communicative approaches had positive washback effects on improving students' oral skills. In addition, the mid-term test results did have washback effects on the teaching at University A. This resulted, however, in negative consequences in Case 1 (see Chapter 9.5), and the students distorted the good intentions of the teacher who just wanted to help them. It is generally accepted that "washback" is a multifaceted issue, which is affected by the five factors: *test*, *prestige*, *personal*, *micro-context*, and *macro-context factors* (Watanabe, 2004), and that it can take a long time to

surface (see Chapter 2.7). Further investigation is required on how to determine precisely how strong the influence of the five factors is on the English listening and speaking classrooms in Taiwanese universities.

10.3 Implications for In-class Listening Materials

In my preliminary interviews with university English listening teachers (see Chapter 1.4, Appendix C.1), it was made clear that individual teachers were free to select the listening materials they wanted to teach. However, none of them evaluated their students' listening ability or asked about their students' needs for learning listening before deciding on the textbooks. When the students themselves were asked, approximately two thirds considered that the in-class listening materials selected by their teachers were practical and related to everyday life, and the remaining students did not have particular expectations of the listening course, but simply accepted what was given. Although the majority of the students were satisfied with the in-class materials which were not based on a needs analysis, it is recommended that at least a formative assessment should be given before the teachers decide on in-class materials. Since formative evaluation or needs analysis have been regarded as an evaluative basis of curriculum design and material selection, in order to match both groups of students' language levels (Rea-Dickins, 1994; Jordan, 1997; see Chapter 2.2.1), the present study suggests that a formative assessment of their English proficiency or a needs analysis of students' learning preferences is necessary.

While teaching, it is essential to use teaching materials for appropriate purposes, and so for example, listening materials should focus on developing listening rather than reading skills. In the textbooks used by University B, the conversations taught were scripted in every chapter (see Appendix F.1 and G.1)

which assumed that students could read while listening, because the two teachers did not ask their students to cover the book before listening to the conversations. Also, it is highly likely that the listening classes turned out to be reading and grammar-based sessions, as both teachers put a heavy emphasis on reading the conversations and completing grammar exercises. My recommendation is that it would be better to use the listening materials without the scripts from the textbooks, and, if the scripts are absolutely required to facilitate students' learning, I would suggest ensuring the students have finished listening tasks or exercises before presenting them with these scripts. Scripts of listening extracts are better used as a reference after listening tasks, thus preventing students relying on what they *see* rather than what they *hear*, as was found with some students in Case 1, who tended to rely on reading the scripts while answering the exercise questions.

Another problem that arose was that the students were not given opportunities to reflect on what they had learned, or on how much they had understood, at the end of each lesson. Student reflections need not always take the form of oral production. As discussed in 10.2, the students said they felt "safer" and "more confident" in writing than in speaking. In the Taiwanese university context, it is suggested that, after finishing listening and speaking tasks, asking students to write a short paragraph regarding the topics that they have just listened to and discussed would not only help them integrate the knowledge they had learned, but also provide the teachers with a guide to their teaching effectiveness. In addition to examining students' reflections on their writing, course instructors are strongly recommended to reflect on their own teaching. Rea-Dickins and Germaine (1998) also support the view that teachers need to "develop a deeper understanding of the tasks in which they are involved." Reflection does not simply improve teaching and learning; it also benefits

curriculum or programme improvement at a more general level.

Lastly, in Chapter 9.2 I argued that neither the language in the textbooks nor the tasks (i.e. textbook exercise questions or skill-based oral production of vocabulary and sentence patterns) were “authentic”, except for the student presentations and English films. Problem-solving tasks were rarely used in class. If it is hard to prepare authentic listening materials for English listeners in Taiwanese universities, it is suggested that tasks that involve using authentic language (e.g. games or problem-solving tasks) are also utilised in order to the students with a chance to use English.

10.4 Implications for Assessment Methods

As communicating and interacting in English are now emphasised in university listening and speaking classrooms, assessment will only be appropriate and meaningful if it assesses students’ communicative ability to listen and speak in real-life or simulated real-life situations. However, in the present study, communicative language testing was not used in either the mid-term or the final exams in any of the four cases. Instead, discrete-point and integrative testing, such as response evaluation, true/false, cloze tests, or dictations, were frequently used. The problem is that while discrete-point testing simply tests students’ ability to recognise correct language forms or answers from decontextualised listening items, and integrative listening items focus on students’ processing of language, neither method takes face-to-face spontaneous communicative language exchanges between people into consideration. Although oral exams *were* used, the test contents remained non-interactive, focusing instead on reproductions of the existing listening transcripts. Thus, the two testing methods neither assessed the students’ communicative ability nor

helped them to develop communicative thinking. Speaking ability can be elicited in a number of ways. Interactive possibilities include stage shows or speech debates that involve groups, as these allow students to interact with each other; alternatively, teachers may have discussions with students on certain topics. Non-interactively, student presentations can also be used. However, no matter how listening and speaking skills are taught, it is suggested that the topics chosen by the teachers should be related to the knowledge the students have learned in class, so that the results can provide a better indication of the extent to which, and how successfully they are able to use the target language.

It was argued in Chapter Nine that using two tests consisting of different formats or involving large jumps in difficulty levels caused a problem when comparing students' progress in class. It is therefore recommended that teachers, as testers, need to carefully *preview* or *pilot* test items before administering them to students, particularly for widely-used test formats, such as multiple-choice and cloze tests that require more time to design, test, and retest. However, it was discovered that the teachers did not have enough time to pilot test items for the two exams or quizzes, so I would suggest that as much as possible, teachers need to *review* and *reflect on* the items that large numbers of students get wrong, in order to have a better idea of item difficulty before choosing or designing the next tests. Alternatively, I suggest that teachers can use clearly-stated short-answer questions with specific short answers to minimize the problem of piloting multiple-choice questions.

While determining the test contents in both the mid-term and final exams, the four teachers had their reasons for using existing tests or for creating their own tests in relation to the lessons they had taught. Nonetheless, not all their reasons were valid or appropriate in the light of the purposes of the English listening and

speaking class. In University B, the test methods and test contents used were invalid, in the sense that the two teachers regarded the reading and grammar tests as measuring the students' listening ability, and took reading tests as measures of students' speaking ability, but they still claimed that they had tested what they taught in class. Such reading and grammar tests seriously invalidate not only the purpose of testing listening ability, but also the interpretation of the students' listening scores. In short, the results of the tests could not even represent how much the students understood the in-class listening materials, but simply represented their reading and grammar comprehension.

More worryingly, the teachers in the present study tended to regard the students' test scores as a reliable indication of how much their students understood in the exams. Testing listening is a complicated issue which involves many different variables and perceptions among individual students. If the assessment methods within a class are not consistent, test scores can be highly unreliable. It is suggested that teachers should minimise potential variables that may affect students' listening comprehension such as the selection of test formats, test rubrics, topics, speech rate, the quality of tape recording, and the test environment, which would improve the accuracy of assessment.

10.5 Implications for Policy and Practice

As discussed in the Introduction and Chapter 1.2, the purpose of dividing the students into smaller groups was to provide each student with more opportunities to get actively involved in communicative tasks in class. However, based on the classroom observation data and students' opinions from the questionnaires, the students themselves did not consider that they were given many chances to speak in class. The student-talk time was frequently confined to answering exercise

questions in the textbooks or the teaching handouts. Although the government encourages understanding and using English in English courses at university level in Taiwan (see Chapter 1.3) and assumes that listening and speaking will be important at the point of classroom communication and interaction, in fact only limited task-based instruction with *communicative* features was discovered in any of the listening classes observed. In the present study, teaching students in smaller groups did *not* give relatively more opportunities for the students to speak English. If the MOE guidelines are to be implemented, students will need to be given more opportunities to speak English and to establish learning autonomy, otherwise it is pointless to divide them in different groups.

This is not to suggest that teaching students in smaller groups is not beneficial for teaching and learning. If teaching in smaller groups is possible, the department head or programme leader should talk to the course instructors together, let them discuss the programme between themselves and agree on what they intend to teach. In the present case studies, as well as in the preliminary interviews with the ten different teachers, the teachers did not know each other's teaching materials or test methods, and it is clear that the teachers did not discuss with each other their students' level of English ability. The freedom to decide on the teaching materials between teachers produced neither appropriate nor fair judgements on the students' performance within and across groups in the two universities, and this can only have reduced the quality of the teaching and learning. It is suggested that the teachers need to reach a consensus on the teaching materials and the test methods based on the students' level of English proficiency in order to minimise the variables involved in judging their performance. As discussed in Chapter 1.3, the HEEACT evaluation emphasises administrative management of curriculum design, students' opinions towards

certain programmes, teacher qualifications, and alumni performance, rather than directly exploring the in-class situation of teaching and assessment. Inasmuch as the data reported in Chapters Five to Eight failed to meet the MOE aim of creating a communicative English teaching environment within universities, the problem may be because the whole topic is not directly included in the HEEACT evaluation. Thus, one recommendation is that, as the proposal for an English teaching environment has implications at a national and an international level, the MOE and the HEEACT evaluation procedures need to be integrated, and the teaching and testing in English listening classrooms need to be evaluated for the purpose of improving teaching to meet students' levels and needs. English major students need this skill while taking other courses, and after graduation.

The survey of Taiwanese universities showed that less than 50% of these higher-education institutions implemented individual self-evaluation of teaching, and where self-evaluation was present, the criteria used for it varied from institution to institution. I suggest that university self-evaluations should both address the English teaching and learning environment and publish the results; government can help promote this. Besides this, the decision-makers of the English listening and speaking programmes, such as the chairpersons or course instructors, can evaluate and reflect on their teaching materials, teaching approaches, teacher-student interactions, and assessment methods to improve the communicative approach they have used in class.

10.6 The Limitations of the Study

In the present thesis, "triangulation", "persistent observation" and "referential adequacy" techniques were used to validate the reliability and validity of the case studies (see Chapter 3.6). However, one limitation was that it was hard to

generalise every TBI characteristic by using a fixed set of TBI criteria, as closed lists inevitably lead one to ignore or miss more subtle aspects of whether the classes were “communicative”. However, in order to code, analyse and compare the data across the four cases systematically and consistently, a compromise category set was needed. Besides this, although I narrowed down the TBI features in the checklist, I was actually in each lesson I observed, noting down important events, and describing them in the findings.

The other limitation was that I was only allowed to observe six lessons plus the two examinations (i.e. eight lessons in total) for each class. My observations occupied just half of the total lessons in a whole semester. I tried to prolong the observation, but observation time was limited by the teachers. Fortunately, the *procedures* of teaching English listening across the six lessons in each teacher’s class were very similar, in the sense that there was no marked change of teaching sequences. However, washback effects were only found in two classrooms. If the observations had been prolonged, it is possible that different types of washback impact would have been discovered in the other two classrooms.

As discussed in Chapter 2.7, the results of tests may or may not have had an impact on teaching and learning, or not at all. It is harder to detect the impact on learning than on teaching, because teaching tends to be explicit and visible, but learning is implicit and not always easy to observe in class. The washback effects on students’ learning may show up important changes in students’ learning behaviour or processes, but it can take time and persistent contact with the students to collect high quality relevant information. Meanwhile, as washback effects may have a different impact on different individuals, it can be very time-consuming to collect all the necessary information from all the students.

In conclusion, despite the fact that many previous studies on the teaching and testing listening comprehension in foreign language classrooms have been carried out by numerous researchers, how English listening and speaking classes are taught, assessed, and evaluated in Taiwanese universities has, until now, remained unexplored. The results of the four case studies in the present thesis have raised a number of important issues about what actually happens in English listening and speaking classrooms in Taiwanese universities with implications for policy and practice at both national and institutional levels. This has been a very small exploratory study, but the results are such that a larger, more comprehensive study is now warranted, perhaps randomly selecting departments and institutions. A final point is that the students in the present study were all English majors. These students were already fluent and exposed to hours of English. The problem of teaching and learning observed may have far greater impact on students majoring in other areas, who are exposed to less English. As quality assurance is now recognised as important in university programme evaluation, there is an urgent need for English listening instructors in Taiwan to inspect the way they teach and test in class, and also pay attention to the materials they use in relation to students' needs and language levels. In other words, because the purpose of educational change in Taiwan implies innovation and progress in higher education, it is important that English listening courses are evaluated via a well-developed self-evaluation system within each university and via the national evaluation project. The evaluation of classroom activities should not be the "end" of a programme, but the "beginning" of planning and innovation of teaching approaches, materials, and assessment methods.

Appendix A

Piloting Version of the Questionnaires, Interview Questions, and Classroom Observation Checklist

A.1 Questionnaire for the Mid-term Exam

Questionnaire on Taiwanese University Students' Opinions towards the English
Listening Course and the Mid-term Exam

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. This questionnaire is intended to investigate the relationship between the English listening course and Taiwanese university students' listening needs, their general preferences concerning the course and in-class tasks, their opinions about the mid-term exam, and their general comments on the course and the test. This questionnaire is for academic research use and your answers will not be revealed other than when reporting the work.

PERSONAL INFORMATION

1. Department: _____
2. Year: _____
3. Gender: M F

Part One – About the course before the mid-term exam

4. Do you think the contents of the in-class English listening materials have been difficult to understand? Please put a tick (✓)

1 2 3 4 5
 never : ____ : ____ : ____ : ____ : ____ : always

If "1" NEVER, go on to Q2

If you tick 2 to 5, what in particular has been hard to understand? Please tick (✓) the appropriate box(es).

- Topical content Vocabulary Accent Speech rate
 The use of colloquial language Sound quality
 Text type (e.g. *news broadcasts*, *lecturette*, and *consultative dialogue*); if you tick this box, please specify which of the three has given you most difficulty in understanding the content? _____ Please explain briefly: _____

5. Are the course contents relevant to your listening needs?
 Yes, please explain. _____
 No, please explain. _____

Please continue overleaf

Part Two – Your general preferences about English listening classes

6. In which mode of answering in English listening comprehension classes do you think you perform better?

Writing

Speaking

Please explain. _____

7. Which type of speech in English listening comprehension classes do you think you can understand better?

Monologue (only one person speaks in the listening passage)

Conversations/dialogues (more than two people interaction)

Please explain. _____

8. What type of comprehension question do you prefer most for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

9. What type of comprehension question do you prefer least for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

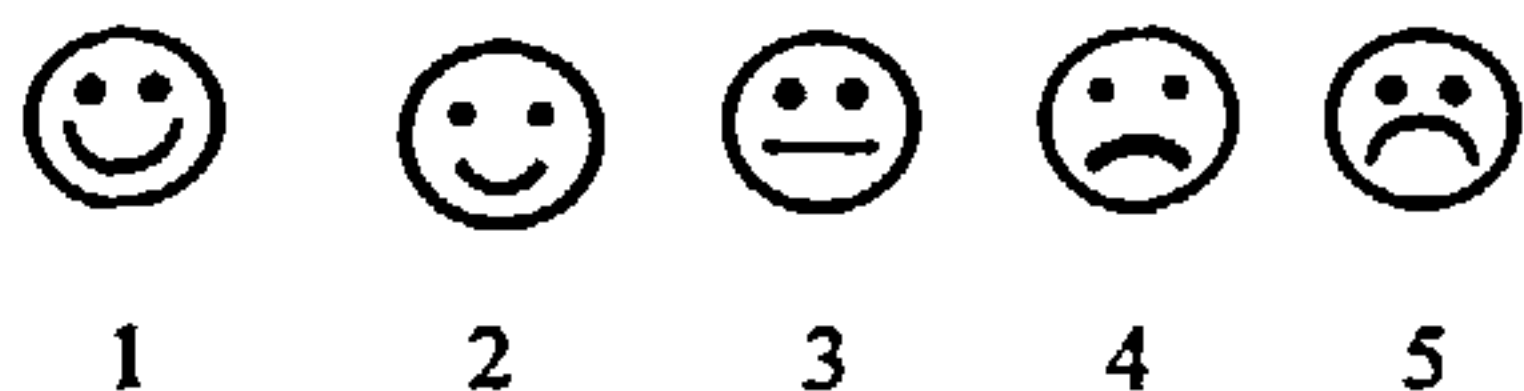
True/false questions

Cloze questions

Please explain. _____

Part Three – About the mid-term exam

10. By and large, were you satisfied with your performance on this test? Please circle ○ the number which best describes your feeling.



Please turn overleaf

11. Listening passages can be difficult for many reasons. **IN THIS TEST**, here are ten possible reasons. Can you say how far each of these proved difficult? Please circle (O) the appropriate number.

1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree

IN THIS TEST,

	SD					SA
	1	2	3	4	5	
a. The topics of the test tasks were difficult.	1	2	3	4	5	
b. The topics of the test tasks were representative of the curriculum taught in class.	1	2	3	4	5	
c. The test tasks were harder than those used in class.	1	2	3	4	5	
d. The accent was too difficult to understand.	1	2	3	4	5	
e. The vocabulary was difficult.	1	2	3	4	5	
f. The sentences were too complicated to understand.	1	2	3	4	5	
g. It was hard to understand what speaker(s) said because they spoke fast.	1	2	3	4	5	
h. It was hard to understand monologue speech.	1	2	3	4	5	
i. It was hard to understand conversations/dialogues.	1	2	3	4	5	

12. **IN THIS TEST**, was colloquial language used?

- Yes No

If YES, did you think that it was difficult to understand the colloquial language?

- Yes No

13. Did you find any of the following to be a problem when you took the mid-term exam? Please tick (✓) the appropriate box(es).

1. The quality of the recording and/or the visual equipment

- a. The quality was so poor that I couldn't hear the text clearly.
 b. The background noise was too loud.
 c. The volume of the tape recorder or audio-visual equipment was too low.
 d. The quality of the recording was clear, and I can hear the texts clearly.
 e. Other: _____

2. Testing time

- a. Time was too limited to answer all the questions properly.
 b. The testing time was sufficient for me to answer all the questions properly.
 c. Other: _____

Please turn overleaf

3. The test/task instructions

- a. The instructions of each test section were not clear.
- b. The instructions were too complicated.
- c. The test/task instructions were clear.
- d. Other: _____

4. The length of the listening texts

- a. The texts in the test were longer than I was used to listening to in class.
- b. Longer texts increased the difficulty of understanding.
- c. The length of texts in the test were similar to what I was used to listening to in class.
- d. Other: _____

14. What type of comprehension question did you find easiest? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

15. What type of comprehension question did you find most difficult in the listening test tasks? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

Part Four – General Comments

16. (a). Have you learned what you expected to learn so far from this class?

- Yes
- No, please explain the reason(s): _____

(b). Were you satisfied with the way the teacher taught so far in class?

- Yes
- No, please explain the reason(s): _____

(c). Were you satisfied with the assessment method the teacher used in the mid-term exam?

- Yes
- No, please explain the reason(s): _____

Thank you very much for taking time to finish this questionnaire!

It is very much appreciated!

7. 你覺得在課堂中，哪種英文聽力內容的表達方式你比較能理解？

- 獨白演說 (只有一個人說話) 對話 (兩個人以上的對話互動)

請詳述原因： _____

8. 哪一種問題類型是你在練習聽力練習時最喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

9. 哪一種問題類型是你在練習聽力作業時最不喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

第三部分 – 有關這次期中考試

10. 整體來說，你對於這次的考試表現滿意嗎？請圈選 ○ 以下最符合你的感覺的笑臉數字。



11. 聽力內容可能因為很多原因而難以了解，在這個考試裡，有十個可能造成聽力內容難以理解的原因，請你指出以下的每一項對你來說有多難理解，請圈選 (○) 適當的號碼。

1 = 強烈不認同; 2 = 不認同; 3 = 沒有認同也沒有不認同; 4 = 認同; 5 = 強烈認同

在這個考試裡，

	強不					強認				
	1	2	3	4	5	1	2	3	4	5
a. 考題主題困難。										
b. 考題主題對課堂所教的課程內容而言是具有代表性的。										
c. 考題比課堂上所做的練習題還難。										
d. 說話者的口音太重以致於內容難以了解。										
e. 某些單字難以理解。										
f. 句型複雜以至於內容難以了解。										
g. 因為說話者速度太快，我很難了解其內容。										
h. 我覺得獨白式的演說難以理解。										
i. 我覺得對話式的内容難以理解。										

請翻頁繼續作答

12. 在這次的考試中，你有聽到過於白話的用語嗎？請勾選 (✓) 下列選項裡的一個答案。

有 沒有

如果『有』，你是否覺得當說話者使用過於白話的用語時讓你很難了解其聽力內容？

有 沒有

13. 在期中考試時，是否有以下的陳述中的問題？請勾選(✓)符合的選項 (選項可多重)。

1. 錄音和/或視覺器材的品質。

a. 錄音帶或視覺器材聲音品質太差以致於我無法清楚聽到內容。

b. 背景噪音太大。

c. 錄音機或視覺器材的音量太小。

d. 錄音帶品質好，我可以清楚的聽到內容。

e. 其他：_____

2. 測驗時間。

a. 時間有限以致於難以回答所有問題。

b. 時間充足可以回答所有問題。

c. 其他：_____

3. 考試或考題指示。

a. 題目的指示不清楚。

b. 指示太過複雜。

c. 題目指示清楚。

d. 其他：_____

4. 聽力內容的長度。

a. 考題聽力長度比我平常在課堂上所聽的還要長。

b. 較長的聽力內容會增加理解上的困難。

c. 考題聽力長度跟我平常在課堂上所聽的差不多。

d. 其他：_____

14. 你覺得哪種類型的問題最簡單？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

15. 你覺得哪種類型的問題最難？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

請翻頁繼續作答

第四部份 – 意見與建議

16. (1). 在這門課裡，你有學到你預期想學的東西嗎？

是

否，請詳述原因： _____

(2). 你對期中考之前授課老師的教學滿意嗎？

是

否，請詳述原因： _____

(3). 你對授課老師的期中考試測驗方式滿意嗎？

是

否，請詳述原因： _____

問卷完，十分感謝您抽空回答這份問卷！

A.3 Questionnaire for the Final Exam

Questionnaire on Taiwanese University Students' Opinions towards the English

Listening Course and the Final Exam

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. This questionnaire is intended to investigate Taiwanese university students' opinion about the English listening course, general preferences concerning the course and in-class tasks after the mid-term exam, their opinions about final exam, and their general comments on the course and the test. This questionnaire is for academic research use and your answers will not be revealed other than when reporting the work.

PERSONAL INFORMATION

1. Department: _____
2. Year: _____
3. Gender: M F

Part One – About the course after the mid-term exam

4. Do you think the contents of the in-class English listening materials after the mid-term exam have been difficult to understand? Please put a tick (✓)

1 2 3 4 5
never : ___ : ___ : ___ : ___ : ___ : always

If "1" NEVER, go on to Q2

If you tick 2 to 5, what in particular has been hard to understand? Please tick (✓) the appropriate box(es).

- Topical content Vocabulary Accent Speech rate
 The use of colloquial language Sound quality
 Text type (e.g. *news broadcasts*, *lecturette*, and *consultative dialogue*); if you tick this box, please specify which of the three has given you most difficulty in understanding the content? _____ Please explain briefly: _____

Part Two – Your general preferences about English listening classes

5. In which mode of answering in English listening comprehension classes do you think you perform better?

Writing Speaking

Please explain. _____

Please continue overleaf

6. Which type of speech in English listening comprehension classes do you think you can understand better?

Monologue (only one person speaks in the listening passage)

Conversations/dialogues (more than two people interaction)

Please explain. _____

7. What type of comprehension question do you prefer most for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

8. What type of comprehension question do you prefer least for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

Part Three – About the final exam

9. By and large, were you satisfied with your performance on this test? Please circle O the number which best describes your feeling.



1



2



3



4



5

Please turn overleaf

10. Listening passages can be difficult for many reasons. **IN THIS TEST**, here are ten possible reasons. Can you say how far each of these proved difficult? Please circle (O) the appropriate number.

1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree

IN THIS TEST,

	SD				SA
a. The topics of the test tasks were difficult.	1	2	3	4	5
b. The topics of the test tasks were representative of the curriculum taught in class.	1	2	3	4	5
c. The test tasks were harder than those used in class.	1	2	3	4	5
d. The accent was too difficult to understand.	1	2	3	4	5
e. The vocabulary was difficult.	1	2	3	4	5
f. The sentences were too complicated to understand.	1	2	3	4	5
g. It was hard to understand what speaker(s) said because they spoke fast.	1	2	3	4	5
h. It was hard to understand monologue speech.	1	2	3	4	5
i. It was hard to understand conversations/dialogues.	1	2	3	4	5

11. **IN THIS TEST**, was the colloquial language used?

- Yes No

If **YES**, did you think that it was difficult to understand the colloquial language?

- Yes No

12. Did you find any of the following to be a problem when you took the final exam? Please tick (✓) the appropriate box(es).

1. The quality of the recording and/or the visual equipment

- a. The quality was so poor that I couldn't hear the text clearly.
 b. The background noise was too loud.
 c. The volume of the tape recorder or audio-visual equipment was too low.
 d. The quality of the recording was clear, and I can hear the texts clearly.
 e. Other: _____

2. Testing time

- a. Time was too limited to answer all the questions properly.
 b. The testing time was sufficient for me to answer all the questions properly.
 c. Other: _____

Please turn overleaf

3. The test/task instructions

- a. The instructions of each test section were not clear.
- b. The instructions were too complicated.
- c. The test/task instructions were clear.
- d. Other: _____

4. The length of the listening texts

- a. The texts in the test were longer than I was used to listening to in class.
- b. Longer texts increased the difficulty of understanding.
- c. The length of texts in the test was similar to what I was used to listening to in class.
- d. Other: _____

14. What type of comprehension question did you find easiest? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

15. What type of comprehension question did you find most difficult in the listening test tasks? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

Part Four – General Comments

16. (a). Have you learned what you expected to learn after the mid-term exam?

- Yes
- No, please explain the reason(s): _____

(b). Were you satisfied with the way the teacher taught after the mid-term exam?

- Yes
- No, please explain the reason(s): _____

(c). Were you satisfied with the assessment method the teacher used in the final exam?

- Yes
- No, please explain the reason(s): _____

Thank you very much for taking time to finish this questionnaire!

A.4 Questionnaire for the Final Exam (Chinese Version)

台灣大學生對於英文聽力課程及期末考之態度問卷調查

您好！我是英國約克大學教育系的博士班學生，周牧璇。這份問卷是調查台灣大學生於期中考後對聽力課程、英文聽力課程與課堂上聽力練習的喜好與習慣的看法、您對期末考的意見以及您對聽力課程及考試的建議。這份問卷將做為學術用途，您的回答將僅作為此次研究，所有資料絕對保密。

個人資訊

1. 科系: _____
2. 年級: _____
3. 性別: 男 女

第一部分 – 有關期中考後的課程

4. 你認為期中考之後課堂上的英文聽力材料內容很難被理解嗎？請勾選 (✓)

1 2 3 4 5

從來不 : ___ : ___ : ___ : ___ : ___ : 總是

如果選“1”從來不，請直接回答第二題

如果你勾 2 to 5, 你覺得是什麼原因讓你覺得聽力內容難以了解？請在框框裡勾選適當的答案 (選項可以多於一項)。

- 主題內容 單字 口音 說話速度
 口語化用詞 聲音品質

聽力內容類型 (例如: 新聞報導、演講跟諮詢式對話); 如果您勾選此欄，請您具體說明以上三項類型中哪種讓你覺得最難理解其內容？_____

請簡述原因: _____

第二部分 – 您對英文聽力課程的一般喜好

5. 你覺得哪一種答題模式會讓你在聽力課堂上回答的比較好？

- 書寫答題模式 口語答題模式

請詳述原因: _____

6. 你覺得在課堂中，哪種英文聽力內容的表達方式你比較能理解？

- 獨白演說 (只有一個人說話)
 對話 (兩個人以上的對話互動)

請詳述原因: _____

請翻頁繼續作答

7. 哪一種問題類型是你在練習聽力練習時最喜歡的？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

請詳述原因： _____

8. 哪一種問題類型是你在練習聽力作業時最不喜歡的？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

請詳述原因： _____

第三部分 - 有關這次期末考試

9. 整體來說，你對於這次的考試表現滿意嗎？請圈選 ○ 以下最符合你的感覺的笑臉數字。



10. 聽力內容可能因為很多原因而難以了解，在這個考試裡，有十個可能造成聽力內容難以理解的原因，請你指出以下的每一項對你來說有多難理解，請圈選 (○) 適當的號碼。

1 = 強烈不認同; 2 = 不認同; 3 = 沒有認同也沒有不認同; 4 = 認同; 5 = 強烈認同

在這個考試裡，

	強不			強認	
	1	2	3	4	5
a. 考題主題困難。	1	2	3	4	5
b. 考題主題對課堂所教的課程內容而言是具有代表性的。	1	2	3	4	5
c. 考題比課堂上所做的練習題還難。	1	2	3	4	5
d. 說話者的口音太重以致於內容難以了解。	1	2	3	4	5
e. 某些單字難以理解。	1	2	3	4	5
f. 句型複雜以至於內容難以了解。	1	2	3	4	5
g. 因為說話者速度太快，我很難了解其內容。	1	2	3	4	5
h. 我覺得獨白式的演說難以理解。	1	2	3	4	5
i. 我覺得對話式的內容難以理解。	1	2	3	4	5

請翻頁繼續作答

11. 在這次的考試中，你有聽到過於白話的用語嗎？請勾選 (✓) 下列選項裡的一個答案。

- 有 沒有

如果『有』，你是否覺得當說話者使用過於白話的用語時讓你很難了解其聽力內容？

- 有 沒有

12. 在期末考試時，是否有以下的陳述中的問題？請勾選(✓)符合的選項 (可多重選項)

1. 錄音和/或視覺器材的品質。

- a. 錄音帶或視覺器材聲音品質太差以致於我無法清楚聽到內容。
 b. 背景噪音太大。
 c. 錄音機或視覺器材的音量太小。
 d. 錄音帶品質好，我可以清楚的聽到內容。
 e. 其他：_____

2. 測驗時間。

- a. 時間有限以致於難以回答所有問題。
 b. 時間充足可以回答所有問題。
 c. 其他：_____

3. 考試或考題指示。

- a. 题目的指示不清楚。
 b. 指示太過複雜。
 c. 題目指示清楚。
 d. 其他：_____

4. 聽力內容的長度。

- a. 考題聽力長度比我平常在課堂上所聽的還要長。
 b. 較長的聽力內容會增加理解上的困難。
 c. 考題聽力長度跟我平常在課堂上所聽的差不多。
 d. 其他：_____

13. 你覺得哪種類型的問題最簡單？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

14. 你覺得哪種類型的問題最難？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請翻頁繼續作答

第四部份 - 意見與建議

15. (1). 從期中考後的這門課裡，你有學到您預期想學的東西嗎？

是

否，請詳述原因：_____

(2). 你對期中考之後授課老師的教學滿意嗎？

是

否，請詳述原因：_____

(3). 你對授課老師的期末考試測驗方式滿意嗎？

是

否，請詳述原因：_____

問卷完，十分感謝您抽空回答這份問卷！

A.6 Interview Questions for the Final Exam

1. Where did the content of the test come from? (Chinese translation: 請問您此次考題內容是從哪裡來的?)
2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them? (Chinese translation: 在次此次期末考，根據您的教學計畫與學生的學習成果，請問您想要達到什麼樣的考試目的？請問您達到了嗎?)
3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria? (Chinese translation: 請問您如何來鑑別與決定期末考試內容及考題的難易度？有什麼標準嗎?)
4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions? (Chinese translation: 您為什麼會選擇這種問題類型來考學生？您認為某種類型的學生會喜歡這樣的測驗方式嗎或者您認為出這種類型的問題會讓這些學生考的比較好嗎?)

A.7 Classroom Observation Checklist

Classroom Observation Checklist

Case No.: _____

The purposes of the Classroom Observation Checklist are to investigate teachers' and students' in-class behaviours as well as the in-class atmosphere, including the interaction between teachers and students and its relation to the curriculum objectives.

No. of Observation: _____ Date: _____

Time Segments in Minutes	5	10	15	20	25	30	35	40	45	50
Q1. Teacher Role										
① Lecturing	①	①	①	①	①	①	①	①	①	①
② Discussing with whole group	②	②	②	②	②	②	②	②	②	②
③ Managing feedback discussion	③	③	③	③	③	③	③	③	③	③
④ Facilitating / coaching	④	④	④	④	④	④	④	④	④	④

Time Segments in Minutes	55	60	65	70	75	80	85	90	95	100
Q1. Teacher Role										
① Lecturing	①	①	①	①	①	①	①	①	①	①
② Discussing with whole group	②	②	②	②	②	②	②	②	②	②
③ Managing feedback discussion	③	③	③	③	③	③	③	③	③	③
④ Facilitating / coaching	④	④	④	④	④	④	④	④	④	④

Time Segments in Minutes	5	10	15	20	25	30	35	40	45	50
Q2. Student Involvement										
① Paying attention to the lecture	①	①	①	①	①	①	①	①	①	①
② Paying attention to other students' speech or presentation	②	②	②	②	②	②	②	②	②	②
③ Discussing with the teacher	③	③	③	③	③	③	③	③	③	③
④ Discussing with each other	④	④	④	④	④	④	④	④	④	④
⑤ Doing listening tasks	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥ Feedback from small groups	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥

Time Segments in Minutes	55	60	65	70	75	80	85	90	95	100
Q2. Student Involvement										
① Paying attention to the lecture	①	①	①	①	①	①	①	①	①	①
② Paying attention to other students' speech or presentation	②	②	②	②	②	②	②	②	②	②
③ Discussing with the teacher	③	③	③	③	③	③	③	③	③	③
④ Discussing with each other	④	④	④	④	④	④	④	④	④	④
⑤ Doing listening tasks	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥ Feedback from small groups	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥

Q3. Task-based Instruction Features	Obs. No.	Observed or not	Comments
1. There is at least one problem-solving task for students to do in class.			
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.			
3. Students report findings of a task to class, in groups or pairs, after problem solving.			
4. The major focus of teaching is on the meaning and then on the form.			
5. Students were given opportunities to reflect on what they have learned and how well they were doing (i.e. reflection period).			

Appendix B

Questionnaires, Interview Questions, and Classroom Observation Form for Main Study

B.1 – Questionnaire for the Mid-term Exam

Questionnaire on Taiwanese University Students' Opinions towards the English
Listening Course and the Mid-term Exam

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. This questionnaire is intended to investigate the relationship between the English listening course and Taiwanese university students' listening needs, their general preferences concerning the course and in-class tasks, their opinions about the mid-term exam, and their general comments on the course and the test. This questionnaire is for academic research use and your answers will not be revealed other than when reporting the work.

PERSONAL INFORMATION

1. Department: _____
2. Year: _____
3. Gender: M F

Part One – About the course

4. Did you think the contents of the in-class English listening materials have been difficult to understand? Please put a tick (✓)

1 2 3 4 5

never : ____ : ____ : ____ : ____ : ____ : always

If "1" NEVER, go on to Q2

If you tick 2 to 5, what in particular has been hard to understand? Please tick (✓) the appropriate box(es).

- Topical content Vocabulary Accent Speech rate
- The use of colloquial language Sound quality
- Text type (e.g. *news broadcasts*, *lecturette*, and *consultative dialogue*); if you tick this box, please specify which of the three has given you most difficulty in understanding the content? _____ Please explain briefly: _____

5. Were the course contents relevant to your listening needs?
 - Yes, please explain. _____
 - _____
 - No, please explain. _____
 - _____

Please continue overleaf

Part Two – Your general preferences about English listening classes

6. In which mode of answering in English listening comprehension classes do you think you perform better?

- Writing Speaking

Please explain. _____

7. Which type of speech in English listening comprehension classes do you think you can understand better?

- Monologue (only one person speaks in the listening passage)
 Conversations/dialogues (more than two people interaction)

Please explain. _____

8. What type of comprehension question do you prefer most for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
 True/false questions Cloze questions

Please explain. _____

9. What type of comprehension question do you prefer least for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
 True/false questions Cloze questions

Please explain. _____

Part Three – About the mid-term exam

10. By and large, were you satisfied with your performance on this test? Please circle O the number which best describes your feeling. Descriptors for the smiley faces are:

1. very satisfied; 2. satisfied; 3. Neither satisfied nor dissatisfied;
4. dissatisfied; 5. very dissatisfied



Please continue overleaf

11. Listening passages can be difficult for many reasons. **IN THIS TEST**, here are nine possible reasons. Can you say how far each of these proved difficult? Please circle (O) the appropriate number.

1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree

IN THIS TEST,

	SD				SA
a. The topics of the test tasks were difficult.	1	2	3	4	5
b. The topics of the test tasks were representative of the curriculum taught in class.	1	2	3	4	5
c. The test tasks were harder than those used in class.	1	2	3	4	5
d. The accent was too difficult to understand.	1	2	3	4	5
e. The vocabulary was difficult.	1	2	3	4	5
f. The sentences were too complicated to understand.	1	2	3	4	5
g. It was hard to understand what speaker(s) said because they spoke fast.	1	2	3	4	5
h. It was hard to understand monologue speech.	1	2	3	4	5
i. It was hard to understand conversations/dialogues.	1	2	3	4	5

12. **IN THIS TEST**, was colloquial language used?

Yes No

If YES, did you think that it was difficult to understand the colloquial language?

Yes No

13. Did you find any of the following to be a problem when you took the mid-term exam? Please tick (✓) the appropriate box(es).

1. The quality of the recording and/or the visual equipment

- a. The quality was so poor that I couldn't hear the text clearly.
- b. The noise outside the testing environment was too loud.
- c. The volume of the tape recorder or audio-visual equipment was too low.
- d. The quality of the recording was clear, and I can hear the texts clearly.
- e. Other: _____

2. Testing time

- a. Time was too limited to answer all the questions properly.
- b. The testing time was sufficient for me to answer all the questions properly.
- c. Other: _____

Please continue overleaf

3. The test/task instructions

- a. The instructions of each test section were not clear.
- b. The instructions were too complicated.
- c. The test/task instructions were clear.
- d. Other: _____

4. The length of the listening texts

- a. The texts in the test were longer than I was used to listening to in class.
- b. Longer texts increased the difficulty of understanding.
- c. The length of texts in the test was similar to what I was used to listening to in class.
- d. Other: _____

14. IN THIS TEST, what type of comprehension question did you find easiest in this test? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

15. IN THIS TEST, what type of comprehension question did you find most difficult in the listening test tasks? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

Part Four – General Comments

16. (a). Have you learned what you expected to learn so far from this class?

- Yes
- No, please explain the reason(s): _____

(b). Were you satisfied with the way the teacher taught in class?

- Yes
- No, please explain the reason(s): _____

(c). Were you satisfied with the assessment method the teacher used in the mid-term exam?

- Yes
- No, please explain the reason(s): _____

Thank you very much for taking time to finish this questionnaire!

It is very much appreciated!

7. 你覺得在課堂中，哪種英文聽力內容的表達方式你比較能理解？

- 獨白演說 (只有一個人說話)
 對話 (兩個人以上的對話互動)

請詳述原因： _____

8. 哪一種問題類型是你在練習聽力練習時最喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

9. 哪一種問題類型是你在練習聽力作業時最不喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

第三部分 - 有關這次期中考試

10. 整體來說，你對於這次的考試表現滿意嗎？請圈選 ○ 以下最符合你的感覺的笑臉數字。笑臉符號的敘述為：

1. 非常滿意；2. 滿意；3. 沒有滿意也沒有不滿意；4. 不滿意；5. 非常不滿意。



1 2 3 4 5

請翻頁繼續作答

11. 聽力內容可能因為很多原因而難以了解，在這個考試裡，有九個可能造成聽力內容難以理解的原因，請你指出以下的每一項對你來說有多難理解，請圈選 (○) 適當的號碼。

1 = 強烈不認同; 2 = 不認同; 3 = 沒有認同也沒有不認同; 4 = 認同; 5 = 強烈認同

在這個考試裡，

	強不			強認	
	1	2	3	4	5
a. 考題主題困難。	1	2	3	4	5
b. 考題主題對課堂所教的課程內容而言是具有代表性的。	1	2	3	4	5
c. 考題比課堂上所做的練習題還難。	1	2	3	4	5
d. 說話者的口音太重以致於內容難以了解。	1	2	3	4	5
e. 某些單字難以理解。	1	2	3	4	5
f. 句型複雜以至於內容難以了解。	1	2	3	4	5
g. 因為說話者速度太快，我很難了解其內容。	1	2	3	4	5
h. 我覺得獨白式的演說難以理解。	1	2	3	4	5
i. 我覺得對話式的內容難以理解。	1	2	3	4	5

12. 在這次的考試中，你有聽到過於白話的用語嗎？請勾選 (✓) 下列選項裡的一個答案。

有 沒有

如果『有』，你是否覺得當說話者使用過於白話的用語時讓你很難了解其聽力內容？

有 沒有

13. 在期中考試時，是否有以下的陳述中的問題？請勾選(✓)符合的選項 (可多重選項)。

1. 錄音和/或視覺器材的品質。

a. 錄音帶或視覺器材聲音品質太差以致於我無法清楚聽到內容。

b. 外在環境噪音太大。

c. 錄音機或視覺器材的音量太小。

d. 錄音帶品質好，我可以清楚的聽到內容。

e. 其他：_____

2. 測驗時間。

a. 時間有限以致於難以回答所有問題。

b. 時間充足可以回答所有問題。

c. 其他：_____

3. 考試或考題指示。

a. 题目的指示不清楚。

b. 指示太過複雜。

c. 题目指示清楚。

d. 其他：_____

請翻頁繼續作

4. 聽力內容的長度。

- a. 考題聽力長度比我平常在課堂上所聽的還要長。
- b. 聽力長度太長造成理解上的困難。
- c. 考題聽力長度跟我平常在課堂上所聽的差不多。
- d. 其他：_____

14. 在這次的考試中，你覺得哪種類型的問題最簡單？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
- 是非題 填充題

15. 在這次的考試中，你覺得哪種類型的問題最難？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
- 是非題 填充題

第四部份 - 意見與建議

16. (1). 在這門課裡，你有學到你預期想學的東西嗎？

- 是
 - 否，請詳述原因：_____
- _____

(2). 你對期中考之前授課老師的教學滿意嗎？

- 是
 - 否，請詳述原因：_____
- _____

(3). 你對授課老師的期中考試測驗方式滿意嗎？

- 是
 - 否，請詳述原因：_____
- _____

問卷完，十分感謝您抽空回答這份問卷！

B.3 – Questionnaire for the Final Exam

Questionnaire on Taiwanese University Students' Opinions towards the English
Listening Course and the Final Exam

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. This questionnaire is intended to investigate Taiwanese university students' opinion about the English listening course, general preferences concerning the course and in-class tasks after the mid-term exam, their opinions about final exam, and their general comments on the course and the test. This questionnaire is for academic research use and your answers will not be revealed other than when reporting the work.

PERSONAL INFORMATION

1. Department: _____
2. Year: _____
3. Gender: M F

Part One – About the course after the mid-term

4. Did you think the contents of the in-class English listening materials after the mid-term exam have been difficult to understand? Please put a tick (✓)

1 2 3 4 5
never : ___ : ___ : ___ : ___ : ___ : always

If "1" NEVER, go on to Q2

If you tick 2 to 5, what in particular has been hard to understand? Please tick (✓) the appropriate box(es).

- Topical content Vocabulary Accent Speech rate
 The use of colloquial language Sound quality
 Text type (e.g. *news broadcasts*, *lecturette*, and *consultative dialogue*); if you tick this box, please specify which of the three has given you most difficulty in understanding the content? _____ Please explain briefly: _____

Part Two – Your general preferences about English listening classes

5. In which mode of answering in English listening comprehension classes do you think you perform better?

- Writing Speaking

Please explain. _____

Please continue overleaf

6. Which type of speech in English listening comprehension classes do you think you can understand better?

Monologue (only one person speaks in the listening passage)

Conversations/dialogues (more than two people interaction)

Please explain. _____

7. What type of comprehension question do you prefer most for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

8. What type of comprehension question do you prefer least for the listening tasks?

Please put only ONE tick (✓) in the appropriate box.

Short-answer questions

Multiple-choice questions

Dictation questions

True/false questions

Cloze questions

Please explain. _____

Part Three – About the final exam

9. By and large, were you satisfied with your performance on this test? Please circle ○ the number which best describes your feeling. Descriptors for the smiley faces are:

1. very satisfied; 2. satisfied; 3. Neither satisfied nor dissatisfied;

4. dissatisfied; 5. very dissatisfied



1 2 3 4 5

Please continue overleaf

10. Listening passages can be difficult for many reasons. **IN THIS TEST**, here are nine possible reasons. Can you say how far each of these proved difficult? Please circle (O) the appropriate number.

1 = strongly disagree; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree

IN THIS TEST,

	SD				SA
a. The topics of the test tasks were difficult.	1	2	3	4	5
b. The topics of the test tasks were representative of the curriculum taught in class.	1	2	3	4	5
c. The test tasks were harder than those used in class.	1	2	3	4	5
d. The accent was too difficult to understand.	1	2	3	4	5
e. The vocabulary was difficult.	1	2	3	4	5
f. The sentences were too complicated to understand.	1	2	3	4	5
g. It was hard to understand what speaker(s) said because they spoke fast.	1	2	3	4	5
h. It was hard to understand monologue speech.	1	2	3	4	5
i. It was hard to understand conversations/dialogues.	1	2	3	4	5

11. **IN THIS TEST**, was colloquial language used?

Yes No

If **YES**, did you think that it was difficult to understand the colloquial language?

Yes No

12. Did you find any of the following to be a problem when you took the final exam?

Please tick (✓) the appropriate box(es).

1. The quality of the recording and/or the visual equipment

- a. The quality was so poor that I couldn't hear the text clearly.
- b. The noise outside the testing environment was too loud.
- c. The volume of the tape recorder or audio-visual equipment was too low.
- d. The quality of the recording was clear, and I can hear the texts clearly.
- e. Other: _____

2. Testing time

- a. Time was too limited to answer all the questions properly.
- b. The testing time was sufficient for me to answer all the questions properly.
- c. Other: _____

Please continue overleaf

3. The test/task instructions

- a. The instructions of each test section were not clear.
- b. The instructions were too complicated.
- c. The test/task instructions were clear.
- d. Other: _____

4. The length of the listening texts

- a. The texts in the test were longer than I was used to listening to in class.
- b. Longer texts increased the difficulty of understanding.
- c. The length of texts in the test were similar to what I was used to listening to in class.
- d. Other: _____

13. IN THIS TEST, what type of comprehension question did you find easiest in this test? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

14. IN THIS TEST, what type of comprehension question did you find most difficult in the listening test tasks? Please put only ONE tick (✓) in the appropriate box.

- Short-answer questions Multiple-choice questions Dictation questions
- True/false questions Cloze questions

Part Four – General Comments

15. (a). Have you learned what you expected to learn after the mid-term exam?

- Yes No

If NO, please explain the reason(s): _____

(b). Were you satisfied with the way the teacher taught after the mid-term exam?

- Yes No

If NO, please explain the reason(s): _____

(c). Were you satisfied with the assessment method the teacher used in the final exam?

- Yes No

If NO, please explain the reason(s): _____

(d). Have your English listening skills been improved after finishing this course?

- Yes No

If NO, please explain the reason(s): _____

Thank you very much for taking time to finish this questionnaire!

B.4 – Questionnaire for the Final Exam (Chinese Version)

台灣大學生對於英文聽力課程及期末考之意見問卷調查

您好！我是英國約克大學教育系的博士班學生，周牧璇。這份問卷是調查台灣大學生於期中考後對聽力課程、英文聽力課程與課堂上聽力練習的喜好與習慣的看法、您對期末考的意見以及您對聽力課程及考試的建議。這份問卷將做為學術用途，您的回答將僅作為此次研究，所有資料絕對保密。

個人資料

1. 科系: _____
2. 年級: _____
3. 性別: 男 女

第一部分 – 有關期中考後的課程

4. 你認為期中考之後課堂上的英文聽力材料內容很難被理解嗎？請勾選 (✓)

1 2 3 4 5

從來不 : ___ : ___ : ___ : ___ : ___ : 總是

如果選“1”從來不，請直接回答第二題

如果你勾 2 to 5, 你覺得是什麼原因讓你覺得聽力內容難以了解？請在框框裡勾選適當的答案 (選項可以多於一項)。

主題內容 單字 口音 說話速度

口語化用詞 聲音品質

聽力內容類型 (例如：新聞報導、演講跟諮詢式對話)；如果您勾選此欄，請您具體說明以上三項類型中哪種讓你覺得最難理解其內容？_____

請簡述原因：_____

第二部分 – 您對英文聽力課程的一般喜好

5. 你覺得哪一種答題模式會讓你在聽力課堂上回答的比較好？

書寫答題模式 口語答題模式

請詳述原因：_____

6. 你覺得在課堂中，哪種英文聽力內容的表達方式你比較能理解？

獨白演說 (只有一個人說話) 對話 (兩個人以上的對話互動)

請詳述原因：_____

請翻頁繼續作答

7. 哪一種問題類型是你在練習聽力練習時最喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

8. 哪一種問題類型是你在練習聽力作業時最不喜歡的？請勾選 (✓) 下列選項裡的一個答案。

- 簡答題 選擇題 聽寫題
 是非題 填充題

請詳述原因： _____

第三部分 - 有關這次期末考試

9. 整體來說，你對於這次的考試表現滿意嗎？請圈選 ○ 以下最符合你的感覺的笑臉數字。笑臉符號的敘述為：

1. 非常滿意；2. 滿意；3. 沒有滿意也沒有不滿意；4. 不滿意；5. 非常不滿意。



10. 聽力內容可能因為很多原因而難以了解，在這個考試裡，有九個可能造成聽力內容難以理解的原因，請你指出以下的每一項對你來說有多難理解，請圈選 (○) 適當的號碼。

1 = 強烈不認同；2 = 不認同；3 = 沒有認同也沒有不認同；4 = 認同；5 = 強烈認同

在這個考試裡，

	強不			強認	
	1	2	3	4	5
a. 考題主題困難。	1	2	3	4	5
b. 考題主題對課堂所教的課程內容而言是具有代表性的。	1	2	3	4	5
c. 考題比課堂上所做的練習題還難。	1	2	3	4	5
d. 說話者的口音太重以致於內容難以了解。	1	2	3	4	5
e. 某些單字難以理解。	1	2	3	4	5
f. 句型複雜以至於內容難以了解。	1	2	3	4	5
g. 因為說話者速度太快，我很難了解其內容。	1	2	3	4	5
h. 我覺得獨白式的演說難以理解。	1	2	3	4	5
i. 我覺得對話式的內容難以理解。	1	2	3	4	5

請翻頁繼續作答

11. 在這次的考試中，你有聽到過於白話的用語嗎？請勾選 (✓) 下列選項裡的一個答案。

有 沒有

如果『有』，你是否覺得當說話者使用過於白話的用語時讓你很難了解其聽力內容？

有 沒有

12. 在期末考試時，是否有以下的陳述中的問題？請勾選(✓)符合的選項 (可多重選項)。

1. 錄音和/或視覺器材的品質。

a. 錄音帶或視覺器材聲音品質太差以致於我無法清楚聽到內容。

b. 外在環境噪音太大。

c. 錄音機或視覺器材的音量太小。

d. 錄音帶品質好，我可以清楚的聽到內容。

e. 其他：_____

2. 測驗時間。

a. 時間有限以致於難以回答所有問題。

b. 時間充足可以回答所有問題。

c. 其他：_____

3. 考試或考題指示。

a. 題目的指示不清楚。

b. 指示太過複雜。

c. 題目指示清楚。

d. 其他：_____

4. 聽力內容的長度。

a. 考題聽力長度比我平常在課堂上所聽的還要長。

b. 聽力長度太長造成理解上的困難。

c. 考題聽力長度跟我平常在課堂上所聽的差不多。

d. 其他：_____

13. 在這次的考試中，你覺得哪種類型的問題最簡單？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

14. 在這次的考試中，你覺得哪種類型的問題最難？請勾選 (✓) 下列選項裡的一個答案。

簡答題

選擇題

聽寫題

是非題

填充題

請翻頁繼續作答

第四部份 – 期中考後的意見與建議

15. (1). 從期中考後的這門課裡，你有學到您預期想學的東西嗎？

是

否，請詳述原因： _____

(2). 你對期中考之後授課老師的教學滿意嗎？

是

否，請詳述原因： _____

(3). 你對授課老師的期末考試測驗方式滿意嗎？

是

否，請詳述原因： _____

(4). 你覺得在修完這門課之後你的英文聽力有進步嗎？

有

沒有，請詳述原因： _____

十分感謝您抽空回答這份問卷！

B.6 Interview Questions for the Final Exam

1. Where did the content of the test come from? (Chinese translation: 請問您此次考題內容是從哪裡來的?)
2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them? (Chinese translation: 在次此次期末考，根據您的教學計畫與學生的學習成果，請問您想要達到什麼樣的考試目的？請問您達到了嗎?)
3. How did you identify and decide on difficulty of the content/items in the final test? What were the criteria? (Chinese translation: 請問您如何來鑑別與決定期末考試內容及考題的難易度？有什麼標準嗎?)
4. Why did you choose a particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe that the students perform better on these types of comprehension questions? (Chinese translation: 您為什麼會選擇這種問題類型來考學生？您認為某種類型的學生會喜歡這樣的測驗方式嗎或者您認為出這種類型的問題會讓這些學生考的比較好嗎?)

B.7 Classroom Observation Checklist

TBI Characteristics	Obs. No.	Observed?	Comments
1. There is at least one problem-solving task for students to do in class.	1 st		
	2 nd		
	3 rd		
	4 th		
	5 th		
	6 th		
2. There are many opportunities for students to practice English orally, including frequent oral interaction among students or with other interlocutors to exchange information and solve problems/tasks.	1 st		
	2 nd		
	3 rd		
	4 th		
	5 th		
	6 th		
3. Students report findings of a task to class, in groups or pairs, after problem solving.	1 st		
	2 nd		
	3 rd		
	4 th		
	5 th		
	6 th		
4. Authentic texts which reflect a real-life situation are used.	1 st		
	2 nd		
	3 rd		
	4 th		
	5 th		
	6 th		
5. The major focus of teaching is on the meaning and then on the form (i.e. whether the teachers mainly focus on the understanding of meaning instead of grammar teaching).	1 st		
	2 nd		
	3 rd		
	4 th		
	5 th		
	6 th		
6. Students were given opportunities to reflect on what they have learned and how well	1 st		
	2 nd		
	3 rd		

they were doing (i.e. reflection period).	4 th		
	5 th		
	6 th		

Appendix C

C.1 Preliminary Interview Data

Question 1: How do you usually evaluate students' English listening skills in the classroom?

Teacher	Interview Answers
1	We have the mid-term and final exams which are required by the university. In my class, I also look at their in-class participation, for example, if they answer the questions in the textbook. Testing them with quizzes is another way I evaluate them.
2	In addition to the mid-term and final exams, I pay attention to the attendance rate, and whether students complete their homework.
3	Two exams – mid-term and final – are compulsory, and I also check if they participate actively in class. I also use quizzes.
4	Mid-term and final exams. I also take their in-class participation into consideration, and performance on quizzes as another way to evaluate their ability.
5	Mid-term and final exams, homework and attendance are also checked in class.
6	Two university regulated exams – mid-term and final exams. In-class participation and quizzes are also used.
7	In addition to the mid-term and final exams. I take students' attendance and homework as part of their total academic performance.
8	Mid-term and final exams accounts for part of the total scores. In-class performances, such as participation and quizzes constitute another two parts.
9	Mid-term and final exams, and also in-class participation and quizzes.
10	Mid-term and final exams, in-class attendance, and homework.

Question 3: What types of listening extract you usually use in class?

Teacher	Interview Answers
1	Conversations, news broadcasts
2	Dialogues or conversations, academic lectures
3	Conversations, news broadcasts
4	Dialogues, news broadcasts
5	Dialogues, news broadcasts
6	Conversations, academic lectures
7	Dialogues, news broadcasts
8	Conversations, news broadcasts
9	Conversations only
10	Conversations, academic lectures

Question 4: Do you teach all students at the same year of study in one class?

Teacher	Interview Answers
1	No, I only taught a group of the students. There was another teacher teaching the other group. (Reason?) It was decided by the department but I didn't really know why they did this, maybe they thought it was too tiring for only one teacher to teach a large group of students. (Do you know who the other teachers are?) I am not very sure. (Do you know what in-class materials they used?) I didn't know.
2	No, I think there were three to four teachers teaching English listening in my department, so I was not the only one who teach all the students in the same year of

	study. (Reason?) I didn't know; this was arranged by the department. (Do you know who the other teachers are?) No, I did not; I think they were part-time teachers from other universities. (Do you know what in-class materials they used?) I did not know.
3	Yes. (Reason?) We didn't have extra teachers to teach this course.
4	No, there were four teachers teaching English listening in my department, so I was not the only one who taught all the students in the same year of study. (Reason?) I was not sure because I did not decide on this, but the chairperson said that it was better to provide our students with more opportunities to practice their English by using smaller classes. (Do you know who the other teachers are?) I only knew some of them. (Do you know what in-class materials they used?) We could choose the textbooks we liked, and I didn't ask them.
5	No, I only taught a group of the students. I guessed there was another teacher teaching the other group. (Reason?) The chairperson said that it was better to provide our students with more opportunities to practice their English by using smaller classes. (Do you know who the other teachers are?) I did not know, maybe it was Dr. X. (Do you know what in-class materials they used?) I did not know.
6	No, I only taught a group of the students, and there was another teacher teaching the other group. (Reason?) I didn't know; this was decided by the department. (Do you know who the other teachers are?) Yes, I knew. (Do you know what in-class materials they used?) I did not ask her.
7	Yes. (Reason?) I don't know; the course was arranged by the department.
8	No, only taught a group of the students. I think there were other teachers teaching other groups. (Reason?) I didn't know; this was decided by the department, maybe they wanted to provide our students with more opportunities to practice their English. (Do you know who the other teachers are?) I only knew some of them. (Do you know what in-class materials they used?) No.
9	No, I only taught a group of the students. I knew there was another teacher teaching the other group. (Reason?) Maybe our chairperson believed that using smaller classes provided our students with more opportunities to practice their English, but in fact I didn't know. (Do you know who the other teacher is?) I heard of her, but I rarely talked to her. (Do you know what in-class materials they used?) No.
10	No, there are three to four teachers teaching English listening in my department, so I am not the only one who teach all students in the same year of study. (Reason?) I didn't know; the department decided this. (Do you know who the other teachers are?) I only knew some of them. (Do you know what in-class materials they used?) I did not ask them.

C.2 Letter of Permission – Higher Authority

Dear (chairperson's name):

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. I am currently in the process of collecting data for my PhD research, and I would like to collect data from *two* English listening classes in the same year of study in your department. The main purpose of my research is to investigate (1) teaching and testing English listening, (2) listening problems that influence Taiwanese students' comprehension, and (3) test impact on the teaching in university English listening classrooms. I will need to observe the lessons, to interview the teachers, to distribute questionnaires, to collect in-class listening materials, test samples, and syllabus notes. It will be very helpful if I am allowed to contact the individual course instructors with your permission.

Yours truly,

Mu-hsuan Chou

C.3 Letter of Permission – Course Instructors

Dear (course instructor's name):

I am a PhD student in the Department of Educational Studies, University of York; my name is Mu-hsuan Chou. I am currently in the process of collecting data for my PhD research, I would like to collect data from your English listening class. As I have contacted the chair of your department about my research needs, she has permitted me to contact the course instructor directly. The main purpose of my research is to investigate (1) teaching and testing English listening, (2) listening problems that influence Taiwanese students' comprehension, and (3) test impact on the teaching in university English listening classrooms.

The information I need from you and your class is as follows:

1. Classroom observations
2. Two questionnaire surveys for your students (I will ask students' permission before the survey)
3. Two interviews with you
4. Mid-term and final test contents
5. A copy of a chapter in the textbook (Copyright will be requested from the textbook publishers)
6. A copy of the in-class materials designed by yourself (if applicable)
7. Your syllabus notes
8. Mid-term and final test marks

The data is for academic research use, and your and your students' names will be kept confidential when reporting the work. If you wish to know the result of your case, it will be presented to you. It will be very much appreciated if you can help me with my research.

Sincerely yours,

Mu-hsuan Chou

C.4 Interview transcription after the mid-term exam

Question 1. Where did the content of the test come from?

Teacher (T): 我是使用現有的考題，因為書商在提供教科書給老師的同時也會附贈一些跟課本相關且輔助的聽力練習題；我就是拿那些練習題考學生。

Interviewer (I): 那學生在考前也能取得這些相關練習題嗎？

T: 不可能，因為這些是書商專門給老師的，所以學生在考前是不會看到這些考題的。

I: 你期中考時有考他們的口語能力嗎？

T: 沒有。

I: 為什麼選擇不考？

T: 嗯....我覺得他們在加強口語能力之前應該先加強聽力，所以我寧願著重在訓練他們的聽力。

Teacher (T): I reused other tests. Because textbook publishers provided practice tests as supplements to their coursebooks, I used the practice tests as the content of the mid-term exam.

Interviewer (I): Was it possible that students accessed the practice tests before the exam?

T: No, it was impossible, since those (practice tests) were only for the teacher, it was impossible for students to obtain the test contents.

I: Did you test their speaking ability in the mid-term exam?

T: No.

I: Could you tell me why you did not want to test their speaking ability?

T: Well....I think that they needed to polish their listening skills first before moving on to the speaking skills, so I'd rather focus on training their listening first.

Question 2. What are the cut-off scores for the mid-term and final examination tests? What percentage does each of the two tests count for in the total final score of the course?

T: 都是 60 分及格。期中考占 30%、期末考占 40%、平常課堂參與表現占 30%。

T: The cut-off score was 60 for the two exams. The mid-term exam counted as 30% and the final one counted as 40% of the total final score. The other 30% was for in-class coursework.

Question 3. What did you expect the students to have learned from your class?

T: 因為這堂課是基礎英文聽力訓練，屬於 beginner's level，所以我會比較著重在聽力單字的發音與正確的拼法，由於學生的年級跟英文程度比較低（五專二年級），所以了解正確的發音與拼字是很重要的！要知道怎麼發音才會聽的懂啊！所以妳旁聽我的課堂時會發現我給學生很多單字發音的練習跟拼法練習，況且大部分的學生都很懶，如果妳不 push 他們背單字他們就不唸書，這樣成績也不好看，學校也會給我們施加壓力，所以我會特別著重在這兩方面。

T: The course was for basic English listening practice at beginner level, so I put more emphasis on the pronunciation and correct spelling of vocabulary in the listening texts. I found that the level of students' English ability did not meet the level it should in their present year of study (the second year of a five-year programme), so understanding correct pronunciation and spelling were important for them at this time! They would, I hoped, learn the vocabulary by pronouncing it. Hence, when you audited my class you would have seen that I asked students to practice the pronunciations and spellings of the vocabulary a lot. Besides, the majority of the students were lazy about studying. If you didn't push them to memorise the vocabulary, they would not do so. Because of this, there was a chance that their marks would be terrible, and then the dean would then put pressure on us (teachers); that's why I tended to focus on the basic memorisation of vocabulary and spelling.

Question 4. What are the mid-term test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Do you think you have achieved them?

T: 我希望學生能熟記課本上所聽到的單字，因為那些單字都是很基礎的，如果他們連基本的也背不起來，那以後會更難理解更長更高階的聽力。學生這次的成績還是有些偏低，因為考的聽力內容都是課堂上教的，只是題目內容問的跟課堂練習不一樣，但是他們還是考不好。

I: 就只因爲問題的方式不同？

T: 是啊！其實內容都差不多。

T: I hoped that the students would be able to memorise the vocabulary they read in the textbook because the vocabulary was pretty basic. If they could not memorise the basic words, it would be more difficult for them to understand advanced listening texts. The marks in the mid-term exam tended to be lower than I had expected. Most of the listening contents in this exam were taught in class, but the way questions were asked was different from the way they had practiced in class; the results were not good.

I: The only difference was the ways in which the questions were asked?

T: Yes! In fact the contents of the listening were almost the same.

Question 5. How do you identify and decide on difficulty of the content/items in the mid-term test? What are the criteria?

T: 難易度是從教科書裡決定的，我在選擇考題時會去考量考題的難易度跟課堂所教的難度是否相似，如果出太難讓他們的成績太低也不好，可能會打擊他們的信心。

T: I decided the level of difficulty based on that in the textbook. When I chose the test content, I considered whether the difficulty of the test content was similar to that taught in class. It was likely that their marks would be lower if I used questions that were too difficult and this might discourage them.

Question 6. Why did you choose the particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

T: 其實也沒有特別去選擇什麼類型的考題，一方面是因為考試時間有限，考聽力又很難只聽一次，所以也不會讓他們寫太多字，大部分都是選項還有一點點簡答，讓他們選擇正確答案，我改起來也方便。

I: 考聽力通常只能聽一次，為什麼您會讓他們聽一次以上？

T: 如果只讓他們聽一次他們通常會考得很爛！

T: In fact I did not specifically choose any types of comprehension question. On the one hand, test time was limited, and listening only once was insufficient for them. In this case, asking them to write a lot of words as an answer would have taken too much time; that's why I used questions with "options" for answering, as well as a few short-answer questions. It was also easier for me to score by using questions with options for answering.

I: Listening contents are usually heard only once in an exam, like TOEFL and IELTS, so why did you allow the students to listen more than once?

T: If they merely heard the content once, my experience was that they usually performed very poorly.

Question 7. Does the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)

T: 因為我覺得他們這次考的並不如我預期那樣好，期中考後會更要求他們發音跟字彙的練習，課堂教材是不會增加、難度也不會加深，這麼簡單他們都考不好了怎麼加深？但是練習的時間會增加。至於期末考題的難度跟期中不會差太多，主要是希望他們的能力會進步，所以考題的難易度就會維持跟期中考一樣。

T: I felt that the students did not perform as well as I expected, so I will demand more as regards practicing the pronunciation and spelling of vocabulary. I won't increase the in-class material or the level of difficulty. Because they could not handle such basic listening contents, how can I increase the difficulty of the material? As for the test content in the final exam, the level difficulty will be similar to that in the mid-term one. I hope that their English ability will increase, so I will maintain the level of difficulty in the final exam.

C.5 Interview transcription after the final exam

Question 1. Where did the content of the test come from?

T: 大部分的題目是我使用現有的考題，就是像期中考用的課本相關聽力練習題，最後有一題是 dictation 的加分題，不是一定要寫的，但是寫對有加分、寫錯不扣分。

I: dictation 的題目從那裡來？

T: 是課本上的一小段課文，上課時就教過了，這次考試只是把它唸出來讓學生填空，很簡單的加分題，只要有上下課有用心就應該能在這題上拿分。

T: I reused other test items for most of the test content, just like the listening test items I used in the mid-term exam. There was an extra point for a dictation test in the last test item. It was not compulsory for students to answer, but there would be a point for a correct answer, but no minus point for wrong answers.

I: Where did the dictation question come from?

T: It was from a small paragraph in the textbook which had been taught in class. It was an easy test item. As long as they worked hard in and after class, the students should have gained point on this item.

Question 2. What are the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Do you think you have achieved them?

T: 還是希望學生能熟記課本上所聽到的單字、發音跟基本的文法句型。學生這次期末的成績比起期中的成績高了一點，雖然還不算高，也沒有達到我希望的成績，雖然我增加了課堂的練習，但是他們上完課也不會很積極的去複習，我只能說我在課堂上盡力了。

I: 您預期他們考到幾分才算達到您教學的標準？

T: 我覺得他們至少要考到 75~80 分，因為題目真的不難。

T: I still expected that the students could memorise the vocabulary and its pronunciation they had heard and the basic grammar sentences in class. The mean of their final marks was slightly higher than the mid-term one, yet more failed, although this was not as high as I expected. Although I increased the amount of exercises done in class, it seemed that the students did not review them after class. What I can say was I have tried my best.

I: What were the marks you expected them to obtain in order to achieve your teaching objective?

T: I felt that they should reach at least 75 to 80 because the test was really not that difficult.

Question 3. How do you identify and decide on difficulty of the content/items in the mid-term test? What are the criteria?

T: 期末考題難易度除了從教科書裡決定還有他們期中的表現，雖然他們期中考的不甚理想，但我在期末考題裡出了加分題就是希望能把分數提高，但是他們拼單字時常拼錯，所以分數還是白白扣掉了。

T: The difficulty of the final exam was decided based on the level of the textbook and the students' performance in the mid-term exam. Although they did not perform well in the mid-term exam, the reason why I added a bonus point in the final exam was to enhance their marks. But as they often failed to spell correctly, they lost marks.

Question 4. Why did you choose the particular type of comprehension questions to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

T: 考題類型跟期中沒有很大的差別，唯一多的是 dictation 的簡答填充，有的填單字有的填片語，就是希望他們不只聽的懂單字還能正確的拼出來。

T: There were no big differences in the types of comprehension question between the mid-term and the final exam. The only difference was the short answer question for dictation, including filling in missing vocabulary and phrases in blanks. I expected that they would be able not only to understand the vocabulary but also spell it out correctly.

Appendix D

D.1 In-class Textbook One

Rost, M. (ed.) (2001) *Impact Listening 3*. Hong Kong: Pearson Education North Asia Limited.

Note: Textbook copyright requested and photocopy permitted.

Unit 1

Old friends, different choices

Different people make different choices in their lives. What will you choose in the future?

choices that affect one's life, such as college, marriage, career, etc.

Vocabulary Task

Connect the first part of each sentence with the second part. Discuss any words or phrases you don't know.

A It's been 10 years since I I was able to travel and And I never got stuck

B Things are going I finally settled Now I work

C Life has sure treated I put myself I was able to work And then I started

see the world.
being a housewife
saw you last.

9 to 5 at a neighborhood store.
great for me these days.
down and got a real job.

through school.
me well.
my own business.
my way up in a company.

to put oneself through = to pay for oneself, without help from others
to work one's way up = to work hard and get promoted in a company (got moved up)

Now listen and check your answers.

Which of these life choices appeal to you? Have you ever run into someone after not seeing them for a long time? What did you talk about?

Listening Task Look at the pictures. Where are these people meeting?

1 First Listening

What did each person do with their life?

1. He went overseas for graduate school and is working at a high-tech company.

2. She was in a few beauty contests, but she never won. She now has two children.

3. He works at the neighborhood post office and has two children.

4. She's the branch manager at the local grocery store.

2 Second Listening

Listen again. How do they feel about their decisions?

check the box with the answer to the question

1. <input type="checkbox"/> He disliked living abroad. <input checked="" type="checkbox"/> He missed his home.	2. <input type="checkbox"/> She feels she's still young enough to be a model. <input checked="" type="checkbox"/> She hopes her daughter will become a model.	3. <input checked="" type="checkbox"/> He loves his family. <input type="checkbox"/> He wants a better job.	4. <input checked="" type="checkbox"/> She's proud of her new job. <input type="checkbox"/> She wishes she had gone to college.
--	---	---	---

8

Real World Listening

1 Predict

Karen and Sharon are meeting for the first time after many years. What do you think they will talk about?

- travel old boyfriends
 family career
 other things _____

Now listen and check your prediction.

2 Get the main ideas

Read each statement. Write T for True or F for False.

- Sharon and Karen haven't seen each other for 30 years.
 Sharon planned to get married to Jim after high school.
 Sharon started her own advertising agency.
 Sharon majored in art in college.
 Karen didn't want to get stuck being a housewife.
 Karen had one child.
 Karen traveled around Europe for a year.
 Karen went to medical school.



3 Respond to the ideas

- Do you think either Sharon or Karen has any regrets about the choices they have made?
- Tell about a major choice you have made in your life and why you chose it. If you had it to do again, would you make the same choice? Why or why not?



{ what was Bill supposed to be?
 { what did Bill end up doing?
 { what is Bill trying to do?

Past and future hopes

Listen and write the missing words.

A: Hey, Julie, how are you? I haven't seen you in a long time!

B: Yeah, Bill, the last time I saw you, you were going to become a TV news anchor. Did that happen?

A: Well, not really. I maored in broadcasting in college, and I was supposed to become famous and travel all around the world. I was trying report on important world events. But instead, I just became a local news reporter. I to

B: That sounds like a good job.

A: Yeah, it is, but I want to travel. That's what I really wish I could do. So I am going to apply for an international news position.

B: Maybe it's not too late!

Put each phrase where it belongs.

PAST HOPES

- were going to
- was supposed to
- was going to

PAST EVENTS

- maored
- became

FUTURE HOPES

- want
- wish
- am going to

INTERACTION LINK
Life Choices

What is the difference between *I was going to* and *I am going to*?

page 49

D.2 In-class Textbook Two

Hanreddy, J. and Whalley, E. (2001) *Mosaic 1: Listening/Speaking*. USA: McGraw-Hill Publishing Co.

Note: Textbook copyright requested and photocopy permitted.

2

Mosaic 1 Listening/Speaking

Did You Know?

- Dr. Harold Williams holds the record for speaking the most languages. He was a journalist from New Zealand who lived from 1876 to 1928. He taught himself to speak 58 languages and many dialects fluently.
- The language with the most letters is Khmer, which used to be called Cambodian. It has 74 letters. The Rotokas of Papua, New Guinea, have the language with the fewest letters. It has only 11 letters (*a, b, e, g, i, k, o, p, r, t, and v*).
- The most complicated language in the world may be the language spoken by the Inuit peoples of North America and Greenland. It has 63 different types of present tense, and some nouns have up to 250 different forms.

*a group of
Estimian
people*

PART 1

Getting Started



Sharing Your Experience

Think about the following questions and make a few brief notes to help you remember your thoughts. Then discuss your answers as a class or in small groups.

- 1 Someone once said that getting to know a person is like peeling an onion. Have you ever peeled an onion? How might this be like getting to know a person?

- 2 Have you ever traveled to a new place or been to a party where you didn't know anyone? Did you find yourself behaving differently than normal? Try to recall an experience like this or try to imagine yourself in this situation. Include answers to these questions.
 1. *Where* were you? _____
 2. *What* did you do? _____
 3. *Why* did you do it? _____
 4. Is it sometimes easier to talk about yourself with people who don't know you? Why or why not?

- 3 Has your study of English changed you in any way? If so, how? Share your answers to the following questions and give specific examples.
 1. How has it made you more or less outgoing?

2. How has it made you more or less critical of how people speak your native language?

3. How has it made you more or less tolerant of other cultures?

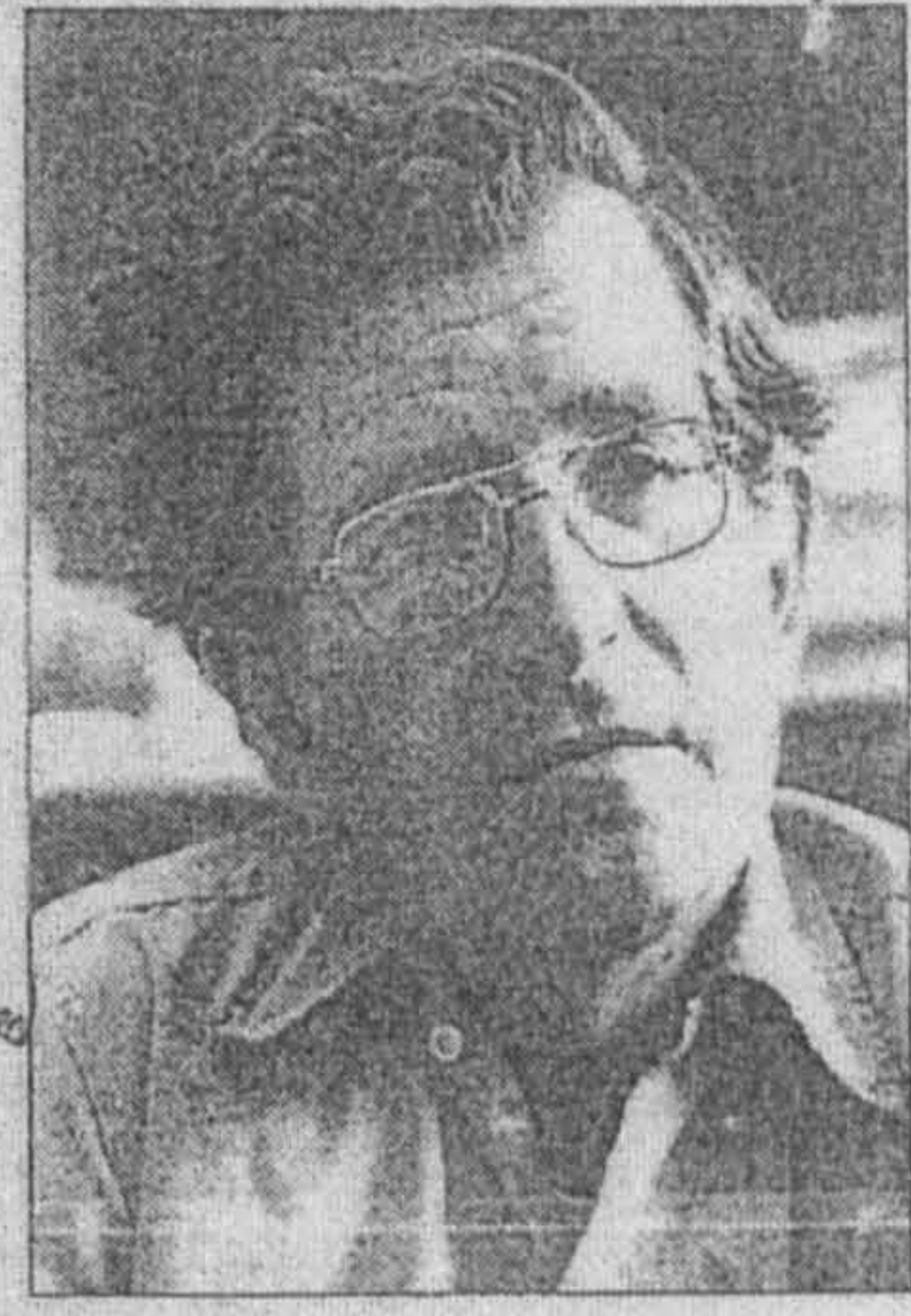
4. How has it changed your understanding or opinion of human nature?

Vocabulary Preview

Pre
Homework
for the 1st week only, otherwise as previous assignment.

4 **Determining Meaning from Context.** You will hear the underlined words in the following sentences in the lecture. Write the letter of the correct definition beside each sentence.

déjà = already
vu = seen } *the illusion of having already experienced st. actually being experienced for the first time.*



Sentences	Definitions
<i>c</i> 1. The professor looked at the <u>collage</u> made of paper, wood, leaves, and glue that was hanging on the wall of his office.	a. connection (in the mind)
<i>f</i> 2. This all looks so familiar. I feel that we've been here before. I guess it must just be <u>déjà vu</u> .	b. to change the nature of something
<i>e</i> 3. As a famous <u>linguist</u> , he is interested in the study of language acquisition.	c. an artistic creation of materials and objects glued onto a surface
<i>h</i> 4. Language presents us with a <u>paradox</u> : it helps us communicate, but communication is not possible if two people speak different languages.	d. accept or agree with someone's point of view
<i>l</i> 5. For a long time, researchers thought we learned language through <u>imitation</u> of others and <u>association</u> of words.	e. person who studies the nature and structure of human language
<i>j</i> 6. If we speak French <u>fluently</u> , we can begin to see the world from a French point of view.	f. something overly familiar; a feeling of having had an experience before
<i>b</i> 7. Learning to speak someone else's language can <u>transform</u> us.	g. present at birth; natural
<i>d</i> 8. I'm not sure I <u>buy</u> that idea.	h. statement/situation that presents opposing views as true at the same time
<i>g</i> 9. Noam Chomsky, a famous linguist, suggested that the ability to learn a language is <u>innate</u> .	i. modeling one's behavior or speech on the behavior or speech of another person
	j. spoken or written effortlessly and naturally

PART 2

Listening to Make Predictions

Surprises can be nice in everyday life, but if they occur frequently in a lecture, the lecture may be difficult to understand. In order not to be surprised too often, it is useful to anticipate what the instructor will say next. Here are two guidelines to help you make predictions:

1. Before you listen to the lecture, think about what you already know and what you want to learn about the topic.
2. As you listen to the lecture, predict what the speaker will say. When the lecturer makes a statement:
 - a. Predict what she or he will say next.
 - b. Judge quickly whether you were right or wrong.
 - c. If you were right, move on to your next prediction.
 - d. If you were wrong, don't worry about it, or you'll miss the next part of the lecture. Just put a question mark in your notes for clarification later and move on to the next prediction.

When you focus your listening in this way, you are less likely to be distracted by thoughts of things such as lunch, your soccer game, or the date you had Saturday night.

Before You Listen

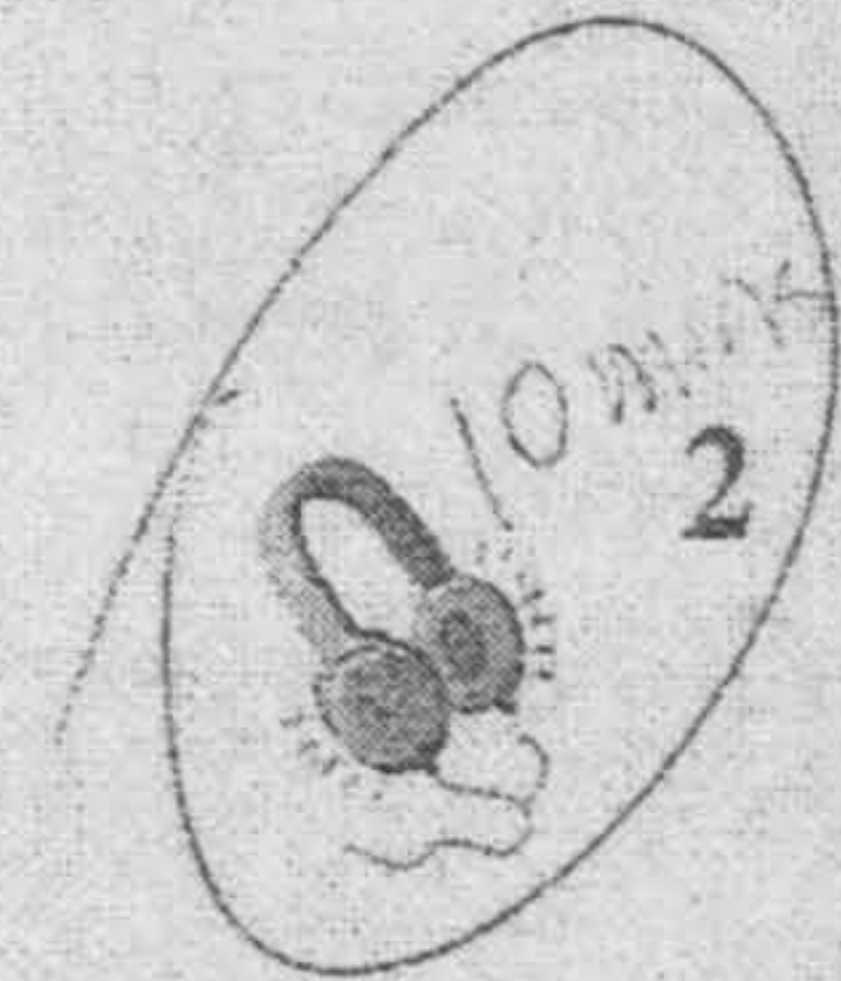


- ① **Discussing the Topic.** Write brief answers to the following questions. Discuss your answers in small groups.

1. What do you already know about the topic "Learning to Speak Someone Else's Language?"

2. What do you think the speaker will discuss?

3. What questions do you have on the topic?



- Listen** *task!*
- ② **Listening to Make Predictions.** Listen to the lecture one section at a time. This will give you the opportunity to understand what has been said already and to predict what will come next. The quotes from the lecture indicate where you should stop the lecture.

Stop 1 *Just call out your questions.*

Predict what questions you think the students will ask.

*Will I ever stop being afraid to make mistakes in English
Why is learning a 2nd language so difficult? How many languages
there in the work*

Stop 2 *Then let's begin with that last question. Can we ever really learn to speak another person's language?*

we Did you predict some of the questions the students asked? What do you think the professor's answer will be to that last question? Why?

b: *He'll probably say both yes and no. He'll talk about why people misunderstand each other.*

Stop 3 *Now this brings us back to the first question on our list: Where does language come from? And how does it develop?*

What do you think the professor's answer will be to this question?

He'll probably talk about how children learn language.

Stop 4 *Chomsky suggested that this accomplishment is possible because human babies have an innate ability to learn any language in the world.*

Have you ever heard of Chomsky? Do you believe that humans have an innate ability to learn language? What will the professor discuss next?

Chomsky is an expert in how people learn language. I think everyone can learn a language. He'll probably discuss how adults learn new languages.

Stop 5 *... our native language actually determines the way we see the world.*

What does this statement mean? What kind of examples do you think the professor might give?

language is a world view

Stop 6 *English sometimes uses words from other languages to express a thought or name a thing in a better way.*

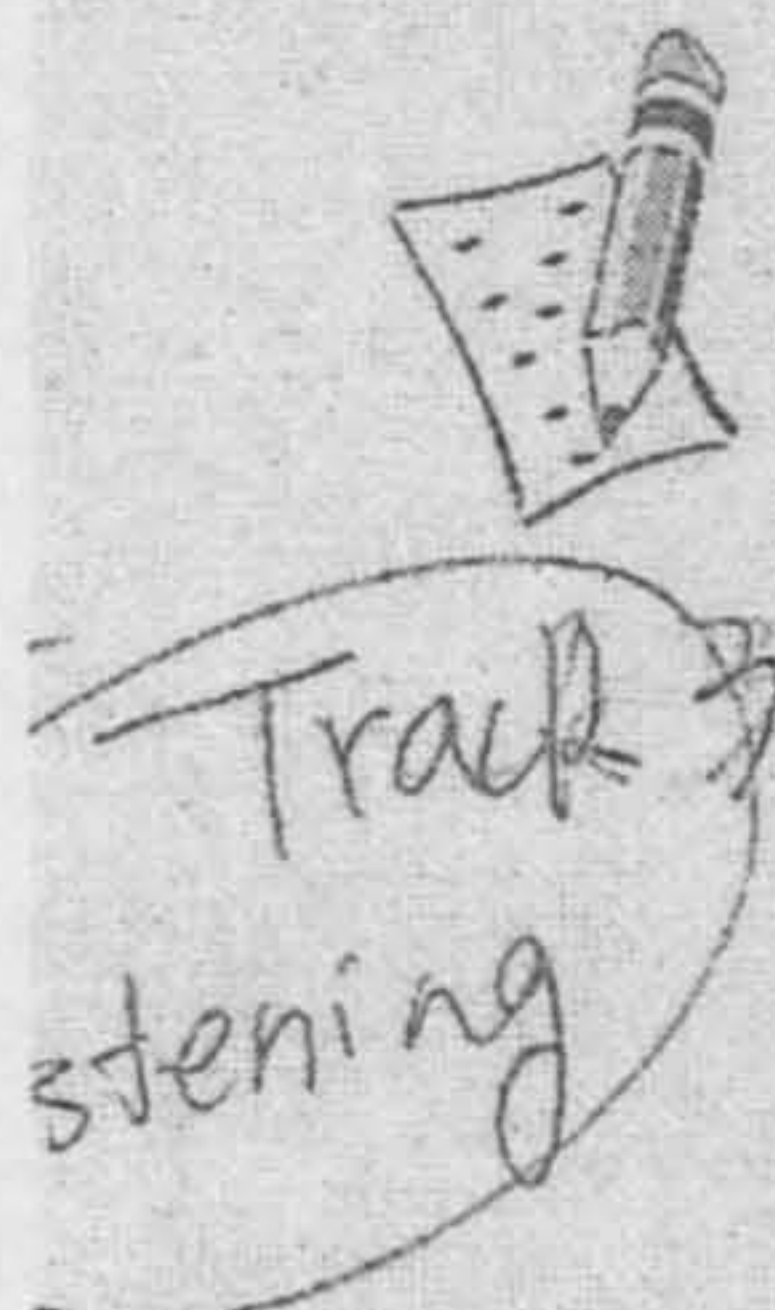
What are some words that the professor might use as examples here?

After You Listen

- 3** **Comparing Predictions.** Listen to the lecture again. At each of the stops, compare your predictions with those of your classmates. Were you able to make accurate predictions? What did you learn from your classmates' predictions?

PART 4

Focus on Testing



Understanding spoken English on standardized listening comprehension tests, such as the TOEFL, is more difficult than listening in most other contexts. During a standardized test, you cannot interact with the speaker to get clarification or rewind the tape to listen again. You get only one chance to listen for the important information. The Focus on Testing exercises in this book will help you practice this type of test.

Listen to the two speakers. After each speaker finishes talking, you will hear a question. Circle the letter of the best answer to each question.

Speaker 1

- why he didn't get the part in the school play
- what Dr. Jackson said yesterday
- what kind of play a pun is
- what Dr. Jackson just said

Speaker 2

- Some languages are more fun to learn than others.
- Adults and children speak the same language.
- Some languages are disappearing from the earth.
- Children shouldn't speak their parents' native language.

being dying out
native language



Video Activities: An Exchange Student

Before You Watch. Discuss these questions in small groups.

- What is an exchange student?
- What problems do you think exchange students might have?

Watch. Circle the correct answers.

- Where is Adáh from?
 - the United States
 - Switzerland
 - Turkey
- Circle the kinds of problems that exchange students and their families sometimes have.
 - money
 - chores
 - studying
 - cultural/language problems

3. What kind of problem did Adáh have?
 - a. Her homestay sister was jealous of her.
 - b. She had to share the computer.
 - c. She didn't have a good social life.
4. Who was Adáh's best friend?
 - a. Jeli
 - b. Corey
 - c. her date
5. What happened to Adáh's best friend?
 - a. She got sick.
 - b. She had a car accident.
 - c. She went home.

Watch Again. Compare answers in small groups.

1. How old is Adáh? 18
2. What are the initials of the exchange student organization?
 - a. EVS
 - b. AFS
 - c. ALS
3. Look at Adáh's report card and answer these questions.
 - a. What languages is she studying?
 - b. What science class is she taking?
 - c. What is her average grade?
4. What percentage of exchange students goes home early or change families?
 - a. 2%
 - b. 12%
 - c. 20%
5. Look at the chart that Adáh made of her "highs and lows." In which month did she feel the best?
 - a. August
 - b. September
 - c. October

After You Watch. Discuss these questions in small groups.

1. Have you ever known any exchange students? What countries were they from?
2. Would you like to be an exchange student? Why or why not? Where would you like to go?

D.3 Listening Scripts from the Textbooks

Selected listening scripts from *Impact Listening 3*

1. A: Jim, I can't believe it – wow! I haven't seen you since we graduated high school! What have you been doing?
B: Well, after I went to college I went overseas for graduate school, and I had a great time. But finally I had to come home and get a real job in a high-tech company.
A: Yeah, I remember you always wanted an international lifestyle.
B: Oh, I had a great time overseas, but I got homesick, too.
2. A: Carol, is that you?
B: Yes, it's me.
A: Oh, it's so good to see you! You know, Carol, I remember you always wanted to be a famous model and travel around the world. Did it happen?
B: Well, no. I was in a few beauty contests, but I never won, and nobody ever offered me a modelling contract.
A: Oh, that's too bad.
B: Yeah, but now I guess I'm too old, and I have two young kids, a boy and a girl. You know, maybe my daughter will get to do it some day.
3. A: What have you been up to over the years, Ed?
B: Well, I have two beautiful children now and work 9-5 at the neighbourhood post office.
A: Hmm, I remember you always wanted to be a professional race car driver – what happened?
B: The closest I've come to that is driving a Porsche over 100 mph on the Autobahn. No, my greatest achievement is my family.
4. A: So, Nora, how have you been?
B: Remember when we were in high school, I worked part-time bagging groceries at the local store?
A: Yeah?
B: So when I graduated it seemed natural to get a full-time job there. I've been working my way up. Now I'm the branch manager. I'm making more money than some people who have college degrees!

D.4 Test Format of the Mid-term Examination

Sophomore Listening Lab

Mid-Term Examination

Chinese Name: _____

Student ID #: _____

I. Listen to each conversation. Then listen to the questions. Answer the questions. Circle the correct answer. 10x4=40

Conversation 1:

- | | |
|--|--|
| 1. a. visited relatives | 2. a. Yes, completely |
| b. went to graduate school | b. No. He wanted to go home. |
| c. worked for a high tech company | c. No. He had to work too hard. |
| d. dreamed of an international lifestyle | d. Yes, but sometimes he got homesick. |

Conversation 2:

- | | |
|--|--|
| 3. a. He goes out dancing every night. | 4. a. She's worried about his health. |
| b. He zips through his homework. | b. She doesn't want him to go dancing. |
| c. He studies late every night. | c. She's impressed by his ability to school. |
| d. He has trouble sleeping. | d. She wants him to take her dancing. |

Conversation 3:

- | | |
|--------------------------------------|--------------------------------------|
| 5. a. He listens to what she says. | 6. a. She cleans his room. |
| b. He feels guilty. | b. She uses his phone. |
| c. He is annoyed by what she does. | c. She takes up too much space. |
| d. He's curious about what she does. | d. She listens to his conversations. |

Conversation 4:

- | | |
|-------------------------------------|---|
| 7. a. He cares a lot about fashion. | 8. a. He really likes designer clothes. |
| b. He worries about money. | b. He wishes he could afford \$500 pants. |
| c. He follows his friend's advice. | c. He thinks his friend spends too much money on clothes. |
| d. He likes to wear cheap clothes. | d. He wants to be more fashionable. |

Conversation 5:

- | | |
|---|--|
| 9. a. He has to learn standard English. | 10. a. Because Mr. Chen is a native speakers of English. |
| b. He can't speak English fluently. | b. Because Mr. Chen can show them how to learn English. |
| c. He has to study Chinese. | c. Because she's always right. |
| d. He doesn't think the teacher will be good. | d. Because Mr. Chen is Chinese, too. |

II. Write the missing words. There could be more than one word in each blank. 20x1=20

A: Hey, what's Sally _____ these days?

B: Didn't you hear? She hit _____ ?

A: What do you mean?

B: She got a part on a daytime _____ soap opera.

A: No kidding. I knew she was _____ acting, but I never thought she'd _____.

B: Yeah, she decided to _____, so she move to L.A.

A: Wow, she's _____, isn't she? She must be getting _____.

B. Yep. Boy, has she _____.

III. Listen to each conversation. Then read the questions. Circle the correct answer.

5x4=20

1. Which of the following is NOT the question from the students?
 - a. Can we learn to speak someone's language?
 - b. Who uses language?
 - c. At what age does language develop?
 - d. What does it make you think about?
2. What is the first question that the professor answered?
 - a. Can we learn to speak someone's language?
 - b. Who uses language?
 - c. At what age does language develop?
 - d. What does it make you think about?
3. Who might be Chomsky?
 - a. a student
 - b. a French
 - c. a child
 - d. a linguist
4. What is the positive side of the paradox mentioned in this lecture?
 - a. Language can be used in different ways.
 - b. Children are born with the ability to learn language.
 - c. Children learn to make requests.
5. What is the negative side of paradox mentioned in this lecture?
 - a. Language is a wonderful way of communication.
 - b. People speak different languages
 - c. We can't learn to speak someone else's language.

IV. Listen to each conversation. Then listen to the questions. Circle the correct answer.

5x4=20

1. a. Are you with me? b. Did you get that? c. Right? d. Is that clear?
2. a. She is offering clarification. b. She is offering confirmation.
c. She is asking for direction.
3. a. She is asking for attention from her son. b. She is scolding her son.
c. She is offering assistance for her son.
4. a. what Dr. Jackson said yesterday. b. what kind of play a pun is.
c. what Dr. Jackson just said.
5. a. Adults and children speak the same language.
b. Some languages are disappearing from the earth.
c. Children shouldn't speak their parents' native language.

D.5 Test Format of the Final Examination

Sophomore Listening Lab	Final Examination
Student ID #: _____	Chinese
Name: _____	
I. Listen to each conversation. Then listen to the questions. Answer the questions. Circle the correct answer. 8x4=32	
Conversation 6:	
11. a. He's single, and had never been married. b. He is currently married. c. He was married before. d. He doesn't say.	12. a. someone who likes quiet evening together b. someone who likes dancing c. someone who likes bodybuilding. d. someone he can trust.
Conversation 7:	
13. a. She's having health problems. b. She works too much in the yard. c. She's retired. d. She feels her home is like a zoo.	14. a. The more people there are, the happier he is. b. He's glad Gloria is there to help. c. He doesn't care. d. It's OK, as long as they do the yard work.
Conversation 8:	
15. a. She likes the restaurants in Costa Rica better. b. She thinks America backpacks are safer. c. She thinks Costa Rica is more secure than the U.S. d. She thinks it's easier to have money in the U.S.	16. a. She likes the variety of restaurants. b. She doesn't worry about things being stolen. c. She worries about walking around with \$20. d. She thinks that everything is too expensive.
Conversation 9:	
17. a. It does your driving for you. b. It fastens your seatbelt for you. c. It keeps you awake while you drive. d. It reads the newspaper to you.	18. a. study b. read c. talk on the phone d. drive
II. Write the missing words. 10x1=10	
Hello. Welcome to the first _____ study skills class of the semester. I'm Lilia Rothman. I'm a _____ in the Education Department and I'll be the _____ for this course. Has everyone found a _____ to sit? There's a chair _____ here, if you need one. Okay. Let's get _____. Did you know that people _____ more than half the time they are awake communicating? They are _____ writing, reading, speaking, or listening. Which do you think they do _____?	

III. Listen to the conversations and the questions. Circle the best answer to each question. 6x3=18

1. a. Frank is not strong enough.
b. Frank is taking too many courses for his semester.
c. First-year students usually take this many courses.
d. Frank has a lot of different interests.
2. a. She is tired and hungry.
b. She doesn't want to use the meal plan on weekends.
c. She thinks the meal plan is too expensive for what she is getting.
d. She can't buy food on Sunday.
3. a. Dad has to help Grandpa and Grandma move this week.
b. Ruth and James have to do their own homework this week.
c. Ruth and James need to practice pitching for the baseball game.
d. Ruth and James need to help around the house more than usual.
4. a. He's explaining why he wants to take the dog along on vacation.
b. He's explaining why they have to leave the dog home this year.
c. He's trying to convince them to visit the Grand Canyon.
d. He's saying that they have to stay home to take care of the dog this summer.
5. a. We should pass more laws on smoking.
b. People have a right to smoke if they want to.
c. People dying of cancer should be allowed to eat where they want to.
d. People should quit smoking at home and smoke in restaurant instead.
6. a. She could win a bet on how much weight she can lose.
b. She is overweight, like most Americans.
c. She should lose 16 pounds.
d. Most people in America read newspaper.

IV. Write T for true or F for false. 10x3=30

- _____ Kristy and Shawna are moving in with Gloria's family.
- _____ Simon got transferred to Florida.
- _____ Lily, Simon's wife, have enough time to take care of her children.
- _____ Marty and Eddie have to sleep in the living room.
- _____ Gloria wants to solve everyone's problems.
- _____ Grandma is feeling a lot better now.
- _____ Eddie is going to spend more time with Marty.
- _____ Nobody asked Marty what he wants.
- _____ Marty's parents used to help his homework.
- _____ Eddie is going to help with Marty's Math homework.

V. Choose the correct answer. 5x2=10

1. How much doesn't Steve have? a. \$30 b. \$300 c. \$3000
2. What product would Steve buy? a. phone b. camera c. watch
3. What can the product do for Steve?
 a. dial for him b. talk to him c. drive for him
4. What was the speaker's major in college?
 a. English b. computer science c. none
5. According to the speaker, what kind of experience does this speaker had in college?
 a. teaching English b. programming c. travelling

D.6 Problems Students Found in the Mid-term and Final Exam

1. The quality of the recording and/or the visual equipment

No.	Characteristics of the Quality of the Recording	Mid-term Exam		Final Exam	
		Freq.	%	Freq.	%
a.	The quality was so poor that I couldn't hear the text clearly.	5	16.1%	1	3.2%
b.	The background noise outside the testing environment was too loud.	6	19.4%	1	3.2%
c.	The volume of the tape recorder or audio-visual equipment was too low.	5	16.1%	2	6.5%
d.	The quality of the recording was good, and I can hear the texts clearly.	18	58.1%	27	87.1%
e.	Other:	0	0%	0	0%

2. Testing time

No.	Characteristics of Testing Time	Mid-term Exam		Final Exam	
		Freq.	%	Freq.	%
a.	Time was too limited to answer all the questions properly.	12	38.7%	5	16.1%
b.	The testing time was sufficient for me to answer all the questions properly.	14	45.2%	26	83.9%
c.	Other: Because I was late. 內容間距速度太短，沒有思考空間 X5。	6	19.4%	0	0%

3. The test/task instructions

No.	Characteristics of Test/Task Instructions	Mid-term Exam		Final Exam	
		Freq.	%	Freq.	%
a.	The instructions of each test section were not clear.	10	32.3%	2	6.5%
b.	The instructions were too complicated.	2	6.4%	2	6.5%
c.	The test/task instructions were clear.	19	61.3%	27	87.1%
d.	Other:	0	0%	0	0%

4. The length of the listening texts

No.	Characteristics of the Length of the Texts	Mid-term Exam		Final Exam	
		Freq.	%	Freq.	%
a.	The texts in the test were longer than I was used to listening to in class.	4	12.9%	0	0%
b.	Longer texts increased the difficulty of understanding.	4	12.9%	0	0%
c.	The lengths of the texts in the test were similar to what I used to listen to in class. 跟課堂上差不多。	27	81.8%	31	100%
d.	Other:	0	0%	0	0%

D.7 Mid-term Interview Transcription

<p>Question 1. Where did the content of the test come from?</p> <p>Dr. N (N): 第一大題是從書上出來的。</p> <p>Interviewer (I): 是課本上的練習題嗎？</p> <p>N: 是書商他們出給老師的考題。</p> <p>I: 有比較難嗎？</p> <p>N: 並不會，裡面的英聽對話跟課本都一樣，但問的方式不一樣。其他的題目是課本上的，但是我自己設計不同的題目。</p> <p>I: 題目問答的方式有變嗎？</p> <p>N: 基本上沒有，但就是問不同的問題，如果學生上課有學起來，他們回答應該沒有問題。</p> <p>I: 在測驗學生前，你有測試這些考題嗎？</p> <p>N: 需要嗎？或許書商已經測試過了。</p>
<p>Dr. N (N): The first part came from the textbook.</p> <p>Interviewer (I): Were those exercises in the textbook?</p> <p>N: The questions were taken from the teacher's manual.</p> <p>I: Were those more difficult than the ones in the textbook?</p> <p>N: Not really, the conversations in the teacher's manual were the same as those in the textbook, but the questions were different. The remaining questions were from the textbook, but I designed different questions.</p> <p>I: Did the types of question change?</p> <p>N: Basically no, I simply asked different questions, but if students paid attention to the lessons, they should have no problems.</p> <p>I: Did you pilot the test items, I mean run a small test, before you test your students in this exam?</p> <p>N: Did I have to? Maybe the book publisher had tested those questions.</p>

<p>Question 2. What are the cut-off scores for the mid-term and final examination tests? What percentage does each of the two tests count for in the total final score of the course?</p> <p>N: 60。其中 25，期末 25。</p> <p>I: 我在觀察課堂的時候聽到你跟學生說要他們去網路上考小考，為什麼小考要在網路上考？他們要怎麼來考試？</p> <p>N: 考題內容是從加州一個成人英語學習網來的，那個網站上有很多主題跟我們目前學的還蠻相近的，他們在期中考前要完成 school 裡面的所有聽力內容，期末考前要完成 going places 裡的所有內容，學生聽完一個主題後有不同類型的題目讓他們做，做完之後會有分數，他們在直接從網路上寄成績給我。這樣會有更多的時間上課。</p> <p>I: 這個網站在使用上有發生過什麼問題嗎？</p> <p>N: 我目前還沒聽說有什麼問題。</p> <p>I: 不在教室裡考試學生有可能會作弊嗎？</p> <p>N: 當然要防止他們在網路考試上作弊是很難的，但如果他們作弊他們的聽力將不會進步，所以他們要為自己的行為負責。</p>
<p>N: "60" is the passing score. Mid-term scores count for 25% of the total marks; final scores count for another 25%.</p> <p>I: During my observations, you did not give the students quizzes directly in class but asked them to complete quizzes afterwards on the Internet. So why did you test your students on the Internet?</p> <p>N: The quizzes were from an adult English learning website in California³. There were many topics which were similar to those we learned in class. They had to listen and complete all lessons in the category of "School" before the mid-term exam, and the categories of "Going Places" after the mid-term. Students did different types of question after they listened to a topic. After finishing the questions, there were be scores on the screen and they had to send their scores directly to me via the Internet. It saved more class time for the teaching.</p> <p>I: Did you find any problems with using this kind of on-line quiz?</p>

N: There have been no technical problems with using the website at all so far.
I: Was it possible for students to cheat because they did not take the quiz in the classroom?
N: Well...of course I could not prevent them from cheating....but if they cheated, their listening ability would not improve. So they had to be responsible for themselves.

Question 3. What did you expect the students to have learned from your class?

N: 我有用兩種教材，一種是比較 general 的聽力，就是一般日常生活會用到的會話，另一種是 academic 的聽力，我一般上課都是第一個小時給他們聽日常生活會話，第二個小時給他們聽 academic 的。

I: academic 的教材是屬於 lecture 的類型嗎？

N: 是啊，裡面的 lecture 都還蠻長的。

I: 那學生聽起來會有問題嗎？

N: 當然有啊！他們聽得很吃力。

I: 那請問老師您覺得您所使用的教材跟學生的程度相符合嗎？

N: general 的還可以，academic 的對他們而言太難了。

N: I use two textbooks. One is for “general” listening, which means conversations we use in everyday life. The other is for “academic” listening. I usually teach everyday English conversations for the first hour, and academic listening for the second.

I: Did the academic listening take the form of a “lecture”?

N: Yes, the lecture in the textbook was quite long.

I: So did students find listening to it problematic?

N: Of course, they found it difficult to understand.

I: Did you think the textbook you used matched students’ English level?

N: I think the “general” book is OK, but the “academic” one is too hard for them.

Question 4. What are the mid-term test objectives you would like to achieve in relation to your teaching plan and students’ learning outcomes? Do you think you have achieved them?

N: 我都有叫他們每個星期上完課都要練習聽課堂上的，考試的目的是想看他們下課有沒有聽上課的內容，看他們有沒有學到課堂上應該聽懂的。

I: 學生都有拷貝聽力錄音帶嗎？

N: 有，general 的這本有附贈 CD，另外一本他們都有一份拷貝。

I: 您覺得這次學生的考試成績如何？有達到你的要求嗎？

N: 我覺得有點偏低，我覺得他們應該考好一些。因為我覺得那些全部都是課本的，可是他們回家都沒有聽。

I: 我有注意到你期中考沒有考學生的口語能力，可以請你告訴我原因嗎？

N: 我不認為這堂課是需要考口語能力的，因為課程名稱叫『英文聽力』，況且我在上課的時候有讓他們練習口語啊，就是 presentation 啊，所以我覺得那就夠了。

I: 所以....你期末考會考口語嗎？

N: 我不會。

N: I did ask them to review and practice the units I taught every week. The purpose of the test was to see whether they had practiced the listening contents after the class and whether they understood what I taught in class.

I: Did all students have a copy of the listening CD?

N: Yes, a CD was included in the “general” textbook; they have a copy of the CD of “academic” listening.

I: What did you think of the students’ performances in the mid-term test? Did you think you achieved your test objective?

N: I think the scores are a little lower than I expected. I think they should have performed better because the test questions were all from the textbooks. It was possible that they didn’t review and practice after the class.

I: I noticed that you did not test students’ speaking ability in the mid-term exam. Could you let me know the reason?

N: I don’t think that testing speaking was necessary for this course, since the title of the course was *English Listening*. But I gave students opportunities to speak English in class, I mean, the

presentations, so I think that it was enough for this course.

I: So...will you use any speaking test for the final exam?

N: No, I won't.

Question 5. How do you identify and decide on the difficulty of the content/items in the mid-term test? What are the criteria?

N: 就從教科書的題材跟進度出題。

I: 那如果學生反應課堂教材太難，請問老師您會將考題出的比較簡單嗎？

N: 目前學生沒有跟我反應這樣的問題，不過由於他們覺得 academic 的聽力比較難，所以這方面的考題我出的比較少。只出了兩題，像有一題就是考 lecture 聽力。

I: 那他們這部份有錯的比較多嗎？

N: 也沒有。反倒是第一大題，也就是 general 那本裡面出的考題他們錯的比較多，其實我也只是將題型稍做變化而已。

I: 那請問老師您覺得他們的單字量足夠應付這堂課的聽力嗎？

N: 嗯...不太夠，因為 general 那本有用一些俚語的語言，如果平常沒有在用他們就聽不懂。不過我不會去考單字，我都是考他們對內容的理解。

I: 所以他們比較不會看到或聽到生字的問題？

N: 對啊。

N: The test content was based on the textbook and the curriculum progress.

I: If the students told you that the in-class teaching materials were too difficult, would you make the test easier?

N: So far I haven't had any reports from the students regarding the difficulty of the materials. But as they felt that "academic" listening was more difficult, I designed fewer questions regarding "academic" listening. Those were in Part Three.

I: Did they score lower in this part?

N: Not really. On the contrary, they scored lower in the first part of the test that was from the "general" listening textbook. In fact, I simply slightly changed the types of questions.

I: Did you think that they could cope in this listening class with their vocabulary?

N: Erm...I am afraid not, because there were some colloquial language usages in the "general" listening textbook. If they didn't use colloquial language frequently, they wouldn't understand it. But I didn't focus on testing the vocabulary; I tested their understanding of the listening passages.

I: Did this mean that they didn't find the vocabulary difficult?

N: Yes.

Question 6. Why did you choose the particular types of comprehension question to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

N: 我個人覺得考填充題他們應該會考的比較好，因為聽到什麼就寫什麼，但是其實不盡然，有些人可以得滿分，但有些人莫名其妙的就 miss 掉了。

N: I personally felt that they would score higher on cloze questions because they could write down what they heard. However it turned out not necessarily so; they did not score higher in this part. Only more than half of the students got full marks in this type of question, but others lost marks for unknown reasons.

Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)

N: 可能 academic 那本書會多花時間讓他們聽，每次聽 academic 那本他們都快睡著了。但另外一本 general 的他們覺得還蠻有趣的，所以在難易度上面就不會多做調整。

I: 那期末考考題呢？

N: 應該不會變更難易度，就跟這次的考題差不多，其實這次的考試有區分出來，有唸書

的就考的比較高，沒讀書的還是那幾個。

I: 您覺得這班學生的程度如何？

N: 我覺得差不多，但有一些學生程度不好又不太認真。

I: 學生會覺得教材太簡單嗎？

N: 沒有聽說他們覺得太簡單。

N: Maybe I would give them more time to listen to the “academic” listening textbook in class, but they seemed to fall asleep when listening to it. On the other hand, they found the “general” textbook more interesting, so I wouldn’t amend difficulties in this part.

I: How about the final exam?

N: I think I wouldn’t adjust the difficulties of the test questions; difficulties would be similar to those in the mid-term exam. In fact, this mid-term did distinguish those who studied harder from those who didn’t. The harder they studied, the higher scores they received. But there were still some students who didn’t study.

I: What did you think the level of your students in this class?

N: I think in general there were not many differences between them, but there were some students whose English was poorer and they didn’t study hard.

I: Did students feel that the teaching materials were too easy for them?

N: I didn’t hear that they felt the materials were too easy.

D.8 Final Interview Transcription

<p>Question 1. Where did the content of the test come from?</p>
<p>Dr. N (N): 跟期中考一樣是從教師手冊來的。只有第二大題是我節錄課本上的一段內容。</p>
<p>Interviewer (I): 那學生在課堂上就有先聽過嗎？</p>
<p>N: 有，所以他們這大題錯的比較少。</p>
<p>I: 您這次期末考的範圍有包含期中考前的範圍嗎？</p>
<p>N: 沒有，這次單純就是考期中考後所教的內容。</p>
<p>I: Did you pilot the test items in the final exam?</p>
<p>N: No. I think the textbook publisher had piloted the questions.</p>
<p>Dr. N (N): It was from the teacher's manual, which was similar to the mid-term exam. But the second part was taken from <i>Mosaic 1</i>.</p>
<p>Interviewer (I): Had the students heard the second part from <i>Mosaic 1</i> in previous class?</p>
<p>N: Yes, that's why they didn't make many mistakes in this part.</p>
<p>I: Did you include what you have taught before the mid-term in the final exam?</p>
<p>N: No. I simply tested what I had taught after the mid-term.</p>
<p>I: Did you pilot the test items in the final exam?</p>
<p>N: No. I think the textbook publisher had piloted the questions.</p>

<p>Question 2. What are the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Do you think you have achieved them?</p>
<p>N: 因為這次考試沒有考期中考前的東西，是考期中考後新教的，考題內容都跟期中考不一樣，所以希望學生有把期中考後的內容學起來。</p>
<p>I: 您覺得他們這次考試表現如何，跟期中考比起來？</p>
<p>N: 嗯...我個人覺得好像有比期中考好一些，雖然也沒特別好。</p>
<p>N: Because I didn't test what had been taught before the mid-term but only after it, the test content was different from the mid-term exam.</p>
<p>I: What did you think of their performance this time, compared with their mid-term performance?</p>
<p>N: Erm...I personally felt that they performed a little bit better than the mid-term this time, but not particularly well.</p>

<p>Question 3. How do you identify and decide on the difficulty of the content/items in the mid-term test? What are the criteria?</p>
<p>N: 其實這次期末考也沒有特別去調整難易度，因為他們期中考也不是那麼的差，所以就說出的跟期中差不多就可以了。</p>
<p>I: 我在觀察您課堂的時候發現到您在給他們聽 academic 那個聽力前有給他們文本看，為什麼您期中考前沒有給他們期中考後才給？</p>
<p>N: 因為他們期中考前好像不是很能了解裡面的內容，我當時怕他們不懂也沒有教很多，這次就想說給他們文本看希望他們能了解內容是什麼。</p>
<p>I: 您有注意到學生聽內容做練習時抄文本上的答案嗎？</p>
<p>N: 我有注意到。我有告訴他們要先聽內容再看文本，我相信文本能更幫助他們了解 lecture, 但他們必須承擔抄答案的後果。如果他們期末考成績不好我也無能為力。</p>
<p>N: In fact I didn't particularly adjust the difficulty of the test content in this final exam. Because they didn't perform very badly in the mid-term, I felt it would be fine if the difficulty of the test content in the final exam was similar to that in the mid-term one.</p>
<p>I: I saw you give students the scripts before listening to the "academic" textbook, so why did you give the scripts to them only after the mid-term but not before it?</p>
<p>N: Because they appeared unable to understand the "academic" listening completely before the mid-term and I was afraid that academic listening was too difficult for them, I didn't teach them too much. This time I thought that giving them scripts might help them understand the listening content.</p>
<p>I: Did you notice that the students copied the answers from the transcript?</p>
<p>N: Well...I did notice their behaviour. I told them that they needed to listen carefully for the first time and then read the transcript. I believed that transcripts could help them understand the</p>

lectures better, but they had to bear the consequences after copying answers from the transcripts...I mean if they scored lower in the final exam.

Question 4. Why did you choose the particular types of comprehension question to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

N: 這次的題型跟期中考的也差不多，因為都是平常聽力課常用的題型，就用了。

N: The types of question were similar to those in the mid-term, because they were common types of question we used in the listening class.

Appendix E

E.1 In-class Textbook One

Lemieux, L. and Dabbs, T. (2004) *Listen Up (kit): Intermediate, with audio program.*
Vancouver: Lynx Publishing Company.

Note: Textbook copyright requested and photocopy permitted.

Unit 1: New Kid in Town

UNIT 1

1a: Preparation - Listen. You will hear six short passages. After hearing each passage, circle the best ending for each sentence below.

b 1. A next-door neighbor lives	a) far away.	b) nearby.	2
b 2. An acquaintance could be	a) your relative.	b) your classmate.	
b 3. To register is to sign up	a) for a school.	b) for a class.	
a 4. A classmate is from the same	a) school.	b) church.	
a 5. A new kid in town is	a) a newcomer.	b) tourist.	
a 6. A soul mate is	a) the best friend.	b) the best person.	

1b: Exercise - Listen to Tom and Megan. After each person speaks, you will hear a statement. Mark the statement True or False.

1.	T	F	3
2.	T	F	
3.	T	F	
4.	T	F	

2a: Preparation - Listen. You will hear six short passages. After hearing each passage, circle the correct definition for each word or phrase below.

a 1. straight ahead	a) forward	b) backward	4
a 2. a lane	a) a small street	b) a big street	
a 3. a curb <i>edge of a road</i>	a) a raised edge	b) a center stripe	
b 4. a pavement <i>sidewalk</i>	a) a dirt surface	b) a concrete surface	
b 5. you can't miss it	a) hard to find	b) easy to find	
a 6. to shade	a) to cover	b) to stand behind	

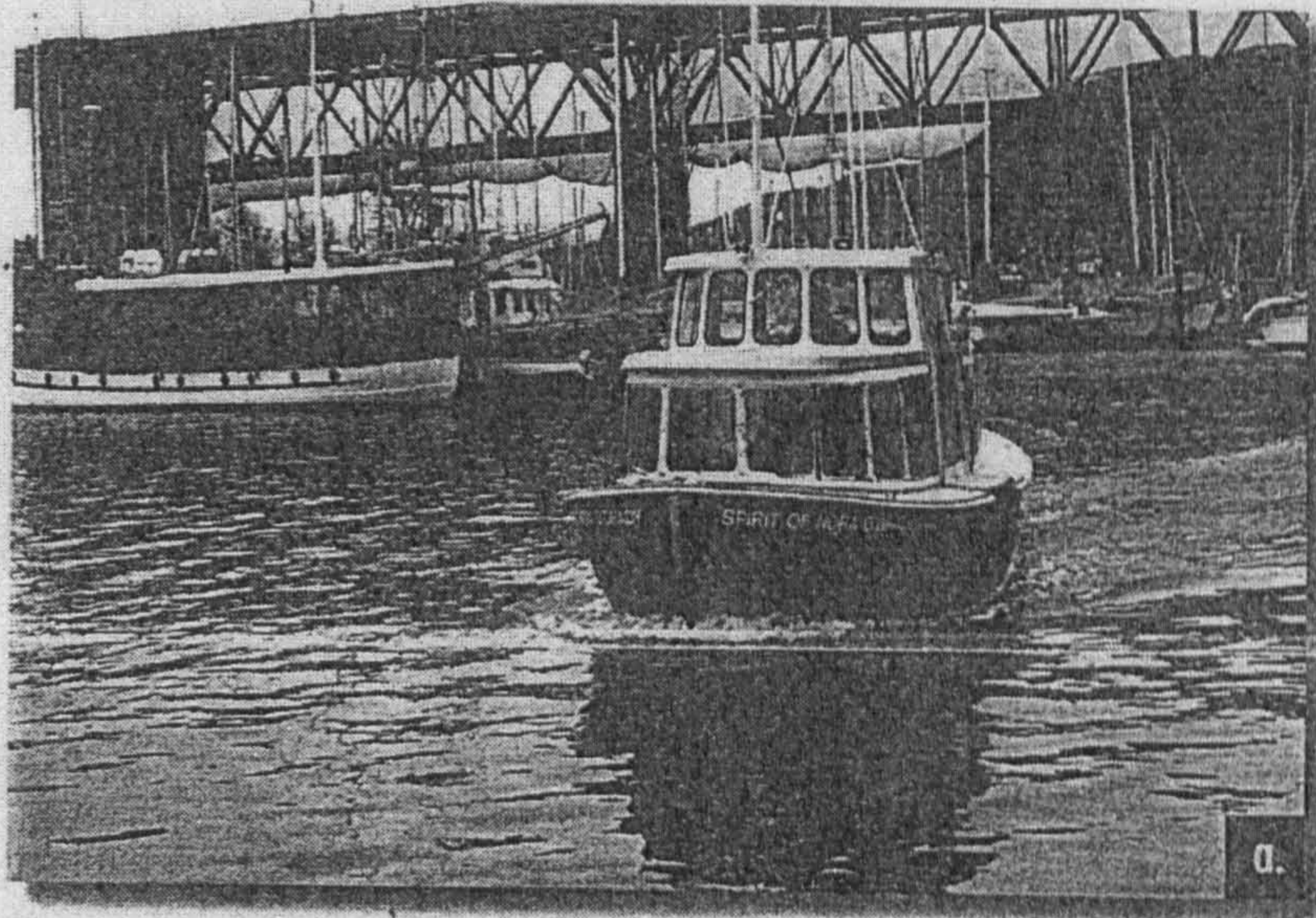
2b: Exercise - Listen to Tom and Lisa. While you are listening, try to put Tom's directions in the correct order (1 to 5).

a) Find a dirt lane.	3	5
b) Walk about five streets.	2	
c) Look for a tree.	1	
d) Walk one kilometer.	4	
e) Find a red brick building.	5	

Now, listen again and double-check your answers.

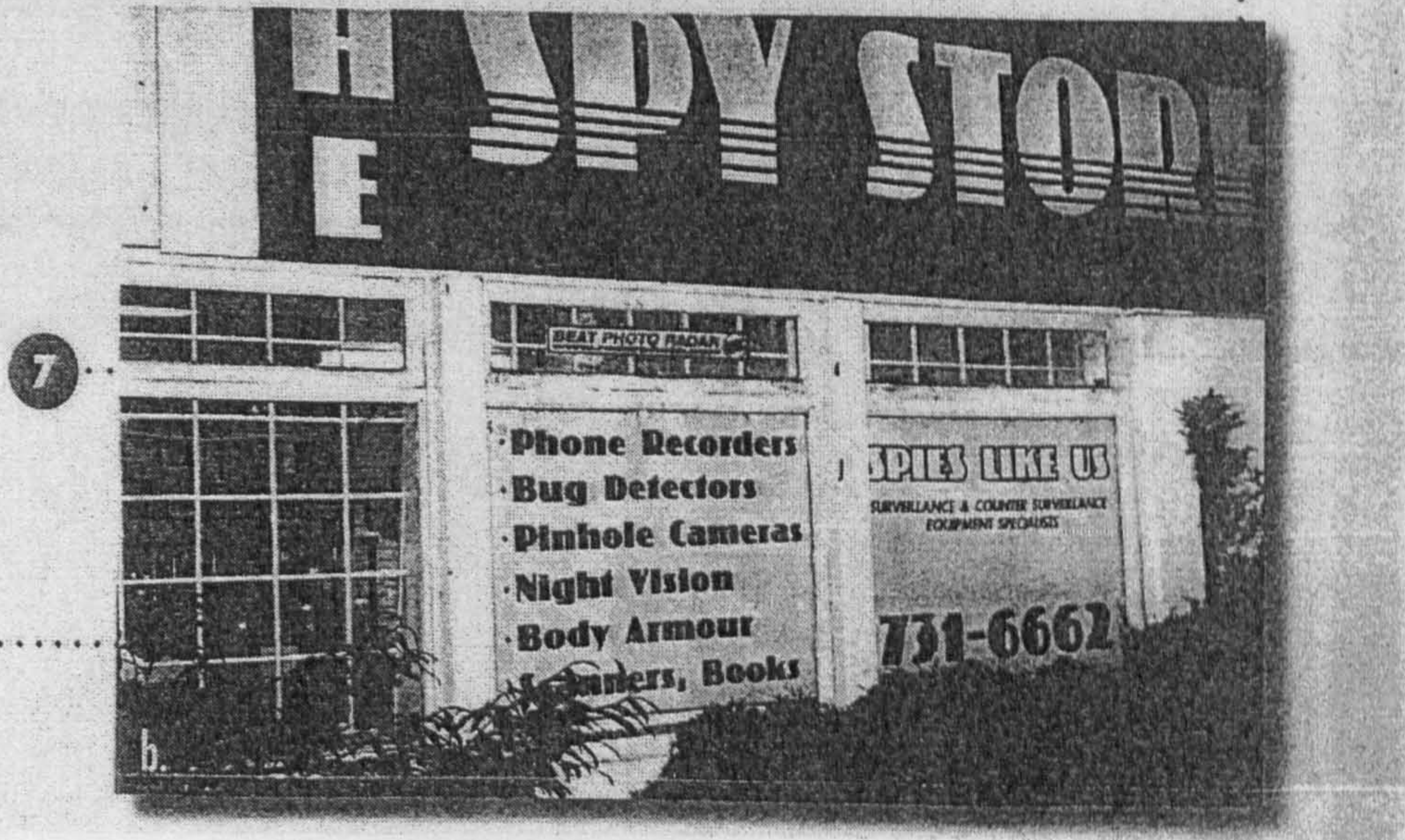
PHOTO GALLERY 1

Look at the pictures and follow the directions you hear.



6

1. 2 2. B 3. C 4. D



7

UNIT 1

3a: Pronunciation - Listen and practice these words with the phonics "ai" sound.

my	high	die
smile	fly	guide
reminder	occupy	deny

8

3b: Listen to the following words. Check the box if the word has the "ai" sound.

- kite sign late dine time flat
 1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12.
 cat say like den hi flight

9

3c: Listen to the sentences and repeat. Underline the words that have the "ai" sound.

1. I sighed when the guy hit me in the eye.
2. My sister likes to dine on time.
3. The diner on Ninth Street is very cheap.
4. Why didn't you turn on the light that night?
5. David tried to file his papers, but he was too tired.
6. She never guided a child up a mountain.
7. My mother doesn't like to make pie.

10

PHOTO GALLERY 2

Look at the pictures and follow the directions you hear.



11

1. b

2. c

3. b



UNIT 1

4a: Preparation - Listen. You will hear six short passages. After hearing each passage, match the word with the correct definition.

- | | |
|-----------------|----------------------------------|
| c 1. Philosophy | a) the mind and its problems |
| b 2. Biology | b) all life |
| d 3. Astronomy | c) thought and beliefs |
| f 4. Chemistry | d) the stars and planets |
| a 5. Psychology | e) social problems |
| e 6. Sociology | f) elements like iron and oxygen |

12

4b: Tom is walking Lisa to the university. Listen carefully to Tom and Lisa's conversation. After the conversation, you will hear five statements. Mark the statements True or False.

1. T F
2. T F
3. T F
4. T F
5. T F

13

4c: Listen carefully to the remainder of the conversation. After the conversation, you will hear four sentences. Circle the best ending for each sentence.

1. a) a psychiatrist. *a doctor trained in the treatment of mental illness*
 b) a psychologist. *someone trained in psychology*
 c) a sociologist.
2. a) in his hometown.
 b) in San Diego.
 c) in San Francisco.
3. a) play tennis with.
 b) teach her tennis.
 c) watch tennis with.
4. a) to a lot of people.
 b) to his sister.
 c) both of the above.

14

E.2 Selected Scripts from *Listen Up*

Selected scripts for *Listen Up* (Teacher's Instruction)

1. When conversation is stale, it is uninteresting.
2. If you are being confrontational, you are speaking in a threatening and pushy manner.
3. Casual conversation is everyday, easy conversation that you wouldn't use with the Queen.
4. If something is inappropriate, it is the wrong thing to say at that moment
5. If something is happening on a monthly basis, it is happening every month.
6. The rule of thumb is the easy and general rule to use in a given situation.

In English speaking countries, small talk is used to break the ice. Don't get too deep into personal affairs at first. Even if an acquaintance is not doing very well, he will say "So-so" or "Okay" when asked "How are you." Confrontational topics, like religion or politics are inappropriate in casual conversation. People in North America start on a first name basis, but Britain is more formal. The senior or older individual should be the first to say "Call me Steve." When talking about family, don't ask if someone is married or why they don't have kids. Instead, start by talking about your own family. When the people you are talking to feel comfortable enough, they will tell you about your family. Two useful topics are the weather and traffic. Both are subjects that everyone can relate to. The rule of thumb is, no matter how stale the conversation, agree and smile. It's only small talk to make everyone comfortable.

Man: So, what have you been waiting long?

Woman: Have I ever! I've been sitting here for 20 minutes already.

Man: The transit is getting worse and worse. But at least the weather is nice.

Woman: Boy, lone line, isn't it?

Man: Yes. Last day before the holiday, it's always so crowded.

Woman: And it's been crawling along like this ever since I arrived. They must have anew teller.

Man: Well, who's in a hurry anyway? It's pouring out there.

E.3 In-class Textbook Two

Hong, H. (2004) *General English Proficiency Test – Intermediate*. Taipei: Bookman.

Part I – Picture Description: Look at the picture, then listen to the question. Choose the best answer.

本測驗分三部份，全為四選一之選擇題，每部份各15題，共45題，作答時間約30分鐘。

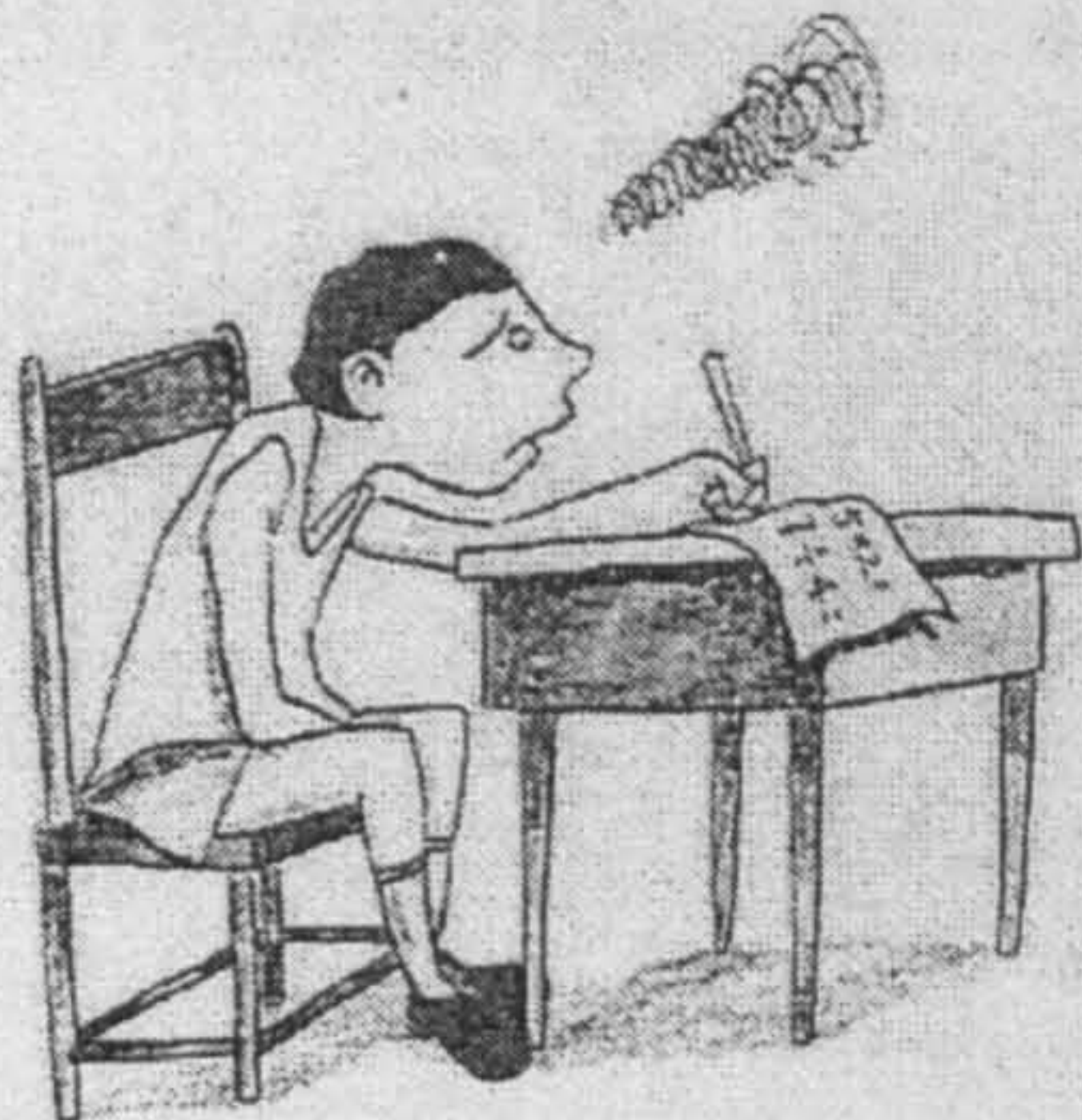


CD1 Track 1

第一部份：看圖辨義

本部份共15題，試題冊上有數幅圖畫，每一圖畫有1~3個描述該圖的題目，每題請聽錄音機播出題目以及A、B、C、D四個英語敘述之後，選出與所看到的圖畫最相符的答案，每題只播出一遍。

例：(看)



(聽)

What is the boy doing?

- (A) He is falling asleep.
- (B) He is doing his math homework.
- (C) He is drawing a picture.
- (D) He is writing a love letter.

正確答案應該選 (B)。

現在開始聽力測驗第一部份。

Part II – Question or Statement Response: Listen to the question and look for the best answer.



CD1 Track 2

第二部份：問答

本部份共15題，每題請聽錄音機播出一個英語問句或直述句之後，從試題冊上A、B、C、D四個回答或回應中，選出一個最適合者作答。每題只播出一遍。

例：

(聽) Did you take the test yesterday?

(看) (A) No, not at all.

(B) No, I didn't go to the picnic.

(C) Yes, it is today.

(D) Yes, it was very difficult.

正確答案應該選 (D)。

現在開始聽力測驗第二部份。

Part III – Short Conversation: Listen to the conversation between two people, choose the best answer.



CD1 Track 3

第三部份：簡短對話：

本部份共15題，每題請聽錄音機播出一段對話及一個相關的問題之後，從試題冊上A、B、C、D四個選項中選出一個最適合者作答。每段對話及問題只播出一遍。

例：

(Man) *How was the party last night?*

(Woman) *I have no idea.*

(Man) *Didn't you go?*

(Woman) *No, I had too much homework.*

Question: What did the woman do last night?

(A) She stayed home and did her homework.

(B) She went to visit her parents.

(C) She did some washing last night.

(D) She went on a trip.

正確答案應該選 (A)。

現在開始聽力測驗第三部份。

E.4 Supplementary In-class Materials

From the *Ez Talk American Conversation Magazine* (published by Heliopolis Culture Group, Taipei)

10) Snow

Recent _____ has been getting more and more unusual. Taiwan normally only gets snow on its very tallest mountains in the winter, but this year had a rare March snow fall on several mountains. This phenomenon wasn't unique to Taiwan; when the warmth of spring should have been taking hold, Mainland China and Texas both had March snowfalls as well. Such strange weather has caused many to fret over whether the _____ that occurred in movie *The Day After Tomorrow* might also occur in real life.

9) Party dues

The _____ for party chairman has caused the issue of KMT party dues to float to the surface. Reputed as being a million members strong, the KMT supposedly has received party dues from less than a third of its members. If the dues must be paid before voting, some _____ there won't be enough votes to be meaningful. But if one can vote without paying the dues, this is unfair to those who did pay up. It looks as though the KMT still has a few things to clarify before election time.

8) Love letter

Can a love letter spark more than just flames of _____ in one's heart? After a teacher took away a Tainan middle school student's love letter from his girlfriend, the student decided to go into the teacher's office to retrieve the letter. But, finding the door locked, the enraged student decided to attack the office with a firebomb. How's that for a perfect example of love driving someone _____!

7) Party flag

You burn my party flag, I'll tear up yours! At the 319 pain-blue march, one participant got _____ and burned a DPP flag. A few days later, in a show of power, DPP legislators in their Legislative Yuan tore up a KMT flag. At the rate things are going, the party flags may _____ abuse for a while longer, as the two parties engage in slash burn.

6) Right to privacy

When a _____ model recently sent her phone in to be repaired, some intimate photos stored in her phone were leaked to the press. Even if the explanation that "it's no big thing for good friends to take _____ together in bathrobes" seemed a little forced, it was, after all, her own private affair. Whether we're talking about the professional integrity of the media or of the cell phone centre, both are in need of repair.

5) The other woman

This time it wasn't that someone was _____. Instead it was that someone published a book that taught the English usage for the Chinese term "the other woman" but taught it wrong. If this writer (actually a musician) had read EZ Talk, she wouldn't have had people laughing at her mistake. EZ Talk is going to help clear up the mess right here – "the other woman" or "the other man" is the way to describe the person who gets involved in someone else's _____.

4) Deposits

In the past, when people _____ for a landline, they had to put up one thousand NT as a deposit. While most people have probably forgotten about this deposit money, legislators are wondering if Chung Hwa Telecom plans to give refunds. The real problem is that, if the refunding process isn't handled correctly, it will be another good _____ for scam organizations to move in for the kill.

3) Teacher-student romance

Is it right for a teacher and a student to fall in love? Those who have been in love before say it is possible, while those who have never been in love say "no way." Students think it could happen, while parents say "not a chance." What do teachers say? If a teacher wants to _____ a student, no matter what, the teacher should wait until the student is old enough and emotionally _____.

2) Water and electricity bills

Savings account _____ rates and salaries never seem to keep up with the rate of inflation. These days gas prices are increasing, and even insurance fees are following the trend. At least the new premier gave everyone a big present upon assuming his office, guaranteeing that water and electricity costs wouldn't _____ during the first half of the year. Now....what about the second half?

1) Anti-Secession Law

The "Anti-Secession Law" has stirred up a lot of _____ responses from the international community, and even the English translation of the name is under debate on both sides of the strait. The Communist Party sees Taiwan as one part of China, _____ the law should be the "Anti-Separation law". But in order to lessen international confusion, Taiwan is calling the law the "Anti-Separation Law" domestically, but calling it the "so-called Anti-Secession Law" in international correspondence. The quotation marks are intentional to demonstrate that the terms are not ones that Taiwan agrees with.

E.5 Test Format of the Mid-term Examination

Listening Comprehension Class (Intermediate) – Mid-term Exam

Student Name: _____ Student No: _____

Part 1: Listen and choose the best answer to the question you have heard. (30%)

- | | | |
|--|---|--|
| 1. A) What a shame.
B) How pitiful.
C) I'm sorry.
D) Don't worry. | 5. A) I'm sorry too.
B) Not very much.
C) Next time.
D) It's nothing. | 8. A) Yes, I like.
B) Yes, I'd like to.
C) No, I don't like.
D) No, I wouldn't like. |
| 2. A) Are you ready?
B) We shall.
C) That's a good idea.
D) Never mind. | 6. A) I don't know who did it.
B) I'm glad to do so.
C) The teacher let us go home early.
D) The teacher is teaching a new lesson today. | 9. A) I said I would consider your proposal.
B) No, you don't have.
C) Never mind about that.
D) No problem with your proposal. |
| 3. A) You are welcome.
B) Please don't say so.
C) With pleasure.
D) I'm pleased. | 7. A) By taxi.
B) By phone.
C) I came by myself.
D) John told me where you were. | 10. A) In ten minutes.
B) After ten minutes.
C) Ten minutes ago.
D) Ten minutes later. |
| 4. A) I don't like it.
B) I'm sorry, I don't know.
C) I don't know the truth.
D) I'm glad to hear that. | | |

Part 2: Listen and choose the best answer. (30%)

- | | | |
|--|--|--|
| 1. A) He shouldn't wear his seat belt.
B) He should wear his seat belt.
C) She's sorry about his loss.
D) She won't wear her seat belt. | 3. A) He wants to watch <i>Oprah</i> with her.
B) He wants to do something else.
C) He's having second thoughts about their relationship.
D) He doesn't know what to think. | 5. A) He fell down.
B) He spilled something.
C) He biked over a spill
D) He own the race. |
| 2. A) She knows the man is right.
B) She thinks the man is crazy.
C) She doesn't know what kind of underpants Bush wears
D) She likes the man's choice of words | 4. A) The man should take a break.
B) The man does a good job with his work.
C) The man needs a new job.
D) The man should continue. | 6. A) totally happy
B) kind of happy
C) very sad
D) angry |

7. A) look at someone
B) take a picture of someone
C) talk to someone
D) get something from someone

8. A) how much money the woman wants
B) why the woman doesn't want to marry him
C) if she made a deal with someone else
D) how he should deal with this problem

9. A) she always agree with Chen
B) she always disagree with Chen
C) something she disagrees with Chen
D) she has no opinion

10. A) actively fix things
B) relax
C) take care of himself
D) take care of her

11. A) the man should change his clothes
B) the man should leave
C) the man can keep the extra money
D) the man should change his attitude

12. A) because it's getting late
B) because she needs to write a note
C) because she doesn't like the man
D) because she noted the man's bad breath

13. A) flowers are worth studying
B) she thinks the man is mean to her
C) she thinks the man isn't mean to her
D) she is very happy to get the flower

14. A) he doesn't know
B) he has a screw driver
C) he doesn't have a screwdriver
D) he wants a date with her

15. A) play football
B) go kick boxing
C) relax
D) get a back message

Part 3: Listen to the short talk and write T (True) or F (False). (28%)

- _____ 1. Her boss demands that she do so.
_____ 2. She is prepared for it.
_____ 3. A news reporter.
_____ 4. She doesn't have time to type the letter.
_____ 5. He is twenty-two.
_____ 6. They are discussing which city they are going to visit.
_____ 7. They are going to New Orleans by train.

Part 4: Question and Answer (12%)

- a. What is the news talking about?
- b. How many Taiwanese people travel to Japan?
- c. According to the news, how long the visitors can stay in Japan?

E.6 Test Format of the Final Examination

Listening Comprehension Class (Intermediate) – Final Exam

Student Name: _____ Student No.: _____

Part 1: Listen carefully to the conversation. Mark the statements True or False. (30%)

1. () 2. () 3. () 4. () 5. ()

_____ 6. They are at the beach.

_____ 7. They would like to meet some girls and make friends there.

_____ 8. They are talking about a beautiful girl in the purple top.

_____ 9. Travis's sister is ugly as a mutant.

_____ 10. They are going to put on their sunglasses in order to check out girls.

Part 2: You are to choose the best answer to each question. (30%)

1. A) she is attracted to Alan because he is still
B) she is attracted to Alan even though he is dating someone
C) she isn't attracted to Alan
D) she doesn't think Alan is serious about the girl he is dating

2. A) she wanted to go to the party
B) she wanted to be invited to the party
C) she didn't go just because she wasn't invited
D) one of the main reasons she didn't go was that she wasn't invited

3. A) he thinks the woman is overreacting
B) he thinks the woman has had too much to drink
C) he thinks the woman is talking too much
D) he thinks someone is trying to carry the woman somewhere

4. A) she wants to be his girlfriend
B) she's going to tell his girlfriend what a great guy he is
C) she wants to meet his girlfriend
D) she is going to tell his girlfriend about what he is trying to do

5. A) he thinks he is very lucky
B) he doesn't believe the woman
C) he can't believe how unlucky he is
D) he thinks he is the only one that has this kind of luck.

6. A) there's no chance she will go
B) she wants to make the man's wishes come true
C) she also wishes what the man wishes
D) she wants the guy to make more wishes

7. A) by the way
B) come over here and I will think of it.
C) why didn't I think of it
D) I forgot

8. A) he hasn't been sleeping well
B) he can think of better things to worry about
C) he can think of better things to dream about
D) he definitely won't be late

9. A) Cathy pushed all the buttons on the elevator and made him mad
B) Cathy got aggressive with him
C) Cathy knows what to say to make the man mad
D) Cathy stole his cell phone

10. A) she thinks the man deserved the get hurt
B) she feels sorry for the man
C) she wants to serve the man something to make him feel better
D) she thinks the man is no good at bike riding

Part 3: Listen to the short talk and answer the question. (15%)

1. What did Michael Jordan do for a living before he retired?
2. What is his second career?
3. What product was the No. 1 launch in this market history? And how much money has brought in?
4. What was the second product, which debuted on Oct. 25, 1999?
5. What did Michael Jordan talk about his second product?

Part 4: Listen to the news and answer the questions. (15%)

1. Where and when "Taiwan Festival" has been held? (6%)
2. What was the theme of "Taiwan Festival"? (3%)
3. What was the festival aimed at showing? (6%)

Part 5: Listen to the lecture and answer the questions. (10%)

1. What is the lecture talking about?
2. What two newspapers don't have this section?

1) Don't stop never _____ your head high and _____ the top

Let the _____ see what you _____

_____ back to you

2) _____ to what you try to be, your _____

When the _____ is on your shoulders, just _____ and let it go

If people try to _____ you _____, just _____ don't turn around

You only _____ to yourself

3) Don't you _____ it's true what they _____ life it _____ easy

But your _____ around. So don't you _____ trying

Chorus

4) 5) 7) Don't stop never _____

_____ your head high and _____ the top

Let the _____ see what you _____ back to you

_____ falling in love. Anything _____ thinking of

When the world _____ to get too _____

_____ back to you (Na na na)

5) Try not to _____ thing. _____ the good time's life can bring

_____ inside you, _____ your feelings show

_____ is the key, cos you are your own _____

You never _____ lonely, when time is _____ your _____

6) Don't you _____ it's true what they _____. Things _____ to try you

But your _____ around. So don't you _____ trying

Repeat Chorus

8) Don't you _____ it's true what they _____

Things _____ for reason

But your _____ around. So don't you _____ trying.

E.7 Quiz Format

Student Name: _____ Student No.: _____

Part 1: Listen to the song and fill in the missing words. (50%)

_____, _____, _____, _____

_____, _____, _____, go go go.

_____, _____, _____, everybody

_____, _____, _____, _____

Let's fun fun together. Let's play the _____.

_____, _____, _____ all together.

_____, _____, _____ all together.

Let's go everybody and play _____.

Part 2: Listen to the news and answer the questions. (25%)

A) Where were they arrested? Why?

B) Where have they been recently?

C) Do they have any previous criminal records?

D) Why did aviation police put them on a special monitoring list?

Part 3: Listen to the conversation and answer the questions. (25%)

A) Where are they? What do they want to do?

B) How big is it?

C) How many people can sleep in there?

D) Why didn't they rent it?

E.8 Problems Students Found in the Mid-term and Final Exam

1. The quality of the recording and/or the visual equipment

| No. | Characteristics of the Quality of the Recording | Mid-term Exam | | Final Exam | |
|-----|---|---------------|-------|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | The quality was so poor that I couldn't hear the text clearly. 某些部份會突然變小聲 X2。 | 11 | 33.3% | 9 | 27.3% |
| b. | The background noise outside the testing environment was too loud. | 5 | 15.2% | 0 | 0% |
| c. | The volume of the tape recorder or audio-visual equipment was too low. | 4 | 12.1% | 14 | 42.4% |
| d. | The quality of the recording was good, and I could hear the texts clearly. | 14 | 42.4% | 13 | 39.4% |
| e. | Other: | 0 | 0% | 0 | 0% |

2. Testing time

| No. | Characteristics of Testing Time | Mid-term Exam | | Final Exam | |
|-----|---|---------------|-------|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | Time was too limited to answer all the questions properly. 答題時間太短，還來不及答題下一題的題目就出來了。 | 12 | 36.4% | 9 | 27.3% |
| b. | The testing time was sufficient for me to answer all the questions properly. 時間太多，聽太多遍了 X5。 | 21 | 63.6% | 24 | 72.7% |
| c. | Other: | 0 | 0% | 0 | 0% |

3. The test/task instructions

| No. | Characteristics of Test/Task Instructions | Mid-term Exam | | Final Exam | |
|-----|---|---------------|-------|------------|-----|
| | | Freq. | % | Freq. | % |
| a. | The instructions of each test section were not clear. | 15 | 45.5% | 1 | 3% |
| b. | The instructions were too complicated. | 0 | 0% | 0 | 0% |
| c. | The test/task instructions were clear. | 18 | 54.5% | 32 | 97% |
| d. | Other: | 0 | 0% | 0 | 0% |

4. The length of the listening texts

| No. | Characteristics of the Length of the Texts | Mid-term Exam | | Final Exam | |
|-----|--|---------------|-----|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | The texts in the test were longer than I was used to listening to in class. | 2 | 6% | 4 | 12.1% |
| b. | Longer texts increased the difficulty of understanding. | 2 | 6% | 3 | 9% |
| c. | I got lost in listening to the longer texts. | 2 | 6% | 3 | 9% |
| d. | The lengths of the texts in the test were similar to those I listened to in class. | 31 | 94% | 29 | 87.9% |
| e. | Other: | 0 | 0% | 0 | 0% |

E.9 Mid-term Interview Transcription

Question 1. Where did the content of the test come from?

T: 從英文雜誌上面來的，還有全民英檢的題庫。

I: 是什麼樣子的雜誌？

T: 就是 Eztalk 的雜誌，因為我覺得他裡面有蠻多實用的東西有關我們周遭的生活，那本雜誌裡有附錄一些考題，就是在聽完裡面的內容可以做考題測驗你的聽力，看看有沒有吸收。

I: 學生知道你用這本雜誌內容當考題嗎？

T: 開始可能不知道，但後來可能就知道了。

I: 可是這本雜誌市面上買的到，有沒有可能學生會在考前接觸到相關考題？

T: 我不會用近期的雜誌內容當考題。

I: 那全民英檢的考題呢？

T: 是從上課用書裡面出來的，因為系主任說希望我們系裡的學生畢業前能通過全民英檢的考試，所以我考試才會出。

I: 所以您這次是自己出題嗎？

T: 對。

I: 學生從來都沒有聽過這次的考試內容嗎？

T: 沒有。我是從教科書裡比較後面的章節選考題，因為全民英檢我們只上了兩回，我這次就從最後一回選題目。

I: 那學生有可能會聽過囉？

T: 基本上機率很低，因為我發現學生都蠻被動的，除非他們真的很認真聽到最後一回，但我還是有改題目。我記得有一次我小考用同樣的聽力內容但不同的問題，學生卻不見得考的好。因為就是全民英檢裡有些 conversation 的問題只有一、兩題，但是其實裡面的內容還有很多可以問，所以我就會改變考試的題型像是 true or false 或簡答題。

I: 學生上課有很主動寫筆記嗎？

T: 沒有，所以我上課的時候讓學生聽完內容就會讓他們找出俚語的部份，然後讓他們分組討論兩三分鐘，然後讓他們講，我都跟他們說你不一定要講對，就算不知道你也要猜對，慢慢培養他們不管對或錯就是講，也就是 autonomous learning.

I: 您覺得您用的教材跟學生程度符合嗎？

T: 我覺得蠻普通的，因為我們這班的學生有些聽力真的蠻好的，可是就是有人聽力很差，中間的比較多，我用的這兩本教材(Listen Up and GEPT) 有些人覺得太簡單、有些覺得太難，但大部份都覺得還好。像是 GEPT 聽力用圖片來回答的他們都覺得太簡單，但像是 short talk 裡面有俚語的或是 news 之類的對他們來講難度就會大很多，平均來說學生都比較難抓到意思。但可能是 GEPT 題目都是選擇題居多，所以他們答對的機率都比較高，但如果我換個題型問得題目不一樣，他們就不一定能答對了。這就是為什麼我會用 Eztalk 的原因，他裡面有很多的俚語，可是 GEPT 裡面的對話感覺上很制式，一聽很容易就知道 conversation 的內容。可是如果我用 eztalk 的東西，很多口語化的東西，我用這樣的給他們，他們都覺得很有挑戰性，比較活。之前第一次小考他們就有跟我反應蠻難的，但我跟他們說不能只學容易的，他們跟外國人溝通不一定都能瞭解他們所說的跟講話速度，所以他們要習慣不同的東西。

I: 那 GEPT 的教材是老師自己想上的嗎？

T: 這個是主任要我上的，可是這個課程內容對我來說我覺得就算你考過了英檢又如何？你只是為了考試去聽聽力，但是既然主任有這樣的想法，我才會搭配英檢的教材。

I: 為什麼您沒有讓學生聽 lecture 的聽力？

T: 因為我發現學生在聽 news 的時候就已經跟不上，你想想 news 都已經很短了，他們都還搞不清楚，那更何況是 lecture 那麼長的內容，基本上我在上課的時候我不喜歡去抹煞學生的學習興趣，要讓他們覺得他們學的東西不會太難也不會太簡單，這樣他們才有動力去學。

I: 在測驗學生前，你有測試這些考題嗎？

T: 沒有，如果我有時間，我會嘗試測試這些考題。

Miss T (T): They were from the English magazine and GEPT test battery.

Interviewer (I): What kind of magazine?

T: It's *EZ Talk* magazine, because I found the topics in the *EZ Talk* magazine were practical and related to our everyday life. There were several test questions included in the magazine which tested listeners' comprehension after listening to the content.

I: Did the students know that you used the test questions from the magazine?

T: They may not have known at the beginning of the course, but they probably did afterwards.

I: But this magazine can be purchased; was it possible that the students procured similar test questions?

T: I didn't use the test questions in recent issues.

I: How about the test questions for GEPT?

T: They were from the in-class textbook. Because the chairperson expected the students in our department would be able to pass the GEPT before graduation, I tested those questions.

I: Did you design the questions yourself?

T: Yes.

I: Have the students ever heard the test content?

T: I think so. I chose the questions from the last chapter. Because I only taught the first two chapters before the mid-term exam, I chose the questions from the final chapter.

I: So it was still possible that students might have heard it before.

T: I think the probability was very low, because I found out that the students were passive about learning, unless they really listened to the last chapter....but I changed the questions. I remembered that I tested the students with the same listening content with different questions before, but the students did not perform as well as I expected. You know...there were simply one or two questions in short conversations in GEPT, but in fact there were other questions to ask based on the conversations. So I would change the type of comprehension question, such as true/false or short-answer questions.

I: Did the students take notes on their own in class?

T: No, so I asked them to find colloquial language in the listening passages, and divided them into groups to discuss the colloquial language for two to three minutes, and asked them to talk about it in class. I told them that they did not have to speak correctly. If they did not know the correct answer, they could still speak out, or guess the answers...I wanted to develop their autonomous learning skills.

I: Did you think that your students' language ability met the level of the listening materials you used in class?

T: I found it OK, because the majority of the students' listening ability was in the middle; only a few students' ability was very good. Some students felt that the two sets of listening materials (*Listen Up* and *GEPT*) I used were too easy, and some felt they were too hard, but the majority said they were OK. For example, they considered that questions in "Picture Description" in the GEPT were too easy, but the colloquial language in the short conversations or news items was more difficult for them; it was harder for them to get the meaning in general. It was possible that GEPT consisted of more multiple-choice questions, so the probability that they answered correctly was higher; but if I'd changed another type of question, they might not be able to answer correctly. That's why I used the *EZ Talk* magazine; there was a lot of colloquial language in there, but the conversations in GEPT were formulaic, because you can understand the contents easily. But I used the questions from *EZ Talk* magazine in the first quiz, the students felt challenged and practical, owing to the colloquial language, but they also told me the quiz was difficult. However, I said to them that they could not always learn something easy; they will not always be able to understand what foreigners say and always follow their speaking speed, so they needed to be familiar with something different.

I: Did you yourself decide to teach GEPT?

T: The chairperson asked me to teach that. But for me, I felt that even though you passed the GEPT, you simply practiced the listening for testing purposes....well...I taught the GEPT since the chairperson asked me to do that.

I: Why did not you teach listening to "lectures"?

T: Because I discovered that the students could not follow the speaking speed while listening to the news....you see, news reports were usually short in length, but they could not cope with those, not to mention the long length of "lectures". Basically I do not like to kill their learning interest in class; they are more likely to feel encouraged when the materials are neither too hard nor too easy.

I: Did you pilot the test items, I mean run a small test, before you test your students in this exam?

T: No, if I had time, I would have tried to pilot those questions.

Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?)

T: 60。期中 30%，期末 40%，quiz 30%。很多期中期末考不及格的學生會要求我期末考簡單一點，或甚至考上課聽過的內容，但是我跟他們說考上課聽過的東西就算你考滿分也根本不是你真正的實力，所以我都不會出的跟課堂上的練習一模一樣的題目，不過就算我用不同的聽力內容，學生程度好的還是拿高分，中間的還是中間，但果我用他們聽過的內容，那不就全部的人都考高分了。如果他們不想被 fail 掉，只能從平常我考的 quiz 上拿分了，因為 quiz 比較簡單。但我還是要求他們要用功而不是一味的要求考題出的簡單。

T: 60. The mid-term score accounted for 30% of the total score. The final exam accounted for 40%. The quiz accounted for 30%. The students who did not pass the mid-term exam asked me to make the final exam easier, or even test them on the listening extracts they had listened to in class. But I told them that, even if you received full scores by being tested on the listening extracts you had listened to in class, it would not constitute your real language ability. So I never used the questions they had practiced in class. And even though I used different listening contents, those who had scored higher still received higher marks, and those with middle scores were still in the middle. So, if I used the listening passages they had listened to before, then everyone would receive high scores. If they did not want to be failed, they would have to try to gain higher scores in the quizzes, because the quizzes were easier. But I still asked them to work harder, instead of them asking me to give them easier tests.

Question 3. What did you expect the students to have learned from your class?

T: 就像我剛才說得，GEPT 的東西比較制式，我希望教他們他們以後會用的到的東西、會接觸到的、或比較生活化的，像是 news，他們不只學到語言，也學到很多目前的知識。如果是語言學習方面，我覺得沒有什麼學習策略，就是多聽多練習多運用，因為就算我今天告訴學生什麼策略來準備考試，學生的實力也不一定顯現出來。所以我跟學生說要他們多聽不同主題，這樣他們的英文聽力才會進步。我跟學生說聽力不像是在考單字，不是前一天晚上聽隔天就能考的好，聽力是累積的。我跟他們說每天至少聽半小時的 radio 或 news 來增進聽力。

T: Just as I said before, the language in GEPT was formulaic. I would like to teach them the language they would use and hear everyday in the future, such as news reports; they could learn the language and current socio-economical information in news reports. As for language learning, I did not think that there were such learning strategies: you simply listen, practice, and use it. Because even though I told the students what strategies they needed for preparing the examinations, it was still hard to test their real ability. I told the students to listen to different topics, so their listening ability would improve. I also told them testing listening is not like testing vocabulary. It is not the case that if you practice listening the day before the examination, you could perform well. Listening ability needs continuous training. I told them to spend at least half an hour listening to the radio or news reports to strengthen their listening ability.

Question 4. What are the mid-term test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Do you think you have achieved them?

T: 我不想把他們訓練成只是爲了通過考試才去聽英聽，我比較希望學生學能夠活用的東西，我覺得教了半個學期以來他們的成績有慢慢的在進步，不像一開始聽不懂，最起碼他們有進步一點，有達到少部分我所期望的，但是因爲每個人的學習狀況都不同，他們的 preference 也不同，學習速度也不同，所以我覺得要完全達到很困難，但是平均說起來他們都還是有在進步。

I: 那這次的考試學生哪個大題考的比較差？

T: 其實沒有說特別差的，但還是有些學生在簡答題上還抓不太到問題的方向。

I: 我注意到你在期中考試裡沒有測驗學生的口語能力？可以跟我說理由嗎？

T: 嗯...基於系上課程的目的，我覺得這個課是專門訓練聽力的，而不是口語，所以我會遵照系上的課程規劃。

I: 那你會在期末考口語嗎？

T: 嗯....我想不會。

T: I did not want them to practice their listening ability for examination purposes. I hope that my students know how to use the language. I felt that they were improving after teaching half a semester. At least they improved a bit, and met a small part of my expectation. However, because each student's learning situations, preferences, and learning speed varied, I found it difficult to completely achieve my goals, but on average, they are improving.

I: On which part of the test did the students perform worst?

T: There were no obvious differences, but there were still several students who did not get the meaning of the short-answer questions.

I: I notice that you did not test students' speaking ability in the mid-term exam. Were there any reasons?

T: Well...based on the course aim in our department, I think this course was designed to train listening rather than speaking. I just followed the departmental curriculum.

I: Will you use any speaking test for the final exam?

T: uh...I don't think so.

Question 5. How did you identify and decide on difficulty of the content/items in the mid-term test? What were the criteria?

T: 會從小考的成績，平常他們課堂的表現，還有課堂補充的資料。如果他們對某些主題沒有興趣或考的不好，我就會依照他們的反應跟表現來調整。

T: I decided on the difficulty of the mid-term exam from the results of the quizzes, their in-class performance, and supplementary materials. If they were not interested in certain topics or they did not perform well, I adjusted the difficulty in the light of their reactions and performance in class.

Question 6. Why did you choose the particular types of comprehension question to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

T: 我之所以會出問答題就是希望他們不要只會回答選擇題，從裡面去猜答案。所以我會用不同類型的考題去考他們，因他們也讀的蠻辛苦的，我也不需要出太難的題目去考倒他們，所以像出簡答題就能知道他們是不是真的聽懂真的有沒有進步。當然學生是比較喜歡選擇題，但我覺得光考選擇題他們可能會猜答案，不能確切的知道他們是不是真的瞭解內容。所以我比較喜歡用不同型態的問題，這樣才能測出他們的實力。我不想 fail 他們，所以如果他們要考過就要靠自己的實力去聽。像 true and false 他們也會猜，要不就全部填 true 或全部填 false，所以我比較喜歡用簡答題，這樣他們才不會去猜。

T: The reason I tested them using short-answer questions was to prevent the students from guessing the correct answers in the multiple-choice questions, so I tested them with different types of comprehension question. I thought that they were busy studying, so I did not need to test them with too difficult questions. If I used short-answer questions, I would know whether they really understood and improved their listening. Well...of course students preferred the multiple-choice questions, but I thought that if I tested them with multiple-choice questions, they might guess the answers, so then I would not be able to know if they really understood the contents. So I preferred to use different types of question to assess their real ability. I did not want to fail them, so if they wanted to pass the exam, it would need to depend on their true ability. They also guessed the answers in the true/false questions; they chose either "true" or "false" for all questions, so I preferred to use short-answer questions to prevent them from guessing the answers.

Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)

T: 我看到他們期中考有進步就會想要在加深難度，所以期中考後 news 的東西會比較多。歌曲跟影片是我蠻想嘗試給他們聽的，可是期中考之前我還沒有給他們嘗試，我想說過了半學期先看看他們的程度在決定要不要給他們聽。因為他們這此考試有進步，我想說過個幾週讓他們聽歌曲跟看影片。我不可能永遠停在原點，不然他們不會進步，雖

然說我進到下一個 level 會對程度不好的同學造成負擔，但我也要考慮程度好的同學，所以就選東西來給他們聽對我而言是個挑戰。

I: 所以您期中考後教科書還是會照用？

T: 對，但是會給他們額外的高階聽力教材，因為他們的 textbook 是最基本的，他們一定要懂得，其他的我就給補充資料。

I: 那您的期末考題也會比較難一點嗎？

T: 我很想考難一點，但是會有很多學生 fail 掉，考的太簡單也不太可能，我覺得應該會比期中考難一點點，希望成績不好的學生在 quiz 上面可以考好一點，不然期末考的比分那麼重，要是連 quiz 都考不好，那我也沒辦法了。

T: I would like to increase the difficulty after seeing their improvement in the mid-term exam, so I will increase the "news report" in class materials. I want to give music and films to the students in English in class, but I have not tried it before the mid-term exam. I was unsure if I wanted to give or not to give them the music in class before I knew their level of English in the mid-term exam. However, because they did improve in this exam, I think I will play music and films in class after several weeks. I could never stay in the same place, otherwise they will not improve. Although moving to the next level might increase the burden on those whose level of English was lower, I still have to consider the students with a higher level of English. Choosing appropriate materials for the students is a challenge for me.

I: Will you still use the two textbooks after the mid-term exam?

T: Yes, I will, but I will give them extra advanced listening materials. Because their textbooks were fairly fundamental, I will give some extra materials.

I: Will you increase the difficulty of test contents in the final exam?

T: I want to increase the difficulty, but many students will fail. But it is not possible for me to give an easier test, so I think it will be a bit harder than the mid-term exam. I hope that those whose English abilities are lower will be able to perform well in the quizzes. The score for the final exam accounted a large percentage of the total score; if they cannot perform well in quizzes, I am not sure what I will do.

E.10 Final Interview Transcription

Question 1. Where did the content of the test come from?

Miss T (T): 也是從 GEPT 的題庫跟 EZ Talk 的雜誌上來的。除了 GEPT 的選擇題，我這次考比較多的簡答題，其中有一題是 short talk，一題是 news，還有一題是 lecture。

Interviewer (I): 您這學期上課有讓學生聽過 lecture 的東西嗎？

T: 沒有，但是他們聽我上課就是在聽 lecture。

I: 可是您在上次的 interview 裡說過，因為學生的程度還沒好到能聽長篇的 lecture，所以你不會讓他們聽，但為什麼這次又考他們？

T: 其實這篇 lecture 並沒有很長，因為我只讓他們聽前三小段，況且我才出兩題蠻簡單的題目，我覺得他們應該沒問題。

I: 那學生這次考的如何？

T: 簡答題得部份還是考的不好，加上題目很多，所以很多程度不好的學生就考砸了。

I: 你這次期末考前，你有測試這些考題嗎？

T: 沒有。期末都比較忙，所以也沒時間準備。

Miss T (T): It was also taken from the test battery of GEPT in the textbook and the extracts from *EZ Talk* magazines. In addition to the GEPT questions in the form of multiple-choice, I tested more short-answer questions this time – one part was a short talk, another was a news report, and the other was a lecture.

Interviewer (I): Had your students listened to “lectures” during the term?

T: No....but they had listened to my lecture, it was a way of listening to “lectures.”

I: In the mid-term interview, you said that you would not ask them to listen to “lectures” because their English ability was not good enough to cope with the long length of “lectures”. But why did you test them with “lectures”?

T: In fact this lecture was not very long, because I simply selected the first three paragraphs to test them, and I only designed two questions regarding the lecture which I thought were easy.

I: What did you think of the students’ performance this time?

T: They did not perform well on the short-answer questions. Because there were too many short-answer questions, the students with lower English ability failed.

Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and students’ learning outcomes? Did you think you have achieved them?

T: 我希望他們能比期中考來的更進步，聽的懂更高一階的聽力內容，也就是希望他們的聽力程度有比期中考來的高。可是我他們這次考試並沒有考的比較好，我想是簡答題的題目太多了，分數比重也高，所以他們才沒拿高分。我原本以為他們這次期末考會考的不錯，因為他們每週上課都有在進步，雖然進步很慢....我想這次出題可能真的出的太難，他們才考不好。

I: 那那首加分歌呢？學生的表現如何？

T: 他們並沒有考的特別好。

T: I hoped that they would perform better than they did in the mid-term exam, and they would be able to comprehend advanced listening content. That is, I hoped that their English ability was higher than it was in the mid-term exam. But they did not perform well in the final exam. It was possible that there were too many short-answer questions which accounted for a large proportion of the total score, so they did not score highly. I hoped that they would perform well this time because they made progress every week, though slowly....I think I may have used too difficult questions this time, so they failed to perform well.

I: How about the song for extra marks? How was the students’ performance?

T: They did not perform very well on that either.

Question 3. How did you identify and decide on the difficulty of the content/items in the mid-term test? What were the criteria?

T: 我是基於他們整學期上課的表現還有期中考的表現來決定期末考題的難易度，但可能我挑的題目過難，學生一時之前還是無法理解吧。

T: I defined the difficulty of the final exam based on the students’ in-class performance this term, but it was possible that I used too difficult questions, so they could not understand the final exam.

Question 4. Why did you choose the particular types of comprehension question to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

T: 我增加了很多簡答題。主要是我覺得考太多選擇題會造成他們猜答案的僥倖心態，簡答題就是真的考驗學生有沒有聽懂題目，結果他們大部分都還不太能抓到題目的重點，要不然就是聽到前面，回答了前面的問題，後面就不知道聽到哪裡去回答不出來。

I: 有沒有可能是題目太長？

T: 我覺得還好耶，我覺得跟在上課聽的長度沒多大的出入。

T: I increased the number of short-answer questions, because I felt that it would result in students guessing the answers if I used too many multiple-choice questions. Short-answer questions really tested whether students understood the questions or not. In fact most of them could not understand the passages, or else they answered the first one or two questions and then got lost in the following questions.

I: Was it possible that the listening extracts were too long?

T: Well, I found the length was OK. I felt that it was similar to what they had listened to in class.

I: Did you pilot the test items before the final exam?

T: No, I was busy at the end of the term, so I did not have time to pilot.

Appendix F

F.1 In-class Textbook

Sampson, N. (1999) *Way Ahead: A listening and speaking course*. Hong Kong: Macmillan.

Note: Textbook copyright requested and photocopy permitted.

CHAPTER THREE

3

A GUIDED TOUR

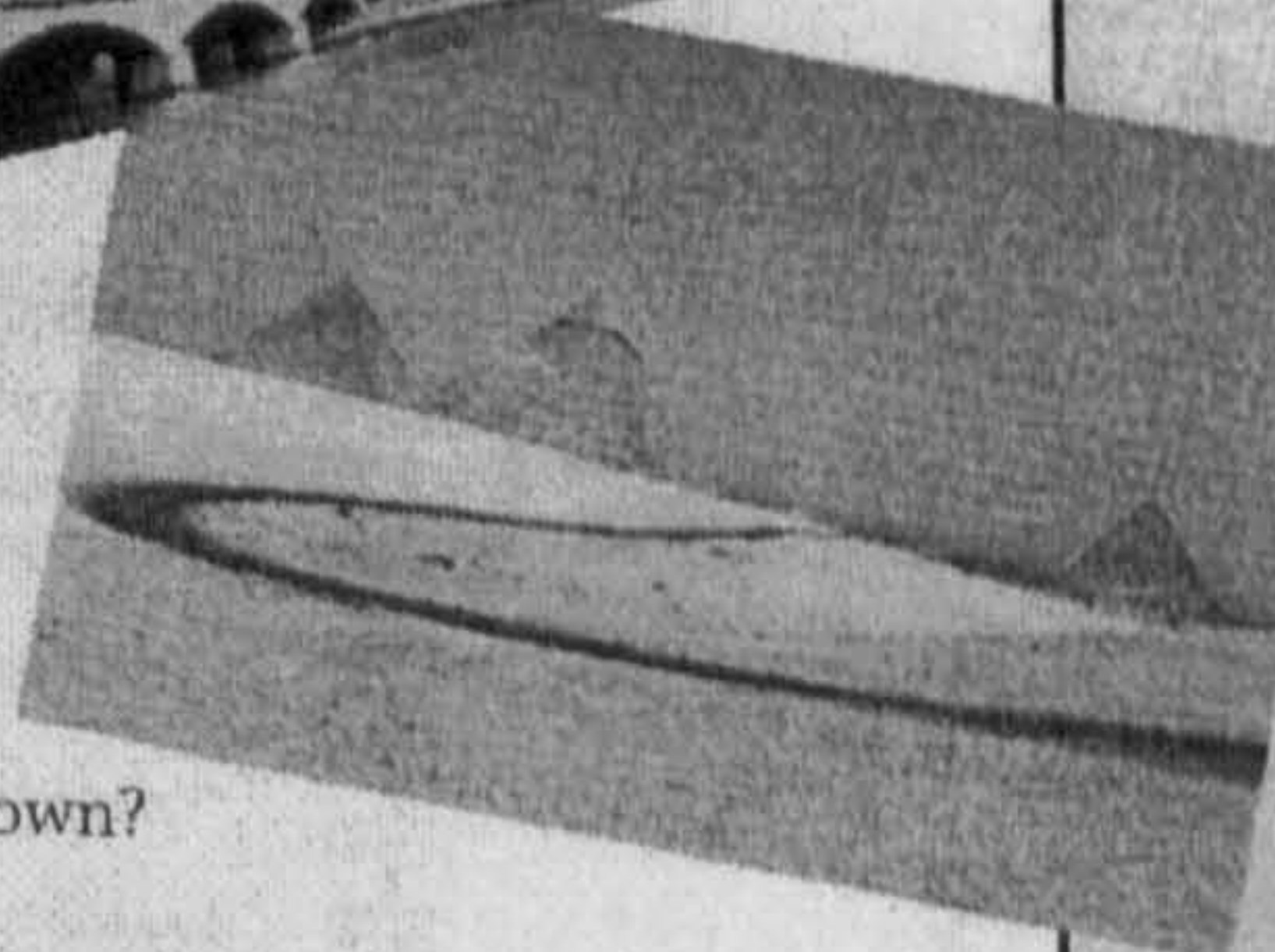
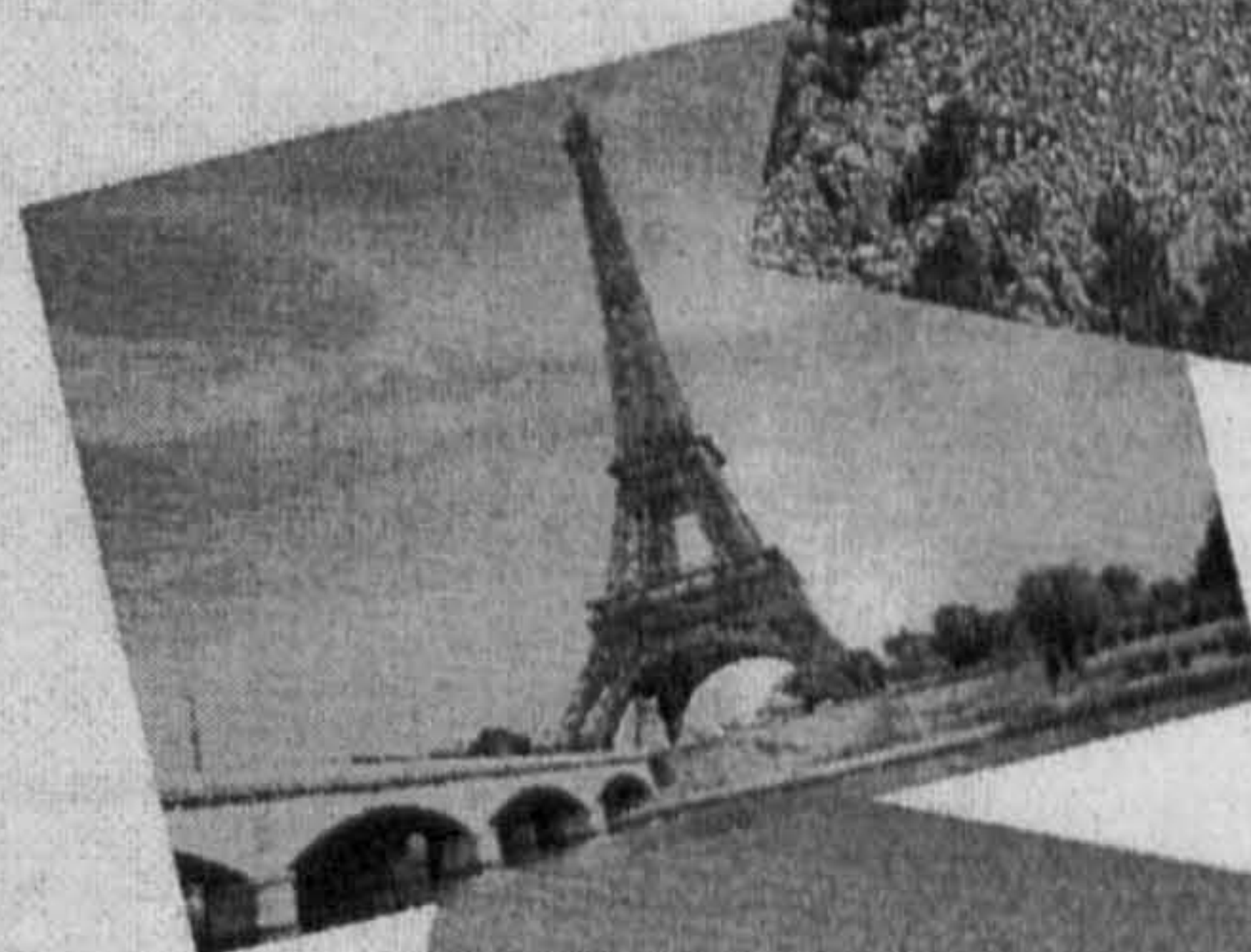
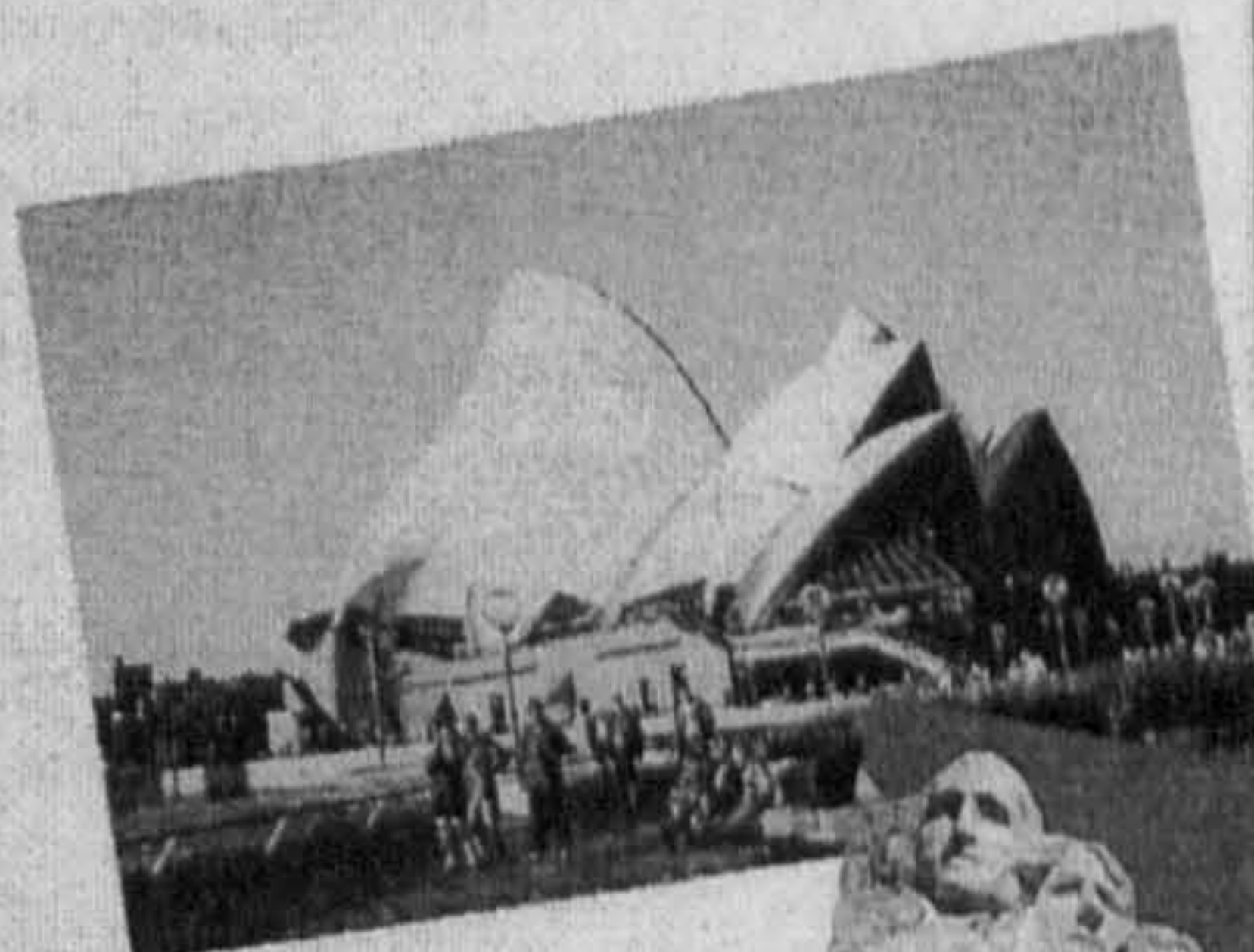
Think about ... Your Hometown

A If you were to show visitors around your hometown, what kinds of places would you take them? Write at least two examples for the attractions. Give reasons for your choices.


- 1 A famous landmark
a) _____
b) _____
- 2 An interesting public place
(such as a park or museum)
a) _____
b) _____
- 3 A good restaurant
a) _____
b) _____
- 4 A local food specialty
a) _____
b) _____
- 5 A shopping district
a) _____
b) _____
- 6 An entertainment area
a) _____
b) _____
- 7 A point of architectural interest
a) _____
b) _____
- 8 A unique feature
a) _____
b) _____

B Discuss these questions with your classmate.

- 1 Do you live in your hometown?
- 2 If not, how often do you visit it?
- 3 Do all of your relatives live in your hometown?
- 4 If not, where do the others live?
- 5 What special memories do you have of your hometown?
- 6 Are you proud of your hometown? Say why or why not.



Listen to it (1)

 Katherine is British. She is in New York for a few days. One evening, she meets her brother's friend, Judith, and arranges plans for the following day. Listen to the dialog then practice it with your classmate.

Judith: If you're not doing anything tomorrow, we could go to Central Park.

Katherine: That's very kind of you, Judith. But it's Saturday and I don't want to take up your time.

Judith: It's no trouble. I'd love to take you there.

Katherine: That would be wonderful. Are you sure you don't mind?

Judith: Of course not! Let's make a day of it. We can do some shopping in the morning, then go to Central Park in the afternoon. Oh, have you been to MOMA yet?

Katherine: No, I haven't. What is it?

Judith: It's short for the Museum of Modern Art. It's fantastic. They have an amazing collection of pop art there, including Andy Warhol's most famous paintings.

Katherine: I'd love to go there.

Judith: OK, we'll meet tomorrow morning and I can take you to Fifth Avenue. That's where all the expensive, upscale boutiques are. We can do some window-shopping there, then go to SoHo. It's also a shopping area, but it has small, unique shops. Do you like Italian food?

Katherine: Yes, very much.

Judith: Great! I know a lovely restaurant in Little Italy. We can go there for lunch, then take a walk in Central Park afterwards. Later on, we can take a cab to MOMA and maybe get some dinner after that. How does that sound?

Katherine: That sounds great!

Judith: I'll meet you at your hotel at 10 o'clock tomorrow morning. Is that OK with you?

Katherine: Yes, that's fine.

Judith: See you tomorrow then!

Katherine: Yes, see you tomorrow.



Work it out Read through the dialog again and fill in Katherine and Judith's itinerary for tomorrow. Follow the example.

| SATURDAY | |
|--------------|---------------------------|
| 10 o'clock | meet at Katherine's hotel |
| Before lunch | _____ |
| Lunch | _____ |
| After lunch | _____ |
| Evening | _____ |

Before making suggestions

| | | | |
|----------|---|-----------------------------------|------------------------|
| Have you | been to Central Park?
visited MOMA?
tried Italian food? | Do you know
Have you heard 'of | Little Italy?
MOMA? |
|----------|---|-----------------------------------|------------------------|

Making suggestions and invitations

| | | | |
|-------|-------------------------|--------------------------------|--------------|
| I can | show you
take you to | Fifth Avenue,
Little Italy, | if you like. |
|-------|-------------------------|--------------------------------|--------------|

| | | |
|--|--------------|------------------------|
| Would you like to
Shall we
Let's
Why don't we
Do you want to | see
go to | Central Park?
SoHo? |
|--|--------------|------------------------|

Checking an invitation


| | | |
|--|---------------------------------|---------|
| Are you sure?
Are you sure you don't mind?
Are you sure it's no trouble?
I don't want to take up your time. | It's no trouble
I don't mind | at all. |
|--|---------------------------------|---------|

Accepting and refusing an invitation

| Accepting | Refusing |
|------------------------------|---------------------------------------|
| That's a good idea. | I'd rather not. |
| Yes, I'd love to (go there). | No, I'm afraid I can't. |
| OK. Let's (do that). | If you don't mind, I'd prefer not to. |
| That sounds good. | I'm sorry, I've already made plans. |

Over to you

Work in pairs. Take turns inviting your classmate to see the sights of your hometown. Use the prompts to plan your itinerary. Think of a few places that you wouldn't take your classmate to, and say why.

| | |
|---|---|
|  | <p style="text-align: right;">Date _____</p> <p>A.M. a meeting time</p> <p style="padding-left: 20px;">a meeting place</p> <p style="padding-left: 20px;">places to visit/things to do before lunch</p> <p style="padding-left: 20px;">a restaurant/local specialty for lunch</p> <p>P.M. places to visit/things to do in the afternoon</p> |
|---|---|

Listen to it (II)



The day after sightseeing in New York, Judith invites Katherine home to meet her mother. Listen to the dialog then practice it with your classmates.



Judith: Mom, do you remember Richard, my friend from England? He used to live in New York and you met him a few times.

Mother: Yes, I remember. A very handsome young man, wasn't he?

Judith: Mom! This is Richard's sister, Katherine. She's in New York for a few days and I've been showing her the sights. Katherine, this is my mother.

Katherine: It's a pleasure to meet you, Mrs ...

Mother: Oh call me Betty, please. And it's lovely to meet you, too, Katherine. Here, let me take your coat. Please make yourself at home.

Katherine: Thank you, er, Betty. That's very kind of you.

Mother: How is your brother, Katherine? He was studying here, wasn't he?

Katherine: Yes, that's right. He's very well. He's continuing his studies back home in London now, but he misses New York a lot.

Mother: That's nice to hear. Now, can I get you anything to drink? Some coffee or tea?

Katherine: Well, er ...

Judith: Some coffee would be nice, Mom. How about you, Katherine?

Katherine: Some tea, if it's not any trouble.

Mother: No trouble at all, dear. And I bought a cinnamon babka at the bakery down the street, I insist you try some.


Judith: Oh, no thanks, Mom. I'm still full from dinner. Have you ever tried babka, Katherine?

Katherine: I'm afraid not. Is it a sweet?

Judith: Yes, it's a delicious Polish cake that is very popular here. Let me get you a piece.

Katherine: Thanks. I would love to try it. One of the things Richard says he misses most about New York is the wide variety of foods available.

Work it out Listen to the questions and write the answers in the spaces provided.

| | | |
|---|--|--|
|  | 1 _____
2 _____
3 _____
4 _____ | 5 _____
6 _____
7 _____
8 _____ |
|---|--|--|

Talk about it (II)

Making offers

| | | |
|----------------------|------|---------------------------------|
| Can I | | take your coat? |
| Why don't you let me | | get you something to eat/drink? |
| May I | | |
| Would you | like | any help? |
| | | me to (do something) for you? |
| | | care for something to drink? |

| | |
|-----------------|------------------|
| Let me | get you a piece. |
| I'll | |
| Leave it to me. | |

Accepting and refusing offers

| Accepting | Refusing |
|--------------------------|---|
| Thank you. | No, thank you. |
| That's very kind of you. | No, it's OK, thanks. |
| OK. Thank you very much. | No, I can manage, thank you. |
| That would be (nice). | No, there's no need. I'm fine, thank you. |

Over to you Work in pairs. Take turns playing the roles of the secretary and the visitor. The visitor has a lunch appointment with his/her brother, who is in a meeting. Use the prompts to conduct the dialog.

Secretary

- Greet the visitor.
- Explain the situation.
- Offer to take the visitor's coat/bag/umbrella, etc.
- Invite the visitor to sit down.
- Offer a drink/magazine.

Visitor

- Introduce yourself. Give a reason for being there.
- Show your understanding of the situation.
- Accept or refuse secretary's offers.
- Accept or refuse.
- Accept or refuse.

Listen up



A It's Friday afternoon. Listen to these people trying to decide what to do on the weekend. Check (✓) the activity chosen in each dialog.

- 1 play touch football
 play basketball
- 2 go to the movies
 watch a video at home
- 3 go swimming
 go kayaking
- 4 go window-shopping
 go shopping
- 5 do research
 go to a music festival
- 6 try out some new computer software
 buy some new computer software



B Listen to the dialogs again and decide if these statements are True or False. If the answer is not given, check (✓) the I column for *Insufficient Information*.

| T | F | I |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

- 1 Tom wants to go to the gym.
- 2 Sarah's friend has bought two tickets to the movie Sarah wants to see.
- 3 Julie hasn't been kayaking before.
- 4 Alison and her friend don't have much money.
- 5 Matt doesn't want to go to the music festival.
- 6 Paul spent a lot of money on computer software.



C Listen to the dialogs one more time and write the phrases that each person uses when inviting, offering or suggesting.

- 1 _____
- 2 _____
- 3 _____
- 4 _____
- 5 _____
- 6 _____



D Listen to these offers. Some offers sound sincere and genuine, and some sound insincere or rude. Put a check (✓) if the offer sounds sincere, and a cross (X) if it sounds insincere.

- 1 I'll take you to the theater if you like.

- 2 Would you like a drink?

- 3 Relax. I'll make dinner tonight.

- 4 Do you need me to pick you up after work tonight?

- 5 Here, I'll help you with the laundry.

- 6 Let me make the coffee.

- 7 Look, I'll drive if you want me to.

- 8 Do you want me to help you with your bags?

- 9 Let's go away for a few days next weekend.

- 10 Why don't you just leave everything to me.

E Choose appropriate replies from the Accepting/Refusing Offers box on page 31 for the offers in Exercise D. Decide whether to accept or refuse the offer depending on how sincere or insincere it sounds. Write the replies in the spaces below the offers.

How to say it

Intonation

The same word can sometimes have different meanings, depending on how it is said. For example, when responding to this question:

Why don't we go to the beach this weekend?

A person could show interest and agreement by replying like this: OK!

The same person could show disinterest by replying like this: OK.

The tone in the second example is much flatter than the one in the first example. Flat tones signal a lack of interest. Tones that rise at the end and which are said with greater emphasis show interest.

Listen to these dialogs and decide whether the second speaker is interested or disinterested. Check (✓) the correct box.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---------------|---|---|---|---|---|---|
| Interested | | | | | | |
| Disinterested | | | | | | |

Around the world

Polite behavior: Invitations, offers, and compliments

Look back at the *Listen to it (I)* dialog on page 28. Note that Katherine refuses Judith's invitation at first. Katherine is being polite. She wants to make sure that Judith is being sincere, and that it won't be too much trouble. Refusing invitations and offers a few times before accepting is a Western custom. Americans may accept the second time, but the British usually refuse three or four times before finally accepting! It is not because the British do not want to accept, but that they need to be absolutely sure that the offer or invitation is genuine.

The British also believe it is polite to reject compliments. Americans, on the other hand, give compliments more often, and accept them with thanks. Both the British and the Americans tend to avoid complimenting their bosses and teachers as they do not want to give the impression that they are 'sucking up' to their superiors.

Discuss the answers to these questions.

- 1 In your culture, what are the rules of polite behavior when refusing invitations and offers?
- 2 How do you feel about giving and receiving compliments?
- 3 Would you compliment your teacher or boss?



F.2 Test Format of the Mid-term Examination

Mid-term Examination

I. Sentence Completion

Match the half-sentences in Column A with those in Column B. Write the letter of the half-sentence in Column B in the spaces provided. Follow the example.

| Column A | | Column B |
|----------------------------------|-------|--|
| 1. I think I met you | [e] | a) most weekends. |
| 2. Please join me for lunch | [] | b) as someone he went to college with. |
| 3. I've studied here | [] | c) together. |
| 4. A teacher helps | [] | d) we've met before. |
| 5. I hope to see you | [] | e) as John's party. |
| 6. I play soccer | [] | f) to get there? |
| 7. Sue and I grew up | [] | g) for two years. |
| 8. Tom works in the same company | [] | h) again very soon. |
| 9. I don't believe | [] | i) at the new restaurant. |
| 10. Does it take long | [] | j) students to learn |

II. Cloze Passage

Complete the passage with the half-sentences. Write the letters in the spaces provided. Follow the example.

Making Small Talk

After you have been introduced to someone, (1) c. When you meet someone for the first time, (2) _____. It may be even more difficult (3) _____ and it's not your first language. You may feel shy because you don't know the other person well. Learning to make small talk can be different but (4) _____, you'll find it easier. Try to find out about the person. What is their job and (5) _____? Maybe you will find that you both have the same interests and (6) _____. Don't ask too many questions though; (7) _____. Try to avoid (8) _____ when you meet someone for the first time as these can be difficult topics of conversation.

- a) you should answer some yourself
- b) talking about religion or politics
- c) you will probably have a chat
- d) then you will have something to talk about
- e) what do they like to do in their free time
- f) if you are trying to speak English
- g) it can be difficult to think of subjects to talk about
- h) if you follow some simple rules

III. Multiple Choice Cloze

Fill in the blanks with the most appropriate word. Follow the example.

Having a part-time job (1) c you are studying has advantages and disadvantages. (2) _____ times you will feel glad that you have it and other times you will wish you had (3) _____ started it.

The advantages are that you (4) _____ money, gain working experience, learn new skills and get (5) _____ know new people. One major disadvantage is that your studies may suffer. Some people don't find the pressure (6) _____ bad, but others find that they just don't have the time (7) _____ work and school. They are tired (8) _____ working and can't concentrate on their studies.

Everyone (9) _____ different and you have to choose which is more important: getting qualifications or gaining work experience. Of course, you may (10) _____ it no problem to do both.

1. a) before
b) after
c) while
d) since

5. a) for
b) ahead
c) to
d) too

9. a) has
b) have
c) is
d) are

2. a) other
b) at
c) any
d) three

6. a) too
b) not
c) rather
d) terrible

10. a) see
b) have
c) seem
d) find

3. a) already
b) never
c) ever
d) sometimes

7. a) for
b) while
c) when
d) before

4. a) have
b) earn
c) save
d) spend

8. a) of
b) because
c) from
d) about

IV. Matching

Match the questions in Column A with the responses in Column B. Write the letter of the response in Column B in the spaces provided. Follow the example.

Column A

Column B

- | | | |
|--|-------|---|
| 1. Where did you study? | [c] | a) Yes, he spent his childhood there. |
| 2. You moved here in July, didn't you? | [] | b) It's hard work. |
| 3. Why do you want this job? | [] | c) In Japan, for three years. |
| 4. Was John brought up in Hong Kong? | [] | d) Just two months ago. |
| 5. He's kind of shy, isn't he? | [] | e) Yes, I've done some community service. |
| 6. How do you find your new job? | [] | f) No, never. |
| 7. Is Sally going to come back when she graduates? | [] | g) To get more experience. |
| 8. When did you move? | [] | h) Yes, the end of July. |
| 9. Have you moved a lot? | [] | i) No, she likes it there. |
| 10. Have you ever been a volunteer? | [] | j) Do you think so? |

V. Cloze Dialog

Complete the dialog with the sentences. Write the letter in the spaces provided. Follow the example.

Robert: Hi, Henry. What have you been doing while I've been at work?

Henry: Catching up on some sleep. I didn't sleep on the flight.

Robert: (1) _____, you've got to see the city! How about if we start with the Grand Palace and then take a walk around the old town?

Henry: (2) _____. I don't want to take up your time.

Robert: Don't be silly. I've taken three days off work so (3) _____.

Henry: Excellent. OK, well I'm paying for dinner. Do you know any good places to eat?

Robert: I certainly do. I hope you like spicy food. (4) _____ in Thailand. I thought we could eat somewhere near Chinatown. There are some great streets and (5) _____. You'll love it.

Henry: Alright, before we go I'll need to change some money.

Robert: No problem. I'll lend you some until we get to a bank. If there's anything you need while you're here just ask.

Henry: (6) _____. I hope I can do the same (7) _____.

Robert: (8) _____.

- a) we can see the sights
- b) there's no time for sleep now
- c) that's very kind of you

- d) it's a specialty
- e) Are you sure?
- f) great architecture
- g) when you visit Taiwan
- h) it's no trouble

VI. Multiple Choice Cloze

Fill in the blanks with the most appropriate word. Follow the example.

It's the weekend again and the weather is going to (1) c fine. The Campbell family (2) _____ ever spend any time together because they are always so busy. Mr. Campbell has the family business to look (3) _____. The children spend all weekend on the computer and Mrs Campbell (4) _____ the weekend cleaning and cooking for her family. Nothing is too much (5) _____, (6) _____ this weekend is different. Mr Campbell has suggested that they take some time to do something (7) _____, as a family. The only problem is choosing something they all want to do. The children want to go to the computer store to try (8) _____ some new software. Mr Campbell wants to (9) _____ window-shopping. He says he never has time when he is at work. And Mrs Campbell wants to (10) _____ something relaxing like going out to the countryside and (11) _____ a picnic. After much discussion, Mrs Campbell finally (12) _____ a decision. "Why don't we (13) _____ it all? And then everyone will be happy. Let's make a day (14) _____ it. We'll get up early, do some shopping and then have a picnic." "That (15) _____ great!" they all shouted.

- | | | | |
|---------------|-------------------|--------------|--------------|
| 1. a) is | 5. a) troublesome | 9. a) have | 13. a) have |
| b) look | b) time | b) go | b) see |
| c) be | c) trouble | c) see | c) do |
| d) make | d) good | d) buy | d) make |
| 2. a) usually | 6. a) and | 10. a) go | 14. a) about |
| b) sometimes | b) so | b) see | b) of |
| c) hardly | c) but | c) do | c) off |
| d) always | d) also | d) make | d) in |
| 3. a) after | 7. a) alone | 11. a) do | 15. a) seems |
| b) before | b) together | b) have | b) was |
| c) in | c) same | c) having | c) sounded |
| d) at | d) usual | d) doing | d) sounds |
| 4. a) does | 8. a) in | 12. a) makes | |
| b) spends | b) on | b) says | |
| c) makes | c) up | c) has | |
| d) has | d) out | d) speaks | |

VII. Rewriting

Below are two passages. Read Passage one and then fill in the blanks to make Passage two the opposite. Follow the example.

Passage one

This electronic dictionary may be cheaper than the rest, but it has all the features we know you want. It is larger than most others but weighs the same and has the same functions. The battery lasts 100 hours, which is a lot longer than most and the dictionary comes in four bright colors: red, yellow, green, and blue. We believe that because it isn't outside your budget and you are getting a quality product, there is no better dictionary to recommend.

Passage two

This electronic dictionary is more (1) expensive than most others and it has (2) _____ of the features people want. It is (3) _____ but weighs (4) _____ than other dictionaries. The battery lasts 50 hours, which is a lot (5) _____ than most. It is only available in two (6) _____ colors: brown and black. Whether this product is (7) _____ your budget or not, there couldn't be a (8) _____ one to recommend.


VIII. Matching

Match the sentence in Column A with the correct response in Column B. Write the letter of the response in Column B in the spaces provided. Follow the example.

| Column A | | Column B |
|--|-------|----------------------------------|
| 1. It's just your style. | [h] | a) Sorry, cash only. |
| 2. Think it over. | [] | b) No, it's too big. |
| 3. Do you think these go together?
one. | [] | c) But it's not as nice as that |
| 4. What do you think of these? | [] | d) Yes, 50 per cent off. |
| 5. I hate shopping on Saturdays.
soon. | [] | e) OK, and I'll let you know |
| 6. I prefer this one. | [] | f) Yes, I need something faster. |
| 7. This printer is quite slow. | [] | g) They're fine. |
| 8. Do you take credit cards? | [] | h) Do you think so? |
| 9. Does this shirt suit me? | [] | i) Yes, it's always so busy. |
| 10. Is there a sale? | [] | j) No, they don't match. |

TEST FOR CHAPTERS 1-5

Listening Comprehension Test

 **A** In this part of the test, you will hear 15 questions or statements. After you hear a question or a statement, read the four possible responses and decide which one is the best response to the question or statement you have heard. Circle the answer. Follow the example.

- | | |
|--|--|
| 1 a) No, of course not.
b) <input checked="" type="radio"/> Sure, no problem.
c) Thank you.
d) Don't mention it! | 9 a) Hi, long time, no see.
b) Pleased to meet you.
c) Why are you here?
d) How have you been? |
| 2 a) No, I'm pleased.
b) I know, it's been ages.
c) How are you?
d) I didn't want to see you. | 10 a) I'll have them both.
b) I do like your car.
c) Which one do you recommend?
d) Do they suit me? |
| 3 a) Yes, I would.
b) No, I don't want to.
c) No, of course not.
d) Here's my coat. | 11 a) Yes, just on Mondays.
b) Yes, rarely.
c) No, but I have some work to finish.
d) Yes, every weekend. |
| 4 a) It was so romantic.
b) I took a flight.
c) I bought a map.
d) I think you would like it. | 12 a) Thank you.
b) You should apologize.
c) It's not your fault.
d) Oh, no, how annoying! |
| 5 a) No, I don't like them.
b) I don't want to go.
c) Can I come with you?
d) If you're sure it's not too much trouble. | 13 a) Because you're boring.
b) That would be great, thanks.
c) Because you're more clever than me.
d) That would be difficult. |
| 6 a) No, I never buy drinks.
b) Sure, I'll look after it.
c) No, thanks. I'm thirsty.
d) No, thanks. I have to be going. | 14 a) Yes, it's near the window.
b) Yes, there's more legroom.
c) Yes, it's near the toilets.
d) Yes, you can see the video screen. |
| 7 a) No, it's next week.
b) I'm sure you'll pass.
c) I do, too. You've worked really hard.
d) I do, too. He's worked really hard. | 15 a) Since 1985.
b) To Ontario.
c) When I was 10 years old.
d) From 1985. |
| 8 a) No, they suit you.
b) Yes, they look fine.
c) Yes, they suit you.
d) They are very tall. | |

X. Bonus: ICRT

One short-answer question: Who are the people that the Taiwanese government would like to help?

Answer: Taiwanese aboriginals.

F.3 Test Format of the Final Examination

Intermediate English Listening and Speaking Practice – Final Examination

Ss ID: _____ Ss Name: _____

Part A: Dictation – Write down the sentences you have heard (40%) _____

1. _____
2. Aren't you annoyed when people don't return things they've borrowed?
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____

Part B: Song dictation (10%) View the questions on the back of the page.

Part C: ICRT News (20%) _____

1. The first news is about: a. entertainment b. weather c. political issues
2. According to the news _____% of the tobacco tax will be devoted to the health insurance program.
3. The US Supreme Court agreed people to plant how many marijuana plants at home to release their illness symptoms? a. 10 b. 12 c. 20
4. The fourth news is from where? a. New York b. Boston c. Washington
5. What's the weather like in Germany? a. heavy snowy b. light snow c. heavy storm
6. The news is about: a. weather b. recycling c. law makers
7. According to the weather report, the weather tomorrow is getting a. hot b. cold c. warm
8. Is there a possibility of snowing in the high mountain in Taiwan? a. Yes b. No
9. The highest degree in Taiwan is around _____ degree.
10. According to the weather report, how many degree in Kaohsiung? _____

Part D: Reading (30%)

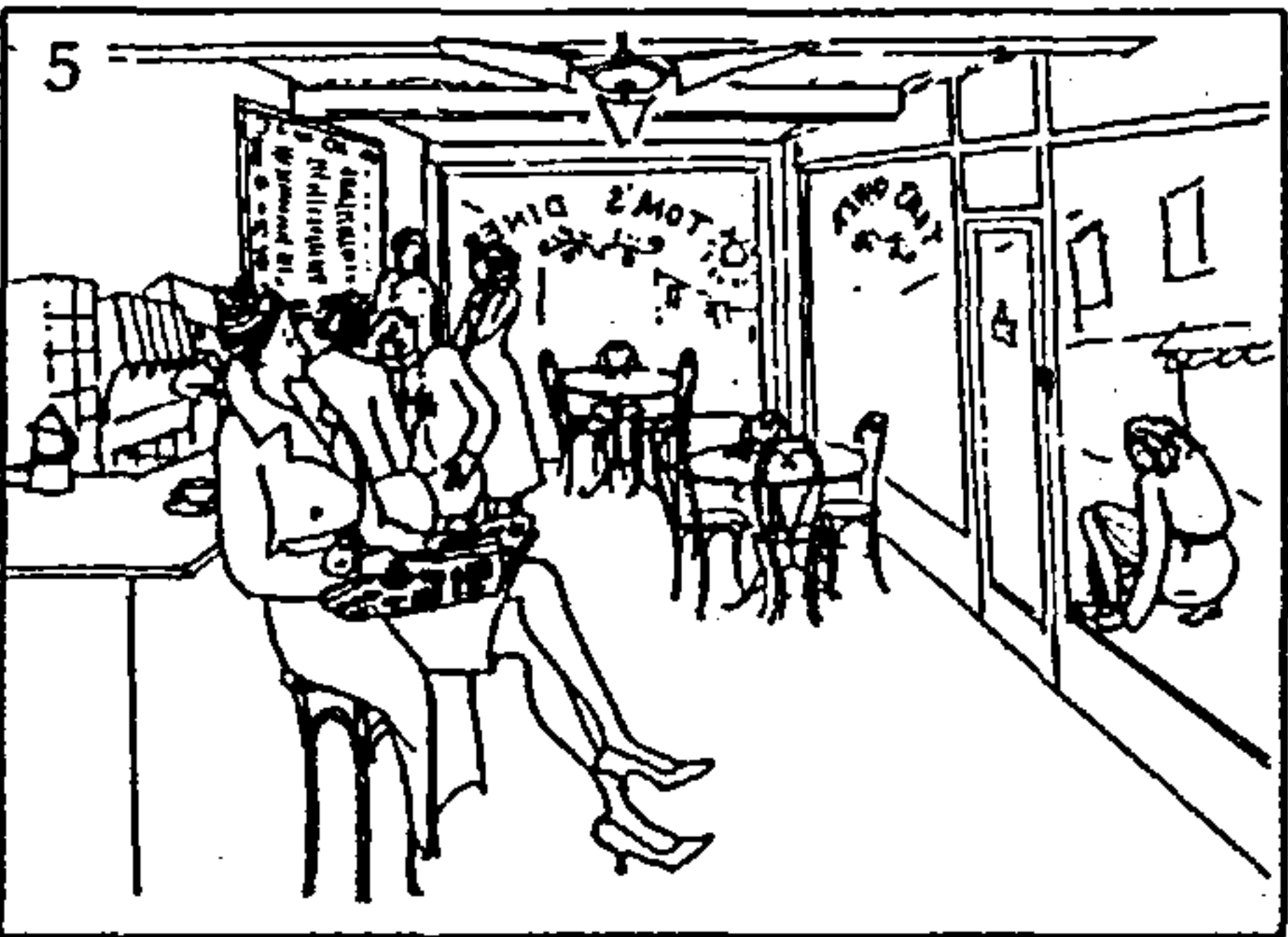
Pronunciation _____ Fluency _____ Intonation _____

Tom's Diner

(by Suzanne Vega)

6

1 Circle the differences.



© Reed Educational and Professional Publishing Ltd 1996.
Published by Heinemann English Language Teaching.

Photocopiable 17

Lyrics of "Tom's Diner" by SUZANNE VEGA

I am sitting in the morning at the diner on the corner
I am waiting at the counter for the man to pour the coffee and he fills it only halfway and
before I even argue
He is looking out the window at somebody coming in
"It is always nice to see you" says the man behind the counter
To the woman who has come in
She is shaking her umbrella
And I look the other way as they are kissing their hellos
I'm pretending not to see them instead I pour the milk
I open up the paper
There's a story of an actor who had died while he was drinking
It was no one I had heard of
And I'm turning to the horoscope and looking for the funnies
When I'm feeling someone watching me and so I raise my head
There's a woman on the outside looking inside
Does she see me?
No she does not really see me cause she sees her own reflection
And I'm trying not to notice that she's hitching up her skirt and while she's straightening her
stockings her hair is getting wet
Oh, this rain
It will continue through the morning as I'm listening to the bells of the cathedral I am thinking
of your voice...
And of the midnight picnic once upon a time before the rain began...
I finish up my coffee it's time to catch the train

F.4 Problems Students Found in the Mid-term and Final Exam

1. The quality of the recording and/or the visual equipment

| No. | Characteristics of the Quality of the Recording | Mid-term Exam | | Final Exam | |
|-----|--|---------------|-------|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | The quality was so poor that I couldn't hear the text clearly. | 5 | 21.7% | 11 | 47.8% |
| b. | The background noise outside the testing environment was too loud. | 13 | 56.5% | 6 | 26.1% |
| c. | The volume of the tape recorder or audio-visual equipment was too low. | 5 | 21.7% | 6 | 26.1% |
| d. | The quality of the recording was good, and I could hear the texts clearly. | 7 | 30.4% | 6 | 26.1% |
| e. | Other: 視聽設備嚴重不足。The audio-visual equipment was seriously inadequate. | 1 | 4.4% | 0 | 0% |

2. Testing time

| No. | Characteristics of Testing Time | Mid-term Exam | | Final Exam | |
|-----|--|---------------|-------|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | Time was too limited to answer all the questions properly. | 11 | 47.8% | 4 | 17.4% |
| b. | The testing time was sufficient for me to answer all the questions properly. | 12 | 52.2% | 19 | 82.6% |
| c. | Other: | 0 | 0% | 0 | 0% |

3. The test/task instructions

| No. | Characteristics of Test/Task Instructions | Mid-term Exam | | Final Exam | |
|-----|---|---------------|-------|------------|------|
| | | Freq. | % | Freq. | % |
| a. | The instructions of each test section were not clear. | 5 | 21.7% | 2 | 9% |
| b. | The instructions were too complicated. | 1 | 4.4% | 1 | 4.4% |
| c. | The test/task instructions were clear. | 17 | 73.9% | 20 | 87% |
| d. | Other: | 0 | 0% | 0 | 0% |

4. The length of the listening texts

| No. | Characteristics of the Length of the Texts | Mid-term Exam | | Final Exam | |
|-----|--|---------------|-------|------------|-------|
| | | Freq. | % | Freq. | % |
| a. | The texts in the test were longer than I was used to listening to in class. | 3 | 13% | 6 | 26.1% |
| b. | Longer texts increased the difficulty of understanding. | 3 | 13% | 5 | 21.7% |
| c. | I got lost in listening to the longer texts. | 2 | 9% | 6 | 26.1% |
| d. | The lengths of the texts in the test were similar to those I listened to in class. | 17 | 73.9% | 11 | 47.8% |
| e. | Other: | 0 | 0% | 0 | 0% |

F.5 Mid-term Interview Transcription

| |
|--|
| <p>Question 1. Where did the content of the test come from?</p> <p>Teacher (C): 從教師手冊印來的。</p> <p>Interviewer (I): 您這次出題除了聽力方面的題目，您還出了閱讀填空的題目，那些也是從教師手冊來的？</p> <p>C: 閱讀填空是在教師手冊裡的，全部都是課本上的內容，但是在習題裡面的。本來是要用來給學生當練習題的，但上課時間不夠用，我就拿來當作期中考的題目，想說驗收一下他們上課所學過的。聽力部份的話是在課本裡面。</p> <p>I: 那他們有可能在考前接觸到這些考題嗎？</p> <p>C: 聽力的部份可以，因為在課本裡，但是他們只能看到題目跟選項，沒有辦法聽到問題。</p> <p>I: 那他們知道您會從課本裡出聽力嗎？</p> <p>C: 不知道，所以他們去注意到的機率也不高。</p> <p>I: 請問那個額外加分的聽力內容是從哪裡來的？</p> <p>C: 從 ICRT 的新聞中擷取出來的。</p> <p>I: 我發現你期中考沒考口語考試。可以請你跟我說為什麼沒有考嗎？</p> <p>C: 嗯...我已經給他們太多閱讀考題，這佔掉太多考試時間，所以沒時間再考口語考試了。</p> <p>I: 在測驗學生前，你有測試這些考題嗎？</p> <p>C: 沒有，我還要準備其他科的考題，沒有時間跑測試考題的事情。</p> |
| <p>Teacher (C): It was taken from the teacher's manual.</p> <p>Interviewer (I): In addition to the listening questions, you also tested students with reading comprehension questions. Were the reading questions also from the teacher's manual?</p> <p>C: Yes, the reading part was taken from the teacher's manual, it was related to the teaching contents in the textbook, but it was in the "Practice" part. I was going to use the reading questions as in-class exercise activities for students, but the class schedule was so tight that I did not have time to use it, that's why I tested them with those questions in the mid-term exam. The listening part was from the textbook.</p> <p>I: Was it possible that they got access to the listening questions before the mid-term exam?</p> <p>C: Yes, they could have done, because the test items were in the textbook, but they could only see the test items without listening to the questions.</p> <p>I: Did they know that you used the questions from the textbook?</p> <p>C: No, they did not, so the probability that they noticed the fact was very low.</p> <p>I: Where did the extra bonus (see Section 7.4) news report come from?</p> <p>C: It was a news excerpt from ICRT (International Community Radio Taipei) radio.</p> <p>I: I found that you did not use any speaking test in the mid-term exam. Could you tell me why you did not test the students' speaking ability?</p> <p>C: uh...I had already given them too many reading questions, which taken up too much testing time, so there was no more time for a speaking test.</p> <p>I: Did you pilot the test items, I mean run a small test, before you tested your students in this exam?</p> <p>C: No, I had to prepare exams for other courses I taught. I did have time to run a pilot.</p> |
| <p>Question 2. What were the cut-off scores for the mid-term and final examination tests? What percentage did each of the two tests count for in the total final score of the course?</p> <p>C: 60 分及格。期中佔 30%，期末佔 30%。</p> <p>I: 有小考嗎？</p> <p>C: 沒有小考，學生每個星期都要做 presentation，算課堂成績。每堂課結束前分配三組學生，每組兩人，下週上課時隨機抽樣一組上去練習課本裡的句型。</p> <p>I: 學生在練習句型時是看著書唸的嗎？</p> <p>C: 可以看著書。</p> <p>C: 60. The mid-term and the final scores all accounted for 30% of the total score.</p> <p>I: Were there any quizzes?</p> <p>C: No, the students needed to do a presentation, which accounted for their in-class participation. I arranged three pairs of students to practice the dialogues in the textbook before</p> |

the end of every lesson. And I chose a pair randomly to present the dialogue they practiced.

I: Could they read the book while presenting?

C: Yes they could.

Question 3. What did you expect the students to have learned from your class?

C: 我希望他們在學聽力的同時也學到口語方面的技巧，因為聽力跟口語是相關的。學生除了教科書裡基本的聽力內容要懂，我也希望他們聽的懂 ICRT radio 的英文。

I: 請問您上課會讓學生聽有關 ICRT 的英聽文本嗎？

C: 不會，上課主要是聽教科書的，不過我有要求他們課後自己要去聽。

I: 您考試會考 ICRT 的相關聽力內容嗎？

C: 會考，不過大概就一兩題而已。

C: I hope that my students learned speaking skills while learning listening, because listening and speaking are connected. In addition to understanding the listening contents in the textbook, I also expect that they can understand the English on ICRT radio.

I: Did you give any ICRT listening materials to the students in class?

C: No, I focused on the textbook in class, but I asked them to listen to ICRT after the class.

I: So, you would include ICRT listening passages in the mid-term or final exams.

C: Yes, but only one or two questions.

Question 4. What were the mid-term test objectives you would like to achieve in relation to your teaching plan and the students' learning outcomes? Do you think you have achieved them?

C: 考試的目的有 80%是要知道他們到底懂多少，另外有 20%是要驗收他們的成果。

I: 您的意思是這次的期中考有兩個目標嗎？

C: 是的。

I: 那請問您如何來決定 80%是來測驗他們理解課本多少，20%是來測驗他們的英文能力？

C: (沉默)我可以從他們的成績得知啊，如果他們成績高一點就代表他們懂得多一些、英文能力也高一些，反之亦然。

I: 所以你相信學生的成績可以告訴你他們理解多少教材內容與他們的英文程度？

C: 是啊，可是學生 ICRT 聽懂的只有 5 成，所以不算達到我的目標。

C: 80% of the test objective was to know how much they understand in terms of the in-class teaching material, the other 20% was to establish their level of English proficiency.

I: Do you mean that your mid-term exam had two different purposes?

C: Yes.

I: How did you decide that 80% was to assess the understanding of the textbook and the other 20% was for their level of English proficiency in terms of the test questions?

C:(silence).....well...I can tell from their marks. If they scored higher, it meant they understood more and their English ability was better, and vice versa.

I: So, you believe that the students' marks told you everything about their understanding of the teaching material and their English level.

C: Yes, but they simply understood 50% of ICRT news report, so I don't think I achieved the test objective completely.

Question 5. How did you identify and decide on the difficulty of the content/items in the mid-term test? What were the criteria?

C: 我都從課本裡出，因為學生上課都聽的懂，所以考課本他們也應該懂，畢竟難度相同。

I: 您選擇教材前有先瞭解他們的英文程度嗎？

C: 一開始上課的時候有給學生做簡短的自我介紹去了解他們的英文程度才選擇教材，我目前用的這本教材比他們的程度來的簡單，因為聽力就是要讓他們聽的懂。

I: 那您覺得他們這次考的如何？

C: 這次考試成績偏低，學生雖然了解課堂的內容，但換一種方式問就不行了，就是閱讀那部份，這次考試學生約了解 7 成。他們聽 ICRT 了解只有 5 成，我雖然要求他們去聽，但成果不好。

C: The test questions were from the textbook. Because they understood the teaching material in class, they should have understood the test questions because they were from the textbook.

I: Did you know the students' level of English before you chose the teaching material?

C: At the beginning of the course, I asked the students to introduce themselves in English in order to find out their level of English and then I chose the teaching material. The textbook I am using right now was easy for them, because if you want to teach listening, you need to choose materials they can understand.

I: What did you think about the students' performance in the mid-term exam?

C: Their marks were lower than I expected; although the students understood the listening contents in class, they did not perform well if I changed to a different type of question, I mean...the reading comprehension questions. They understood approximately 70% of the test content. They could only understand the ICRT news report partially – approximately 50%. Though I asked them to listen, the outcome was not satisfactory for me.

Question 6. Why did you choose the particular types of comprehension question to test the students? Did you think that these test methods might favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

C: 沒有刻意去選，我就只用了教師手冊跟課本裡的考題，就聽力部份而言就是選擇題，因為他們只能聽一次，聽完就要在時間內看四個選項回答出正確答案。

I: 為什麼不用其他類型的題目？

C: 因為考試時間不夠。

C: I did not choose the questions in particular. I simply used the test items in the Teacher's Manual and the textbook. For the listening part, I used multiple-choice items. They were allowed to listen to the conversations once, and they had to choose the best answer immediately after listening to the questions.

I: Why did not you use other types of question?

C: Because there was not enough testing time.

Question 7. Will the result of the mid-term examination impact on the teaching of the following second half term (i.e. will the perceived difficulty in the contents of teaching materials be increased or decreased?), or influence how you design the final examination test (i.e. will the test be designed to be more difficult or easier?)

C: 期中後的上課內容會更難，因為他們這次考的偏低，就更需要訓練他們。

I: 期末考題呢？

C: 也會更難，ICRT的比重會更高。

I: 您不擔心考題變難學生不能 pass 嗎？

C: 不擔心，不能 pass 就重修吧！

C: I think I will increase the difficulty of the teaching content. Because they scored lower this time, they needed to be trained.

I: How about the final exam?

C: Perhaps it will be harder; the number of ICRT news questions will be increased.

I: Aren't you worried that the students might not be able to pass the course due to the harder test items?

C: I am not worried about it. If they can't pass, I am afraid they will just have to re-take the course.

F.6 Final Interview Transcription

Question 1. Where did the content of the test come from?

Dr. C (C): 聽寫部份是從 Teacher's Manual 上面來的，歌曲是從一本英文教材裡來的，ICRT 的新聞是我從 radio 上面錄下來的。

Interviewer (I): 那口語考試的內容呢？

C: 就是讓他們各自唸一段課文對話，他們平常上課練習的。

I: 大家唸的對話都一樣嗎？

C: 看我翻到哪一課他們就唸哪一課，但基本上會重複到。

I: 您是從哪些方面去評分？

C: 我考卷上有寫，從 pronunciation, fluency, 還有 intonation。

I: 您有用任何的評分量表來敘述學生的這三方面的程度嗎？比如說 5 分代表他們的發音很標準沒有錯誤、4 分代表有些許的錯誤但大致上表達都正確....之類的量表嗎？

C: 我沒有這種表格....用這個評分會很花時間。

I: 所以您都怎麼評分？

C: 就發音越正確、講的越流暢的給分越高。

I: 你期末考前有測試考題嗎？

C: 沒有，就像我期中說過的，我需要準備其他科的考題。

Dr. C (C): The "Dictation" was taken from the Teacher's Manual, the English song was taken from an English teaching textbook, and the news reports were recorded from the ICRT radio by me.

Interviewer (I): How about the oral test?

C: I asked them to read a dialogue from the textbook which they had practiced in class.

I: Did every student read the same dialogue?

C: They read the dialogue I chose randomly from Chapter Six to Ten in the textbook.

I: What dimensions did you look at while assessing their speaking ability?

C: I put it in the final exam answer sheet (see Part D, Appendix F.3). I looked at three dimensions – pronunciation, fluency, and intonation.

I: Did you use any rating scales with specific descriptions to mark the students' proficiency in the three dimensions? For example, point 5 meant that their pronunciation was correct, point 4 assumed that they made a few mistakes on pronunciation but they were reasonably correct in general.

C: I did not use a rating scale with specific descriptions....because it would have taken a lot of time to score it.

I: How did you score their speaking ability?

C: The more accurate their pronunciation was and the more fluently they spoke, the higher the scores they could get.

I: Did you pilot the test items before the final exam?

C: No. As I have told you for in the mid-term exam, I needed to prepare exams for other courses I taught.

Question 2. What were the final test objectives you would like to achieve in relation to your teaching plan and students' learning outcomes? Did you think you have achieved them?

C: 這次的期末考是一個整合的考試，要考他們這一學期所學到的。

I: 期中考前的課程內容有包含在期末考裡面嗎？

C: 這個倒沒有，但是像考歌曲，他們一整個學期都在做 presentation，對聽英文歌也應該很熟悉....ICRT 的新聞也一樣，我都有叫他們回家自己聽 radio，也聽了一學期。

I: 那您覺得他們這次的考試表現如何？

C: 我覺得還好耶，我以為他們會在新聞考題那部份會考得很糟，但可能我出題並沒有很難，蠻多學生都答對了，但還是有一些學生沒有考好。不過他們口語部份都考的不錯啦，畢竟上課練習了很多次，也讓他們做 presentation，這部份拿高分的人很多。

I: 所以妳對他們的口語考試還算滿意？

C: 對啊！畢竟這些 dialogues 都不難理解，他們也練習很多次，沒有理由考不好吧！

C: The final exam was an integrated test which tested what they had learned this term.

I: Were the teaching contents before the mid-term exam included in the final exam?

C: No. Take the English song in the final exam as an example, the students had presented

different songs in class for the whole term, so they should have been familiar with listening to English songs. Similarly, I asked them to listen to the ICRT radio after class, and they should have listened for a term.

I: What did you think about their performance in the final exam?

C: I found it OK. I thought they would perform poorly in the ICRT news part, but many students answered the questions correctly – maybe I did not design too difficult items. I think they performed well in the speaking part. After all, they had practiced so many times in class and in presentations; many students scored highly in this part.

I: So you mean you are satisfied with their oral test scores?

C: Yes, those dialogues were not difficult for them to understand, and they had practiced them so many times, there should not be any reasons why they would not perform well!

Question 3. How did you identify and decide on the difficulty of the content/items in the mid-term test? What were the criteria?

C: 由於聽寫部份是從 Teacher's Manual 出的，裡面的題目跟課本程度差不多，而課本裡的東西大部分的學生都能理解，所以覺得出跟課本相似程度的他們應該可以答對。不過我這次有出比較多的新聞題，我知道對他們而言會比較有挑戰性，所以我考題有出簡單一點，讓他們挫折感不會那麼大。

C: Since the "Dictation" part was from the teacher's manual, the difficulty of the questions was similar to that in the textbook, and the majority of the students could understand the content; I think they could answer this part correctly. But this time I tested them with more ICRT news questions, I knew that it was challenging for them, so I designed easier test items so that they would not feel frustrated.

Question 4. Why did you choose the particular types of comprehension question to test the students? Did you consider that these test methods favour particular types of student, or did you believe the students perform better on these types of comprehension questions?

C: 考聽寫是因為我之前沒考過這樣的題目，想說換一種題型，不要總是用同一種。

I: 那你不擔心學生會不熟悉這種考題然後考不好嗎？

C: 聽寫主要是聽到什麼寫什麼，這種基本的事情他們應該不會覺得陌生吧！

I: 那新聞考題呢？

C: 就像我剛剛說的，讓他們聽新聞已經很挑戰了，如果考題又出的很難他們可能會考得很差，所以我出選擇題，這樣他們答對的機率會比較高。

C: The reason I tested them with dictation was that I have never used this type of test item, and I wanted to change to another type so that they would not be tested by the same types of questions all the time.

I: Weren't you worried that they might lose marks because they might not be familiar with dictation?

C: Dictation is the type of question where they write down what they listen to. This is a very easy type of question which I thought they would be familiar with.

I: How about the test question for the ICRT news reports?

C: Well...as I said before, it could be challenging to test their listening ability with news reports. If I designed too difficult test items, they might perform poorly. That's why I used multiple-choice questions, so the probability that they answered correctly would be higher.

Appendix G

G.1 In-class Textbook

Richards, J. C. (1998) *New Interchange: Student's Book 3*. Cambridge: Cambridge University Press.

Note: Textbook copyright requested and photocopy permitted.

2 Career moves

1 SNAPSHOT

Unique Jobs

| | | | | |
|--|---|--|---|---|
|
personal shopper
Does people's shopping for them |
gossip columnist
Writes about famous people's lives |
chocolate taster
Eats candy and gives opinions |
menu writer
Chooses the right words to describe a restaurant's food |
toy tester
Decides if new toys are fun and safe |
|--|---|--|---|---|

Source: *The New York Times*

Complete the task and talk about the questions.

Put the jobs in order: from the most interesting (1) to the least interesting (5).

Which job did you rank number 1? Why?

What are three jobs in your culture that might seem unusual to a person from another culture?

2 CONVERSATION Job fair

A Listen and practice.

- Tim: Wow! There are so many jobs to choose from! What do you think?
- Diane: Working in the media could be fun – there's TV, newspapers, the Internet. . . .
- Tim: Well, let's look. Hmm. How about this? You could be a TV news director.
- Diane: Are you kidding? Directing the news would be nerve-racking!
- Tim: Well, writing for a magazine must be exciting. How about that?
- Diane: No. I'm really more interested in working with computers. Hey, look. Designing interactive media. I'd like that!
- Tim: Designing interactive media? It sounds interesting, but what is it?

B Listen to the rest of the conversation. What does an interactive media designer do? Does it sound interesting to you? Why or why not?



3 GRAMMAR FOCUS

Gerund phrases

Gerund phrases as subjects

Working in the media could be fun.
 Directing the news would be nerve-racking.
 Designing interactive media seems challenging.
 Working with computers sounds interesting.

Gerund phrases as objects

I'd love working in the media.
 I would hate directing the news.
 I wouldn't like designing interactive media.
 I'm interested in working with computers.

A Would you like doing any of the jobs in column A? First, check (✓) the jobs you would like. Then write your opinion of each job by choosing information from columns A, B, and C.



- A**
- 1. doing medical research
 - 2. working as an archaeologist
 - 3. writing for a newspaper
 - 4. teaching physically challenged children
 - 5. working on a movie set
 - 6. being a politician
 - 7. conducting an orchestra
 - 8. being wealthy and not having to work

- B**
- seems
 - sounds
 - must be
 - could be
 - would be

- C**
- pretty difficult
 - fascinating
 - nerve-racking
 - kind of boring
 - fantastic
 - pretty awful
 - really rewarding
 - very challenging

1. Doing medical research would be really rewarding.

B Pair work Give your opinions about the jobs in part A.

A: For me, **doing medical research would be really rewarding** because it would help save people's lives.

B: I agree! **I'd like doing medical research**, too. It would be very challenging.

B: Really? **I wouldn't like doing medical research.** I think it sounds pretty difficult.

useful expressions

For me, . . .
 As far as I'm concerned, . . .
 In my opinion, . . .

4 WORD POWER Collocation

A Find three phrases in the list that are usually paired with each verb. Then think of one more phrase for each verb.

- with animals a secret agent freelance writing
- a criminal lawyer as an artist an astronaut
- public speaking on a cruise ship interviews with famous people

| | |
|---------|--|
| being | |
| working | |
| doing | |

B Put the occupations in order: from the most interesting to the least interesting.

5 UNUSUAL CAREERS

Group work Describe three unusual careers you would like to have. Use information from Exercises 1–4 and your own ideas. Other students ask follow-up questions.

- A: I'd like doing interviews with famous people on TV.
- B: Why is that?
- A: Talking to people about their lives would be fascinating.
- C: Who would you interview?
- B: Anybody famous – politicians, movie stars, authors.



"I'm interested in interviewing Nelson Mandela."

6 WRITING What a job!

A Choose one of the jobs you talked about in Exercise 5. Make a list of the advantages and disadvantages of the job. Then write two paragraphs about the job. In the first paragraph, describe the advantages. In the second, describe the disadvantages.

| |
|---|
| <i>Working as a TV journalist would be a fascinating job. You would get to travel all over the world to cover important events.</i> |
| <i>In addition, you would meet many famous people, and</i> |
| <i>On the other hand, being a TV journalist could be difficult.</i> |
| <i>You could be in dangerous situations. For example,</i> |

useful expressions

In addition,


Further,

On the other hand,

For example,

B Pair work Take turns reading your papers. Then briefly summarize your partner's topic and ideas. Could you remember all the major points?


7 CONVERSATION Summer jobs

A  Listen and practice.

Tracy: Good news! I've found a summer job!
 Mark: That's great! Anything interesting?
 Tracy: Yes, working at an amusement park.
 Doesn't that sound fantastic?
 Mark: Sure, it does.
 Tracy: So, have *you* found anything?
 Mark: Nothing yet, but I've got a couple of leads.
 One is working as an intern for a record company – mostly answering phones.
 Or I can get a landscaping job again.
 Tracy: Being an intern sounds more interesting than landscaping. And it's probably not as hard!
 Mark: Yeah, but a landscaper earns more money than an intern. And you get a great tan!



CLASS AUDIO ONLY

B  Listen to the rest of the conversation.
 What is Tracy going to do at the amusement park?

8 GRAMMAR FOCUS

Comparisons

| | |
|--|--|
| A landscaper earns more than an intern. | An intern doesn't earn as much as a landscaper. |
| An intern has better hours than a landscaper. | A landscaper has worse hours than an intern. |
| A landscaper is better paid than an intern. | An intern is not as well paid as a landscaper. |
| Being an intern is more interesting than landscaping. | Landscaping is less interesting than being an intern. |
| Landscaping is harder than being an intern. | Being an intern is not as hard as landscaping. |

A Match the information to make sentences. Then compare with a partner.


- | | |
|--|---|
| <p>A</p> <ol style="list-style-type: none"> A counselor at a summer camp has worse hours Selling popcorn in a movie theater is not as rewarding A part-time tutor doesn't earn Working on a construction site is more dangerous A tour guide is not as well paid | <p>B</p> <ol style="list-style-type: none"> as a lifeguard. than working on a cruise ship. as working with the elderly. than a dog walker. as much as a housepainter. |
|--|---|

B Rewrite each sentence from part A in a different way.


1. *A dog walker has better hours than a counselor at a summer camp.*

C Add your own information to the clauses in column A of part A.
 Then compare with a partner.

9 PRONUNCIATION *Sentence stress*


A  Listen and practice. Notice the stress in these sentences.

Working at an **amúsement** park is **móre fún** than being a **báby**-sitter.
Báby-sitting is **nót** as **wéll páid** as **tútoring**.
 Being a **tútor** is just as **hárd** as working as a **cóunselor**.

B  Listen again to the sentences from the grammar box in Exercise 8. Mark the stressed words and then practice the sentences.

10 LISTENING

LASS
AUDIO
ONLY

A  Listen to Carlos, Paul, and Julia talking about their summer jobs. Where does each person work? Write the correct name under each picture.




1.

2.

3.

LASS
AUDIO
ONLY

B  Listen again. Do Carlos, Paul, and Julia like their jobs? Why or why not? Take notes.

interchange 2

The best and the worst

What kinds of summer or part-time jobs have you had? Turn to page IC-3.

11 PROS AND CONS

A Group work Choose two summer jobs from the list. Then use the questions to compare the jobs.

| Summer jobs | |
|--------------------------|----------------------|
| a baby-sitter | a chef's assistant |
| a dance instructor | a park ranger |
| a hiking trail guide | a dog walker |
| an assistant in a museum | a telephone operator |

Which job do you think pays more?
 Which one has better hours?
 Which one is more interesting? harder?
 more challenging? more rewarding? Why?
 What are the advantages and disadvantages of each job?

B Class activity Which job does your group prefer? Tell the class why.

1. INTRODUCTION

2. OBJECTIVES

3. SCOPE

4. METHODOLOGY

5. RESULTS AND DISCUSSION

6. CONCLUSION

7. REFERENCES

8. APPENDICES

9. BIBLIOGRAPHY

10. GLOSSARY

11. INDEX

12. LIST OF FIGURES

13. LIST OF TABLES

14. SUMMARY

15. EXECUTIVE SUMMARY

16. ACKNOWLEDGEMENTS

17. DECLARATION

18. CERTIFICATE

19. CURRICULUM VITAE

20. DECLARATION OF ORIGINALITY

21. DECLARATION OF NON-COPYRIGHT

22. DECLARATION OF NON-PLAGIARISM

23. DECLARATION OF NON-CONFIDENTIALITY

24. DECLARATION OF NON-CONFLICT OF INTEREST

25. DECLARATION OF NON-CONFLICT OF OBLIGATION

26. DECLARATION OF NON-CONFLICT OF LOYALTY

27. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY

28. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY

29. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY

30. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM

31. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT

32. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY

33. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST

34. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION

35. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY

36. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY

37. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY

38. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM

39. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT

40. DECLARATION OF NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY AND NON-CONFLICT OF INTEREST AND OBLIGATION AND LOYALTY AND CONFIDENTIALITY AND ORIGINALITY AND NON-PLAGIARISM AND NON-COPYRIGHT AND NON-CONFIDENTIALITY

G.2 Format of the Mid-term Examination

Mid-term Test – Units 1 - 4

Name: _____ Ss ID: _____

B. Circle the best word to complete these sentences.

1. Ted promised to meet me after school, but he didn't come. And he did the same thing last week. He's not very (moody/reliable/sociable).
2. My boss likes everyone to get along with each other. She doesn't like it when people are (patient/easygoing/upset).
3. My father likes being the school counsellor because he can help many children. It's a (rewarding/boring/dangerous) job.
4. Being a doctor is a (part-time/challenging/fascinating) job. You have to work long hours, face a lot of pressure, and make very difficult decisions.
5. Pat is a very (ambitious/generous/modest) person. She plans to have her own business and buy a house and a new car by the time she is 25.

C. Complete these sentences with your own information.

Example: It annoys me when people call me late at night.

1. I like it when friends _____
2. I hate it when someone _____
3. I can't stand neighbours who _____
4. I like a teacher who _____
5. I like people who _____
6. I don't mind it when friends _____

D. Write sentences that have the same meaning.

Example: A teacher earns more than a typist.

A typist doesn't earn as much as a teacher.

OR

A typist earns less than a teacher.

1. Being a tour guide is not as dangerous as being an astronaut.
Being an astronaut _____
2. A tutor earns less than a dance instructor.
A dance instructor _____
3. Working as a journalist is usually more stressful than working as an artist.
Working as an artist _____
4. A politician usually does more public speaking than an author.
An author _____
5. A baby-sitter usually has fewer new job leads than a dog walker.
A dog walker _____

F. Circle the incorrect word and write the correct word in the blank.

1. Maria made an apology on the teacher. _____
2. Chen received an impression to a party in the mail. _____
3. Nancy offered an accusation for her bad behaviour. _____

G. Make polite requests using the phrase given.

1. You want to use your friend's computer.
Could I _____
2. You want to borrow \$20.
Would you be able _____
3. You want someone to type a letter for you.
I wonder if you'd mind _____
4. You want to use a friend's phone.
Would it be OK _____

H. Check (✓) the correct phrase to complete each request.

1. Could you ask Dean when does the party start?
 when is the party start?
 when the party starts?
2. Would you ask the teacher what we should bring tomorrow?
 what should we bring tomorrow?
 if we should bring tomorrow?
3. Would you ask Simon please call me at five o'clock?
 to call me at five o'clock, please?
 please to call me at five o'clock?
4. Could you ask Diana does she have Rita's telephone number?
 that she has Rita's telephone number?
 if she has Rita's telephone number?

I. Read these conversations. Complete the questions.

1. A: When you worked at the bank, did you ever _____ to work on the freeway?
B: Yes, I used to drive to work on the freeway every day.
2. A: Did you happen to see Helen as you _____ work last night?
B: No, I didn't. By the time I left, she was already gone.
3. A: Did you _____ to lock the door when you went out to pick up the pizza?
B: No, I didn't. It was locked when I came back with the pizza.
4. A: Did you discover any surprises while you _____ the shipwreck?
B: Yes. Just look at this picture! I got it all on film.

J. Complete these sentences using one verb in the simple past and one verb in the past perfect.

Example: Just after I had come out (come out) of my house, I ran into (run into) an old friend from school.

1. Luckily, it wasn't until after the plane _____ (land) that the fire _____ (start).
2. After I _____ (finish) my swim in the ocean, I _____ (see) a shark in the water.
3. Julie and Brian _____ (decide) to get married after they _____ (take) a trip to Australia.
4. We _____ (turn down) the volume on the TV because we _____ (hear) a knock on the door.

G.3 Format of the Final Examination – Written Exam

Final Test – Units 5-8

Name: _____ Ss ID: _____

A. Listen to people talking. Check (✓) the correct answer.

1. Donna's grandfather and grandmother _____
 did not talk to each other until they were married.
 met only with a family member in the room.
 had to stay with relatives till they got married.
2. One of Tina's complaints to her landlord is that _____
 the refrigerator doesn't work. the air conditioner needs to be fixed.
 the floor is stained.
3. To stop depleting the ozone layer, Bob recommends _____
 cutting down the trees. reducing air pollution. using hair spray.
4. Sally is going to take _____
 a music appreciation course. a landscape photography course.
 an auto-repair course.

B. Circle the best word to complete each sentence.

1. It's a (culture/custom/reason) to take off your shoes before you enter someone's home in Japan.
2. Hotels are very (comfortable/curious/uncertain) here, and most have good service.
3. The crime rate is very high in this city, but this is a safe neighbourhood, so don't be (embarrassed/calm/nervous) walking around here.
4. Unfortunately, I wasn't dressed appropriately for the (church/hotel/hospital). I didn't have a hat, so I wasn't allowed in.
5. I hate flying. I always feel (anxious/calm/secure) and frightened.
6. When you visit a temple, keep in mind that (smoking/praying/marrying) is not allowed.

C. Answer these questions about the customs in your country. Write complete sentences.

1. What should you take with you when you are invited to someone's home?
When you _____
2. When you meet someone for the first time, what should you do?
When you _____
3. If you eat in a restaurant, what should you leave for a tip?
If you _____
4. What do you do if a friend gets engaged?
If a friend _____
5. What does your family usually do when a relative graduates from high school?
When a relative _____

D. Write a sentence describing a specific problem with each of these things.

Example: VCR I can't get a clear picture. OR The cord needs to be fixed.

1. air conditioner _____
2. telephone _____
3. television _____
4. oven _____
5. refrigerator _____

E. Complete these sentences using another form of the boldface word.

1. These jeans have a **tear** in the knee. They are _____.
2. These sunglasses are **scratched**. They have a _____.
3. Albert's best shirt is **stained**. His shirt has a _____.
4. There is a small **leak** in the plastic bag. The bag is _____.
5. The living room carpet has some **damage**. The carpet is _____.
6. Did you notice the **dent** in the lampshade? The lampshade is _____.
7. I can't afford such a **well-made** jacket. This jacket is very _____.

G. Write complete sentences about your preferences. Say why you prefer each thing.

Example: go to a public or a private school

I'd rather go to a public school than a private one because it's cheaper.

OR

I'd prefer going to a private school to going to a public one because it has better facilities.

1. study music appreciation or poetry

2. learn the grammar or the vocabulary of a new language

3. play the guitar or the violin

4. take an auto-repair class or an art class

5. date a competitive person or a person with good communication skills

6. perform in a choir or ride a motorcycle

7. spend money on new software or on a sport

8. be able to read faster or do math faster

H. Complete these sentences with *by (not) + gerund*. Use your own information.

Example: You can make new friends *by joining a club or taking a class.*

1. A good way to enjoy the weekend is _____

2. The best way to save money is _____

3. You can learn to dance better _____

I. Read the passage and circle T (true) or F (false).

Developing Good Study Habits

Good study habits can help make doing homework feel less stressful. Here are some ideas that other have found helpful for planning time, for managing space and study materials, and for learning better.

- Spend one to two hours of preparation and review for each hour you spend in class. Review class notes before and after class. Keep track of learning successes and problems. What kinds of problems did you have in your classes? How did you try to solve those problems? Do you see any new solutions or strategies that work well for you?
- Start working on major assignments as soon as they are given. Divide these large assignments into sections and complete a section each day. Don't wait to begin studying for a major exam until the night before the exam. Instead, study a little bit each day. Don't spend all your time studying, though! Take a ten-minute break after each hour of studying or when you change subjects. Make a schedule every week for each day's study activities, but be sure to leave time for rest and recreation.
- Make important ideas in your textbooks with a highlighter or make a light pencil mark in the margin. Write down all homework assignments, test dates, and assignment due dates on a calendar. Get ten or twenty file folders to keep in your study area. You can use a separate folder for each class, important personal papers, and even financial receipts. If you get folders in different colours, you can find your papers more easily.

If you can discover what techniques help, you can set goals, identify problems, and find appropriate solutions. The most important thing to remember is to use the techniques that work best for you.

1. T F The best time to study for a final exam is the night before the test.
2. T F You should include time for rest and recreation when you plan your week.
3. T F Never mark in your textbook.
4. T F Coloured folders can help you organise your materials.

G.4 Format of the Final Examination – Oral Exam

New Interchange III – Unit 5a

S1: I hear <the name of one of your classmates> is going to work in India.

S2: India! Wow! I hear it's a beautiful place, but I don't think I could ever live there.

S1: Why not?

S2: Well, it's too far from home. I'd miss my family.

S1: I don't think I'd mind moving to a foreign country. The language is the only thing that I'd be worried about.

S2: Yea, but wouldn't you miss your friends?

S1: Sure, for a while, but I'd make new ones.

S2: You certainly sound very confident.

S1: You know, actually, there is one thing I'd miss.

S2: What's that?

S1: My dog!

S2: You wouldn't have to, you know.

S1: What do you mean?

S2: Many moving companies can help families move their pets, too. It's also fairly easy to get cats and dogs on airplanes now.

S1: You mean I could buy an airline ticket for my dog? Isn't that kind of expensive?

S2: No. Of course, dogs do not fly as normal *passengers*! Most large airplanes have special areas that are completely safe for pets. People fly their pets all the time these days.

S1: Sure, I know this already. But there's still one thing I'd be concerned about.

S2: What's that?

S1: I think my dog prefers to fly business class!

New Interchange III – Unit 5b

S1: Guess what! I just got invited to my teacher's house for dinner.

S2: Oh, how nice.

S1: Yes, but what do you do when you're invited to someone's house here?

S2: Well, it's the custom to bring a small gift.

S1: Really? Like what?

S2: Oh, maybe some flowers or desert.

S1: And is it all right to bring a friend alone?

S2: Well, if you want to bring someone, you're expected to call first and ask if it's OK.

S1: I see. I don't want to be rude, especially to my teacher!

S2: Remember, your teacher is probably planning dinner for a certain number of people.

S1: So I need to make sure there will be plenty of food, right?

S2: It's not just the food. You have to think about space and other things, too. When is the dinner?

S1: Friday. Tomorrow night.

S2: That's not much time. Actually, it's customary to ask about bringing extra guests *when* you are invited, not several days later.

S1: What do you think I should do? I want to bring my new friend, because I'm afraid I will feel a little nervous attending a dinner party with strangers.

S2: I suggest that you go alone, as invited. Your teacher knows you are foreign student, so she is probably planning a very small dinner party to help you feel relaxed and welcome.

New Interchange III – Unit 6a

S1 (clerk): Can I help you?

S2 (customer): Yes, I'd like to return this jacket.

S1: Is there something the matter with it?

S2: Yes. I didn't notice when I bought it, but there are a few problems. First, it has a tear in the lining.

S1: Hmm. Actually, it's torn in several places.

S2: And some of the buttons are very loose. This one came off, in fact. And there's a stain on the collar.

S1: I'm really sorry about this. Would you like to exchange it for another one?

S2: Well, to be honest, I don't think this jacket is very well made. I'd rather get a refund.

S1: I understand. Do you have the receipt?

S2: Um, no I don't. The jacket was a birthday gift. But I do have the original store tags.

S1: I see. I'm afraid I can't give you a refund without the receipt. Do you know if the jacket was purchased with cash or credit card?

S2: I really don't know. As I said, the jacket was a gift.

S1: Well, there are two options. Since you still have the store tags, I can let you exchange the jacket for another one. But if you really want a refund, you will have to ask your friend to bring the receipt.

S2: I understand. Ok, I guess an exchange will be fine, then.

New Interchange III – Unit 6b

S1 (tenant 1): (*Knocks on building manager's door....*) Hi, I'm <S1>.

S2 (manager): Uh, <S1>....in Apartment 205?

S1: No, in Apartment 305.

S2: Oh, yes – that's right. What can I do for you? Does your refrigerator need fixing again?

S1: No, it's the *oven* this time. I think the temperature control needs to be checked. Everything I try to cook gets burned.

S2: Really? Ok, I'll have someone look at it right away.

S1: Thanks a lot.

S2: Uh, by the way, <S1>, are you sure it's the oven and not your cooking?

S1: That's funny – but, yes, I'm sure it's the oven.

S3 (tenant 2): (*Walks up as S1 leaves....*) Oh, I'm glad I caught you, <S2>. I'm <S3>, from Apartment 216.

S2: Yes. How can I help you, <S3>?

S3: I'm having a problem with the electricity in my apartment.

S2: What sort of problem, exactly?

S3: Well, I don't seem to *have* any electricity!

S2: Hmm, that's strange. Do you mean for the lights, or is it the appliances, too?

S3: Let me check. I'll be back in a minute. Well, the refrigerator is OK, so it must be the electricity for the lights. I think something might be wrong with the fuse box (*pronounced "fuze"*).

S2: You're probably right. I'll come to your department this evening and check.

S3: Ok, great! I'll be waiting for you....in the dark.

New Interchange III – Unit 8a

S1: Do you want to take a class with me at the community college?

S2: Maybe. What are they offering?

S1: Well, here's the course catalog. Take a look.

S2: Hmm. They've got a lot of language classes: Chinese, German, Japanese. Would you rather learn an Asian language or a European one?

S1: Actually, I think I'd rather take an art class. They have one on landscape photography and another on making videos.

S2: That sounds OK. But I think I'd prefer studying video to learning about photography.

S1: Oh, wait. It says here that you need to provide your own video equipment.

S2: Oh, I'd rather not spend a lot of money. Let's see what else they're offering.

S1: Hey, this sounds fairly interesting: *the art of bonsai*. (pronounced like bonz-eye)

S2: What's bonsai?

S1: You know, those tiny trees they grow in little tubs. It says here they have a course on how to grow them, and how to develop a business selling bonsais. Apparently, it's possible to sell them from \$500 each!

S2: Wow! That does sound good. Is the course expensive? Do you have to buy any special equipment?

S1: Nope. Nothing is required except plant containers and some young trees.

S2: Alright, let's check it out, then.

New Interchange III – Unit 8b

S1: So how's your French class going?

S2: Not bad, but I'm finding the pronunciation difficult.

S1: Well, it takes a while to get it right. You could improve your accent by listening to tapes.

S2: That's a good idea. But how do you learn new vocabulary? I always seem to forget new words.

S1: I learn new words by writing them on pieces of paper and sticking them on my bedroom wall. I look at them every night before I go to sleep.

S2: Maybe I should try something like that.

S1: So how do you usually study your French vocabulary?

S2: I keep a record of new words, and then prepare study cards.

S1: Study cards?

S2: They're just pieces of paper with the words on one side, and meanings on the other side. I go through the cards whenever I have free time.

S1: Oh, you mean *flash cards*! Well, using flash cards is supposed to be one of the best ways to learn new vocabulary. So, I'm surprised it's not working for you.

S2: I'm sure the problem isn't the cards. The problem is I don't go through the cards often enough. I've got to find more free time for studying them.

S1: I suggest you keep those cards with you all the time. Five minutes on the toilet....10 minutes waiting for the bus....Get my point?

G.5 Problems Students Found in the Mid-term and Final Exam

1. The quality of the recording and/or the visual equipment

| No. | Characteristics of the Quality of the Recording | Mid-term Exam | | Final Exam | |
|-----|--|---------------|----|------------|-----|
| | | Freq. | % | Freq. | % |
| a. | The quality was so poor that I couldn't hear the text clearly. | 0 | 0% | 0 | 0% |
| b. | The background noise outside the testing environment was too loud. | 0 | 0% | 5 | 20% |
| c. | The volume of the tape recorder or audio-visual equipment was too low. | 0 | 0% | 0 | 0% |
| d. | The quality of the recording was good, and I could hear the texts clearly. | 0 | 0% | 20 | 80% |
| e. | Other: | 0 | 0% | 0 | 0% |

2. Testing time

| No. | Characteristics of Testing Time | Mid-term Exam | | Final Exam | |
|-----|--|---------------|------|------------|------|
| | | Freq. | % | Freq. | % |
| a. | Time was too limited to answer all the questions properly. | 0 | 0% | 0 | 0% |
| b. | The testing time was sufficient for me to answer all the questions properly. | 25 | 100% | 25 | 100% |
| c. | Other: | 0 | 0% | 0 | 0% |

3. The test/task instructions

| No. | Characteristics of Test/Task Instructions | Mid-term Exam | | Final Exam | |
|-----|---|---------------|------|------------|-----|
| | | Freq. | % | Freq. | % |
| a. | The instructions of each test section were not clear. | 0 | 0% | 3 | 12% |
| b. | The instructions were too complicated. | 0 | 0% | 0 | 0% |
| c. | The test/task instructions were clear. | 25 | 100% | 22 | 88% |
| d. | Other: | 0 | 0% | 0 | 0% |

4. The length of the listening texts

| No. | Characteristics of the Length of the Texts | Mid-term Exam | | Final Exam | |
|-----|--|---------------|------|------------|-----|
| | | Freq. | % | Freq. | % |
| a. | The texts in the test were longer than I was used to listening to in class. | 0 | 0% | 1 | 4% |
| b. | Longer texts increased the difficulty of understanding. | 0 | 0% | 0 | 0% |
| c. | I got lost in listening to the longer texts. | 0 | 0% | 1 | 4% |
| d. | The lengths of the texts in the test were similar to those I listened to in class. | 25 | 100% | 23 | 92% |
| e. | Other: | 0 | 0% | 0 | 0% |

Glossary

- assessment** It is often used interchangeably with testing (Davies *et al.*, 1999), but also used to encompass the gathering of test contents and methods, including test results, for the purpose of evaluation and making decisions.
- authenticity** For the purposes of the present study and its focus on listening, the target language will be regarded as authentic when discourses are produced by a real speaker for a real audience.
- criterion-referenced test** Tests that examine a specific domain of knowledge or skill which testers should have mastered. In contrast to norm-referenced tests, a cut-off score (e.g. pass mark) is set for criterion-referenced tests to judge whether testees can meet the criterion.
- direct test** It measures ability directly in an authentic context and format, as opposed to an indirect test that requires performance of a contrived task from which inference is drawn about the presence of the ability concerned (Henning, 1987: 191). Direct testing is considered to establish greater predictive validity as it provides information on the test-taker's language ability in real-life situations.
- programme evaluation** In an academic environment, a systematic gathering of teaching, learning, and assessment information from a programme or a course. The purposes of evaluation are to inform decisions and ensure quality.
- task-based instruction** It relates to language learning procedures conceptualised as a series of formal tasks with preparation and follow-up activities. Tasks require

students to use the target language to solve the problems, and spontaneous and interactive oral communication is frequently emphasised.

test input

In language testing, input can be verbal (from a single word to a discourse) or non-verbal (from a picture, to body language via a live interlocutor) (Davies *et al.*, 1999: 83).

test rubric

The instructions written on a test paper which explain the aim of the test, the instructions for the tasks, time allocation, response format, and possibly the relative weighting of test sections.

test validity

In a broad definition, a test is valid when test items correspond to the teaching and test objectives (Henning, 1987). The most important quality of validity lies in the interpretation of test scores so that any inferences or decisions made on the basis of test scores are meaningful, appropriate, and useful (American Psychological Association, 1985).

language testing

Instruments that consist of specified tasks to measure language ability or aptitude for specific purposes.

washback effect

It is sometimes called “backwash” and means the influence of test results on teaching and learning. Positive washback means that students’ language skills are improved after the test and teaching is modified to benefit students’ learning, not simply teaching or learning for test purposes. Negative washback comes from testing students with a narrow definition of test content or format, and this constrains the teaching and learning goals in ways held to be undesirable.

References and Bibliography

- Adair-Hauck, B., Glisan, E., Koda, W. K., Swender, E. B. and Sandroek, P. (2006) The Integrated Performance Assessment (IPA): Connecting assessment to instruction and learning. *Foreign Language Annals*, 39 (3): 359-382.
- Alder, P. A. and Alder, P. (1994) Observational techniques. In N. K. Denzin and Y. S. Lincoln (eds.) *Handbook of Qualitative Research*, 377-392. London: Sage Publications.
- Alderson, J. C. (1992) Guidelines for the evaluation of language education. In Alderson, J. C. and Beretta, A. (eds.) *Evaluating Second Language Education*, pp. 274-304. Cambridge: Cambridge University Press.
- Alderson, J. C. (2000) *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. and Banerjee, J. (2001) State-of-the-Art Review: Language testing and assessment (Part 1). *Language Teaching*, 34, 213-236.
- Alderson, J. C. and Banerjee, J. (2002) State-of-the-Art Review: Language testing and assessment (Part 2). *Language Teaching*, 35, 79-113.
- Alderson, J. C., Clapham, C. and Wall, D. (1995) *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C. and Wall, D. (1993) Does washback exist? *Applied Linguistics*, 14 (2): 115-129.
- Aldridge, A. and Levine, K. (2001) *Surveying the Social World: Principles and practices in survey research*. Buckingham: Open University Press.
- Allen, J., M. Fröhlich, and N. Spada (1984) The communicative orientation of language teaching: an observation scheme. In Handscombe, J., R. Orem, and B. Taylor (eds.) *On TESOL '83: The Question of Control*. Washington D.C.: TESOL.

- Allwright, D. (1984) Why don't learners learn what teachers teach? The interaction hypothesis. In Singleton, D. M. and Little, D. G. (eds.) *Language Learning in Formal and Informal Contexts*, pp. 3-18. Dublin: IRAAL.
- Allwright, D. and Bailey, K. M. (1991) *Focus on the Language Classroom: An Introduction to Classroom Research for Language Teachers*. London: Longman.
- Allwright, R. L. (1984) The importance of interaction in classroom language learning. *Applied Linguistics*, 5 (2): 156-171.
- Anastasi, A. and Urbina, S. (1997) *Psychological Testing*, 7th ed. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Anderson, A. and Lynch, T. (1988) *Listening*. Oxford University Press.
- Anderson, D. F. (1953) Test of achievement in the English language. *ELT Journal*, 7 (2): 37-69.
- Anderson, R. C. (1972) How to construct achievement tests to assess comprehension. *Review of Educational Research*, 42 (2): 145-170.
- Arnold, J. (2000) Seeing through listening comprehension exam anxiety. *TESOL Quarterly*, 34, 777-786.
- Bachman, L. F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004) *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F. and Palmer, A. S. (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
- Bailey, K. M. (1996) Working for washback: A review of the washback concept in language testing. *Language Testing*, 13, 257-279.
- Baker, T. L. (1994) *Doing Social Research*, 2nd ed. New York: McGraw-Hill Inc.

- Banks, S. R. (2005) *Classroom Assessment: Issues and Practices*. Boston, Massachusetts: Allyn and Bacon.
- Bax, S. (2003) The end of CLT: a context approach to language teaching. *ELT Journal*, 57 (3): 278-287.
- Beglar D. and Hunt, A. (2002) Implementing task-based language teaching. In J. C. Richards and W. A. Renandya (eds.) *Methodology in Language Teaching: An Anthology of Current Practice*, pp.96-106. Cambridge: Cambridge University Press.
- Bell, J. (2005) *Doing Your Research Project: A guide for first-time researchers in education, health and social science*, 4th ed. Maidenhead: Open University Press.
- Bennett, J. (2003) *Evaluation Methods in Research*. London: Continuum.
- Biggs, J. (1995) Assumptions underlying new approaches to educational assessment: implications for Hong Kong. *Curriculum Forum*, 4 (2): 1-22.
- Blackwell, R. and Backmore, P. (eds.) *Towards Strategic Staff Development in Higher Education*. Buckingham: Open University Press.
- Bleiklie, I. (2002) Explaining change in higher education policy. In P. R. Trowler (ed.) *Higher Education Policy and Institutional Change: Intentions and Outcomes in Turbulent Environments*, pp. 24-45. Buckingham: Open University Press.
- Blumberg, H. H., Fuller, C. and Hare, A. P. (1974) Response rates in postal surveys. *Public Opinion Quarterly*, 38, 113-123.
- Blythman, M., MacLeod, D., and Ciesla, M. (1989) *Classroom Observation from Inside*. Edinburgh: The Scottish Council for Research in Education. (<http://www.scre.ac.uk/pdf/spotlight/Spotlight16.pdf>)
- Boyle, J. and Suen, D. L. K. (1994) Communicative considerations in a large-scale listening test. In J. Boyle and P. Falvey (eds.) *English Language Testing in Hong Kong*. Hong Kong: The Chinese University Press.

- Bradburn, N. M., Seymour Sudman, E.B. et al. (1979) *Improving Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- Breen, M. (1985) Authenticity in the language classroom. *Applied Linguistics*, 6 (1): 60-70.
- Breen, M. (1987) Learner contributions to task design. In C. Candlin and D. Murphy (eds.) *Language Learning Tasks*. Englewood Cliffs, NJ: Prentice-Hall.
- Brennan, J. and Shah, T. (2000) *Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change*. Buckingham: Open University Press.
- Brindley, G. (1998) Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Brindley, G. (2002) Issues in language assessment. In R. B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics*, pp. 459-470. Oxford: Oxford University Press.
- Brindley, G. and Slatyer, H. (2002) Exploring task difficulty in ESL listening assessment. *Language Testing*, 19 (4): 369-394.
- Brown, G. (1987) Twenty-five years of teaching listening comprehension. *English Teaching Forum*, 25 (4): 11-15.
- Brown, S. (1990) *Getting the Most from Observing Teaching*. Edinburgh: The Scottish Council for Research in Education.
(<http://www.scre.ac.uk/pdf/spotlight/Spotlight32.pdf>)
- Brown, S. and Glasner, A. (eds.) (1999) *Assessment Matters in Higher Education: Choosing and Using Diverse Approaches*. Buckingham: Open University Press.
- Brown, J. D. and Hudson, T. (2002) *Criterion-referenced Language Testing*. Cambridge: Cambridge University Press.
- Buck, G. (1988) Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*, 10, 15-42.

- Buck, G. (1992) Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42 (3): 313-357.
- Buck, G. (1990) *The Testing of Second Language Listening Comprehension*. Unpublished PhD Dissertation, University of Lancaster.
- Buck, G. (1991) The testing of listening comprehension: an introspective study. *Language Testing*, 8 (1): 67-91.
- Buck, G. (2001) *Assessing Listening*. Cambridge: Cambridge University Press.
- Burger, S. and Doherty, J. (1992) Testing receptive skills within a comprehension-based approach, pp. 299-318. In R. J. Courchene, J. I. Glidden, J. St. John and C. Therien (eds.) *Comprehension-based second language teaching*. Ottawa: University of Ottawa Press.
- Carroll, B. J. and Hall, P. (1985) *Make Your Own Language Tests*. Oxford: Pergamon.
- Canale, M. and Swain, M. (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Canale, M. (1984) Considerations in the testing of reading and listening proficiency. *Foreign Language Annals*, 17 (?): 349-357.
- Canale, M. (1988) The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67-84.
- Cannell, C. F., Miller, P. V., and Oksenberg, L. (1981) Research on interviewing techniques. *Sociological Methodology*, 12, 389-437.
- Chapelle, C. A. (1998) Construct definition and validity inquiry in SLA research. In L. F. Bachman, and A. D. Cohen. *Interfaces between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Chapelle, C. A. (1999) Language testing: Methods. In B. Spolsky (ed.) *Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier.

- Chapelle, C. A. (1999) Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254-272.
- Che, L.-C. (2004) *Educational Research of High Educational Development of Our Country of Liberalization Policy*. Unpublished MA Dissertation, National Sun Yat-Sen University.
- Chen, M.-Y. (2003) *The Challenge of Higher Education under the Impact of WTO*. Unpublished MA Dissertation, Chung-Hua University.
- Cheng, L. (1999) Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15, 253-271.
- Cheng, L. and Curtis, A. (2004) Washback or backwash: A review of the impact of testing on teaching and learning. In Cheng, L., Watanabe, Y., and A. Curtis (eds.) *Washback in Language Testing: Research Contexts and Methods*. London: Lawrence Erlbaum.
- Chiang, C. S. and Dunkel, P. (1992) The effect of speech modification, prior knowledge, and listening proficiency on EFL lecture learning. *TESOL Quarterly*, 26 (2): 345-374.
- Chiang-Fan, Y-C. (2005) *The Effectiveness of Implementing Task-Based Instruction in a Primary School in Taiwan*. Unpublished MA Dissertation, Yuan-Ze University.
- Clark, H. H. and Schober, M. F. (1992) Asking questions and influencing answers. In J. M. Tanur (ed.) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.
- Cohen, A. D. (1980) *Testing Language Ability in the Classroom*. Rowley, Massachusetts: Newbury House Publishers, Inc.
- Cohen, A. D. (1994) *Assessing Language Ability in the Classroom*, 2nd ed. Boston, Massachusetts: Heinle & Heinle.
- Cohen, A. D. (1998) *Strategies in Learning and Using a Second Language*.

London: Longman.

Cohen, A. D., Larsen-Freeman, D. and Tarone, E. (1991) The contribution of SLA theories and research to teaching language. Paper presented at the Regional Language Centre Seminar on Language Acquisition and the Second/Foreign Language Classroom. Singapore, 22-26 April.

Cohen, L., Manion, L., and Morrison, K. (2000) *Research Methods in Education*, 5th ed. London: RoutledgeFalmer.

Cohen, A. D. and Macaro, E. (eds.) (2007) *Language Learner Strategies: Thirty years of research and practice*. Oxford: Oxford University Press.

Coulthard, M. (1977) *An Introduction to Discourse Analysis*. London: Longman.

Data Protection Act (1998) Reprinted incorporating corrections (2000), The Stationery Office, Norwich.

Davies, A. (1978) Language testing: survey articles (Part 1). *Language Teaching and Linguistics Abstracts*, 11, 145-159.

Davies, A. (1978) Language testing: survey article (Part 2). *Language Teaching and Linguistics Abstracts*, 11, 215-231.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. (1999) *Dictionary of Language Testing* (Studies in Language Testing 7). Cambridge: Cambridge University Press.

Davies, A. and Elder, C. (2005) Validity and validation in language testing. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp. 795-813. New Jersey: Lawrence Erlbaum Associates, Inc.

Davies, J. and Brember, I. (1994) The reliability and validity of the "smiley" scale. *British Educational Research Journal*, 20 (4): 447-454.

Denscombe, M. (2003) *The Good Research Guide for Small-scale Social Research Projects*, 2nd edn. Buckingham: Open University Press.

- Derwing, T. and Munro, M. J. (2001) What speaking rate do non-native listeners prefer? *Applied Linguistics*, 22 (3): 324-337.
- Derwing, T. M. and Munro, M. J. (2005) Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39 (3): 379-397.
- De Vaus, D. A. (2001) *Surveys in Social Research*, 5th ed. London: Routledge.
- Diamond, I. and Jefferies, J. (2001) *Beginning Statistics: An Introduction for Social Scientists*. London: Sage Publications.
- Dochy, F., Moerkerke, G, and Martens, R. (1996) Integrating assessment, learning and instruction: Assessment of domain-specific and domain-transcending prior knowledge and progress. *Studies in Educational Evaluation*, 22 (4): 309-339.
- Dörnyei, Z. (2001) *Motivational Strategies in the Language Classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Z. (2003) *Questionnaires in Second Language Research: Construction, Administration, and Processing*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Douglas, D. (1998) Testing methods in context-based second language research. In L. F. Bachman, and A. D. Cohen. *Interfaces between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Earl, L. M. (2003) *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Thousand Oaks: Corwin Press, Sage Publications Inc.
- Elder, C. (1997) What does test bias have to do with fairness? *Language Testing*, 14 (3): 261-277.
- Ellis, Rod (1990) *Instructed Second Language Acquisition*. Oxford: Basil Blackwell.

- Ellis, Roger (ed.) (1993) *Quality Assurance for University Teaching*. Buckingham: Open University Press.
- Ellis, Rod (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, Rod (1997) The empirical evaluation of language teaching materials. *ELT Journal*, 51 (1): 36-42.
- Ellis, Rod (2003) *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.
- Ellis, Rod (2005) Instructed language learning and task-based teaching. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp. 713-728. New Jersey: Lawrence Erlbaum Associates, Inc.
- Farrall, S., Bannister, J. Ditton, J. and Gilchrist, E. (1997) Open and closed question. *Social Research Update*, 17. Guildford: University of Surrey. (<http://www.soc.surrey.ac.uk/sru/sru17.html>)
- Field, J. (1998) Skills and strategies: towards a new methodology for listening. *ELT Journal*, 52 (2): 110-118.
- Field, J. (2002) The changing face of listening. In J. C. Richards and W. A. Renandya (eds.) *Methodology in Language Teaching: An Anthology of Current Practice*, pp.242-247. Cambridge: Cambridge University Press.
- Field, J. (2003) *Psycholinguistics: A resource book for students*. London: Routledge.
- Flowerdew, J. and Miller, L. (2005) *Second Language Listening: Theory and Practice*. Cambridge: Cambridge University Press.
- Freedle, R. and Kostin, I. (1999) Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16 (1): 2-32.

- Fulcher, G. (1999) Assessment in English for Academic Purposes: Putting content validity in its place. *Applied Linguistics*, 20 (2): 221-236.
- Fulcher G. and Davidson, F. (2007) *Language Testing and Assessment: An Advanced Resource Book*. New York: Routledge.
- Gillham, B. (2000) *Developing Questionnaire*. London: Continuum.
- Gilmore, A. (2007) State-of-the-art article: Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40 (2): 97-118.
- Gorard, S. and Taylor, C. (2004a) *Combining Methods in Educational and Social Research*. Maidenhead: Open University Press.
- Gorard, S. and Taylor, C. (2004b) What is “triangulation”? *Building Research Capacity*, 7-9.
- Gourlay, L. (2005) Directions and indirect actions: learner adaptation of a classroom task. *ELT Journal*, 59 (3): 209-219.
- Grbich, C. (1999) *Qualitative Research in Health*. London: Sage Publications.
- Green, C. F., E. R. Christopher, and J. Lam (2002) Developing discussion skills in the EFL classroom. In J. C. Richards and W. A. Renandya (eds.) *Methodology in Language Teaching: An Anthology of Current Practice*, pp.225-233. Cambridge: Cambridge University Press.
- Grinyer, A. (2002) The anonymity of research participants: assumptions, ethics and practicalities. *Social Research Update*, 36. Guildford: University of Surrey. (<http://www.soc.surrey.ac.uk/sru/sru36.html>)
- Gronlund, N. E. (1968) *Constructing Achievement Tests*. Englewood Cliffs, N. J.: Prentice-Hall, Inc.
- Guba, E. G. and Lincoln, Y. S. (1987) Naturalistic inquiry. In M. J. Dunkin (ed.) *The International Encyclopedia of Teaching and Teacher Education*. Oxford: Pergamon.
- Hamp-Lyons, L. (1997) Washback, impact and validity: ethical concerns.

Language Testing, 14 (3): 295-303.

Harrison, A. (1983) *A Language Testing Handbook*. London: Macmillan Press Ltd.

Hasan, A. S. (2000) Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, 13 (2): 137-153.

Hedge, T. (2000) *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press.

Henning, G. (1987) *A Guide to Language Testing: Development, Evaluation, Research*. Boston, Massachusetts: Heinle & Heinle Publishers.

Hiep, P. H. (2007) Communicative language teaching: unity within diversity. *ELT Journal*, 61 (3): 193-201.

Hildyard, A. and Olson, D. (1982) On the comprehension and memory of oral versus written discourse. In Tannen, D. (ed.) *Spoken and Written language*, pp. 19-32. Norwood: NJ Ablex.

Ho, Y-K. (2003) Audiotaped dialogue journals: an alternative form of speaking practice. *ELT Journal*, 57 (3): 269-277.

Ho, U-S. (2006) *Elementary School English Teachers' Beliefs and Practices toward Communicative Language Teaching*. Unpublished MA Dissertation, Da-Yeh University.

Hogg, M. A. and Vaughan, G. M. (2005) *Social Psychology*, 4th ed. London: Pearson Education Limited.

Hsu, C-H. (2003) *Impacts of English Teachers' Perceptions of Communicative Language Teaching on Classroom Practices in Senior High School in Taiwan*. Unpublished MA Dissertation, National Kaohsiung Normal University.

Hughes, A. (2003) *Testing for Language Teachers*, 2nd ed. Cambridge: Cambridge University Press.

Hulstijn, J. H. (2003) Incidental and Intentional Learning. In Doughty, C. J. and

- Long, M. H. (eds.) *The Handbook of Second Language Acquisition*, pp. 349-381. Oxford: Blackwell.
- Jenkins, J. (2004) Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics*, 24, 109-125.
- Jennings, M., Fox, J., Graves, B. and Shohamy, E. (1999) The test-takers' choice: an investigation of the effect of topic on language-test performance. *Language Testing*, 16 (4): 426-456.
- Jones, R. H. (2002) Beyond "listen and repeat": pronunciation teaching materials and theories of second language acquisition. In J. C. Richards and W. A. Renandya (eds.) *Methodology in Language Teaching: An Anthology of Current Practice*, pp.178-186. Cambridge: Cambridge University Press.
- Jordon, R. R. (1997) *English for Academic Purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kelly, R. (1999) Foreign language testing. In B. Spolsky (ed.) *Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier.
- Kiely, R. (1998) Programme evaluation by teachers: issues of policy and practice. In P. Rea-Dickins and K. P. Germaine (eds.) *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, pp. 78-104. Harlow, Essex: Addison Wesley Longman.
- Kiely, R. and Rea-Dickins, P. (2005) *Programme Evaluation in Language Education*. New York: Palgrave Macmillan.
- Kumaravadivelu, B. (1991) Language learning tasks: teacher intention and learner interpretation. *ELT Journal*, 45: 98-107.
- Kunnan, A. J. (2000) Fairness and justice for all. In A. J. Kunnan (ed.) *Fairness and Validation in Language Assessment: Selected Papers from the 19th Language Testing Research Colloquium, Orlando, Florida* (Studies in Language Testing 9), pp. 1-14. Cambridge: Cambridge University Press.

- Kuo, P.-Y. (2005) *Higher Education Policy Analysis – A Case Study in Taiwan*. Unpublished MA Dissertation, Hsuan-Chuang University.
- Kuo, Y-H. (2002) *The Effect of Communicative Language Teaching on English Oral Communicative Competence of the Elementary School Students*. Unpublished MA Dissertation, National Hualien University of Education.
- Krashen, S. (1981) *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.
- Krashen, S. (1982) *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.
- Krathwohl, D. R. and Payne, D. A. (1971) Defining and assessing educational objectives. In R. L. Thorndike (ed.) *Educational Measurement*, pp. 17-45. Washington, DC: American Council of Education.
- Lado, R. (1961) *Language Testing: the construction and use of foreign language tests*. London: Longman.
- Larsen-Freeman, D. and Long, M. H. (1991) *An Introduction to Second Language Acquisition Research*. London: Longman.
- Lee, R. M. (2000) *Unobstrusive Methods in Social Research*. Buckingham: Open University Press.
- Leung, C. (2005) Classroom teacher assessment of second language development: construct as practice. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp. 869-888. New Jersey: Lawrence Erlbaum Associates, Inc.
- Lewkowicz, J. A. (1991) Testing listening comprehension: a new approach? *Hong Kong Papers in Linguistics and Language Testing*, 14, 25-31.
- Lewkowicz, J. A. (1997) *Investigating Authenticity in Language Testing*. Unpublished PhD Dissertation, University of Lancaster.
- Lewkowicz, J. A. (2000) Authenticity in language testing: some outstanding

- questions. *Language Testing*, 17 (1): 43-64.
- Lightbown, P. M., Spada, N., and Wallace, R. (1980) Some effects of instruction on child and adolescent ESL learners. In Scarcella and Krashen (eds.) *Research in Second Language Acquisition*, pp. 162-172. Rowley, Mass.: Newbury House.
- Lin, J-J. (2004) *Implementing Task-Based Activities into an EFL Fourth-Grade Class in Taiwan: A Case Study*. Unpublished MA Dissertation, National Taipei University of Education.
- Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic Inquiry*. Beverly Hills: Sage Publications.
- Littlewood, W. (1981) *Communicative Language Teaching*. Cambridge: Cambridge University Press.
- Lo, H.-C. (2004) *Changing Governance in Higher Education: A Comparative Study of Hong Kong and Taiwan*. Unpublished MPhil Thesis, City University of Hong Kong.
- Long, M. H. and Crookes, G. (1992) Three approaches to task-based syllabus design. *TESOL Quarterly*, 26 (1): 27-56.
- Long, R. L. (1990) What you don't know can't help you. *Studies in Second Language Acquisition*, 12: 65-80.
- Low, G. (1996) Intensifiers and hedges in questionnaire items and the lexical invisibility hypothesis. *Applied Linguistics*, 17 (1): 1-37.
- Low, G. (1999) What respondents do with questionnaires: accounting for incongruity and fluidity. *Applied Linguistics*, 20 (4): 503-533.
- Lun, L-I (2004) *A Study of Teachers' and Students' Perceptions of Task-Based EFL Instruction in Vocational High School Setting*. Unpublished MA Dissertation, National Chung Cheng University.
- Lynch, T. (2002) Listening: Questions of level. In R. B. Kaplan (ed.) *The Oxford Handbook of Applied Linguistics*. Pp. 39-48. Oxford: Oxford University Press.

- Lynch and Mendelsohn (2002) Listening. In Schmitt, N. (ed.) *An Introduction to Applied Linguistics*, pp.193-210. London: Arnold.
- MacDonald, M. Badger, R. and White, G. (2000) The real thing? Authenticity and academic listening. *English for Specific Purposes Journal*, 19: 253-267.
- Mackay, R. (1994) Undertaking ESL/EFL programme review for accountability and improvement. *ELT Journal*, 48 (2): 142-149.
- Mackay, R., Wellesley, S. and Bazergan, E. (1995) Participatory evaluation. *ELT Journal*, 49 (4): 308-317.
- Macklin, M. C. and Machleit, K. A. (1989) Measuring preschool children's attitude. *Marketing Letters*, 1 (3): 253-265.
- Malamah-Thomas, A. (1987) *Classroom Interaction*. Oxford: Oxford University Press.
- May, T. (1997) *Social Research: Issues, Methods and Process*, 2nd ed. Buckingham: Open University Press.
- Mason, J. (2002) *Qualitative Researching*, 2nd ed. London: Sage Publications.
- Masters, G. N. and J. Keeves (eds.) (1999) *Advances in Measurement in Educational Research and Assessment*. Oxford: Pergamon.
- McCarthy, M. and O'Keeffe, A. (2004) Research in the teaching of speaking. *Annual Review of Applied Linguistics*, 24, 26-43.
- McNamara, T. (1999) Language testing: Users and uses. In B. Spolsky (ed.) *Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier.
- McNamara, T. (2000) *Language Testing: Oxford Introductions to Language Study*. Oxford: Oxford University Press.
- McTear, M. (1975) Structure and categories of foreign language teaching sequences. In Allwright, R. (ed.) *Working Papers: Language Teaching*

- Classroom Research*. University of Essex, Department of Language and Linguistics.
- Mehan, H. (1979) *Learning Lessons: Social Organization in the Classroom*. Cambridge, Mass.: Harvard University Press.
- Mendelsohn, D. J. (1994) *Learning to Listen: A Strategy-Based Approach for the Second-Language Learners*. San Diego: Dominic Press.
- Mendelsohn, D. J. (1998) Teaching listening. *Annual Review of Applied Linguistics*, 18, 81-101.
- Messick, S. (1989) Validity. In R. L. Linn (ed.) *Educational Measurement*, 3rd ed. New York: Macmillan. Pp.13 – 103.
- Messick, S. (1994) The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13-23.
- Messick, S. (1996) Validity and washback in language testing. *Language Testing*, 13 (4), 241-256.
- Miles, M. B. and Huberman, A. M. (1994) *Qualitative Data Analysis*, 2nd ed. Thousand Oaks: Sage Publications.
- Morrow, K. (1977) Authentic texts and ESP. In S. Holden (ed.) *English for Specific Purposes*. London: Modern English Publications, 13-17.
- Morrow, K. (1979) Communicative language testing: revolution or evolution? In C. J. Brumfit and K. Johnson (eds.) *The Communicative Approach to Language Teaching*, pp. 143-157. Oxford: Oxford University Press.
- Mortimore, P., Sammons, P. Stoll, L. Lewis, D. & Ecob, R. (1986) *The Junior School Project Report Part D Technical Appendices*. London: Inner London Education Authority, Research and Statistics Branch.
- Muijs, D. and Reynolds, D. (2001) *Effective Teaching: Evidence and Practice*. London: Paul Chapman Publishing.

- Munn, P. and Drever, E. (2004) *Using Questionnaires in Small-Scale Research: A Beginner's Guide*. Edinburgh: the SCRE Centre of the University of Glasgow.
- Munro, M. J. and Derwing, T. M. (1995) Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.
- National Research Council (1979) *Privacy and Confidentiality as Factors in Survey Response*. Washington, D. C.: U.S. Government Printing Office.
- Nunan, D. (1989) *Designing Tasks for the Communicative Classroom*. Cambridge: CUP.
- Nunan, D. (1991) *Language Teaching Methodology: A Textbook for Teachers*. Hemel Hempstead: Prentice Hall International.
- Nunan, D. (2002) Listening in language learning. In J. C. Richards and W. A. Renandya (eds.) *Methodology in Language Teaching: An Anthology of Current Practice*, pp.238-243. Cambridge: Cambridge University Press.
- Nunan, D. (2004) *Task-based Language Learning*. Cambridge: Cambridge University Press.
- Oller, J. W. (1971) Dictation as a device for testing foreign-language proficiency. *ELT Journal*, 25 (3): 254-259.
- Oller, J. W. (1979) *Language Tests at School: A Pragmatic Approach*. London: Longman.
- Opie, C. (ed.) (2004) *Doing Educational Research: A guide to first-time researchers*. London: Sage Publications.
- Oppenheim, A. N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement*. London: Pinter Publishers Ltd.
- Patton, M. Q. (2002) *Qualitative Research and Evaluation Methods*, 3rd ed. London: Sage Publications.

- Payne, S. (1951) *The Art of Asking Questions*. Princeton: Princeton University Press.
- Perlesz, A. and Lindsay, J. (2003) Methodological triangulation in researching families: making sense of dissonant data. *International Journal of Social Research Methodology*, 6 (1): 25-40.
- Perrott, E. (1982) *Effective Teaching: A Practical Guide to Improving Your Teaching*. Harlow Essex: Longman.
- Peterson, R. A. (2000) *Constructing Effective Questionnaires*. Thousand Oaks: Sage Publications.
- Pica, T., Young, R. and Doughty, C. (1987) The impact of interaction on comprehension. *TESOL Quarterly*, 21 (4): 737-758.
- Pollitt, A., Hutchinson, C., Entwistle, N., and De Luca, C. (1985) *What Makes Exam Questions Difficult? An Analysis of "O" Grade Questions and Answers*. Edinburgh: Scottish Academic Press.
- Read, J. (2002) The use of interactive input in EAP listening assessment. *Journal of English for Academic Purposes*, 1, 105-119.
- Rea-Dickins, P. (1994) State-of-the-art article: Evaluation and English language teaching. *Language Teaching*, 27 (2): 71-91.
- Rea-Dickins, P. and Germaine, K. P. (1998) The price of everything and the value of nothing: trends in language programme evaluation. In P. Rea-Dickins and K. P. Germaine (eds.) *Managing Evaluation and Innovation in Language Teaching: Building Bridges*, pp. 3-19. Harlow, Essex: Addison Wesley Longman.
- Rea-Dickins, P. (2004) Understanding teachers as agents of assessment. *Language Testing*, 21 (3): 249-258.
- Rea-Dickins, P. and Scott, C. (2007) Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education*, 14 (1): 1-7.

- Reynolds, J. and Saunders, M. (1987) Teacher responses to curriculum policy: beyond the “delivery” metaphor. In J. Calderhead (ed.) *Exploring Teachers' Thinking*, pp. 195-214. London: Cassell Educational Limited.
- Richards, J., J. Platt and H. Weber (1986) *Longman Dictionary of Applied Linguistics*. London: Longman.
- Rivers, W. M. (1981) *Teaching Foreign Language Skills*. Chicago: University of Chicago Press.
- Reamer, F. G. (1979) Protecting research subjects and unintended consequences: the effect of guarantees of confidentiality. *Public Opinion Quarterly*, 43, 479-506.
- Rost, M. (1990) *Listening in Language Learning*. London: Longman.
- Rost, M. (2001) *Teaching and Researching Listening*. Harlow: Longman.
- Rost, M. (2005) L2 listening. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp. 503-527. New Jersey: Lawrence Erlbaum Associates, Inc.
- Rubin, J. (1994) A review of second language listening comprehension research. *Modern Language Journal*, 78 (2): 199-221.
- Sabatier, P. A. and Jenkins-Smith, H. (eds.) (1994) *Policy Change and Learning: An Advocacy Coalition Framework*. Boulder, CO: Westview Press.
- Salkind, N. J. (2004) *Statistics for People Who (Think They) Hate Statistics*, 2nd ed. Thousand Oaks: Sage Publications.
- Savignon, S. J. (1971) *A Study of the Effect of Training in Communicative Skills as Part of A beginning College French Course on Student Attitude and Achievement in Linguistic and Communicative Competence*. Unpublished Doctoral Dissertation, University of Illinois, Urbana-Champaign.
- Savignon, S. J. (1972) *Communicative Competence: An Experiment in Foreign Language Teaching*. Philadelphia, PA: Center for Curriculum Development.

- Savignon, S. J. (2005) Communicative language teaching: strategies and goals. In E. Hinkel (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp. 635-651. New Jersey: Lawrence Erlbaum Associates, Inc.
- Schmidt-Rinehart, B. C. (1994) The effects of topic familiarity on second language listening comprehension. *The Modern Language Journal*, 78 (2): 179-189.
- Schrafnagl, J. and Cameron, D. (1988) Are you decoding me? The assessment of understanding in oral interaction, pp. 88-97. In P. Grunwell (ed.) *Applied Linguistics in Society*. London: Centre for Information on Language Teaching and Research.
- Schuman, H. and Presser, S. (1981) *Questions and Answers in Attitude Survey: Experiments on Question Form, Wording and Context*. New York: Academic Press.
- Schumann, F. M. and Schumann, J. H. (1977) Diary of a language learner: An introspective study of second language learning. In Brown, Yorio, and Crymes (eds.) *On TESOL '77, Teaching and Learning English as a Second Language: Trends in Research and Practice*, pp. 241-249. Washington, DC: TESOL.
- Schwartz, B. (1993) On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15, 147-163.
- Setter, J. and Jenkins, J. (2005) State-of-the-Art Review Article: Pronunciation. *Language Teaching*, 38 (1): 1-17.
- Shaw, M. E. (1966) Social psychology and group processes. In J. B. Sidowski (ed.) *Experimental Methods and Instrumentation in Psychology*. Pp. 607-43. New York: McGraw-Hill.
- Shizuka, T., Takeuchi, O., Yashima, T., and Yoshizawa, K. (2006) A comparison of three- and four-option English tests for university entrance selection purposes in Japan. *Language Testing*, 23 (1): 35-57.

- Shohamy, E (1998) Critical language testing and beyond. *Studies in Educational Evaluation*, 24 (4): 331-345.
- Shohamy, E. (1998) How can language testing and SLA benefit from each other? The case of discourse. In L. F. Bachman, and A. D. Cohen. *Interfaces between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Shohamy, E. (1999) Language testing: Impact. In B. Spolsky (ed.) *Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier.
- Shohamy, E. (2001) *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. Harlow, England: Pearson Education Limited.
- Shohamy, E., Donitsa-Schmidt, S. and Ferman, I. (1996) Test impact revisited: washback effect over time. *Language Testing*, 13 (3): 298-317.
- Shohamy, E. and Inbar, O. (1991) Validation of listening comprehension tests: the effect of text and question type. *Language Testing*, 8 (1): 23-40.
- Simpson, M. and Tuson, J. (2003) *Using Observations in Small-Scale Research: A Beginner's Guide*. Edinburgh: the SCRE Centre of the University of Glasgow.
- Sinclair, J. and M. Coulthard (1975) *Towards an Analysis of Discourse*. Oxford: Oxford University Press.
- Singer, E. (1978a) Informed consent: consequences for response rate and response quality in social surveys. *American Sociological Review*, 43, 144-162.
- Singer, E. (1978b) The effect of informed consent procedures on respondents' reactions to surveys. *Journal of Consumer Research*, 5, 49-57.
- Singer, E. (1984) Public reactions to some ethical issues of social research: attitudes and behavior. *Journal of Consumer Research*, 11, 501-509.
- Singer, E. (1993) Informed consent and survey response: A summary of the empirical literature. *Journal of Official Statistics*, 9 (2): 361-375.

Singer, E. and Frankel, M. R. (1982) Informed consent procedures in telephone interviews. *American Sociological Review*, 47, 416-426.

Skehan, P. (1988) State-of-the-Art: Language testing (Part 1). *Language Teaching*, 211-221.

Skehan, P. (1989) State-of-the-Art: Language testing (Part 2). *Language Teaching*, 1-13.

Skehan, P. (1996) Second language acquisition research and task-based instruction. In J. Willis and D. Willis (eds.) *Challenge and Change in Language Teaching*, pp.17-30. London: Macmillan Heinemann.

Slimani, A. (1992) Evaluation of classroom interaction. In Alderson, J. C. and Beretta, A. (eds.) *Evaluating Second Language Education*, pp. 197-221. Cambridge: Cambridge University Press.

Smith, J. K, Smith, L. F. and De Lisi, R. (2001) *Natural Classroom Assessment: Designing Seamless Instruction and Assessment*. Thousand Oaks, CA: Corwin Press, INC.

Spada, N. M. (1997) Form-focused instruction and second-language acquisition: A review of classroom and laboratory research. *Language Teaching*, 30, 73-87.

Spence-Brown, R. (2001) The eye of the beholder: authenticity in an embedded assessment task. *Language Testing*, 18 (4): 463-481.

Spolsky, B. (1999) Language testing. In B. Spolsky (ed.) *Concise Encyclopedia of Educational Linguistics*. Oxford: Elsevier.

Stern, H. H. (1983) *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.

Suchman, L. and Jordan, B. (1992) Validity and the collaborative construction of meaning in face-to-face surveys. In J. M. Tanur (ed.) *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation.

- Tauroza, S. and Allison, D. (1990) Speech rates in British English. *Applied Linguistics*, 11, 90-105.
- Taylor, L. (2005) Washback and impact. *ELT Journal*, 59 (2): 154-155.
- Thompson, I. (1995) Assessment of Second/Foreign Language Listening Comprehension. In D. J. Mendelsohn, and J. Rubin (eds.) *A Guide for The Teaching of Second Language Listening*. San Diego, CA: Dominic Press.
- Thurstone, L. L. (1970) Attitudes can be measured. (Reprinted from the *American Journal of Sociology* (1928), 33, 529-554.) In G. F. Summers (ed.) *Attitude Measurement*. Pp. 127-41. Chicago: Rand McNally and Company.
- Tierney, R. D. (2006) Changing practices: influences on classroom assessment. *Assessment in Education*, 13 (3): 239-264.
- Trosborg, A. (1984) Stimulating interaction in the foreign language classroom through conversation in small groups of learners. In Singleton, D. M. and Little, D. G. (eds.) *Language Learning in Formal and Informal Contexts*, pp. 177-189. Dublin: IRAAL.
- Tseng, C-Y (2006) *A Study of the Effect of Task-based Instruction on Primary School EFL Students*. Unpublished MA Dissertation, National Chung Cheng University.
- Tsui, A. B. M. and J. Fullilove (1998) Bottom-up or top-down processing as a discriminator of L2 listening performance. *Applied Linguistics*, 19 (4):, 432-451.
- Ur, P. (1984) *Teaching Listening Comprehension*. Cambridge: Cambridge University Press.
- Vandergrift, L. (2004) Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3-25.
- Vandergrift, L. (2007) State-of-the-art article: Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40 (3): 191-210.

- Van Lier, L. (1988) *The Classroom and the Language Learner: Ethnography and Second Language Classroom Research*. London: Longman.
- Van Lier, L. (1991) Inside the classroom: the learning processes and teaching procedures. *Applied Language Learning*, 2, 29-69.
- Van Teijlingen, E. R. and Hundley, V. (2001) The importance of pilot studies. *Social Research Update*, 35. (<http://www.soc.surrey.ac.uk/sru/sru35.html>) Guildford: University of Surrey.
- Wall, D. (1997) Impact and washback in language testing. In C. Clapham and D. Corson (eds.) *Encyclopedia of Language and Education, Volume 7: Language Testing and Assessment*, 291-302. Netherlands: Kluwer Academic Publishers.
- Wang, C. (2002) Innovative teaching in foreign language contexts: the case of Taiwan. In Savignon, S. J. (ed.) *Interpreting Communicative Language Teaching: Contexts and Concerns in Teacher Education*, pp. 131-153. New Haven: Yale University Press.
- Watanabe, Y. (1997) Washback effects of the Japanese university entrance examination: Classroom-based research. Unpublished Doctoral Dissertation, Lancaster University, UK.
- Watanabe, Y. (2004) Methodology in washback studies. In Cheng, L., Watanabe, Y., and A. Curtis (eds.) *Washback in Language Testing: Research Contexts and Methods*. London: Lawrence Erlbaum.
- Watt, J. (1995) *Ethical Issues for Teacher Researcher*. Edinburgh: The Scottish Council for Research in Education.
(<http://www.scre.ac.uk/pdf/spotlight/Spotlight49.pdf>)
- Webb, E. J., Campbell, D. T., Schwartz, Z. D. and Sechrest, L. (1966) *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.
- Webb, E. J., Campbell, D. T., Schwartz, Z. D., Sechrest, L., and Grove, J. B. (1981) *Nonreactive Measures in the Social Sciences*. Dallas: Houghton Mifflin.

- Weems, G. H., Onwuegbuzie, A. J. and Collins, K. (2006) The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation and Research in Education*, 19 (1): 3-20.
- Weir, C. J. (1990) *Communicative Language Testing*. New York: Prentice-Hall.
- Weir, C. J. (1993) *Understanding and Developing Language Tests*. Hemel Hempstead, UK: Prentice Hall.
- Weir, C. J. (2005) *Language Testing and Validation*. New York: Palgrave Macmillan.
- Weir, C. and Roberts, J. (1994) *Evaluation in ELT*. Oxford: Blackwell Publishers.
- Wesche, M. (1987) Second language performance testing: the Ontario Test of ESL as an example. *Language Testing*, 4 (1): 28-47.
- Wesche, M. and Skehan, P. (2002) "Communicative, task-based, and content-based language instruction", in Kaplan, R. (ed.) *The Oxford Handbook of Applied Linguistics*, pp. 207-228. Oxford: Oxford University Press.
- Widdowson, H. G. (1978) *Teaching Language as Communication*. Oxford: Oxford University Press.
- Widdowson, H. G. (1979) *Explorations in Applied Linguistics*. Oxford: Oxford University Press.
- Wiggins, G. P. (1998) *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass Publishers.
- Willis, J. (1996) "A flexible framework for task-based learning". In J. Willis and D. Willis (eds.) *Challenge and Change in Language Teaching*, pp.52-62. London: Macmillan Heinemann.
- Willis, D. and Willis, J. (2007) *Doing Task-based Teaching*. Oxford: Oxford University Press.
- Wu, Y. (1998) What do tests of listening comprehension test? – A retrospection

study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15(1): 21-44.

On-line Sources:

California Distance Learning Project (2005) *Adult Learning Activities*.
<http://www.cdldonline.org>. 15 November 2006.

Center for Teaching Effectiveness, University of Texas at Austin (2004)
Classroom Observation Worksheet.
<http://www.utexas.edu/academic/cte/ClassWorksheet.pdf>. 15 October 2005.

Chen, M.-L. (2007) The significance of evaluating social science and humanity subjects. *Evaluation Bimonthly*, 5, 38-41.
<http://epaper.heeact.edu.tw/archive/2007/01/03/150.aspx> 25 December 2007.

Council on Education Reform (1996) *The Overall Report on Education Reform Consultation*. <http://www.edu.tw/>. 17 October 2007.

Du, C.-S. (2004) *Report of the Education and Cultural Committee*.
<http://www.edu.tw/>. 30 June 2006.

ETS (2006) *TOEFL Test data 2006*.
<http://www.etsliteracy.com/Media/Research/pdf/TOEFL-SUM-0506-CBT.pdf>
13 December 2007.

HEEACT (2007) *Bulletin of Evaluation Results of Taiwanese Universities*.
<http://www.heeact.edu.tw/2007-eva/bulletin.htm>. 10 January 2008.

HEEACT (2008) *Higher Education Evaluation and Accreditation Council of Taiwan*. <http://www.heeact.org.tw/>. 10 January 2008.

IELTS (2006) *IELTS Test Data 2006 – Academic*.
<http://www.ielts.org/teachersandresearchers/analysisoftestdata/article382.aspx> 13
December 2007.

Kuang, H.-Y. (2006) The announcement of evaluator list. *Evaluation Bimonthly*, 3:
37-38. <http://epaper.heeact.edu.tw/archive/2006/09/15/69.aspx>. 4 December

2007.

Liu, W.-C. (2007) What is the criteria of evaluation for academic department evaluation. *Evaluation Bimonthly*, 7: 4-5. <http://epaper.heeact.edu.tw/archive/2007/07/06/288.aspx>. 4 December 2007.

Maine Center for Meaningful Engaged Learning (MCMEL) (2005) *Laptop Learning Observation Form*. http://www.mcmel.org/MLLS/eval/Observation_Checklist_v4.pdf. 15 October 2005.

Ministry of Education (MOE) (2002) *Education Commentary in October 2002*. http://www.edu.tw/EDU_WEB/EDU_MGT/SECRETARY/EDU0835001/91publicnews/9111/B39.htm. 5 April 2005.

Ministry of Education (MOE) (2005) *Internet Newspaper – Implementation of Evaluating University Programme*. <http://epaper.edu.tw/news/940301/940301d.htm>. 5 April 2005.

Ministry of Education (MOE) (2005) *Internet Newspaper – First Meeting of University Evaluation Programme*. <http://epaper.edu.tw/news/940301/940301c.htm>. 5 April 2005.

Ministry of Education (MOE) (2008) *History of the Minister of Education*. <http://history.moe.gov.tw>. 5 January 2008.

Wang, P.-C. (2007) The process of university department evaluation. *Evaluation Bimonthly*, 8: 7-10. <http://epaper.heeact.edu.tw/archive/2007/07/06/290.aspx>. 4 December 2007

Wu, P.-C. and Chang, M.-J. (2007) The analysis of current evaluation for Taiwanese university teacher. *Evaluation Bimonthly*, 9: 9-15. <http://epaper.heeact.edu.tw/archive/2007/09/03/330.aspx>. 4 December 2007.