

An investigation into the effect of the Skills for Life
Strategy on assessment and classroom practices in ESOL
teaching in England.

Tania Caroline Horák

MA, Lancaster University

BA (Hons), Hull University

This thesis is submitted in partial fulfilment of the requirement
for the degree of
Doctor of Philosophy

Department of Linguistics and English Language
Lancaster University

November 2011

**An investigation into the effect of the Skills for Life Strategy on
assessment and classroom practices in ESOL teaching in England.**

Tania Caroline Horák, MA, Lancaster University; BA (Hons), Hull University

This thesis is submitted in partial fulfilment of the requirement for the degree of
Doctor of Philosophy, Department of Linguistics and English Language, Lancaster
University, UK.

November 2011

ABSTRACT

This study investigates the impact of the Skills for Life Strategy (2001) on assessment practices in ESOL teaching in England, and whether these assessments resulted in any washback. In this qualitative study, the Henrichsen (1989) model of the diffusion of innovation acted as the framework to explore the assessment of ESOL students in 3 further education colleges in the UK, using interviews and observations. The research found that due to the Strategy, assessment became considerably more standardised, with the focus falling on a range of external exams, although the effects of internal measures such as Independent Learning Plans was also noted. While washback was detected, mostly in the form of changes in staff-student relationships, the 'double accounting' of students preparing to sit the Skills for Life exams as well as other exams, and to some degree more of a focus on accuracy in classroom work, the washback was not particularly strong. This was attributed to the timing of the study, being relatively close to the introduction of the new range of exams. The washback was also noted to be differential, namely, that the washback was not uniform across the sites studied. Factors to explain this were investigated, including the variability of the stakes for various stakeholders, features of the teachers themselves, the quality and nature of the communication of the changes and finally other factors, as suggested by the Henrichsen model. The latter suggested some distortion of the aims of assessment caused by perceived pressure to reach targets to secure funding. The results suggested there was considerable

variability, leading to the conclusion that washback studies, which are vital for monitoring exams, need to avoid being simplistic and thus missing key factors which illuminate contextual detail. The nature of washback can easily be masked by superficial investigation.

DECLARATION

I declare that this thesis is my own work and has not been submitted for the award of a higher degree elsewhere.

ACKNOWLEDGEMENTS

Thank you to my friends who kept me going through this long process, especially those who bore the brunt of the ranting and complaining. Thank you for waiting for me to emerge at the other end of the process.

Thanks to my sister, Jana, for all the encouragement along the way.

Thanks to the study participants who gave their time generously.

Thanks also to the tea-pickers of the world. I may have single-handedly sustained the tea production industry during my studies through the copious quantities of tea I have consumed.

And thank you to Marjorie Wood in LAEL, and of course to my supervisors.

TABLE OF CONTENTS

TITLE	i
ABSTRACT	ii
DECLARATION	iii
ACKNOWLEDGEMENTS	iii
1 INTRODUCTION.....	- 1 -
1.1 The topic of this study.....	- 1 -
1.2 Assessment practices in ESOL in the UK	- 3 -
1.3 Impact and washback	- 5 -
1.4 The study location.....	- 5 -
1.5 High stakes testing.....	- 7 -
1.6 The research questions	- 10 -
1.7 The methodological approach	- 12 -
1.8 Study aims.....	- 12 -
1.9 Study outline	- 13 -
2 BACKGROUND TO ESOL	14
2.1 Chapter overview	14
2.2 What is ESOL?	14
2.3 The nature of ESOL students in the UK	17
2.4 Where is ESOL?	21
2.5 ESOL teaching in other English speaking countries	23
2.5.1 New Zealand	23
2.5.2 Canada	25
2.5.3 Australia.....	27
2.5.4 United States of America	28
2.5.5 Overview	30
2.6 Changes in ESOL provision	30
2.6.1 A history of volunteerism.....	30
2.6.2 Funding.....	31
2.6.3 Changes in the culture of education	32
2.7 Time for reform.....	33
2.8 Summary.....	34
3 BACKGROUND TO THE SKILLS FOR LIFE STRATEGY	35
3.1 Chapter overview	35
3.2 Skills for Life	35
3.3 The five strands of the Strategy	37
3.3.1 ESOL Core Curriculum	37
3.3.2 Teacher Development – qualifications and training	40
3.3.3 Teaching Materials.....	40
3.3.4 Research.....	42
3.3.5 Inspection	43
3.3.6 Assessment, including the national exams	44
3.4 Funding for the Strategy	48
3.5 Summary.....	50
4 BACKGROUND TO WASHBACK STUDIES	51
4.1 Chapter overview	51
4.2 What is washback?.....	51
4.3 The evolution of washback studies	52
4.4 Key frameworks in washback studies	53
4.5 Washback direction/parameters	55

4.6	Measurement driven instruction	59
4.6.1	Supporters of MDI	59
4.6.2	Non-supporters of MDI	60
4.7	Key components of washback studies.....	61
4.8	The contribution of educational innovation studies.....	63
4.9	The role of the teacher in washback	63
4.10	Washback and the role of stakes.....	64
4.11	The purpose of washback studies	66
4.12	Summary.....	67
5	METHODOLOGY –THEORETICAL ISSUES IN USING QUALITATIVE	
	METHODS.....	69
5.1	Chapter overview	69
5.2	The qualitative paradigm	69
5.3	Qualitative research in the field of language testing	72
5.4	Methods used to ‘capture’ washback and impact	75
5.5	Data collection - interviews	76
5.5.1	What is interviewing?	76
5.5.2	Why interview?	77
5.5.3	Dangers	77
5.6	Data collection – observations.....	79
5.6.1	Format.....	79
5.7	Key considerations in qualitative research	80
5.7.1	Sampling	80
5.7.2	Evaluating the quality of qualitative research – reconsidering the traditional criteria.....	81
5.7.2.1	Generalisability	82
5.7.2.2	Reliability.....	82
5.7.2.3	Validity	83
5.8	Other considerations.....	86
5.8.1	Researcher-practitioner issues	86
5.8.2	Ethics in qualitative research.....	86
5.9	Stages in the research process	88
5.9.1	Transcription	88
5.9.2	Analytic procedures	89
5.9.3	Data organisation.....	91
5.9.4	Description through coding - search and re-search	92
5.9.4.1	Terminology	92
5.9.4.2	How to code	93
5.9.4.3	Use of software	93
5.9.4.4	The source of codes	94
5.9.5	Interpretation	95
5.9.6	Presentation of the research	97
5.10	Summary.....	98
6	METHODS – INSTRUMENT DESIGN, DATA COLLECTION AND ANALYSIS	99
6.1	Chapter overview	99
6.2	Theoretical framework.....	99
6.3	General methodological considerations	101
6.4	Stage 1: Exploratory study	102
6.5	Institutional and teacher profiles	104
6.5.1	Institutional profile.....	104
6.5.2	Teacher profiles.....	104
6.6	Observations	105
6.6.1	Format.....	105
6.6.2	Recording.....	107
6.7	Interviews.....	107

6.8	Opportunistic data	108
6.9	Analysis	109
6.10	Outcomes of the exploratory study	109
6.11	Emerging themes	110
6.12	Stage 2: Main study.....	112
6.12.1	Revised research questions.....	112
6.13	Sources of data	113
6.13.1	Teacher profiles.....	114
6.13.2	Directors of Study Profiles	116
6.13.3	Institutional profiles	117
6.14	Interviews: content	118
6.15	Interviews: format	119
6.16	Considerations entailed in the methods employed	121
6.16.1	Two way benefit	121
6.16.2	Researcher practitioner issue	121
6.16.3	Recording.....	122
6.16.4	Interview management	122
6.17	Observation.....	123
6.18	Ethics	123
6.19	Hawthorn effect	124
6.20	Convenience Sampling.....	125
6.21	Counteracting bias	127
6.22	Personal – researcher profile.....	127
6.23	Data collection	128
6.24	Data preparation.....	131
6.24.1	Transcription	131
6.24.2	Data analysis	131
6.24.3	Coding.....	132
6.24.4	Observation data.....	134
6.25	Making meaning – interpretation	135
6.26	Summary.....	135
7	FINDINGS AND DISCUSSION: ASSESSMENT PRACTICES IN THE ESOL CLASSROOM	136
7.1	Chapter overview	136
7.2	A student’s ‘assessment journey’	137
7.3	Internal Assessment.....	139
7.3.1	Initial Interview/ screening	139
7.3.2	Placement Test.....	140
7.3.2.1	Function	140
7.3.2.2	Tests used.....	141
7.3.2.3	Procedure	143
7.3.2.4	Problems with the ‘official’ test.....	144
7.3.2.5	Spiky profiles and the problem of placement	146
7.3.2.6	Placement testing for ESOL v. literacy students	147
7.3.2.7	Administrative v. pedagogical concerns	149
7.3.2.8	Conclusions.....	149
7.3.3	Diagnostic Testing	150
7.3.3.1	Function	150
7.3.3.2	Procedure	151
7.3.3.3	Conclusion	154
7.3.4	Independent Learning Plans (ILPs)	155
7.3.4.1	Function	155
7.3.4.2	Description of ILPs	156
7.3.4.3	Old tool: new use	156

7.3.4.4	Problems in producing the ILPs.....	157
7.3.4.5	Literacy model	159
7.3.4.6	Compulsory or not?.....	161
7.3.4.7	Attitudes to ILPs	163
7.3.4.8	Conclusion	165
7.3.5	Progress Tests	166
7.3.5.1	Function	166
7.3.5.2	The nature of progress tests	166
7.3.5.3	Objections to progress testing.....	168
7.3.5.4	Conclusion	169
7.3.6	Portfolio leading to a College Certificate	169
7.3.6.1	Function	169
7.3.6.2	The nature of the qualification.....	170
7.3.6.3	Problems	171
7.3.6.4	Conclusion	172
7.3.7	Mock Exams.....	172
7.3.7.1	Function	173
7.3.7.2	Conclusion	173
7.4	External Assessment.....	174
7.4.1	Function.....	174
7.4.2	Assessment preference	175
7.4.2.1	Choice of Exam Boards	175
7.4.2.2	Portfolio v. external exams	179
7.4.3	Currency.....	181
7.4.4	What is being measured?.....	185
7.4.5	Administrative conflict	186
7.4.6	Summary.....	186
8	FINDINGS AND DISCUSSION: WAS WASHBACK EVIDENT?	189
8.1	Chapter overview	189
8.2	Defining washback.....	191
8.2.1	Key parameters in discussing washback.....	191
8.3	What is washback? What it isn't?	191
8.3.1	Student's test anxiety	192
8.3.2	Teachers' anxiety.....	193
8.3.3	Test-wiseness	194
8.4	Evidence of washback	196
8.5	Effects of internal assessment	199
8.5.1	Independent Learning Plans (ILPs)	199
8.5.2	Coping mechanisms in use of class time.....	201
8.5.3	Summary of washback from internal assessment	203
8.6	Effects of external assessments.....	203
8.6.4	Content.....	204
8.6.4.1	One class: two goals.....	204
8.6.4.2	Focus on accuracy	209
8.6.4.3	Effect intensity	212
8.6.4.4	Materials used in class	213
8.6.5	Methodology	215
8.6.5.1	Interaction patterns.....	215
8.6.5.2	Exam practice.....	217
8.6.6	Content or methodology, or both?.....	221
8.6.6.1	Narrowing of the learning experience.....	221
8.6.6.2	Atomistic approach	224
8.6.6.3	Change in attitude towards ESOL classes	226

8.6.7	Summary of washback found from external assessments	228
8.7	Impact	228
8.7.1	Raised profile of ESOL.....	228
8.7.2	Increased opportunity.....	229
8.8	Issues.....	230
8.8.1	Differential effects.....	230
8.8.2	Washback from the Skills For Life exams or from the curriculum?..	233
8.9	Reflection on the washback identified.....	234
8.10	Summary.....	235
9	FINDINGS AND DISCUSSION: FACTORS AFFECTING WASHBACK: THE ROLE OF STAKES, COMMUNICATION, RECIPIENTS AND OTHER FACTORS ..	237
9.1	Chapter overview	237
9.2	Stakes	239
9.3	Chain of pressure: the funding imperative	242
9.4	Stakes for the institution.....	243
9.5	Stakes for ESOL departments.....	246
9.6	Stakes for directors of studies.....	249
9.7	Stakes for teachers	251
9.7.1	Vicarious success	252
9.7.2	'Payment by results'	253
9.7.3	Targets	255
9.8	Stakes for the students.....	257
9.9	Differential washback	263
9.10	The Receivers: the teachers, their experience and attitudes.	265
9.11	Experience.....	267
9.12	Parameters	268
9.13	Tutor and DoS descriptions	270
9.14	Teachers' attitudes towards assessment.....	271
9.15	Advantages of the new external exams	272
9.16	Disadvantages of the new external exams	273
9.17	Evaluation of Skills for Life.....	276
9.18	Communication	279
9.18.1	Official sources.....	282
9.18.2	Professional bodies	283
9.18.3	Management	284
9.18.4	Peers.....	285
9.18.5	Exam boards	287
9.18.6	Training.....	288
9.18.7	Issues with the dissemination of information	289
9.18.7.1	Uneven dissemination pattern.....	289
9.18.7.2	Inconsistent messages	290
9.18.1	Awareness of the rationale for Skills for Life.....	293
9.19	Factors which may hinder effective change.....	295
9.19.1	Inappropriate candidacy.....	296
9.19.2	Dual functionality	300
9.19.3	Dual functionality and validity	303
9.20	Summary.....	305
10	CONCLUSION	307
10.1	Chapter overview	307
10.2	Review of the aims	307
10.3	Review of the findings.....	308
10.3.1	Assessment practices	308
10.3.2	Washback.....	309
10.3.3	Stakes	310
10.3.4	Receiver factors.....	311

10.3.5	Communication	311
10.3.6	Factors hindering effective change	312
10.3.7	Summary	313
10.4	Implications of the study	314
10.4.1	Theoretical contribution	314
10.4.2	The value of washback studies	315
10.4.3	Dual functionality : washback to impact	316
10.5	Close	320
11	LIMITATIONS AND POSSIBLE FUTURE RESEARCH	321
11.1	Limitations	321
11.1.1	Time frame	321
11.1.2	Diachronic approach	321
11.1.3	Elicitation techniques	322
11.1.4	Student data	322
11.1.5	Sole researcher	323
11.2	Possible future research	324
	REFERENCES	326
	APPENDICES	342

LIST OF FIGURES

Figure 1 The potential reasons for ESOL students being in the UK and potential reasons for attending English classes	20
Figure 2 A basic model of washback direction (Green 2007).....	55
Figure 3 Matrix of washback conditions.....	56
Figure 4 Henrichsen hybrid model of diffusion/ implementation of innovation (1989)	101
Figure 5 The relationship of the research questions and sources of data	113
Figure 6 Chain of pressure.....	243
Figure 7 Adopters, adapters or resisters	278
Figure 8 The communication network - sources of information drawn on in this study	281
Figure 9 The dual functionality of exam results and its inherent tensions	300

LIST OF TABLES

Table 1 Basic Skills levels within the National Qualifications Framework (NQF) compared to other qualifications frameworks.....	36
Table 2 Exam boards offering S4L exams accredited by QCA at various levels	47
Table 3 Range in reported amounts of funding for ESOL	49
Table 4 Stages of analysis	90
Table 5 Teacher profiles.....	105
Table 6 Description of the observation schedule.....	106
Table 7 Key themes emerging from the exploratory study.....	110
Table 8 Topics covered by the interview schedule.....	120
Table 9 Data collected – Observations	128
Table 10 Data collected – Interviews	129
Table 11 Coverage of observations of students at different NFQ levels.....	130
Table 12 The ESOL Students’ Assessment Journey: the range of assessments ESOL students’ experience and the function of each type.....	138
Table 13 The profile of assessment practices across sites.....	139
Table 14 Components of the DfES ‘official’ diagnostic tests.....	152
Table 15 Interview questions specifically focusing on washback effects	196
Table 16 Level of exam visibility	197
Table 17 Observed examples of washback	198
Table 18 Teacher profiles.....	267
Table 19 DoS Profiles	268

LIST OF APPENDICES

Appendix 1: Alderson & Wall's (1993: 120-121) 15 washback hypotheses.....	342
Appendix 2: Observation Sheet	343
Appendix 3: Student Interview Schedule	349
Appendix 4: Site Profiles.....	350
Appendix 5: Interview Schedule	351
Appendix 6: Letter of Consent	354
Appendix 7: The Quantity of Data	355
Appendix 8: Chronology of Data Collection	356
Appendix 9: Observation Details	357
Appendix 10: Interview details	358
Appendix 11: Nationalities, gender and language ability level of student informants	359
Appendix 12: Transcription Conventions	360
Appendix 13: Sample Coding (using Atlas –ti).....	361
Appendix 14: Complete List of Codes	362
Appendix 15: Example ILP (from DfES ESOL Tutor Manual 2001)	364
Appendix 16: Information on the exam boards offering Skills for Life qualifications	367
Appendix 17: Acronyms used in this study	368

1 INTRODUCTION

1.1 The topic of this study

Like gunpowder, another Chinese invention, the examination is far from an unmixed blessing, capable of dangerous misuse as well as useful for good ends
(Spolsky 1994: 64)

This thesis is a study of the influence of examinations in society. While they may not prove as fatal as gunpowder can be, the consequences of misuse of examination results can have far reaching effects. As with any tools, although they may be professionally produced and made with good intent, there is no guarantee they will always be used as intended. Examination is such a part of modern educated society that it is a topic many people have some knowledge and experience of, if not first hand, then at least via the press or via anecdote. The effect of National Curriculum assessments (commonly known as SATs) in the UK in recent years which reputedly caused severe narrowing of the curriculum and excessive exam stress for pupils, teachers, head-teachers and parents, is a case in point.

Assessment and testing, it seems, tend to have rather negative associations amongst the general public, probably because it touches on the primal human principle of fairness and on the vulnerability of being judged. As Carlsen suggests:

‘When journalists, parents and teachers discuss language testing, they often tend to draw a picture with which few professional language testers would agree. Language testing is often described as a technical enterprise that has little or nothing to do with language learning, and test constructors are

described as psychometricians who, in the search for satisfying numbers, sacrifice concern for pupils. In modern approaches, however, language testing is just as much a matter of ethics as it is a matter of mathematics' (2007: 97).

There has been in recent years a steady increase in the field of language testing in the investigation of fairness and ethics alongside the quality and properties of exams (see recent conference themes for the Association of Language Testers in Europe (ALTE) (see <http://www.alte.org>) and the European Association of Language Testing and Assessment (EALTA)(see: <http://www.ealta.eu.org> for details). LTEST-L, (<http://lists.psu.edu/TEST-L>) a professional language testers' e-discussion group, has also hosted various discussion topic threads on the topic including the Test of English debate.

This could be said to have been precipitated by Messick's (1989) proposal of the concept of consequential validity, whereby the 'societal influences of tests' (Shohamy 2001: 47) are considered (although he had been drawing attention to the impact of exams as far back as the early 1980's). Other writers such as Kunnan (2009) and McNamara (2001) have also subsequently carried on this theme.

The introduction of a new assessment regime lends itself to questioning what the effect of such a change will be on the main stakeholders: the teachers, students and educational institutions. This was the basis for this study. In 2001 in the UK, such a new regime began in the field of English For Speakers of Other Language (ESOL), i.e. English for non-native English speaking students who are learning English to enhance their lives here in the UK, when the government introduced the Skills for Life strategy. This strategy had profound consequences for the stakeholders, the repercussions of which were being felt even as this study reached completion in 2011.

Assessment of ESOL students in the UK is an area which is underrepresented in the literature. I am particularly interested in the effect that the 2001 Skills for Life strategy, which aimed at enhancing adult basic skills education, had on those assessment practices. In this study I draw on data collected at three local Further Education (FE)¹ colleges to investigate the practices in assessment of ESOL students' English language abilities.

1.2 Assessment practices in ESOL in the UK

There exists a rich body of research on foreign language ability assessment, and due to the predominance of English teaching and learning throughout the world, much of the research concerns English assessment, but it primarily concerns the EFL (English as a Foreign Language) situation which is the term generally used in the UK (but not necessarily in other English-speaking countries) for teaching English to learners who are not ordinarily using English in an English-speaking environment. These studies, due to the global role of English, are situated worldwide. There are subtle but significant differences between the two types of students: ESOL and EFL, for example being potentially surrounded by English in their lives, its everyday application and their need and opportunity to communicate with English native speakers are factors which affect motivation and choice of class content. (For further discussion of relevant differences in terms of ESOL and EFL (see Section 2.3, of Background to ESOL).

On the introduction of the Skills For Life strategy, when ESOL was grouped with Adult Literacy and Numeracy for funding and development purposes (jointly referred to as ALLN: Adult Literacy, Language and Numeracy), the injection of funding

¹ Further education is post-compulsory education at pre-degree level, which may include (the opportunity to take) qualifications also available at the level of compulsory schooling. It is distinguished from Higher Education which is usually viewed as education leading to at least a bachelor's degree or equivalent. Source: <http://www.qualityresearchinternational.com/glossary/index.htm#h>

resulted in a proliferation of publications. There is a wealth of work, for example, on adult literacy in the UK, largely thanks to the New Literacy Studies practitioners (Barton et al 2000; Baynham 1995; Hamilton & Merrifield 2000; Papen 2005; Street 1993; Tusting 2005). However, generally the ESOL aspect of the trio has been under-represented, also noted by Ward (2007) even though the numbers of students involved are significant.

In searching the literature on assessment practices in ESOL programmes at the beginning of my study it became clear that there are a great deal of studies concerning the situation in the USA, in Canada, in Australia and a certain amount on the New Zealand context but a distinct paucity of studies relating to the UK situation. The Barton and Pitt (2003) ESOL bibliography, one of the surveys of literature on ALLN commissioned by NRDC (National Research and Development Centre), confirmed this finding. In subsequent years some further works were produced (e.g. Derrick et al 2007; Lavender et al 2004; Schellekens 2009, Schellekens 2011), but it remains an under-researched field.

I was curious as to the imbalance of research on UK ESOL assessment at that time considering the fact that the sector was experiencing some potentially very interesting times as regards assessment practices. For example, the increase in funding resulting from the new strategy, which had not been seen on such a scale previously in ESOL, potentially allowed significant changes in practice in teaching and learning, and thus assessment, and also the market for new ESOL-tailored examinations opened up because of the requirements for formalised proof of student achievement entailed in the new system. (See Section 3.4, in Background to Skills for Life, for further explanation of this).

1.3 Impact and washback

In addition to curiosity about this lack of information, having worked already in the area of exam impact², my ‘research antennae’ were tuned towards what the effects of the new exam regime might be in the ESOL classroom. This phenomenon, the effect of an exam (or other form of assessment) on prior teaching and learning, is commonly known as washback (or backwash)³.

Introducing change into a system is rarely smooth and fully successful. Change in educational systems is particularly fraught with difficulties due to the interplay of various groups of actors, usually known as stakeholders, within that system: the students, the teachers, the management and the instigators of the change which may or may not be one of these groups but is often a government level (local or national) group, removed from the reality of the day to day effects of the change. Educational innovation literature (Henrichsen 1989; Waters 2001; Rogers 1995; Wall 1999; Fullan 1982; Markee 1993; to name but very few) aims to analyse the chief relevant components contributing to success or failure. An understanding of the nature of such change is helpful to make sense of any washback identified when the innovation concerns educational change in the form of a new exam.

1.4 The study location

In this section I will mention some of the work to date undertaken into washback of language exams, to provide the context for my research interests regarding location, firstly in terms of geographical location and then regarding institutional type.

² On an IELTS impact study (see Hawkey 2006) and a TOEFL iBT impact study (Wall & Horak 2006; Wall & Horak 2008; Wall & Horak 2011).

³ The term washback is more commonly used in the British applied linguistics community but is synonymous with backwash.

The vast majority of investigations into washback have been undertaken in either the North American context (e.g. Gordon & Reese 1997; Herman & Golan 1993; Loofbourrow 1992; Paris et al 1991; Smith 1991; Stecher et al 2004) or the Asian context (e.g. Andrews 1994a; Andrews et al, 2002; Cheah 1998; Chen 2002; Cheng 1998; Lam 1994; Lii-shih 1991; Qi 2004; Robb & Jay 1999; Roberts 2002; Shih 2007). A variety of studies across Europe have also been carried out, for example in the Netherlands (Wesdorp 1983), Ireland (Kellaghan et al, 1982), Romania (Gosa 2004), Italy, (Hawkey 2006), Turkey (Hughes 1988), Finland (Huhta et al 2006), in multi-locations across Europe (Wall & Horak 2006; Wall & Horak 2007; Wall & Horak 2008; Wall & Horak 2011) and also further east, in Israel (Shohamy et al 1996; Ferman 2004). Scott's work (2005) in the field of English as an Additional Language (EAL) as it is referred to in the UK (namely language specific study support offered to school age pupils with languages other than English as their L1) is one of the few studies focussing on the UK situation, however, the exams involved were not L2 language exams but the National Curriculum assessments which all school children were then taking. My study aims to contribute to the field of washback of EFL/ ESOL language exams, where very few of the studies have been situated in the UK.

What is more, the vast majority of the studies have investigated the washback of exams in, typically, two types of situation. The first is the study of the effect of exams taken within the state school systems, i.e. the National Curriculum assessments in the case of Scott (2005) already mentioned, or Kellaghan et al (1982), Shohamy et al (1996) and Cheng (1997; 1998). These concern students of school age who are completing their compulsory education. The second main type of washback study concerns investigations into the effects of major international English language exams, for which preparation classes are generally held in private language schools. These studies typically concern the Cambridge exams designed for non-native

speakers of English such as IELTS⁴ (see Green 2003; Hayes & Read, 2004), FCE⁵ (Tzagari 2006) but also the Educational Testing Systems (ETS) exams e.g. TOEFL⁶ (Alderson & Hamp-Lyons 1996; Blewchamp 1994; Roberts, 2002; Wall & Horak 2006, 2007, 2008, 2011). These exams are not compulsory, as they are not being taken within a state education system, but the exams are usually taken for reasons of future career or study plans or indeed for fulfilling immigration requirements in some cases.

The situation obtaining in my study however does not lie with either of these two groups of exam types. While the Further Education (FE) sector falls within the description of state education, attendance is not compulsory as it caters for students above the age of 16 in the UK⁷. The ESOL classes offered at FE provision are thus not compulsory but, rather, are taken, as already mentioned, to enhance the students' lives in the UK. The only other studies I found at the time of beginning my research which focussed on washback in the context of ESOL, specifically adult migrants, were by Burrows (1998, 2001, 2004). This was located in the Australian context and investigated a newly introduced system of classroom-based teacher-led assessments, and so differed from the Skills for Life approach which instigated a centralised exam system, (although some classroom based assessment was still permissible). There proved to be no studies of a similar ESOL situation with which to compare.

1.5 High stakes testing

For the UK ESOL students already referred to, their reasons for taking classes varies enormously from socialising, to confidence boosting, to securing employment or securing continuance in an already established career e.g. as a doctor. They are a

⁴ International English Language Testing System

⁵ First Certificate in English

⁶ Test of English as a Foreign Language

⁷ Compulsory education ends at 16 years old in the UK

mixed group with mixed aims. Their need for qualifications varies as much as they do and the gravity of the consequences of gaining these qualifications, or not, is commonly referred to as the stakes involved.

I have been involved in the machinery of high-stakes testing having acted as an examiner for IELTS in situations where candidates needed a certain grade for immigration purposes or to travel abroad for studies. Both are life- changing decisions and therefore can be described as high-stakes. High-stakes is a term often used in the testing literature but deserves some investigation. I suggest it is more complex than is often assumed to be, not necessarily fixed but liable to variation during the exam preparation period, and not of equal measure for all candidates.

High stakes testing is not a new phenomenon. An example of a high stakes test is related in the Bible in the Book of Judges, and it recounts the fate of one of two rival tribes who were distinguished by their enemy via a fatal pronunciation test (McNamara 2000: 68). This account relates how the Gileadites wishing to wipe out their enemy, the Ephraimites, after a particular battle set up a block on a strategic crossing of the River Jordon which the Ephraimites, who were fleeing back to their homelands, would need to use. Everyone crossing there was ordered to pronounce the word 'shiboleth' (which is usually understood to mean 'grain of corn') and since the Ephraimites could not pronounce 'sh' (/ʃ /) as the Gileadites did (cf. 'th' in English and 'ř' in Czech, which cause difficulties for most non-native speakers), were immediately recognised and slaughtered. This could be said to be the ultimate high-stakes test (for many further examples of such use of language as a test in critical situations see McNamara & Roever (2006, chapter 6).

Another illustration of a language learning situation involving stakes of the highest extreme can be found in an account, which though fictional, is highly acclaimed for its

research-base, namely the account of the Shogun period in Japan by Clavell. The account is as follows:

‘He – Lord Yabu told them [the villagers] you [a captured Englishman] are his honoured guest here. That you are also Lord Taranaga’s very honoured vas-retainer. That you are here mostly to learn our tongue. That he has given the village the honour and responsibility of teaching you. The village is responsible, Anjin-san. Everyone here is to help you. He [Lord Yabu] told them [the villagers] that if you have not learned satisfactorily within six months, the village will be burnt, but before that every man, woman, and child will be crucified’ (Clavell 1975: 506).

These two arresting examples where the stakes are as high as they can be, may be dismissed as fictional but yearly press reports on the phenomenon of the effects of high stakes exams proves its modern day very real continuance. Suicides as a result of pressure on students in Asian cultures, notably Japan, particularly those studying for university entrance exams, but also affecting even younger students, is used as an example of the unfortunate consequences of the pressure they feel under to succeed, since university places are highly competitive and the societal pressures to succeed are intense (Locastro 1990; Simmons 1988; Zeng & Le Tendre 1988).

High stakes situations are, thankfully, not usually so extreme; they may however be life-changing. The outcome of an exam may mean access to future work. In some societies such as in the developing world the every-day scrabble for economic survival has very real consequences not only for the candidate but potentially also their family, and thus qualifications equate with hope. Nevertheless, probably the most commonly encountered high stakes exams around the world come in the form of university or school entrance exams (or other equally prestigious training opportunities or

promotion). The stakes represent the chance to enhance one's social capital and life chances in general, even if a negative outcome is not as drastic as those cited above.

In most studies concerning high stakes exams, the stakes are usually described in terms of the stakes for the candidate (e.g. Green 2003). Sometimes the stakes for the teachers are also discussed (e.g. Cheng 2004). The issue of the nature of the stakes for other stakeholders, beyond the candidates and teachers, is rarely discussed and was therefore worthy of investigation.

Stakes are generally described in terms of the purpose for taking the test e.g. in England in the days of the 11+ exam, access to a grammar school and the effect of exam pass rates. For example, in the UK concerning National Curriculum assessments again, the pass rates were used to form informal, though highly influential school league tables (Black 1998; Salmi & Saroyan 2007). In addition, in many studies the candidate group is relatively homogenous (e.g. regarding age, length of education in a certain culture, access to teaching of a particular type and within a relatively predictable quality range) and will be taking the exams for similar reasons, and thus the stakes will be similar. The situation regarding the stakes of the exams in the UK ESOL teaching situation is, in contrast, somewhat different, the students generally being noticeably heterogenous in terms of education, goals, and motivations, amongst other aspects. This is discussed in further detail in Section 2.3, in Background to ESOL.

1.6 The research questions

Having considered the issues which formed the basis of this piece of research, I will now outline the research questions which formed the backbone of the main study. They divide into three sets of questions.

First of all, since, as with many educational reforms, there was not a completely clean sweep of old practices and a comprehensive introduction of a completely new set of practices, it needed to be established what the assessment practices in ESOL were prior to examining their nature and effects further. This led to the first set of questions:

RQ 1.a) What is the range, nature and function of assessment practices in UK ESOL teaching?

RQ 1.b) How are these practices linked to the Skills for Life strategy?

Secondly, while washback has become a much more widely discussed and understood topic in recent years in the field of testing, particularly since Alderson and Wall's seminal study (1993), there is a danger of familiarity with the issues leading to assumptions about the influence of exams, especially if of high-stakes. It is important to investigate whether there really is evidence of any washback, how strong that evidence is and how strong the 'evidential link' is, (Messick, 1996) i.e. the proof that there is indeed a causal relationship between the exam and certain behaviours.

Therefore the second set of research questions consist of:

RQ2.a) To what extent is there evidence of washback from the assessment practices?

RQ2.b) Is any washback related only to the assessment practices resulting from Skills for Life?

Lastly, in the literature, various aspects of washback are discussed and it is generally assumed the stakes of the exam are what cause the washback (e.g. Alderson & Wall 1993). In other words the consequences of assessment results are such that the assessment influences teaching and learning prior to the exam. As well as further investigating the nature of stakes, namely whether the stakes are differential (affect individuals differently) or not, I believe it is worth trying to investigate what other types of factors may influence whether washback occurs or not, and how these factors interact with the stakes. In short this research question is encapsulated as:

RQ3) What are the factors which may drive washback?

Having outlined the research question this chapter now sets out the chosen methodology.

1.7 The methodological approach

The approach of this study is very much from the perspective of the social responsibility of formal assessment, and it is not a study of the nature and psychometric properties of the various exams and assessments used. Because the study aims to investigate the nature of the inter-relations between various stakeholders, their goals and beliefs and purposes for their actions, a qualitative approach has been taken. Further details of the rationale for this and the methods used can be found in Chapters 5: Methodology and 6: Methods, respectively.

1.8 Study aims

This study thus aims to redress the imbalance in the amount of research into assessment practices in ESOL within the UK, with a focus on washback and the role of stakes, specifically, how they operate in this population. I hope that by researching

an under-studied educational location and investigating the nature of the possible driving mechanisms behind washback, some insights into the nature of washback may be gained, as well as contributing to our understanding of the effects of the Skills for Life strategy.

1.9 Study outline

I have outlined here the starting position of my study in the form of the research questions around which the study took shape, and where they originated. Before detailing the methodology, as mentioned above, I will discuss three key topics of significance for this study: ESOL in the UK, The Skills for Life strategy and, finally, washback in the field of language learning. As I proceed to discuss what the findings revealed, I will detail how (and why) the study developed through the process of analysis and interpretation. The study draws to a close with the overall conclusions I reached, followed by a critique of this piece of research. This work is a description of my research journey.

2 BACKGROUND TO ESOL

2.1 Chapter overview

This chapter is concerned with providing the background to the specific field of education which this study focuses on, namely teaching English to students residing in the UK who are non-native speakers of English. It outlines the main types of students who fall into this category, sets out the key factors which affect this field currently, as well as making a brief comparison with the situation in other English-speaking countries to highlight options and trends and finally provides some historical context to make sense of the recent reforms.

2.2 What is ESOL?

In the UK a distinction has traditionally been drawn between two sorts of English language classes for non-native speakers of English. Firstly, there are those where English is taught to those who do not usually live within an English-speaking environment, whether studying in their homeland or temporarily here in Britain, namely English as a Foreign Language (EFL) classes. English as a second language (ESL), in contrast, has been used to refer to those learners of English who have come to live in an English speaking country, or where English is the lingua franca (Richards et al 1992: 269). ESL, in the UK at least has largely been superseded by the term English for Speakers of Other Languages (ESOL) partly in recognition of the fact that students at such classes often already possess proficiency in two or more languages due to the linguistic diversity of their native environment; English for them is frequently not their second language, but third or fourth, if not more. It is not clear when the term ESOL took precedence. It has been in operation to a certain extent, according to Schellekens (2007), since the 1960s, used as an umbrella term for all

types of English language learning, both ESOL and EFL. Cambridge ESOL, for example, became the new name for UCLES⁸ (the language exam board associated with Cambridge University), and they produce suites of exams for EFL candidates (e.g. PET, KET, FCE, CAE etc) as well as a suite of exams for 'ESL' candidates (e.g. the Skills for Life suite). The change in terminology can be confusing when consulting the literature on the subject since it is not immediately clear what type of student is being discussed. While there are various topics for which differentiation is not relevant, such as basic good teaching principles, in other cases their situations are quite diverse, for instance learning environment and use of English. In addition, their learning goals and their needs might be quite different.

The same terminology is not necessarily used in other English-speaking countries; for example ESL is generally used as an umbrella term for all teaching of English to non-native speakers in the USA, Canada and New Zealand, whether referring to migrants or short-term visitors. The academic literature in this field can be confusing in its usage due to recent changes.

Another term which on first viewing would seem to be synonymous with ESOL is EAL (English as an Additional Language) but this has come to be used in the UK for English language support for school-age students within the state school system (as described in Scott's (2005) work on the washback of SATs, amongst others). This is another area which has its own particular practices, methodological trends, policy changes and difficulties, as recognised by the existence of its own specific professional body: NALDIC⁹.

The primary subject of this study however is exclusively the students who enter the system of post-compulsory education in the UK, namely the world of Further

⁸ University of Cambridge Language Exams Syndicate

⁹ National Association for Language Development in the Curriculum.

Education, and thus the study is located exclusively in the world of ESOL, not EAL. In my study, in addition to ESOL, the terms EFL and ESL will be used where the distinctions between student types, as outlined above, need to be referred to.

So why is this distinction important? We can view EFL from the perspective of an ‘added value’ approach in that most students are adding English to a range of other skills to enhance themselves personally and financially as well as boosting their social capital. On the other hand, ESL is typically viewed in terms of a deficit model, i.e. these students lack the skills needed to function fully in their new homeland. (See Cook (1999) for discussion of traditional EFL discourse concerning L2 learners, revolving around what they cannot do in comparison with the native speaker.) For these individuals, learning English may be a matter of necessity rather than choice. What is more ESOL provision is often grouped with Adult Literacy and Numeracy in a Basic Skills Unit (or similar) and this again supports the deficit model since ‘[a] qualification carrying a title of basic skills labels the user as having low basic skills rather than as having improved them’ (Hamilton & Hillier 2009: 135). At school and university level in the UK learning an additional language is not viewed as a basic skill in any way, so the reason for this grouping does not seem rational, and again reinforces the deficit view, which can hardly be justified.

Another way of viewing the two groupings is in terms of income generation and expenditure for the UK. EFL and the associated activities of teaching materials production, as well as the exam business, is a multi-million pound industry (not only in the UK but elsewhere in the world also). On the other hand, ESL students incur (financial) costs to the UK due to provision of language classes. These points of view are put in simplistic terms to highlight some key distinctions.

2.3 The nature of ESOL students in the UK

In the UK a wide range of students attend classes which come under the broad label of ESOL. These students can be roughly categorised according to their circumstances, which highlights the variety of students and their motivations and needs in an ESOL class.

One group consists of the migrants who intend to be only short-term residents in the UK. This group could include, for example, au pairs, who fully intend to return to their home and their very reason for being in the UK may indeed solely be to enhance their English language skills. These would previously mostly have been categorised as EFL students. However, other students are in the UK purely to pursue employment opportunities, especially due to the possibilities resulting from recent enlargement of the European Union (EU) which allows EU citizens access to the UK workplace and equally many of this group, but not all necessarily, intend to return to their native countries at some point.

There are also the long-term migrants, sometimes known as the 'settled' migrants, who have come to the UK to make it their home. Even within this group we have the whole gamut of types of learners: those who need better English and perhaps also qualifications to prove their language ability in order to gain better positions in the workplace and at the other end there are those, very typically spouses of migrant workers, who have little actual need of English due to the nature of their home environment but for whom the classes may provide a social function. There are learners who have already lived in the UK for many years and those who have just arrived, those who are highly literate in their heritage language and others with few or no literacy skills (KPMG 2005; Rosenberg 2009; Schellekens 2004; Schellekens 2007). For example, Baynham et al (2007) found 12% of 500 ESOL students surveyed reported lack of reading and writing skills in both their L1 and English.

Refugees and asylum seekers constitute another group of students who are accessing assistance to help them settle in a new homeland and are in the UK by force of, often quite traumatic, circumstances. In lay terms they are often seen as synonymous but an asylum seeker is a refugee who is awaiting official permission to remain in the country.

A final, probably smaller, group are the spouses of those in the UK for the short-term, such as for higher studies. Their reasons for learning English may align with any of the above groups, and their own language background and skills may be equally varied.

These categories help clarify who may be attending ESOL classes, but it must not cloud the fact that these groups are fluid; an au-pair from an EU country who intended to come to the UK only for a short-time may decide to stay for good. A refugee is not only a refugee but also a migrant, having to find employment and establish a life-style in a foreign environment. A spouse within the 'settled' migrant community who has little access to, and need for, English may find circumstances change once children have grown up and left, and with them support mechanisms, so learning English becomes more of a survival strategy than a social occasion.

Schellekens (2004) discusses the EFL versus ESOL distinction in other terms. In her terms, key differences between the two student types are: 1) EFL students can go home, whereas ESL students are resident in the UK for various reasons. 2) In business terms, who is the client? In EFL the student is clearly the client. For ESOL, in contrast, the government is the client. This difference has a profound effect on how classes are managed. 3) ESOL classes encompass more than just language; they involve much pastoral care, due to the nature of the students and their personal

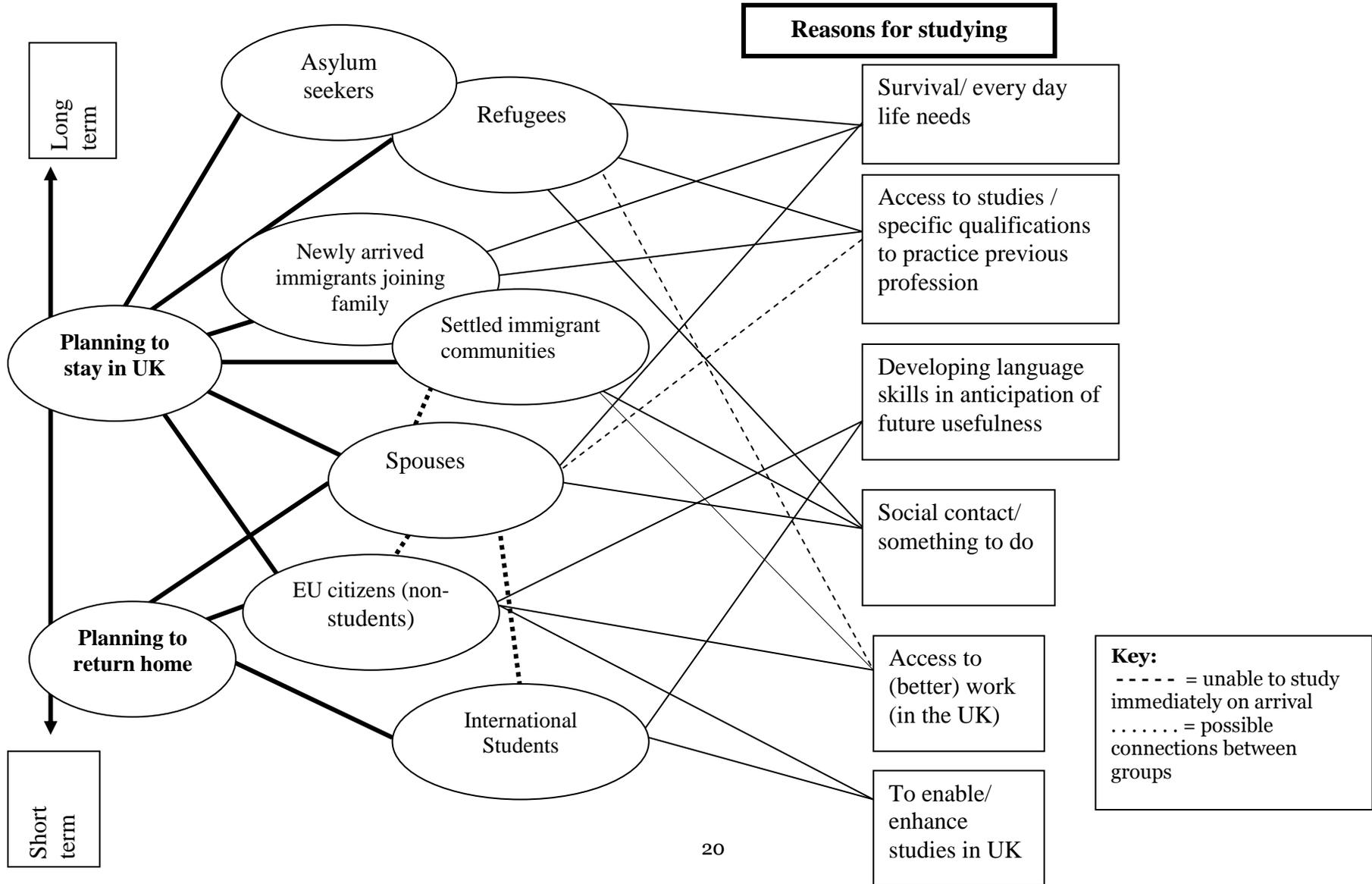
circumstances. Despite these differences, she poses the question of whether the two groups are increasingly merging. I would suggest that with reference to points 2) and 3), finance and class content, they do still remain distinct.

In terms of considering effective teaching practice in ESOL classes this categorisation is useful for evaluating plausible student motivation and needs. As Pitt (2005) points out, early SLA research focussed on cognitive ability but recent moves in the field better recognise the role of the social dimension of SLA. Fig. 1 (below), sets out the web of varying permutations of student types and reasons for learning, thus highlighting the complex nature of the ESOL student motivation and experience, the role of which to language acquisition has been highlighted by the work of Block (2007) and Norton (2000).

Breaking the Language Barriers (Moser 2000) categorised the main groups of ESOL learners as: settled communities, refugee and asylums seekers, migrant workers, and partners and spouses of learners. This describes the profiles prior to the Skills for Life funding changes which altered their nature. This will be discussed in later chapters.

The predominant learner type in an ESOL class will change from location to location around the country depending on local circumstances such as presence of, for example, a strong community of a certain minority ethnic group (e.g. Italians in Worthing, Yemenis in Cardiff) or institutions where asylum seekers are housed, or a university where visiting academic staff and their spouses might need classes. This thus reflects a wide variety of needs: in course type, in course length, in accommodation of student personal circumstances, such as women-only classes for certain groups of Moslem women or timetables which can accommodate students' work commitments.

Figure 1 The potential reasons for ESOL students being in the UK and potential reasons for attending English classes



While the ESOL teaching community is aware of this diversity there is little data on actual student numbers, or characteristics. Schellekens reports that from the 2001 census 6.2% of the UK population was born in countries where English is not the national language (2004). As she says, this does not provide any detail of their language ability needs. The data is not available since language need is not systematically recorded in data collection concerned with migration (Ward, 2007). Estimates of ESOL students on courses for 2001 to 2003 were provided by KPMG as part of their review of ESOL by citing enrolment figures, but as Schellekens has pointed out (2004) one enrolment does not represent one student necessarily since they can move around and also be enrolled on more than one course at once.

2.4 Where is ESOL?

The majority of ESOL provision is from Colleges of Further Education or Adult Education Colleges but other providers do exist. The Armed Services and Prisons, for instance, offer ESOL classes, as do some providers in the private, the voluntary and charity sectors (Baynham et al 2007; Schellekens 2007). These are in the minority however. In 2005 it was reported only 16.7% of ESOL delivery was through franchise or partnership arrangements (KPMG 2005).

The FE sector covers such a wide range of subjects and qualifications that it is useful to see within which niche ESOL lies. Adult Basic Education (ABE) can be viewed as the area which deals with ensuring that the adult population of the UK has a sufficient level of skills in the use of English and numeracy to allow them to fulfil their potential in the workplace. Within colleges ESOL is often, administratively at least, if not physically, located alongside literacy and numeracy provision in one Basic Skills unit rather than being located alongside modern foreign language teaching, with which it

can be argued to have more in common. ESOL, as mentioned, is one of the three partners of ABE, alongside adult literacy and numeracy.

The ESOL sector, however, appears to have experienced a pattern of being treated differently from literacy and numeracy for Adults within ABE and one reason for this might be that originally, in England and Wales at least, ESOL was developed under the Home Office Department of Internal Affairs as a response to immigration (Hamilton & Merrifield 2000). 'Unlike ABE, which was addressed as an educational problem, ESOL was treated as a social problem resulting from immigration' (Hamilton & Merrifield 2000 np).

Another reason why ESOL may not have been paid the attention that literacy and numeracy initially received, was that it was not included in the 'Right To Read' Campaign launched in 1973, the first adult literacy campaign in Western Europe (Hamilton & Merrifield 2000). The needs in adult literacy were in the public eye far earlier and understandably since the ESOL teaching needs lie within a minority of the population. Another interpretation may be that elements of the indigenous population (though that is a spurious concept in the UK being a nation consisting primarily of waves of migrants, albeit from pre-history onwards) may not react well to news of the resources spent on new arrivals, as evidenced by news stories about the common myth of social housing being allocated to 'newcomers' before locals in need and other discriminatory treatment of foreign newcomers to the UK. The role of the British press in shaping views towards newcomers to the UK has been highlighted in the work of Baker et al, (2008) and Greenslade (2005, cited in Wade 2007) who show how the dominant discourse generally prejudices against refugees and asylum seekers in much popular press.

2.5 ESOL teaching in other English speaking countries

Having discussed the UK ESOL position, in the following section, in order to set the UK situation within the context of immigration and refugee language provision in other English speaking countries, I will briefly describe four examples and thereby illuminate points of commonality and discrepancy, with which I aim to illustrate further the nature of how UK ESOL is currently situated.

2.5.1 New Zealand

The position of ESOL in New Zealand (usually referred to there as ESL) has been markedly different to that in the UK until recently as regards language ability of migrants. Due to immigration policy, migrants have for many years needed to demonstrate proof of language qualifications before entry. Any subsequent language classes which they choose to take are their own financial responsibility (though some help is available for refugees (Watts 2001). The purpose of this decision on language requirements is quite blatant:

‘a lack of language skills can impose [a cost] on New Zealand’ (New Zealand Immigration Service 1995:10, cited in Watts 2001).

Since 1995 International English Language Testing System (IELTS) has been used as the pre-entry test of proof of language ability for those entering under the General Skills and Business Investor category (in addition to paying a bond, which in effect paid for the language course needs of any dependents) (Watts 2001). The required scores were raised in 2002, which was criticised by education and business groups who felt that desirable immigrants would thus be deterred from entering (de Lotbiniere 2002). As Watts says:

‘English language has been, and still is, a gate-keeping device that discriminates against people from non-traditional (i.e. non-Western) immigrant countries’ (2001:1).

Studies by Massey University as part of the New Settlers Programme (Trlin et al 1998) were commissioned to identify the gaps and needs in ESOL provision,

‘since there has been a “growing realisation [] at the official level that the ‘hands-off’ approach to the ESOL development of new settler is unsatisfactory” and “that there is a need to develop a strategy to cater for their English learning needs in ways that will assist them to participate more fully in the social, cultural and economic life of the country”’ (Watts 2001: 2).

The role of English skills in the success of immigrants is thus, as in the UK, clearly acknowledged.

Several findings of the Massey University studies, of relevance to the transitional UK situation, are that the ESOL providers (i.e. colleges) felt that there was general support for the national curriculum and the assessment guidelines as well as the quality control and accountability mechanisms. Secondly,

‘a major concern expressed in the questionnaires and interviews [of the Massey University studies] was the lack of cooperation between institutions, imposed market-driven ideologies had led to increased competition between providers resulting in a wastage of time and resources’ (Watts 2001:9).

In addition, the providers felt ‘there should be more consistency in assessment across the different institutions’ (Watts 2001). The latter are issues which needed to be explored in the UK context since, as with so many areas related to ESOL, we had little documented information on this.

2.5.2 Canada

As in New Zealand and also Australia (see below), in Canada ‘there are strong links between immigration policy and economic labor policy’ (Van Duzer & Berdan 2000). Prior to 1996 there existed no standardisation of levels for assessing the English language skills of immigrants in order to assign them to language programmes (Pierce & Stewart 1997). Consultations of experts discovered that no single tool or set of benchmarks was being used and that those used were not always appropriate to the needs (CCLB 1981). Nowadays however, benchmarks are in use, which were produced by the Center for Canadian Language Benchmarks (CCLB), their aim being to try ‘to promote coherence, effectiveness and consistency’ (CCLB 2000: np) of adult ESOL provision (referred to as ESL here too) across Canada by linking all stakeholders, maintaining standards in training, and facilitating the use and implementation of the Canadian Language Benchmarks.

These benchmarks were produced in collaboration with ESOL learners, teachers, administrators, immigrant service providers as well as government representatives (Van Duzer & Berdan 2000). The whole project from which the benchmarks emerged was funded by the Government of Canada as a move to enhance language training which began in 1992. The result was:

‘[i]n small, but increasing numbers, immigrants are able to demonstrate to employers, using the Canadian Language Benchmarks, that they have the language skills needed for available jobs and to demonstrate to registrars that they have the language skills needed to succeed on non-ESL/EFL courses’ (CCLB 2000: v).

The Canadian situation differs from the UK in that these Benchmarks do not claim to provide a curriculum guide or ‘dictate local curricula and syllabuses’ and that they do not provide ‘descriptions of discrete elements of knowledge and skills that underlie

communicative proficiency' (CCLB 2000: np). The UK has provided exactly the opposite: a common core curriculum consisting of the 'packages' of language which should be 'mastered'.

As regards assessments, the options for how a benchmark standard is to be assessed and/or reported are:

- 'A score on an externally developed task-based proficiency assessment test or achievement test
- A rubric that describes various levels of knowledge and skills and usually provides more specific information than the test score
- An evaluation portfolio
- A variety of frequent evaluating techniques in the classroom, including checklists of outcomes and anecdotal records
- A combination of non-test evaluation techniques and an externally delivered test.

The external test may be applied selectively to a sample of learners in an ESL program, or to all learners in the program (CCLB 2000). From this it is clear that a variety of approaches are flexibly applied.

Students can work through 3 levels of proficiency, stages I to III: beginners, intermediate and advanced, each constituting four Benchmarks. The 'adequate mastery criterion' which proficiency is judged against is not that of the educated native speaker but one which has been 'pragmatically established by a sampling of performance of competent language users' (CCLB 2000: xi). 'Each benchmark describes a person's ability to use English to accomplish a task [], performance conditions, situational conditions [] and sample tasks' (Van Duzer & Berdan 2000).

The guidelines tell us that at the end of each section can be found information on monitoring, evaluation and reporting. It states that there is a suggested performance

criterion at the end of each of the three stages which are for use as points of reference to ‘monitor progress’. However it is stated *‘this is neither to be confused with nor used as an assessment test or evaluation test’* (my emphasis) (CCLB 2000: np). A focus on progress as well as product is a message which the guidelines explicitly project.

2.5.3 Australia

As elsewhere, immigration policy and labour policy are closely inter-related and traditionally Australia has welcomed immigrants. After World War II the main source of migrants was no longer the UK and Ireland, namely English-speaking countries, and therefore the AMEP, Adult Migrant Education (now renamed English) Programme, was formed, which it is claimed has become the largest government funded English language training programme in the world (Burns 2009). The AMEP programme from inception has been explicitly and closely associated with immigration policy as opposed to adult education in general (Burns 2009), unlike in other English speaking countries, such as the UK.

Having had one central curriculum, then later state-level devolved powers regarding the curriculum, there are now ‘a number of competency-based curriculum documents that are nationally or state accredited.’ (Van Duzer & Berdan 2000: 17). One of the most widely used is the Certificate in Spoken and Written English (CSWE). Competencies describe what learners can do at three proficiency stages (beginning, post-beginning and intermediate), which are based on Australian Second Language Proficiency Ratings (ASLPR).

The CSWE acts as a curriculum framework only and thus does not specify any programme outlines but it does specify criteria under which competencies have to be

assessed. The permissible techniques include teacher observation, interviews, role-plays, learner self-assessment, amongst others (Van Duzer & Berdan 2000). A certificate is issued only after Stage 3 had been completed. It can be seen that, as in Canada, a range of permissible techniques for evaluating progress are permissible.

According to Van Duzer & Berdan, on the introduction of these measures, teachers showed concerns about reliability and validity across programmes as well as the time needed to perform the assessments. However it seems that a positive outcome prevailed and they found that the measures enabled them to give more explicit feedback about learner progress and added clearer direction to their teaching. Issues of reliability and validity still plague use of such assessment measures, especially where they are increasingly used for accountability purposes (Van Duzer & Berdan 2000). This has been a major area of concern for Brindley (1998), Brindley & Hood (1994), Burrows (2001, 2004) and Burns (2009).

2.5.4 United States of America

Accountability is the watchword in ESOL assessment within the United States. All states need to include in their five-year plans details of how they establish levels of performance which the programmes must meet (Van Duzer and Berdan 2000). According to Van Duzer and Berdan (2000: 6) in nine states, standardised tests are used because of the ease of administration, minimal teacher training and that they purport to have construct validity and scoring reliability. The standardised tests widely used in the USA are BEST (Basic English Skills Test), EFF (Equipped for the Future) Assessment Framework, CASAS (Comprehensive Adult Student Assessment System) Assessment System, and REEP (Arlington Education and Employment Program) Writing Assessment (Van Duzer and Berdan 2000), which test different skills as some of their names suggest.

Reports are drawn up following guidelines from the National Reporting System (NRS) established by the Department of Education (ED). Programmes thus decide on a student's placement level and then any further gains with the aid of a standardised assessment procedure, approved by the ED (Van Duzer and Berdan 2000). As Stites reports:

'Accountability is why the adult literacy field [within which ESL sits] can't take its time with standards. While it is possible to have accountability without explicit content and performance standards, defining standards through a broad-based consensual process provides an opportunity for many voices to inform key decisions about who needs to be held accountable, how they should be held accountable, and for what. At a minimum, we need to have performance standards and test results to show how many learners are making enough progress to be counted as success stories. Of course, the usual success stories include more than test scores. That's fine as far as it goes, but in policy and funding circles these days, it doesn't go far enough' (1999:6).

Proof of progress, as in the UK, is the key to funding.

The Workforce Investment Act of 1998 'reflects a priority towards intensive, higher-quality services rather than rewarding the number of students served. It also puts a much greater emphasis on learner outcomes, and therefore on accurate measurement and reporting' (Balliro & Bickerton 1999, cited in Garner, 1999). Van Duzer & Berdan however point out that Adult ESOL providers in the USA have long since experienced problems with measurement and reporting the range of desired outcomes satisfactorily to the stakeholders (2000). Faced with this tension teachers have voiced their concerns (Van Duzer 2002; Wrigley 1992).

2.5.5 Overview

There are several points of similarity and contrast with the situations in these English speaking countries and the UK situation, regarding ESOL teaching. In Australia and the USA, as in the UK post-Skills for Life, literacy and ESOL have been aligned in terms of how they are considered at policy level. The USA most resembles the UK in terms of the development of assessments and the use of scores for accountability purposes. In Canada and Australia assessment has primarily remained classroom-based rather than centralized and standardized, as it is currently in the UK, although a great deal of work has been put into the development of frameworks to guide student development levels. New Zealand stands in contrast in that due to its stringent immigration policies the provision, in terms of its funding, purpose and thus student assessment, does not seem to need the measures put in place in the other four examples described here. It highlights how the role of policy, not just in education, affects the ESOL classroom.

2.6 Changes in ESOL provision

2.6.1 A history of volunteerism

Until recently ESOL in the UK could be said to have developed primarily in the spirit of volunteerism (Khanna et al 1998). Recognition of a need for some kind of enhanced provision was forced by the wave of immigrants to the UK from Idi Amin's Uganda, and as a result a literacy campaign began in the early 1970's. The ESOL strand was led by Ruth Hayman who launched the Neighbourhood English Classes (NEC) (Rosenberg 2007). The teaching teams consisted mostly of middle-class women: 'the liberal strand of the British social fabric' (Khanna et al 1998:10) who focussed more on the well-being of the immigrants than previous programmes provided by the government, which had concentrated solely on attempts at

assimilation, according to Khanna et al (1998). ESOL teaching was ‘characterized by dependence on volunteers, one-to-one tutoring, and ad hoc, often creative approaches to teaching and learning’ (Hamilton & Merrifield 2000:1). Rather than waiting for students to come forward to join up at teaching institutions, the volunteers sought them out, especially women students (Khanna et al 1998). There was thus at that time a lack of system, tutor support or professionalization in ESOL.

2.6.2 Funding

As Rosenberg reports, not until 1967 was central funding made available for ESOL (in England and Wales), under Section 11 of the Local Government Act of 1966. It was only available for certain groups, namely immigrants from the New Commonwealth, who had been in the UK less than a decade (2007). She points out that glitches in the system of funding soon became apparent, such as that funding could only be claimed where more than 2% of the constituency of a local authority (LA) qualified. The breadth of the funding remit did not change until 1993, extending to learners of English from elsewhere, despite significant changes in the immigrant profile (Rosenberg 2007: 92). LEAs could claim 50% of their costs to accommodate such learners for whom they had to make ‘special provision’ since their ‘language and customs differ from those of the rest of the community’ (Bagley 1992: 1, cited in Rosenberg 2007:90). The policy at that point in time aimed at maintaining separate cultural identities, supporting heritage language classes.

Further sources of funding subsequently came on stream courtesy of the Urban Programme in 1968, European Social Fund (once the UK joined the EU) from 1973, and employment–related courses funded by the Manpower Services Commission, although none of these were ESOL specific. Section 11 remained the only consistent source of resources for LAs until it was cut in 1998 (Rosenberg 2007). Provision was

thus uncertain until the injection of funding from the Skills for Life programme arrived in 2001. Its importance in terms of funding cannot be underestimated.

2.6.3 Changes in the culture of education

Hamilton and Merrifield report that in 1996 Alan Wells, the Director of the Basic Skills Agency, acknowledged that 'no real policy has emerged and that there are no universal opportunities for new citizens to learn English' (Wells 1996). However, since then, ESOL provision has moved from the 'patchwork of community programs [] with diverse funding streams to ever-greater [] accountability, documented performance and systematic standards' (Hamilton & Merrifield 2000:1).

It could be said that ABE was following, somewhat belatedly, the trend set by the move towards a national curriculum for schools instigated by the Education Reform Act of 1988. This was an innovation which standardised learning via a core curriculum and introduced a set of checks in the form of Key Stage National Curriculum Tests (commonly known as SATs - Standard Assessment Tests) designed to assess pupil achievement and now widely used to compare schools in unofficial but widely consulted league tables and also to inform funding decisions, as already mentioned. This was one of the profoundest changes to the British education system since the Education Acts of 1870. Although there has always been an element of accountability in any state education system, due to its inherent social function (Nicholls 1983), the 1988 Act catalysed a shift to centralisation, and standardisation and thus greatly increased accountability to a level not previously witnessed in UK state education. This shift to centralisation and standardisation was what Skills for Life brought to ESOL, and the next section will outline how ESOL finally was aligned with the rest of the state education system.

2.7 Time for reform

The move from a situation of varied and ad hoc funding to multi-million pound injection into Adult Basic Education was a considerable leap for ESOL. It is thus worth considering what happened to instigate this change. An important turning point when considering the provision for ESOL in the FE sector is often cited as being the Moser Report (1999), titled 'A Fresh Start'. The report highlighted that an unacceptable percentage of the unskilled or under-skilled labour force was potentially being wasted due to poor basic skills. (This report was the source of the much bandied about figure of 7 million adults having poor levels of literacy).

A subsequent report by Moser, commissioned by the DfEE, titled 'Breaking the Language Barrier' (2000) focussed specifically on ESOL issues. The now well-used figure of between half and one million people living in the UK with insufficient proficiency in English was first reported in this publication.

'All the evidence suggests that lack of fluency in English is a very significant factor in poverty and under-achievement in many minority ethnic communities, and a major barrier to employment and workplace opportunities and further and higher education' (Moser 2000: 1).

The report's main recommendations included setting up a long-term strategy to address the needs identified, namely a common curriculum, enhanced teacher training provision, national tests, a new inspection system, and a research programme to support the strategy. Surviving on the spirit of volunteerism was no longer acceptable. Improved professionalism was seen as the way to raise general standards in ESOL provision, or at least this was the explicit agenda.

2.8 Summary

This chapter covered the differences between what have been labelled EFL and ESL students, both coming under the umbrella of ESOL. The situation regarding ESOL teaching in certain other English-speaking countries was also described in order to locate the current situation in the UK in the wider context. This chapter also outlined developments in the field of ESOL teaching in the UK, and charted its increasing professionalism, organisation and increased input in funding terms as the attitudes towards L2 learners of English in the UK has changed over the years.

3 BACKGROUND TO THE SKILLS FOR LIFE STRATEGY

3.1 Chapter overview

This chapter discusses the reasons for the instigation of the Skills for Life strategy and sets out its five key components. It aims to highlight what a novelty this centralised approach was for ESOL. It also introduces some of the areas of criticism of the strategy to date focussing on those which later proved of relevance to this research, namely the genesis of the curriculum which is pivotal to the strands of the strategy. The emphasis is on the assessment strand of the strategy since this is the focus of this study.

3.2 Skills for Life

The Moser Report (1999), A Fresh Start, as already discussed, was the catalyst for action and in 2001 the Skills For Life strategy was born, to be implemented by the newly formed Adult Basic Skills Strategy Unit (ABSSU)¹⁰. It was handed the remit of overseeing development of five areas: curriculum, teacher training, testing, inspection and research, which the Moser report had recommended. Adult literacy and numeracy were tackled first. In fact ‘the mandate for the development of the ESOL curriculum came as an afterthought in the wake of the literacy curriculum’ according to Ade-ojo (2004:23) and indeed the Moser report had not paid much heed to the needs of learners of English in the UK. It was only due to lobbying indeed that ESOL was included (Cooke & Simpson 2008).

The goal of ABSSU was to produce the means to develop adult basic skills through a centralised approach. The first step for each of the three needs areas within ALLN

¹⁰ It must be noted that the strategy is only implemented in England and Northern Ireland. Wales has a different strategy, Words Talk – Numbers Count, in which both English and Welsh are target languages. (See www.elwa.ac.uk). Scotland which has historically differed educationally from the rest of the UK offers a different programme and set of qualifications for learners of English. (See www.sqa.org.uk/sqa).

was to produce a curriculum. These were structured according to levels set out by the National Qualifications Framework (a Qualifications and Curriculum Authority (QCA) initiative¹¹) and Basic Skills covers five levels within this: Entry One, Entry Two, Entry Three, Level One and Level Two (see Table 1)¹², although Entry 1-3 are seen as sub-levels. This is significant in terms of funding in that Entry level qualifications are not counted towards the public service agreement (PSA) targets for student progress as set by the Skills for Life strategy and implemented by LSCs (Merrifield 2006: 3). No funding is available for Level 3 since ESOL students move onto other mainstream courses at this level if they have not already done so.

Table 1 Basic Skills levels within the National Qualifications Framework (NQF) compared to other qualifications frameworks

NQF	Basic Skills (Adult Literacy, Numeracy & ESOL)	Key Skills/ NVQs	National Curriculum	GCSE	Languages Ladder	CEFR
Entry	Entry 1		Key Stage 1		Breakthrough 1-3	A1
	Entry 2		Key Stage 2			A2
	Entry 3					B1
Level 1	Level 1	Level 1	Key Stage 3	Grade G-D	Preliminary 4-6	B2
			Key Stage 4			
Level 2	Level 2	Level 2	Key Stage 4	Grade C- A*	Intermediate 7-9	C1

Key: NVQ National Vocational Qualifications

GCSE General Certificate of Secondary Education

CEFR Common European Framework of Reference

¹¹ According to QCA it is a framework which helps learners make informed decisions on the qualifications they need by comparing the levels of different qualifications and identifying clear progression routes to their chosen career (<http://www.qca.org.uk/493.html>).

¹² NB: The Framework has since been revised to cover 8 levels up to Doctorate level.

Sources:

- a) RaPAL Bulletin. No. 41 Spring (2000)
- b) DCSE (2003). The Languages Ladder – Steps for Success from University of Cambridge: www.cambridgeesol.org/sfl/levelsmt.htm

In the next section I will describe the development of the five areas which the Moser report recommended in order to describe the context of change which the strategy instigated.

3.3 The five strands of the Strategy

3.3.1 ESOL Core Curriculum

As a result of the Moser recommendations the Department for Education and Skills (DfES), through ABSSU, commissioned the Basic Skills Agency (BSA) to produce a new curriculum for ESOL (DfEE, 2001). The London Language and Literacy Unit (LLLU), who Ade-Ojo (2003: 22) claims, ‘are strongly linked to the sponsors’ (i.e. ABSSU) were responsible for producing the new curriculum with input from Prof. Ronald Carter of Nottingham University, as well as representatives from several colleges, and consultation with ESOL practitioners via the National Association for Teachers of English and Community Languages to Adults (NATECLA). Other stakeholders such as the Learning and Skills Development Agency (LSDA) and, of course, the Qualification and Curriculum Authority (QCA) as the chief accreditation body in England were also involved. However, Ade-ojo claims that from his study of the development of this curriculum only the stakeholders who had control over resources and responsibility for translating government policy were represented in the actual development, namely DfES and BSA (Ade-Ojo 2003: 22).

The structure and content of the curriculum is pivotal since it acts as the skeleton for the other strands of the strategy, namely, teacher training, teaching materials, quality assurance in the form of inspection, and assessment tools. The nature of the ESOL curriculum has for the reason of its importance been the cause of discussion (e.g. Schellekens 2004, Ad-Ojo 2003). The model of language it assumes is a direct result of 'its origins in a skills-based literacy curriculum' (Cooke & Simpson 2008: 8) and Fowler confirms its origins in the Literacy Curriculum (Fowler 2005). Language in the curriculum is deconstructed in terms of word, sentence, and text level. This is, as Simpson & Cooke (2008) point out, in contrast to a whole-text and genre-based approach which the Australian ESOL curriculum adopts, for example. As Ade-ojo notes:

'the DfES acknowledge the similarity between the structure of the curriculum and those of vocational courses. This, in their opinion, was because it was "designed to achieve specific ends" and was "developed on the basis of a skills audit"' (Ade-Ojo 2003: 25).

Ward (2007) questions whether the nature of language learning has been taken fully into consideration for the production and expected use of the ESOL curriculum. She says:

'ESOL straddles the curriculum area of Skills for Life and language teaching but does not align wholly with either. The linguistic theories and pedagogy of language teaching are almost entirely absent from Skills for Life, and the social and cultural and political elements influencing language acquisition and use are often missing from modern foreign language teaching' (Ward 2007:33).

Likewise, Roberts et al (2004) reported that a great deal of the learner language evident from their studies (in five case studies) 'cannot be neatly tied into curricular objectives' (p12).

It assumes a linear progression in terms of acquisition of specific structures through the levels, and takes no account of the iterative and cyclical nature of language learning or that as Breen says supporting what students and teachers will know, that acquiring another language is complex (Breen 2001). Neither is there acknowledgement of the social nature of learning (Barton & Papen 2005). Another aspect of note is that it takes Standard English exclusively as its model, as Cooke & Simpson (2008) highlight, without referring to the wide range of English variation students are likely to encounter in their everyday lives.

Although the curriculum, as stated, is central to the implementation of the Strategy, its use is not unproblematic. Some insecurity about the exact role of the curriculum, for example, was expressed by practitioners in the study by Davies (2005) who found uncertainty about whether the Core Curriculum should be used as a rigid set of guidelines or could be simply a looser guiding document on which to build. It must be noted that that study was the first stage in an impact study of Skills for Life and no distinction between Literacy, Numeracy or ESOL was made, so whether concerns were particularly about the ESOL curriculum is not clear.

Nevertheless, the NIACE ESOL enquiry (2006) found that overall the curriculum was generally welcomed by ESOL staff, in providing a comprehensive framework and standardised skills descriptors across ability levels, as noted in the New Zealand context. It was found to be useful for guiding novice teachers and serving as a reminder of consolidated knowledge for experienced teachers (NIACE 2006).

3.3.2 Teacher Development – qualifications and training

Another of the key recommendations of the Moser Report was to enhance ABE teacher training and as a result teaching qualifications became compulsory for new Further Education teachers, which had not been the case previously (Jones 2004).

Subsequently, the ‘Teacher Subject Specifications (Standards) for level 3 & 4’ were published in 2002 by the Further Education National Training Organisation (FENTO) (who are responsible for teacher qualifications at FE level) and the DfES¹³. They set out the knowledge and understanding to be included in qualifications for those teaching ESOL, including the personal language skills required for this work. Teacher training courses were made available at eleven universities and colleges around the country (Jones 2004) e.g. Diploma in Adult Basic Education: Literacy Numeracy and ESOL offered at Lancaster University, amongst many others. Hughes (2004) reports that the Government wanted all FE teachers (except new entrants) to be fully qualified or enrolled on a course by 2010 and aimed at 90% of full-time teachers and 60% of part-time teachers to be qualified by 2006. There was a considerable amount of confusion over the exact requirements, however, especially regarding which previously gained qualifications could be considered for Accreditation of Prior Learning (APL) (Hughes 2004).

3.3.3 Teaching Materials

Further to the previous discussion of how ESL and EFL in the UK differ, another way is that the majority of commercially produced materials are still aimed at the EFL

¹³ Department for Education and Science – the department then responsible for FE. The Department for Business, Innovation and Skills now covers FE.

market and ESOL teachers have to work at adapting materials or writing their own (Williams 2004), although more materials have certainly been published since Skills for Life began.

In 2003, however a set of teaching materials was produced and widely distributed by the DfES specifically for ESOL purposes. They 'were professionally produced and distributed free to all colleges who requested them' (Cooke & Simpson 2008: 55).

These materials consist of files containing learner materials, teacher notes and accompanying CDs of audio materials, all aligned to the ESOL curriculum. They are available in two sets, for Entry Levels and another for Level One and Two.

The language, style, and theoretical approach to language closely mirrored that of the core curriculum while their subject material 'reflected current concerns with multiculturalism, integration and social cohesion.' (Cooke & Simpson 2008: 55). The elements of survival English seen in material of previous decades was still prominent in terms of the language for typical daily situations covered, such as accessing health services, transport and day to day life whereas the anti-racism agenda of the 1980's has been replaced largely by workplace rights materials (Cooke & Simpson 2008).

Subsequently other more specific materials have been distributed. For example, one set has been produced for the teaching of Citizenship to ESOL learners (2004) in order to prepare students who wish to sit the test produced by the University for Industry (Ufi), introduced in October 2005, which is a requisite for application for UK citizenship for students at a level lower than Entry 3. The teaching materials are based on the 'The New and The Old', the report by the Life in the United Kingdom Advisory Group (2003).

Complementary materials are now also gradually being produced such as ‘Police ESOL’, again produced by the Basic Skills Agency (2005: 2), with the aim of ‘improving community safety and communications between the police and minority linguistic communities’. Other materials, for the higher level ESOL students have been produced by the Centre for British Teachers (CfBT) amongst others, which aim at provision where there is embedding of English into various vocational courses that are available at FE level, such as in mechanics, nursing or catering. Embedding English was part of the Skills for Life policy strategy to ensure students reaching an appropriate level of English return to, or join, the workforce.

The development of materials has generally been welcome (Ward 2007) as studies have shown (Roberts et al 2004) that materials aimed at EFL learners are not always relevant to the needs of ESOL learners, due to the work and life context they find themselves in. For example,

‘[L]earners themselves often navigate the course of learning and are continuously recontextualising the often bland and invented worlds of the course materials so that they can make themselves meaningful to their own lives’ (Roberts et al 2004: 12).

Appropriate materials are a key component of effective teaching.

3.3.4 Research

Shortly after the Moser 2000 report was produced, further ESOL specific research was published including ‘English Language as a Barrier to Employment, Training and Education’ (Schellekens 2001), which highlighted the gross lack of detailed information about ESOL students. This somewhat reinforced the Cinderella image of ESOL, being the overlooked party, never receiving the focus of attention and resources. This report focussed on the same main issue as that of the Moser reports,

namely that, in basic terms, improved ABE was needed for economic reasons. The potential UK workforce needed maximising by facilitating the entry of the under-skilled into the labour market. However, despite this political aim, little data had been collected on the exact numbers of students requiring ESOL classes, or the exact nature of their need, as already reported in the previous chapter.

This lack of knowledge began to be addressed by the creation of the National Research Development Centre (NRDC)¹⁴ whose remit was to produce research to inform policy. An early study by Barton & Pitt highlighted that little relevant research up to then had been undertaken (2003). As Roberts et al (2004) pointed out, in this respect, the UK lagged far behind Australia and the USA. The NRDC commissioned a variety of studies, which initially seemed to focus far more on adult literacy and numeracy, but this imbalance was later redressed to a certain extent (e.g. Roberts et al 2004, Baynham et al 2007). Such research is welcome since as Ward says, ESOL in the UK is, on the whole, insufficiently researched (2007).

3.3.5 Inspection

Quality assurance is carried out via inspections from two bodies, either the Office for Standards in Education (OFSTED) or the Adult Learning Inspectorate (ALI) depending on the nature of the institution. OFSTED has traditionally dealt with school inspection and covers provision for 16-19 year olds. ALI undertakes inspections where 19+ year old learners and work-based learning for 16+ year olds are catered for. Codes of practice have recently been put in place and are aligned to the standards as exemplified in the ESOL curriculum.

¹⁴ A consortium led by the Institute of Education, together with teams at Lancaster, Sheffield and Nottingham University.

Working to the new Common Inspection Framework (CIF) which was revised in April 2005 (Pearson 2005), teams are now supposed to specifically evaluate basic skills provision whereas previously this did not happen. The inspection system is of interest in that part of their remit is to find out how well learners achieve, this being one of the three main areas they are expected to report on¹⁵. Pearson reported that at that point (April 2005) 70% of LSC funded learners were not taking qualifications and it was therefore ‘very hard to judge their achievements’ (2005: 5).

3.3.6 Assessment, including the national exams

The final strand of the strategy to be drawn into alignment with the curriculum, and the focus of this research, is the range of assessment tools for teachers to use in order to evaluate their students. Initial assessment tools were available in pre-Skills for Life days from the Basic Skills Agency (BSA), designed to be used for placement and diagnostic purposes when learners first join a college, but Schellekens (2001) reports that these were generally viewed as not being appropriate for the special needs of ESOL learners. A new Initial Assessment instrument was produced by the BSA and made widely available to ESOL providers, along with the Skills for Life curriculum and teaching materials. There is no compulsion for it to be used and it has not been universally endorsed, as will be made clear in Chapter 7.

Achievement testing, on the other hand, had not at all been standardised and provided centrally, in the pre-Skills for Life period. Colleges were free to choose whichever assessment tools they wished to use. There were in general terms, four options. First (but not in terms of priority) were the exams designed for EFL students. These typically include Cambridge FCE, CAE, IELTS or Trinity exams. Secondly,

¹⁵ The others being a) Quality of Education and Training, and b) Leadership and Management

there were qualifications accredited by bodies such as the National Open Credit Network (NOCN) or Open College of the North West (OCNW) (now known as Ascentis), which were often portfolio-based assessments or thirdly, there were those developed by colleges themselves (such as the Criteria for Assessment of Language Skills produced by LLLU) (Rees & Sunderland, 1995). Lastly there were qualifications designed for native speakers of English who faced literacy problems (e.g. Wordpower).

Despite the range of assessments available not all language learning was accredited, however. For example, in 1997/8 FEFC¹⁶ reported that 40,000 learners were on courses without external accreditation (Schellekens 2001: 51). The importance of accreditation increased with the change of regime when Skills for Life was introduced, in that funding was linked to targets, attainment being assessed in terms of students gaining qualifications via the QCA accredited exams.¹⁷

An interim period existed between the launch of Skills for Life in 2001 and September 2005 in which new Skills for Life assessment tools were developed by the exam boards. From this date only those qualifications which had been granted accreditation as a consequence of having been demonstrably aligned to the ESOL core curriculum to an acceptable standard have been accepted by the LSC as fundable evidence of achievement. Colleges still retain their choice in terms of which of the boards' accredited exams or portfolio assessments they select. Ward reports, based on DfEE reports (1999 and 2000), that the rationale for culling and streamlining of the ESOL exams was the inconsistency in assessment standards and confusion due to the great number of exams available (2007).

¹⁶ Further Education Funding Council

¹⁷ A new performance framework was introduced in 2007. This was thus not in operation at the time of the data collection.

Whereas in typical EFL curricula and in most of the exams from the main exam boards the four main skills (readings, writing, speaking and listening) are described, taught and assessed as separate facets of overall language ability, the Skills for Life Entry Level exams have followed the model of the literacy (and ESOL) Curriculum, which it is to be remembered is designed for native speakers of English. In this, speaking and listening are combined in the curriculum and thus also in the exams. Ward notes there appears to be a lack of theoretical rationale for this (2007). Level One and Level Two exams, although now available for ESOL candidates specifically, were not available at the time of data collection, and ESOL students only sat the exams designed for literacy students. These consisted (and still consist) of a test of reading and indirect test of writing. Although speaking and listening feature in the Adult Literacy Curriculum, literacy students were not examined in these skills. At that point, one single exam board, City & Guilds, provided the literacy exams, whereas a variety of exam boards provided the accredited ESOL exams.

Parity of ESOL students with other students is reduced since native speakers (literacy students) have to accomplish far less to achieve a Level One or Level Two qualification than an ESOL student (Ward 2007) since ESOL students are examined in four skills, while literacy students are examined in only reading directly, and writing indirectly. Provision for these learners is more costly in covering four skills, rather than one (reading) in their tuition and thus it is in providers' interests to direct such students to literacy classes rather than ESOL classes (Ward 2007).

A market for Entry level ESOL specific exams aligned to the Skills for Life ESOL curriculum thus opened up and these exams were made available from September 2004. From January 2005 only exams from accredited exam boards have counted towards PSA targets (Mallows 2009). Table 2 lists the exams, and the modules

available, at the time of data collection, and highlights the range available to colleges and other ESOL providers.

Table 2 Exam boards offering S4L exams accredited by QCA at various levels

Board	modules available				Assessment Format
	Combined	Speaking & Listening	Reading	Writing	
City & Guilds ¹⁸	All	All			Assignment. Externally set and marked. On demand. Spk/Lst and Wrt by portfolio
University of Cambridge ESOL Examinations	All	All	E3, L1, L2	E3, L1, L2	Exam. Externally assessed. On demand.
Edexcel	All	Entry			Exam
EDI (LCCI)	All	All	All	All	Exam
English Speaking Board (ESB)		All (+ Pre-Entry level)	n/a	n/a	Internally examined, externally moderated. On demand
National Open College Network (NOCN)	E1-E3, L1				Portfolio
Open College of the North West (OCNW)	Entry	Entry			Portfolio

¹⁸ Pitmans which offered widely used exams in the ESOL context was subsumed by City & Guilds in 1990 although its separate name was retained until recently.

Trinity	All	All			Exam
---------	-----	-----	--	--	------

Key: All = all 5 levels: Entry 1, Entry 2, Entry 3, Level 1 & Level 2.

Source: QCA website (as of June 2005).

It must be noted that within the range of qualifications available a variety of formats exist. Not all assessment is conducted by external examiners, for example. Equally, some, such as Cambridge ESOL, offered different modes at different levels which clearly accommodated students' spikey profiles, i.e. different levels of ability in different skills. Also, speaking and listening for ESOL students at Level 1 and Level 2 were not initially part of most exam boards' range of exams. These have subsequently become available.

3.4 Funding for the Strategy

In the same year as Skills for Life came into being, the Learning and Skills Council (LSC) took over responsibility for funding adult education programmes. The mechanism by which funding is now received is perhaps not essentially different from that of the FEFC, however, which had previously allocated financial resources 'through a complex formula based on the number of students who are enrolled, attending class, and gaining credentials' (Hamilton & Merrifield 2000: 13). When the FEFC took over from Local Education Authorities (LEA) in 1993, as a result of the 1992 Further and Higher Education Act which instigated a change in funding regime (Ward 2007), funding was significantly reduced and accountability and cost-effectiveness increased (Say, 1997). The change at that time was probably more noticeable than the change from FEFC to LSC control. Once Skills For Life funding was released however, the largest change of all for ESOL was witnessed in terms of funding in that the amounts of money available outweighed any previous funds (Cooke & Simpson 2008: 54).

The major shift in practice and significant change experienced by ESOL may be indicated by the amount of funding poured into supporting the Strategy’s activities.¹⁹ Until September 2007 ESOL classes were free to eligible students, paid for by the LSC (Ward 2007). In addition there was the attractive situation of eligible learners attracting uplift of 12% and ESOL learning aims were weighted at 1.4 (LSC 2006) resulting in ESOL being well funded in recognition of the extra resources needed for the staff and support required (Ward 2007). Table 3 outlines estimates of the amount of funding involved²⁰.

Table 3 Range in reported amounts of funding for ESOL

	2001/2	2002/3	2003/4
Lavendar*(2007)	£185 million	£235 million	£267 million
Ward (2007)	£103 million	Not reported	£270 Million
KPMG (2005)	£170 million	£212 million	£256 million

*Based on information from Hansard

Ward reminds us that the funding was not from one source alone, or one funding stream, and confusion has been reported amongst providers due to a variety of eligibility rules (2007). For example, funding has been derived only for students working towards an accredited, nationally recognised, Skills for Life ESOL exam. The LSC introduced a benchmark of an 80/20 split between learners working on these qualifications, and those who are not, amongst the reasons being because they are pre-entry level students for which no accredited exams exist. As Ward says, it is not clear on what basis this split has been decided on (2007).

¹⁹ While this funding of ESOL was generous at the time of the study it has subsequently undergone drastic cuts, and fees were introduced from September 2007.

²⁰ Lavendar (2007) reports that between 2004 and 2007 £1.7 billion was spent on Skills for Life.

Given the lack of research into ESOL in the UK, alluded to above, the vast amounts of money pumped into it may be seen as rather surprising. It is probably that it was the result of, as Cooke & Simpson (2008) remark, the politically high-agenda issue of immigrant integration in that period, where high profile government ministers were publicly linking lack of good English language skills with terrorism and inter-cultural unrest.

3.5 Summary

It can thus be seen that Skills for Life has had a significant impact on the realm of ESOL teaching in the UK. It instigated an even closer alignment between ESOL and adult literacy and saw the injection of an unprecedented amount of funding to support teaching, and new ESOL qualifications which were closely aligned to the new ESOL curriculum.

4 BACKGROUND TO WASHBACK STUDIES

Principle 2: The Principle of Beneficence: A test ought to bring about good in society (Kunnan 2003)

4.1 Chapter overview

Washback is a field within language assessment which has grown rapidly as the growth in understanding of the social consequences of assessment has developed. This has involved an ever-increasing consideration of the value of understanding the exam environment, the ‘actors’ and the ‘theatre of operation’, and the consequences of examination. In this chapter I will set out some of the key basic concepts of this field, starting with definitions and establishing the parameters of the term as used in this research. Then some background to the growth in interest in washback will be considered. Various facets of the phenomenon will be explored before finally addressing the question of why washback is worthy of continued study.

4.2 What is washback?

As mentioned already in Chapter 1 (Introduction) the phenomenon of washback refers to the influence of an assessment on teaching and learning. In most studies to date the assessments in question have been specific exams. Since teaching and learning are seen to be the key relevant factors in washback, in many studies it is pinned down only to effects in the classroom. However not all researchers have worked within these parameters. For instance, (Cohen 1994, cited in Bachman & Palmer 199: 30) speaks of washback in terms of influencing ‘educational practices

and beliefs' also. Various studies have researched the effect within an arena wider than just the classroom, for example the effect on educational policy (Wall, 2005), and this is generally, in contrast, referred to as *exam impact* rather than *exam washback* to distinguish the spheres of influence (Weir & Milanovic 2003; Wall 1996). Weir & Milanovic (2003) distinguish the two by referring to them as working at the micro-level (washback) and the macro-level (impact). In educational research the terms *consequences* and *effects* of testing are more widely used, acting as an umbrella term covering both.

Spolsky suggests the term 'backwash²¹ is better applied only to accidental side-effects of examinations, and not to 'those effects intended when the first purpose of the examination is control of the curriculum' (Spolsky 1994: 55). In contrast, Cheng's working definition of washback is 'an intended direction or function of curriculum change on aspects of teaching and learning by means of a change of public examinations' (2005: 28). Washback will be used in this study to cover all effects, whether intended or not. What is more it will be used for all effects on the classroom, whether intended, unintended, planned or incidental, negative or positive, and whether the exam has been introduced to instigate change or is an extant exam.

4.3 The evolution of washback studies

Although exams have been known to have been used for a long time, maybe more than a thousand years if we consider the Civil Service exams in China, the effect of exams has only more recently been seriously studied (see Cheng 2005). The profile of washback has been raised considerably since at least the mid 1980's when Morrow (1986) for example raised the issue of 'washback validity', and in the same period Alderson (1986) suggested washback as an area for fruitful research and in a similar

²¹ The term washback is more commonly used in the British applied linguistics community but is synonymous with backwash.

vein Messick (1994) proposed the concept of 'consequential validity', as already noted, which states that the consequences of a test or exam should be taken into consideration when evaluating the validity.

Until the mid 1980's, washback seems to have been viewed primarily in terms of being a negative phenomenon; the detrimental consequences of the effects of testing on learning were focussed on (e.g. Kellaghan et al 1982, Frederiksen 1984). Swain (1985) changed this perception however, by her discussion around 'working for washback'. She proposed that the effects of an exam could indeed be beneficial if addressed correctly. Nevertheless the majority of the studies of that era dealt with these ideas at a purely theoretical level, ideas which had not yet been demonstrated in practice.

Alderson & Wall (1993) in their seminal paper asked the testing community the basic question of whether washback really exists, the reason being that up to that point few empirical studies (i.e. that which is verifiable by observation, rather than proposed from mere conjecture, theorising or supposition) had been undertaken, a point raised by Bailey, a few years later (in 1996). By empirical, in this particular context, is meant that rather than relying on accounts of classroom practice described by teachers to researchers, the researchers undertook classroom observation themselves to align this evidence of washback with that from other sources such as interviews. In other words a more systematic, academic approach was taken rather than resorting to commonsense and anecdote.

4.4 Key frameworks in washback studies

A significant catalyst for various washback studies were the 15 hypotheses which Alderson & Wall (1993) proposed in their paper (see Appendix 1); they have acted as the basis for many of the plethora of studies into washback which have blossomed since then. The two most common themes pursued arise from hypotheses number 3

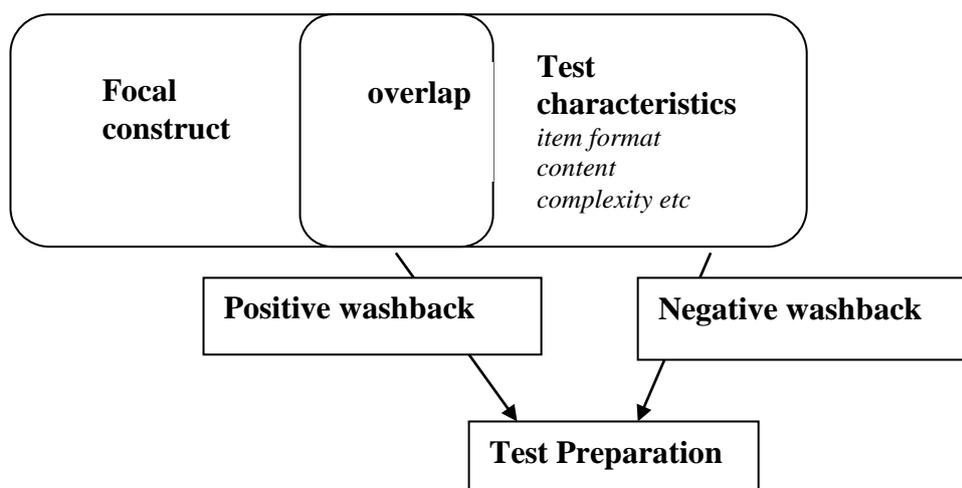
and 4 dealing with whether, and how, both *content* and *methodology* are influenced by exams/assessments. Other studies, along with the (Wall & Alderson 1993) Sri Lanka study, have provided evidence for, for example hypothesis #3: that a test will influence what teachers teach (see Cheng 1997). Other studies (Glover 2006) have backed up evidence from Alderson & Hamp-Lyons (1996) which contradicts hypothesis #4: a test will influence how teachers teach, by suggesting that teachers take a quite individual approach, and different teachers, working with their students towards the same exam do not necessarily teach in the same way.

While many studies have focussed on these two areas, there are still many such questions and many further corners of the vagaries of washback to be investigated. Few studies for example have investigated the role of motivation in washback. Another area in which little progress has been made to date relates to Alderson & Wall's (1993) hypothesis #14: Tests will have washback on all learners and teachers and its converse, #15: that tests will have washback effects for some learners and some teachers, but not for others. (However see Andrews et al (2002) for their conclusions regarding the effect on students and Cheng (1997) for her findings concerning teachers). In other words washback may be differential, affecting different groups involved in the examining process, in different ways, and affecting individuals within these groups in different ways.

What arose from the interest in washback was a drive for frameworks to aid further research. Hughes contributed to the field of washback one of the first models of washback, which has since been widely used as at least a starting point for other studies (1993). His tripartite model focussed on three key aspects of washback: the Participants (including students, learners, administrators and materials producers), the Process (the learning) and the Product (amount and quality of learning) of washback and began the search for the components and causes of washback. Bailey's model which has also been influential can be seen as a development of Hughes' model (1993) and also built on the work of Alderson & Wall (1993). It added the

dimension of consideration of washback to the learner and washback to the programme. However it had limitations in not considering the processes by which such washback may occur. Another more recent models proposed by Green (2003) adds the dimension of test characteristics and the recognition that washback may turn out to be positive as well as negative.

Figure 2 A basic model of washback direction (Green 2007)



Saville's (2010) model embraces the interplay of actors and activity beyond the classroom environment by moving to the effects of assessment on a wider stakeholder group. This work represents more recent studies which have acknowledged the fact that testing and examining affect more than just the teachers and students. This model was not available at the time of the development of my study but in any case while interesting and one of few to encompass more than washback, i.e. going beyond the classroom context, it lacks concrete direction and detail.

4.5 Washback direction/parameters

It is too simplistic to suggest, as it seems early work on washback assumed, that a good test would promote good washback. As McNamara (2000: 74) says, 'ethical

language testing practice [...] should work to ensure positive washback from tests'. This reminds us that the effects can exert either a negative or positive effect, or indeed a mixture of both, as 'all assessment has consequences, some of which are intended, others unintended' (Stobart 2003). While much of the literature suggests washback relates only to the intended effects of an exam on prior teaching and learning, this need not be the case. Washback relates to any effect, whether intentional, programmed, or not, and this is a central point for the current research. The matrix in Figure 3 simplifies the possible and improbable combinations of the consequences. The common goal of exam reform lies in intended positive outcomes, in the top left quarter of the diagram, but the other dimensions remind us of the other variations of outcomes which need monitoring, for possible replication, or avoidance, as relevant.

Figure 3 Matrix of washback conditions

	Positive consequences	Negative consequences
Intentional	Possible	✘
Unintentional	Possible	Possible

It would be hoped there are no education systems aiming to achieve negative washback; there is plenty of evidence of various systems planning for positive washback however. Qi, for example, reports on the situation in China where the NMET²² university entrance exam was introduced in 1985 with the specific goal of reforming teaching and learning (2004). By positive washback is meant, as Messick puts it:

²² National Matriculation English Test

‘for optimal positive washback there should be little, if any, difference between activities involved in learning the language and activities involved in preparing for the test’ (1996: 241).

It promotes the alignment of teaching and examining and avoids behaviour which simply increases scores, without enhancing language development.

An example of negative washback would be inclusion of items in an exam which mirror language use far (production or reception) from Bachman & Palmer’s (1996) Target Language Usage (or in other words real-life language usage). This is felt to thus encourage teachers and learners to spend time on skills which do not relate to probable actual usage of the language. A commonly referred to example would be time spent on mastering how to tackle multiple choice items over learning to compose a letter. One of the frequently arising concerns is that multiple choice format does not easily capture higher cognitive processes and thus encourages surface rather than deep learning, which leads to what Frederiksen refers to as ‘the real test bias’, a move against teaching higher level skills (1984: 210). However it must be noted that a study by Wesdorp found that the purported negative effects of using multiple choice testing were unfounded (1983). This underlines the need for research in washback to establish the effects and actual causes.

Other examples of negative washback include what is referred to as narrowing of the curriculum whereby subjects which are to be tested dominate classroom time at the expense of those which are not (James 2000; Smith 1991). This is not a recent phenomenon; even Latham as far back as 1877 mentions that ‘narrowing the range of learning’ occurred as a result of the content of the oral exams which he was discussing being known by tutors and students (1877, cited in Cheng 2005: 35).

'Teaching to the test' is another widely reported behaviour associated with testing. By this is meant a programme of teaching which focusses on content and methods mirroring closely that to be found in the impending exams, at the exclusion of wider learning. Test preparation practice, as Mehrens & Kaminski point out, is however, in itself neither positive or negative (1989). The crux of the matter is the nature of the test preparation. Having to teach to the test is an indicator of a poor teaching situation. In language teaching, if the test does not match target language usage (TLU) namely, teaching does not cover what students require for their communication needs (see Schellekens 2007) then a mismatch is likely and teachers may feel under pressure to focus on material and skills, not in the curriculum (which should cover the TLU), which it is felt students need to pass the exams. However, equally, teaching to the test may be beneficial if the test aligns closely with the curriculum, which also in turn aligns with students' language needs. The test and its relationship to the curriculum and language needs are what is pivotal.

Attempts to instigate change in an education system, where there was a mismatch between the curriculum and the exams has led to failure of the initial intentions of some educational projects. This was seen in the ELTIP project (1994 to 2003) managed by the British Council in Bangladesh (personal experience). This project's main aim was to improve secondary level teaching of English through the development of a nationwide network of resource centres and both teacher training and trainer training (Khan 2002). Since within the national education system the English exams were not revised, the implementation of changes in the curriculum, although backed up by revised specially designed materials and extensive teacher training, was not fully effective. In a society in the developing world where school leaving qualifications are of extremely high-stakes because future job opportunities hinge on them, pressure from parents meant that all efforts on the teachers' part had to be seen to be assisting the children in achieving their full potential in the exams,

which in this case did not align with the newly proposed teaching methodology. Other similar examples are cited by Kellaghan & Greaney (1992) regarding fourteen African countries, Cheng (2004) concerning Ethiopia, and Somerset (1983) in relation to Kenya.

4.6 Measurement driven instruction

The introduction of new exams is frequently used to effect 'targeted pedagogic change' (Andrews 1994b), as already mentioned. Assessment reform is a widely used tool for curriculum reform, or teaching reform (McNamara 2001). There is a plethora of studies examining this phenomenon especially from the USA, where the introduction of standardised exams (where they were not previously a feature of school life) took hold nationwide. This has caused much debate in recent years (Mitchell 1992) as to the success, or otherwise, of such a move. The exams, as the publicly visible measure of success, are used as the catalysts for change, or as Pearson has referred to it, as 'levers for change' (Pearson 1988: 98). When an assessment is used as an instigator of change in teaching and learning, this is known as 'measurement driven instruction' (MDI). This means of attempting to raise educational standards has both its supporters and its opponents.

4.6.1 Supporters of MDI

Frederiksen & Collins (1989), who discuss the issue in terms of 'systemic validity', support the notion of measurement driven instruction. They focus on desired outcomes. However their work is purely theoretical in nature and, being based on no empirical research, is of limited value. Popham et al (1985) review five conditions from the literature in the field under which measurement driven instruction will be

beneficial. They claim when assessments display the following features the chances of positive washback are enhanced:

1. criterion referenced
2. defensible content (i.e. important knowledge)
3. manageable number of targets
4. provides instructional illumination
5. instructional support for teachers provided

Popham (1987, cited in Cheng 2005:36) suggests the rewards and sanctions of exams encourage teachers to focus on the objectives embodied by the exams. However there must be other ways, through adequate teacher training to highlight the benefits of a new approach. If teachers do not want to change it is discouraging to think that only a 'stick' without a 'carrot' method focussing on outcomes rather than focussing on process of learning is administered. This is why various academics, see below, have condemned this kind of approach to instigate change.

4.6.2 Non-supporters of MDI

Amongst academics who oppose MDI are Smith (1991) and Madaus (1988) who oppose the notion that assessment can and should be used to drive alterations in classroom practices through the introduction of new examination or assessments. They cite problems associated with this approach such as narrowing of the curriculum, as already mentioned, and that the teachers' own judgments of student ability are demeaned, leading to a sense of de-professionalisation. These are criticisms also heard of the UK situation regarding the effect of SATs (James 2000, Warshauer 1995).

Another criticism directed at MDI is termed ‘test score pollution’ (Haladyna et al 1991) which means that by introducing test preparation into the classroom to focus on raising scores, which often happens when a focus is placed on scores over the learning process, these scores will no longer represent the exam candidates’ true capability. What is more, some researchers have doubted the assumption that test preparation will even raise scores in the ways assumed (e.g. Koretz et al 1991). Herman & Golan (1993), for example, studied the situation at secondary level education at various locations across the USA, investigating how far teaching to the test affected scores. They found there was no guarantee that it helped.

Chapman & Snyder (2000) report five ways testing is purported to improve instruction through exam reform e.g. that a test can be used to ‘pull’ or shape pedagogical practices in desirable ways, in line with the MDI approach. They warn however that the desired outcomes are not ensured, even when a curriculum and exams are in alignment, suggested by Popham (1987) above. For it to be successful, there must be an understanding of the ‘intermediate conditions’ e.g. the resources needed to enact the desired reforms, and the assurance that the skills and knowledge to best use these resources are available to ensure educational change; the provision of the resources alone is insufficient. They claim, that all too often, the ‘intermediate conditions’ are overlooked. This stresses the need for understanding the complexity of change in educational systems, taking into consideration a wide range of factors which may contribute to success or otherwise.

4.7 Key components of washback studies

Alderson & Wall describe washback as the extent to which a test influences language teachers and learners to do things ‘*they would not necessarily otherwise do*’ that promote or inhibit language learning (1993:117) (my emphasis). It has already been

discussed that washback could relate to an extant exam, not just a newly introduced one. However Messick (1996) makes an important point that there is a need to search for the causal connection to be sure a certain behaviour is indeed the result of an exam, and not due to some other cause, namely to establish an evidential link, as he terms it. Such detail is not often made explicit in the washback studies to date.

The means to explore evidential links have been attempted by either baseline studies (Wall & Horak 2006) establishing the nature of teaching prior to a new exam, or via a comparative element to the research, such as by Alderson & Hamp-Lyons (1996) by considering the nature of teaching in classes leading towards the exam in question, and teaching of non-exam classes by the same teachers. It must be noted, that these methods are not always possible, for example if a washback study concerns not a newly introduced exam but an extant one, a baseline is not possible e.g. in the case of studies of IELTS (Green 2003) and FCE (Tsagari 2006).

There is now general agreement that washback does indeed exist (cf the title of Alderson & Wall's seminal 1993 paper: Does washback exist?), as a result of numerous empirical studies. The range of washback studies has snowballed in recent years. Alderson & Wall (1993) were amongst the first to state we must beware of seeing it as a simple phenomenon, since the range of studies to date have shown 'washback is a highly complex rather than monolithic phenomenon' (Watanabe, 2004: 19). (See for example Tsagari's (2009) study: The Complexity of Test Washback, for more on this issue). As Alderson & Wall (1993) propose, corroborated by Cheng (2004), in this field we need to move on from the 'what' onto the 'why' of washback, to examine which factors are pivotal.

4.8 The contribution of educational innovation studies

The study of washback fits neatly within the field of educational innovation studies which have proved useful in this search for pivotal factors. While change is simply a new way of 'doing business', innovation entails planned change with a specific intended aim. Because washback has lately been recognised as so complex, there is a need to study the wider context than just the test/ exam and associated observed classroom activity. Alderson & Wall (1993) recognised early on the need for a widening of the research field to better understand washback. Insights from educational innovation studies can usefully inform washback research as Wall (2005; 1999) in particular has shown.

The contribution of innovation studies, and particularly those from education, is the understanding that it cannot be presumed all actors in the network of activity related to examination and its washback will be affected in the same way, or in the desired way, by planned changes to a system. Rogers & Shoemaker (1971), Kennedy (1987) and Markee (1993) amongst others have tried to model the introduction of innovations to better understand how they may be managed to best effect. Henrichsen (1989), on examining the introduction of communicative language teaching into the Japanese context, drew on a number of these models to draw up a detailed composite model to describe the diffusion of this innovation. The model he composed aims to account for the time frame from pre- introduction to post- introduction, implementation and also aims to account for the complexity of factors which influence outcomes.

4.9 The role of the teacher in washback

A number of the Alderson & Wall (1993) hypotheses focus on the behaviour especially of teachers, e.g. A test will influence what teachers teach, a test will influence how

teachers teach, a test will influence the rate and sequence of teaching, a test will influence the degree and depth of teaching (p121). Hughes (1993) and Bailey (1996) in their work also look beyond simply the role of the student in washback. It makes sense to study the teachers' role as they act as the mediator of the curriculum and/or assessments; teachers are among the main stakeholders, they are pivotal in the design and control of the classes. The students experience the teaching implemented by the teachers, and management provide resources and policy guidance, but the teachers have the pivotal role between the two other key stakeholders in the exams.

Several studies have already focussed their attention on teachers, such as Lam (1994) who found, amongst other things, that the classes of more experienced teachers were susceptible to negative washback and vice versa. He concluded that there is a need to change the teacher culture not just the exam, which builds on aspects of Alderson & Wall (1993) where they stress the importance of teacher training in the efforts to attain positive washback. Andrews (1994a) also examining teachers in Hong Kong, questioned the teachers' awareness and practices, having looked at experienced and novice teachers and teachers' strategies and their match or mismatch with the test designer's intentions. Watanabe's (1996) work studying the significance of individual teacher style found three factors at work in determining the washback effect: the proximity of the exam, the teacher's educational background and their methodological beliefs. Teacher profiles are thus a useful contribution to exploring washback.

4.10 Washback and the role of stakes

Throughout washback studies the role of the consequences of the exam results is recognised as an attributive factor. For example, Alderson & Wall (1993) list among the washback hypotheses: 'Tests that have important consequences will have

washback' (#12). There is a common sense aspect to this that is easily accepted, that if the results matter, more effort and energy will go into preparing to do well on an exam, and practices may consequently change to try to maximise success. One reason may be as Chapman & Snyder (2000) say:

‘High stakes tests are one of the few elements of an education system that are controlled at the central level of the system, but which have direct impact at the classroom level’ (p 458).

The power of high-stakes exam results is too strong to counteract desired changes in teaching method or content, even if recognised to be beneficial in other terms. In the example, of the ELTIP project in Bangladesh, already mentioned, methodological innovation was extremely hard for teachers to introduce when the results of such changes were not directly applicable to the exam requirements, the exam, which was still based primarily on a traditional demonstration of grammatical knowledge, not being in alignment with the new curriculum or new text books. The urgent social need for good exam results to increase one’s employability, in a society where the consequences of lack of work can be extreme, result in stakes of an extremely high order. Intense pressure from students’ parents add to the pressure on teachers to prepare for exams, not to adhere to a new curriculum, however much the latter may better equip the students to effectively use the English they learn in a real-life work situation. ‘[P]arents and teachers may not always be natural allies in efforts to raise educational quality, at least if there is short-term risk to their children’ (Chapman & Snyder 2000: 463). As London (1997, cited in Chapman & Snyder 2000: 460) states, even if these teachers wish to innovate regarding material or their teaching, they are often handicapped by concerns that their actions will disadvantage their students on the exam. Thus all aspects of an educational innovation: materials, methods, teacher training, and vitally also the assessments, must be made to fit the curriculum and

materials if positive washback is to have a chance (Woods 1988) though it can never guaranteed.

The presence of stakes in the washback 'equation':

$$\text{exam} + \text{high stakes} = \text{washback}$$

appears to be generally accepted but the nature of the stakes is generally less examined. For example, the study by Kellaghan et al's (1982) longitudinal study, concerning IQ and reading comprehension, was insightful in terms of its contribution to the canvas of effects of exams but there were no consequences resulting from exam results, i.e. no stakes, which suggests the role of stakes also is complex, and merits further investigation.

4.11 The purpose of washback studies

To conclude this chapter I will turn to a question central to this research: what is the point of studying washback? As McNamara says, 'washback is often rather unpredictable' (2000). Washback is hard to study but is important for monitoring purposes to avoid negative consequences. In brief, it appears to be necessary in order to ensure that when planning educational programmes, which almost inevitably involve some kind of assessment, that obvious unintended negative consequences can be avoided by surveying the breadth of research already undertaken in this field to date. Washback studies, although complex, should be included as good practice as a matter of course in all exam development programmes. Washback study data should inform a decision to revise an exam (Hawkey 2006), the development of a new one, as well as on-going monitoring of an exam.

As stated already, assessment is sometimes but not always seen as the driver of a system, whether it is an innovation or an extant programme. On the contrary in many

cases the assessment element of a programme is simply taken for granted; the effects are not necessarily considered. Washback studies can help raise the profile of assessment and encourage ‘thinking around’ learning from other situations.

Washback should be studied as a reminder not to let exams become the ‘be all and end all’ of an education programme but to continually monitor the quality of students’ learning, as Gipps stresses (1994). It must be remembered not to lionise the scores (or other measurements of ability) but remember they are simply a means to obtain an approximation of candidates’ abilities. It must always be remembered they have no actual meaning in themselves.

Returning to the Kunnan quote at the beginning of the chapter, it reminds us of the role of exams in society and it also reminds us of the role washback can take in ensuring the quality of exams to inform good practice. We will not know how to avoid negative washback without more extensive studies in a wider range of settings. While it is hard to imagine a situation where negative washback would be the goal of the design of a certain exam, the point is that without a thorough understanding of washback, unintended consequences can lead to unfavourable classroom behaviours which may detract from pedagogical aims. Therefore in discussing washback we study the specific influence of one exam (or set of exams or other assessments) in a certain specific situation and can use this to good effects in other situations, what might be termed ‘washforward’. This is the goal of washback studies.

4.12 Summary

The aim of this chapter was to outline key studies in washback and to highlight how far these studies have brought our understanding of the possible scope, parameters and direction of washback. It also highlighted the key concept of the evidential link

which focuses on the causal link with classroom behaviour and exams. The role of the various stakeholders, as well as the importance of the stakes involved to the occurrence of washback was discussed, as well as consideration of the system within which exams operate. Thus the complexity of studying washback was highlighted, but reinforced the point that it must be attempted as the value of washback is to monitor for positive consequences of exams, and to identify and thus act to remedy negative consequences.

5 METHODOLOGY –THEORETICAL ISSUES IN USING QUALITATIVE METHODS

Not everything that counts can be counted and not everything that can be counted counts. (Einstein)

5.1 Chapter overview

The aim of this chapter is to set out the theoretical considerations behind the methods utilized to design the data collection instruments, to collect the data and to analyse it. The reason for going into such depth here is that qualitative data must counteract the criticisms made against it in terms of rigour. One way this can be achieved is by the clarity and transparency of the methods drawn upon. Thus the rationale for these decisions is set out before going on to describe, in the following chapter, the actual methods used in this study.

5.2 The qualitative paradigm

There has been a tendency in our modern society to lend more weight to that which can be reduced to numbers: that which is the more easily measurable. It is argued by Kvale & Brinkmann (2009: 308) that this is not a new phenomenon but has been extant since medieval times, influenced heavily by the Greeks by way of Pythagoreans and Christian traditions via St Augustine. Statistics tend to give credence to report findings and much academic research, as well as being widely used in journalism to lend gravitas (even though sometimes poorly or even manipulatively utilized), whereas the techniques of qualitative research methods have often been criticized as ‘journalistic and impressionistic’ (Kelle 2004: 444). Thus if a study is not to measure

in some way its topic of investigation (i.e. take a quantitative approach), then quality and credibility in particular need addressing to counterbalance the prevailing authority of statistics.

Many quantitative methods have developed clear guidelines regarding, for example, specific procedures. In order to undertake certain statistical tests the requisite conditions and nature of the data are prescribed in order to be able to successfully run the tests. This renders such procedures easily describable. On the other hand, qualitative analysis has suffered from a great deal of poor description of the analytic process and thus this presents problems for the novice researcher to follow the procedures undertaken in previous studies and to know what to do with one's own data.

In addition in qualitative research, analysis entails a process of uncertainty (Kelle 2004: 445) as there are no fixed techniques. There is much 'feeling one's way' to discover the concrete methods to use, since much qualitative research is written up in abstract terms, focussing it seems on context and outcomes (the interpretation) rather than method. Methods employed in a piece of research are not consistently accounted for in detail, probably because of the lack of a shorthand which can be used as in quantitative data analysis techniques when established statistical tests are applied to a data set, for example. 'Methodological rules applied by different schools of thought are often not explicated but form a folklore of research passed on verbally from teachers to pupils' (Kelle 2004:445). Indeed Kelle has called for further description of the 'actual processes of data management and data analysis' from real instances from research studies (Kelle 2004: 457). The word limits of most journal articles will not permit detailed description of the specifics of every stage in a qualitative study and consequently the methods employed can appear fuzzy,

especially to a novice researcher who has little or no previous experience to draw on to understand the procedures referred to.

One of the key differences between approaches in the positivist domain and those in the constructivist domain is the very approach to research taken from the inception of the study. 'Grounded theory' was first posited by Glaser and Strauss (1967) whereby the researcher entered the new field of study with as much of a tabula rasa mentality as is possible. It has since been widely critiqued as a study method and many researchers (e.g. Dey 2004, Mason 2002), will refute that this is even possible.

Grounded theory it is said was based on an unfeasible premise 'there are and can be no sensations unimpregnated by expectations' (Lakatos 1982: 15, cited in Kelle 2004: 449). We cannot help but bring our extant ideas and world-view to a study which will inevitably affect the lens through which we see the subject of the research. What is more, it is often a particular opinion about a certain aspect of the field we work in, which triggers the research in the first place, so we cannot pretend to be working from the 'blank slate' position. What we can do however is try to be fully aware of what our own position is and not allow this to skew our exploration of the data.

The different perspectives of positivism and constructivism can be summed up as looking 'at' versus looking 'for'. These two methods are not dichotomous in my view but simply lie on a cline. The post-positivist arguments do not supersede positivism; they simply complement it. In any case we should not fall into the trap of believing that even quantitative methods are without bias and that they are entirely neutral interpretations of whatever is under study. All research relies on human design, choices, and decisions and is thus subjective to some degree. My position is that of Marshall and Rossman: 'all research is interpretive' (2006: 4), so we must always remember the human element in all research.

5.3 Qualitative research in the field of language testing

Qualitative studies in general aim to reflect the world as it is and make sense of it. Qualitative approaches to research, used as the main method as opposed to a supporting role for quantitative methods, have only recently (within the last 20 years or so) begun to be generally accepted within the field of language testing. Smith (1991), for instance, was one of the first major studies to take an entirely qualitative approach. The majority of research in this field to date has been dominated by the quantitative paradigm, where primarily the psychometric properties of exams have been explored through statistical analysis. Whereas previously the qualities of the test itself and efforts to ensure its technical precision had predominated as the focus of the research in this field, the role of social context began to be taken more widely into consideration as worthy areas of research. From the early 1990's the increased interest in washback (and then impact) encouraged mixed methods approaches. There is currently an increasing awareness of the importance of studying, for example, the effects of testing on candidates, as well as the wider stakeholder population (Green 2003; Hawkey 2006; Saville 2010; Wall 2005). Shohamy (2001) and McNamara (2000) have further pushed the boundaries of investigating the social context in what has come to be called critical language testing (McNamara 2000: 132). Yet few studies take an entirely qualitative approach.

The changes witnessed in the spectrum of research methods used in the area of language testing could be said to mirror developments within language testing itself, in the West at least. The need for large-scale testing operations in the US for army recruitment purposes in the early 20th century precipitated the wide use of tests composed of discrete items, mostly in multiple-choice format. Such a format facilitated efficient testing and processing of candidates. What was tested in terms of language ability however was influenced heavily by the methods the situation necessitated, namely, knowledge *of* language rather than ability to use it.

The benefits of easily administered tests caught on and the discrete item test, not only within language testing, became a very popular format. The high reliability afforded by such methods was attractive. As time went on, once the communicative language movement became so influential, there occurred a move beyond discrete item testing to embrace the complexities and challenges of integrated testing. The phenomenon of written and oral exams was already centuries old. Spolsky (1995) writes about exams used in ancient China to become a Mandarin, and Oxbridge examining systems based on the oral interview technique. However, these involved relatively small numbers. The mass testing which was facilitated by multiple choice techniques was growing as interest in language qualifications grew but needed methods addressing communicative approaches. A concern for fairness in terms of improving test quality turned the focus on validity and reliability in particular. The modern subject specialism of language testing thus emerged.

Despite a broadening out from a predominantly psychometric approach to encompass test impact for example, the qualitative paradigm appears to still not be as highly valued in the field of language testing as psychometric methods, as surveying the chosen methodologies for studies included in key journals such as *Language Testing* will demonstrate. Various recent studies have however embraced a mixed methods approach (see for example Green 2003; Tsagari 2006; Wall 2005). A qualitative approach can indeed act as a useful foil to quantitative approaches, especially research based on experimental design which tends to be of a procedural nature, in other words, a series of stages to be described. As stated however, it is not the trend to undertake research in language testing drawing entirely on qualitative data collection and also on qualitative analytical methods, with no complementary quantitative aspect to the study. The pervasive attitude in the field seems to remain

that a quantitative element adds weight and credibility to research; numbers are concrete and can be trusted.

The basic premise of looking at what the phenomena offer up in the way of data rather than setting up an experimental situation and analysing the resulting data is a profound shift in perspective which is, I believe, not truly understood by many researchers in language testing who include qualitative methods in their work. I am not criticizing a mixed methods approach as I believe it brings great benefits to a study and affords a whole (360 degree) view. Done well, it provides a healthy overview and balance to any research. Yet few studies have the time and resources to undertake both approaches thoroughly and a truly mixed methods approach requires two very different perspectives on the research process which demands a fundamental epistemological shift. Most often it seems, one method is in effect used as a mere support for the other. It requires considerable experience to use each method well and is often 'likely to result in bastards of low quality, according to quantitative as well as qualitative logics' (Kvale & Brinkmann 2009: 118).

While the more recent research into the wider social aspects (i.e. effects on and influence of the world on the test) of testing has been undertaken in a variety of ways (Shohamy 2001; Wall & Horak 2006; Wall & Horak 2008; Wall & Horak 2011) the predominant methods have included at least some kind of qualitative aspect which stands in stark contrast to the psychometric techniques of 'pure' testing research. Interviews and observations, for example, have become common place in such studies. However little research is entirely free of the positivistic yoke, meaning that although many studies collect stretches of discourse, the studies usually rely on frequency of instances of a particular behaviour or use of language for example, to be counted as significant enough to be reported, and thus adhere to the norms of quantitative paradigm to some degree at least (e.g. Green 2007; Tsagari 2006).

The choice of research paradigm will come down to, at worst, a) ignorance of the range of possible available methods or, at best, b) an epistemological stance which influences the researcher's personal philosophy, depending on the researcher's position on the cline between a positivist and a social-constructivist view of the world. Nevertheless, not all researchers have fully explored and determined their personal position prior to commencing an independent research project. The research process is often a way of doing so.

5.4 Methods used to 'capture' washback and impact

A mixed methods approach is prevalent in recent washback studies (e.g. Green 2003; Qi 2005; Tsagari 2006). Generally in studies to date the most common tools used in washback research have been questionnaires, interviews and observation. Watanabe is one of the few researchers to explicitly focus on discussion of the research methods used in washback studies so far (2004). Watanabe calls for further qualitative studies and such a review of the techniques used to date is timely.

However, Watanabe's work is not comprehensive; some other methods are now being used to collect evidence of washback from differing perspectives. One example is diary studies (Gosa 2004; Tsagari 2006) which offer insights into the effects of exams/assessment on learners, a perspective which only a minority of the washback studies address. This method is particularly time-consuming and requires a longitudinal approach and a level of trust between researcher and journal writer (usually exam candidates) which is beyond the scope of many research scenarios. This is a welcome opening up of investigation into the type of data which different methods may yield and thus add to our body of understanding of this phenomenon.

In deciding which method to use, the advantages and disadvantages found in the methods employed in previous studies is of great value and so the next section details the two main methods mentioned by Watanabe, and most widely used to date, which pertain particularly to a qualitative methodology: interviews and observation. There follows an outline of key characteristics to substantiate the methodological choices made, as described in the next chapter: Methods.

5.5 Data collection - interviews

5.5.1 What is interviewing?

Interviewing can range from the highly scripted and structured, as those typically used in market research, to very open ended, unstructured conversational-type interviews. They range on a cline of control from high to barely at all. The only core features in common with all types are that there is an interlocutor (researcher) and at least one interviewee and that the interaction is conducted orally. Beyond that the resemblance may vary enormously and will be a matter of the researcher's aims and objectives.

It was with the arrival of easily manageable recording devices that the interview could provide information capable of multiple examinations and re-examinations and it could be argued it was only then the material captured became 'data', a term taken from the quantitative research arena (Kelle 2004:444). It turned an 'otherwise informal style of inquiry to a standard model of research' (Kelle 2004:444).

5.5.2 Why interview?

The choice of research methods ultimately reflects the researcher's epistemological position. A belief that the realities of all participants (or informants as they are often referred to) are part of their personal versions of the 'truth', are 'meaningful properties of the social reality' (Mason 2002:63), suggests the appropriacy of qualitative methods to achieve a rich, 'thick description' (from the term used by Geertz 1973) drawing on those various 'truths', rather than a wider-scale study, as afforded via questionnaires. The latter would produce more generalisable data (this issue will be discussed below) but less personal, exploratory accounts. There is the danger our own extant ideas, beliefs and, unfortunately also, biases may be replicated, and limit what we can find by means of the questions we pose in this way.

The advantage of interviews is thus the potential to explore beyond one's own boundaries. To truly explore, the researcher needs techniques to go beyond one's own realm of experience and even maybe imagination. We can in interviews attain the 'unknown unknowns' (c.f. Rumsfeld 2002) whereas in a questionnaire one is more likely only to access the 'known unknowns' (Rumsfeld 2002)).

5.5.3 Dangers

While interviews are thus often chosen for the rich data they can provide, it has been pointed out that the interaction which is the mainstay of interviewing is 'the source of both its advantages and disadvantages as a research technique' (Cohen et al 2000: 269). As they say, interviews allow greater depth than most other methods but are prone to interviewer bias (p269). By this is meant that the researcher can consciously or sub-consciously influence the interviewee into proffering information which they feel the interviewer wishes to hear. The nature of the interview will probably influence

this, namely, in a very open interview, interviewees will more likely respond within their own frame of reference whereas in an interview which more closely resembles a 'verbal questionnaire' with a very tightly structured method they will less be able to.

In addition, another danger is that 'the interview method is heavily dependent on people's capacities to verbalize, interact, conceptualize and remember' (Mason 2002: 64). We often do not know as researchers how articulate the interviewee is and if they are not forthcoming, for whatever reason, valuable time and effort may not be put to best use. However, in the case of interviewing teachers we should be able to assume this is not a problem as long as they feel comfortable with the interview situation. It is after all a requirement of their job to verbalise and explain.

Nevertheless memory is indeed an issue relevant to any interviewee group. Desiring spontaneous responses in the hope these will better represent the informants' current views, as opposed to a prepared, possibly rehearsed answer if cues / prompts were given in advance, means the researcher has to sacrifice a certain amount of certainty concerning veracity of recall when the interviewee is recalling events. We can however be more confident if perceptions and attitudes are being elicited since they represent current states of mind as opposed to stored factual information, which is more prone to attrition over time. Nevertheless, capacity to recollect is notoriously unreliable so extra thinking time is no guarantee of better, more accurate recall (Engle 2002).

Therefore when precisely to use interview as a technique, to gather what kind of data, should be a key consideration in qualitative studies.

5.6 Data collection – observations

5.6.1 Format

The use of observation as a valued research technique probably originated in anthropology. The detailed ethnographic methods used to describe and explain the lives of others, usually in remote parts of the world, and then more recently of sub-cultures closer to home have now permeated the research landscape of the social sciences. The term ‘ethnography’ has indeed taken on a familiarity in various fields, which belie the thoroughness and longitudinal nature typical of anthropological studies.

The tradition of classroom observation in teacher education and in language studies is well established. To provide systematicity to this research activity, various methods have been devised, from those resulting in data of a more quantitative nature such as event sampling or interval recording (Cohen et al 2000: 308 – 9) through to naturalistic observation (p312). As a classroom is, usually, quite dynamic it is necessary to have such a framework or guidelines to facilitate a systematized observation which can be repeated across various locations and occasions and facilitate comparison of informants within a study. Morrison (1993, cited in Cohen et al 2000:305) suggests there are four key parameters to observational data:

- 1) Physical setting i.e. the environment
- 2) Human setting e.g. characteristics of the individual or group being observed
- 3) Interactional setting e.g. ‘formal, informal, planned, unplanned’ etc.
- 4) Programme setting e.g. resources and their organization, ‘pedagogic styles’

The physical setting is of interest in that it may well influence both the students and teacher and help illuminate features of the classroom dynamics. A bright modern well- equipped room may have a very different effect on a group than a dark, cramped, uninviting classroom would, for instance. Equally, the characteristics of both teacher and students can inform the research

since attitudes towards each other and towards the learning environment, for example, will affect behaviour. The formality or informality of a situation, and whether a certain interaction is planned or spontaneous can inform the observed outcomes. What resources the teacher uses, and how, is also worth noting since this tells us something of their teaching skills and expertise.

These are very general parameters but they can inform a framework for observing classes. This framework then allows a picture of the context to be built, within which any critical incidents which occur can be situated, and hopefully better understood, minimising chances of misinterpretation or false assumptions. An unsystematic, unordered observation procedure or record is more prone to observer bias and inconsistent data.

5.7 Key considerations in qualitative research

5.7.1 Sampling

Sampling traditionally is a key stage in any research whose purpose is to achieve representativeness (or generalisability) of outcomes. As this is not one of the goals of qualitative research (see section 5.7.2 below), sampling does not need to follow the same stipulations as for quantitative studies (King & Horrocks 2010). In qualitative methodology sampling decisions are often ‘criticized for being ad hoc and vague’ (Mason 2002: 137). Sampling has often also been dismissed in the past due to claims that a sample needs to be representative so therefore must be based on probability models (see Gobo 2002: 409), which is not the case in most qualitative studies. However, it must be noted that a truly probabilistic sample is in any case rarely achieved (Gobo 2002: 405). Qualitative research while non-probabilistic is not unsystematic and random. Certain clear parameters of the sample should be sought to ensure credibility of the outcomes. Such considerations may include whether

typical or extreme case samples will be selected for instance. As Mason (2002: 137) states:

‘given that theoretical and purposive sampling are not based on a notion of empirical representativeness, the issue of how one substantiates the relationship between the sample and the wider universe is not so well rehearsed, and it is, therefore, even more important for researchers to specify exactly what they see this relationship to be.’

Convenience sampling is often employed because finding informants who match certain key desirable criteria is, it could be argued, harder than taking a group from a random sample of a population. Cohen et al’s (2000) description of convenience sampling suggests an element of choice: ‘the researcher simply chooses the sample from those to whom she has easy access’ (2000: 102), typically drawing on a ‘captive audience’. However this overlooks that most convenience sampling operates within informant profiles parameters and is not a totally open selection. In addition, one particular strength of this self-selection mode of sampling is that the informants’ candour can be relied on more than if they are coerced in any way.

5.7.2 Evaluating the quality of qualitative research – reconsidering the traditional criteria

Three key criteria traditionally used to assess the quality of research are generalisability, reliability and validity. These terms have been much discussed within the qualitative paradigm and are generally deemed to now lie on a cline from *in need of reinterpretation* to *utterly redundant and inappropriate measures*, depending on whose version of methodological advice is being followed.

5.7.2.1 Generalisability

Amongst researchers working within the qualitative paradigm there is a difference in views on the role of generalisability. On the one hand, King & Horrocks (2010:160) suggest that instead of generalisability we should consider transferability. Strauss & Corbin (1990:91 cited in Gobo 2004:421) take the stance that such research should not be aiming to generalize but to specify. There is no claim to be applicable to other studies. Mason's (2002) point of view is that qualitative researchers need to make their own stance clear rather than working for generalisability, and this is the view I align myself with most closely.

Gobo points out that representativeness of the case should not be confused with representativeness of the characteristics (2004: 422). I take this to mean that the nature of social situations and relationships is so complex, characteristics and influencing factors so diverse, you could never compare one with another. However to atomise somewhat and talk about specific aspects of the research focus which can be identified elsewhere may be fruitful and lessons can be learned from this. This is also a stance I find fruitful.

Gobo goes on to argue, supported by Lincoln & Guba (1985) and Hammersley & Atkinson (1983), that there is in any case a lack of rigour in social research sampling when quantitative techniques are being used, which renders the results questionable; 'statistical inference in social studies is problematic' (see Gobo 2004: 421 in Seale et al 2004). Generalisability is in fact a chimera that has taken on weight and respectability through familiarity.

5.7.2.2 Reliability

Instead of discussing the term 'reliability', which is a preoccupation of research within the quantitative paradigm, I will outline instead two other parameters considered to

be as important, as discussed by Denzin & Lincoln (2003). Firstly there is confirmability. This is what the King & Horrocks (2010: 160) term ‘trackable variance’ deals with. The second is transferability.

First of all, a PhD thesis has the benefit, due its permissible length and likely audience, of being able to include data which can be referred to by future researchers, whereas most pieces of research can only appear in summary form. As regards transferability, Mason says:

‘qualitative research is particularly good at constituting arguments about how things work in particular contexts rather than representing the full range of experience’ (Mason, 2002: 136).

I intend to ensure there is sufficient ‘thick description’ of the context (referred to already), within the limits of a study of this length, to allow other researchers to decide whether my conclusions may pertain to their own context, and thus exhibits transferability.

5.7.2.3 Validity

King & Horrocks (2010: 160) suggest credibility should be attended to instead of validity which is a key consideration in much research but especially that within the quantitative paradigm. Ways to achieve credibility are suggested by Cohen et al, who advise that

‘in qualitative data validity might be addressed through the honesty, depth, richness and scope of the data achieved, the participants approached, the extent of triangulation and the disinterestedness or objectivity of the researcher’ (2000: 105). (See Section 6.22 below for how this was addressed).

King & Horrocks (2010) also recommend expert panels, thick description and audit trails. Member checking also (known as respondent feedback) is another widely

advised technique (see Mason 2002). Member checking needs some preparation in terms of warning the interviewees of the nature of transcribed data since they can often feel uncomfortable faced with the detailed visual representation of the actual spoken word, with its false starts, incomplete sentences and peculiarities of spoken grammar etc. which is something we rarely encounter.

King & Horrocks (2010) also advise independent coding. Independent coding (as used by Tsagari 2006; Wall 2006) can also be problematic however. The idea may simply be taken as a means of checking inter-rater reliability of coding. If this were the case the process would be quite complex. The second coder, assuming it is not the researcher as would usually be the case for a PhD study, does not have the in-depth knowledge of all the informants, and their environment. It is unlikely that without the 'intimacy' with the data which the main researcher achieves that high co-occurrence of codes would occur. This need not be seen as a methodological failure. Where both coders have good knowledge of the whole study (as in Wall & Horak 2006, 2008, 2011) then a relatively high co-occurrence can occur but the main value of the procedure is clarification of the codes rather than achieving reliability itself. This is the aim that I find more convincing as a useful tool, and by checking the code boundaries thoroughly and regularly and iterative checking of the data that this can be achieved and it is not vital to achieve this through independent coding.

Triangulation is frequently discussed regarding the assessment of quality in terms of validity (e.g. Lincoln and Guba (1985, cited in Cohen et al 2000: 108); King & Horrocks (2010) and most qualitative methods research manuals). Since Alderson & Wall's (1993) call for (more) observational data in washback research, the use of more than one method of data collection has become common practice in this field, for this reason. The most compelling reason to compare the results from observations with those from other data sources is because, as Markee remarks, 'what we think happens

and what actually happens in “our” classrooms are often different’ (1993: 13). The impartial outside observer probably will describe the classroom differently to the teacher.

According to Seale (1999: 53), Denzin was the first to use triangulation in qualitative research over thirty years ago (1970). Denzin sets out four types of triangulation, although other researchers have since added to this list (e.g. Cohen et al 2000).

Denzin (1978) refers to:

- Data triangulation
- Investigator triangulation
- Theory triangulation
- Methodological triangulation

In each case, more than one source is drawn on to ensure rigour. For example in Hayes & Read’s (2004) study of washback in an IELTS preparation course the collection of data was from both students and tutors. Since two researchers were working on the study, they also employed investigator triangulation. Their methodological triangulation was the use of interviews, questionnaires and observations.

Many research manuals advocate triangulation as a means of increasing the validity of studies. However, it is a much used term and thus in danger of being widely, and perhaps erroneously, interpreted. Gorard & Taylor (2004) for example, warn that metaphors can confuse as easily as clarify and they cite various examples of how the original concept, derived from the field of surveying, has become muddled. Silverman corroborates this in claiming that ‘triangulation exercises can deepen understanding as a part of some fallibilistic approach to field work but are themselves no guarantee of validity’ (1993, cited in Seale 1999:58). Further, Flick (1998: 230 cited in Denzin & Lincoln 2003) states ‘[t]riangulation is not a tool or a strategy of validation, but an

alternative to validation'. In other words, we should not employ triangulation simply to judge whether our data is true or false but to offer a clearer picture of the phenomenon by viewing it from various angles. This method alone is far from sufficient to deal with validity, and should be used in conjunction with others mentioned above.

5.8 Other considerations

5.8.1 Researcher-practitioner issues

Another aspect of data collection which must be considered, especially in light of the call for 'honesty' to enhance validity, is that of the role of the researcher. Opinion differs somewhat as to exactly how a 'researcher practitioner' might be defined (see Burgess (1984) for example). Gold (1958, cited in Cohen et al 2000: 305) describes a cline depicting the range of possible positions a researcher may take in a research situation. They range from 'complete participant', to 'complete observer'. The relevance of identifying this position is the belief that this role will influence the data collection and analysis. An example might be that a teacher-researcher in a certain research scenario may not have the necessary distance to see issues in that teaching situation but their advantage on the other hand lies in depth of knowledge of the context and possibly informant acceptance. Due to the fact the effect on interpretation is evident, the researcher's role regarding the research situation needs to be made explicit.

5.8.2 Ethics in qualitative research

The anonymisation of data is now almost standard practice in social research. Among its advantages is that it assures confidentiality where the airing of personal views on sensitive topics may have very real personal consequences such as stigma. Research

is conducted with the aim of publication so work must be carefully scrutinized not just for identifying names but also for any other identifying features, such as place of work or involvement in specific public events. Although it is by no means compulsory for all research settings, even when not dealing with particularly sensitive matters the practice tends to be followed. A key rationale for this is that it may not be obvious to the researcher, or it may not indeed be obvious at the time of the research, what the attribution of ideas and beliefs to certain named informants may lead to. Erring on the side of caution is a sound guiding principle.

The extent to which we are prepared to question the potential consequences of our research signifies the depth of our ethical conscience as individuals. Effects for example arising during the data collection process can include stress and even 'changes in self- understanding' (Kvale and Brinkman 2009: 63). We must consider whether we wish to be responsible for such changes in other people. What is more, an ethical researcher should consider possible effects not only arising during but also after the data collection event (Kvale and Brinkmann 2009). To sum up:

'from a utilitarian ethical perspective, the sum of potential benefits to a participant and the importance of the knowledge gained should outweigh the risk of harm to the participant and thus warrant a decision to carry out the study' (Kvale and Brinkmann 2009: 73).

The effects on an informant may arise out of what we decide to do with their 'voice'. Do informants for example get a say in how their words are interpreted in the writing up of the research. As researchers we wield great power in altering others' words to suit the needs of research, both knowingly and also unknowingly and it is incumbent upon us in working ethically to minimize this.

Another consideration is the extent of the anonymisation. This can cause a dilemma since context is everything to achieve understanding of the interviewer's meanings but over zealous efforts at anonymity can distort the picture which the reader is trying to process. Consideration of the agreement regarding confidentiality comes into effect here. It is perfectly normal for, firstly, it to be agreed that the research team will see all the data and that secondly, anonymisation of any data which will subsequently be in the public arena will be strictly adhered to. The level of sensitivity of the topic under research will affect the extent to which either of these stages is realized. These are issues which must be considered when producing transcripts.

5.9 Stages in the research process

5.9.1 Transcription

Given that transcribing is such a time-consuming process and that much of the data does not feature in the final product it may rationally be queried why it is undertaken. The simple response is that it allows a much richer interaction with the data being in such an accessible, complete format. It facilitates the iterative nature of analysis typical of qualitative data, which is in itself facilitated by the use of data analysis software (see next section), which has traditionally worked with the written word rather than the original spoken format, although tagging of audio material is possible in certain programs (such as Atlas-ti).

While transcription is generally seen as necessary, there are a series of important considerations which are too easily overlooked. As King & Horrocks (2010: 144) point out, in qualitative research to date, little emphasis has been put on the transcription process in comparison to other stages of analysis, with the exception of (Poland 2003). Poland, for instance, lists three key threats to transcription quality and they are: recording quality, missing context and tidying up talk. It must be

realized that the production of a transcript is already an interpretative process; the transformation of oral to written form necessitates ‘judgments and decisions’ (Kvale and Brinkman 2009: 177-178) which results in an ‘abstraction which helps to deliver theory’ (Kvale and Brinkman 2009:177).

5.9.2 Analytic procedures

There are many differences between quantitative and qualitative research methods but fundamentally the goal of both is ‘to identify clear and consistent patterns of phenomena by a systematic process’ (Marshall & Rossman 2006: 192). In qualitative research it is often the case that lack of clarity or thoroughness regarding the analysis of data leads to criticism of bias as it is difficult to judge the quality of the analysis or its outcomes without such detail. Therefore description of this messy process, which is the reason why it is so often poorly reported, is important.

There are several key considerations which need to be borne in mind when analyzing data, particularly so when taking a qualitative approach. One of the most important, as Mason (2002: 148) reminds us is that ‘cataloguing or indexing systems are not analytically neutral’. As Kvale & Brinkmann add:

‘There are multiple questions that can be posed to a text in an analysis, with different questions leading to different meanings. A researcher’s presuppositions enter into the questions he or she poses to a text and thus codetermine the subsequent analysis’ (2009: 212).

We as researchers shape the analysis and so need to be clear about what has influenced that shape.

While procedures may differ widely in their detail, some researchers have tried to clarify the main stages of qualitative research. The aim is to move beyond the literal

to interpretive and reflexive reading of the data (Mason 2002: 149). Thus data is coded and re-coded until saturation is reached (i.e. until all meaningful ‘episodes’ are accounted for). Kvale & Brinkmann remind us that ‘meaning condensation entails an abridgement of the meanings expressed by the interviewees into short formulations’ (2009: 202) i.e. we summarise and paraphrase to extract meaning. This allows us to, ideally, move from description to theory (Kvale & Brinkmann 2009). These descriptions offer an overview which leads to a framework, the means of arriving at an interpretation of the data as a whole.

Data, particularly qualitative data, is often messy, and a clear procedure for analysis is vital. King & Horrocks (2010) for instance set out three steps: description of ‘episodes’ in the data, then interpretation of these descriptions which leads, finally, to identification of key overarching themes. Such ‘episodes’ could constitute anything: actions, events, meanings, norms etc (Kvale & Brinkmann 2009). The three steps may not, consequently, be distinctly sequential. Kvale & Brinkmann (2009) suggest another means of interpreting the analysis of an interview, which has 5 steps, which builds on King & Horrock’s stages by adding an initial overview stage. Finally, an even more elaborate seven stage analytic procedure is outlined by Marshall and Rossman (2006). These stages are compared in Table 4 below.

Table 4 Stages of analysis

King & Horrocks (2010:153)	Kvale & Brinkmann (2009:205-7)	Marshall and Rossman (2006:156)
		organising the data
	read whole interview for ‘sense of the whole’	immersion in the data
description	researcher determines the ‘meaning units’	generating categories and themes
	dominating themes from meaning units restated ‘as simply as possible’	coding the data
interpretation	‘interrogate’ the meaning units ‘in terms of the specific purpose of the study’ (p207)	offering interpretations through analytic memos searching for alternative

		understandings
overarching themes	essential, 'nonredundant' themes from whole interview 'tied together into a descriptive statement'	
		writing the report to present the study

The three descriptions of the process are not contradictory; while King & Horrocks focus on the outcomes, Marshall and Rossman, in contrast, focus on the process. All descriptions add constructive pointers to the researcher to help focus on the means and objectives of the analysis. It is such subtleties of procedure, highlighted by the variation in the stages suggested, which are developed with experience.

5.9.3 Data organisation

Data organisation and immersion, stages 1 and 2 in Marshall and Rossman's (2006) schema, may be viewed as essential preparatory stages. The organization of the data is essential if the data is to be efficiently navigated over and over again as is necessary with such means of analysis. The immersion which happens as a result of the organization, which entails the transcription of the data, is also necessary in order to become acquainted with the data. This step helps give a sense of the whole study before diving into a process akin to atomization in the coding stage.

The process begins with the category of organising and indexing. Marshall and Rossman (2006) emphasise the importance of good data management to be effective. It must be remembered above all that it is not the software which organises the data. Any organizational system is only as good as the researcher who loads the data.

5.9.4 Description through coding - search and re-search

One of the most common methods in this stage of the research process when dealing with qualitative data is the sorting stage. By this point inevitably description is already becoming clear and themes are emerging. However, to be systematic in one's approach to the analysis, necessary for thorough, rigorous interrogation of the data, one must examine the data sequentially and describe it in an orderly manner. This does not preclude numerous iterations of this stage. It is in fact necessary.

There are various decisions to be made in the description process given the freedom of the exploratory approach seated in this type of research. Amongst these are 'how to code' and 'where to draw the codes from' (Kelle 2004: 445). These issues will be discussed in turn below once some key terminology has been clarified.

5.9.4.1 Terminology

There are various terms used in the qualitative research literature which are not used consistently so at this point I aim to clarify how I intend to use the terms and clarify the points of reference. Firstly *codes* and *coding* is widely used. This applies to attaching a label to a chunk of text (of whatever size) to allow for retrieval and later examination and comparison with other chunks of text. The terms *tags* and *tagging* are sometimes used synonymously, and the issues around this choice are discussed in section 6.24.3 below). A stretch of language which has been coded can be termed *chunk*, as used already, or *episodes* or *instances*. All three terms will be used here, synonymously.

Category is used in the literature to mean various aspects of the description process. Mason appears to use categories synonymously with *codes* (see Mason 2000: 151). I

have chosen however to use the term category to refer to groups of similar codes.

Themes is the term I have used exclusively for description of emerging interpretation based on codes and/ or categories, identified at the description stage. I have separated them out, although category and theme may be easily confused, for that very reason to try to clarify the interpretation process.

I have avoided the term sub-code, as found in the literature since in my analysis, code is the basic unit, which may group into categories in the hope that the codes and categories lead to themes.

5.9.4.2 How to code

The sorting can be undertaken by the method known as ‘categorical indexing’ (as well as categorizing, assigning nodes or simply coding (Mason 2000: 250)) which involves applying labels to chunks of text which display certain characteristics, as perceived by the researcher. For example, within a set of interviews the informants may discuss their understanding of the concept of teacher beliefs. This section of the text (transcribed interview) would thus be labelled indicatively with an easily retrievable code such as ‘Teachers’ Beliefs’. This process enables themes to emerge, leading us to interpretation, as we see these labels reoccurring, for example, across informants, across locations, across specific situations. It must be remembered that the aim of this stage is description and that the codes are the means, not the ends in themselves.

5.9.4.3 Use of software

It could be argued that the quality of qualitative studies has grown since the availability of CAQDAS (computer aided qualitative data analysis) software. CAQDAS software programmes such as Atlas-ti, NVivo, MAXQda, or Data Miner allow multiple codes, ‘[r]elated and interrelated categories and subcategories’ (Mason 2000: 151),

and also facilitate multiple iterations of naming and re-naming where necessary, retrieval of instances of certain codes for checking and rechecking, and finally also filtering of the data, however large the data set. It has enabled ever more complex data sets to be successfully and more creatively analysed. We are saved from the literal cut and paste techniques and witness minds free from the drudgery of such laborious, rather inflexible methods to better see patterns and themes (Kelle 2004: 456).

We can better manage the complexity yet the basic analytical techniques remain the same and it is perhaps erroneously named analytic software since, as Kelle says, the software is merely a 'clerical tool' for 'data administration and archiving' not 'data analysis' (Kelle 2004: 456). It can only help organize and manage the data (Kelle 2004: 446); it cannot itself interpret. As Mason explains: 'the function of the categories is to focus and organise the *retrieval* of section of text, or elements of data' (2002:151); it does not produce the codes and certainly cannot make meaning from the coding.

5.9.4.4 The source of codes

Codes can be described as being of two basic types depending on their provenance. One group consists of those which we have in mind before we examine the data. These are labeled variously 'concept-driven' (see Kvale & Brinkmann 2009: 202), predetermined, preconceived, or *a priori*. The codes which the researcher discovers only on examining the data are of the second type: 'data-driven' (Kvale & Brinkmann 2009: 202) or emergent codes. It is useful to bear in mind how the codes have emerged when subsequently examining the interpretation. Intentions, biases and misunderstandings can affect the coding process and knowing the type of code can help put right misinterpretations. The use of memos, a facility in most CAQDAS software, to track this is a useful way to track and monitor code stability and change.

5.9.5 Interpretation

As King & Horrocks (2010) remind us, themes are not there within the data waiting to be uncovered, like fossils in a rock bed. They arise as a result of the combination of the data, and a variety of factors such as the researchers' aims and own particular knowledge, background, and philosophical standpoint. As themes arise it is necessary to return to the data and check the coding and where necessary re-code using any newly emerged codes to try to achieve saturation.

Once the 'episodes' have been identified (or labelled with codes) they can then be examined, through inter- and intra-case examination of what these episodes may mean, to extract themes. Themes, what they are and how they are identified, are rarely clearly explained beyond the commonsense meaning of patterns in the data that highlight areas of interest relevant to the research topic (King & Horrocks 2010: 149). King & Horrocks (2010: 150) note that a theme is identified as:

'recurrent and distinctive features of participants' accounts, characterizing particular perceptions and/ or experience, which the researcher sees as relevant to the research question'.

This may be seen as not particularly more specific than the general definition given above. It is important therefore to be clear about the process of identifying them and of how that identification leads to interpretations so that it is clear within each study what is meant by them. It is necessary to 'formulate explicitly the evidence and arguments that enter into an interpretation' (Kvale & Brinkmann 2009:212) so that its quality may be evaluated.

The identification of themes is most usually a highly iterative process, with one theme uncovering the hint of others which then must be searched for in the data; the emergent themes must be examined to see what is revealed. The interpretation process should indeed be iterative since the researcher needs to return to and

interrogate the outcomes of an interpretation to avoid impulse and instinct. Some themes will be highly interactive and present a strong case by consisting of information on a certain topic from various angles. Others will not be substantial enough to merit further inclusion in the analysis: 'the researcher searches for those that have internal convergence and external divergence, i.e. internally consistent but distinct from one another' (c.f. Guba 1978 cited in Marshall & Rossman 2006: 159).

A further analytical technique, to test the robustness of themes, is negative case analysis. This simply entails selecting the key identified themes, and returning the data to check for counter-evidence. This is best accomplished on an on-going basis to most efficiently use research time. For example, if a promising emergent theme is examined in this way and soon found not to be robust in light of the evidence found it can then be abandoned.

To some extent the quantitative influence cannot be avoided since an occurrence of a particular 'episode' is generally seen as notable or sufficiently weighty if it occurs frequently enough (Kelle 2004: 55), as outlined already in discussing theme identification. Trying to find strands of coherent ideas which may lead us to a useful overarching interpretation without this method would most likely be very hard, and result in incoherence. Whereas other methods such as content analysis rely on quantification of such codes, grounded theory (to use this term to contrast with other non-qualitative methods, even if not all are based on 'hard-core grounded theory') does not rely on quantification to such an extent and data is analysed by looking at relationships between codes and their context.

5.9.6 Presentation of the research

Depending on one's own writing habits, this final stage could be viewed as merely the tangible outcome of the study, achieved only once all other stages are complete and is simply the physical manifestation of all the prior analyses. Alternatively, it may be viewed as the stage at which the final analysis is achieved, via the writing process. Those who advocate the writing as a thinking process (see Ivanič 1998) will approach this as also being a stage in the interpretation.

It is often overlooked, as in King & Horrocks' three stages, that the formulation of the interpretation into a form which is suitable and accessible for 'public consumption', even if it is a public knowledgeable and interested in the area of the research, is itself another layer of the research process, where the choices concerning inclusion and omission to achieve clarity are made. There are also choices to be made in the style of how to write up the interpretations. 'I interpreted this event' or 'the data revealed' convey very different approaches to the research (Marshall & Rossman 2006: 164). Writing up is often overlooked, or underplayed at least, but communication of the research outcomes should be viewed as just as important as prior stages since if the outcomes are not made clear, accessible and open to evaluation then the aim of contributing to the body of knowledge in academia, is diminished.

In summary, the four key stages of analysis can be reduced to Preparation, Description, Interpretation and Presentation. It must be noted that these are key components of the analytical procedure but are by no means distinct. There are a multitude of decisions to be made at each step of the research process from design, collection to analysis and presentation. A key decision is also a matter of when to stop analysing and make meaning.

5.10 Summary

Returning to the Einstein quote at the start to this chapter, I wish to interpret this as a reflection of the main rationale for the choice of methodology in this study. The qualitative approach avoids quantification in an effort to better explore and represent the nature of the object of our research. It is a complex, vulnerable process which requires rigorous documenting to approach a robust version of events, and welcomes and manages the detail in the research as it is unclear 'what counts' until this emerges.

In the next chapter I will discuss how I applied the theoretical consideration into practical terms and undertook the design, data collection, analysis and interpretation.

6 METHODS – INSTRUMENT DESIGN, DATA COLLECTION AND ANALYSIS

6.1 Chapter overview

This chapter demonstrates the methods used in this study, which arose from the theoretical considerations discussed in the previous chapter. I have proceeded on the basis that transparency is more important than representativeness, seeing as the latter is often not feasible in small-scale studies such as this. I aim to offer a rich enough description of the three cases I chose to study for other researchers to subsequently judge, how or whether lessons from this study can be transferred to other situations.

6.2 Theoretical framework

In the previous chapter the role of frameworks to guide a piece of research was discussed. A model, which is a diagrammatic representation of a complex system, often has both relational and procedural dimensions. In other words it seeks to explain the relationship between the constituent parts of the system and also to represent how the system's procedures operate. These can thus be invaluable in shaping the form and scope of research. The Henrichsen (1989) model (see Figure 4), as used by Wall (2000) in the study of the impact of the 'O' level exam in Sri Lanka, and Wall & Horak (Wall & Horak 2006, 2008, 2011) studying the impact of the introduction of the iBT TOEFL²³ was chosen for use in this study. It proved to be a useful guide to considering washback as it goes beyond a description of washback to the effect (impact) of an innovation in an educational system. It is a hybrid of various previous models and studies in innovation, not necessarily solely within the field of

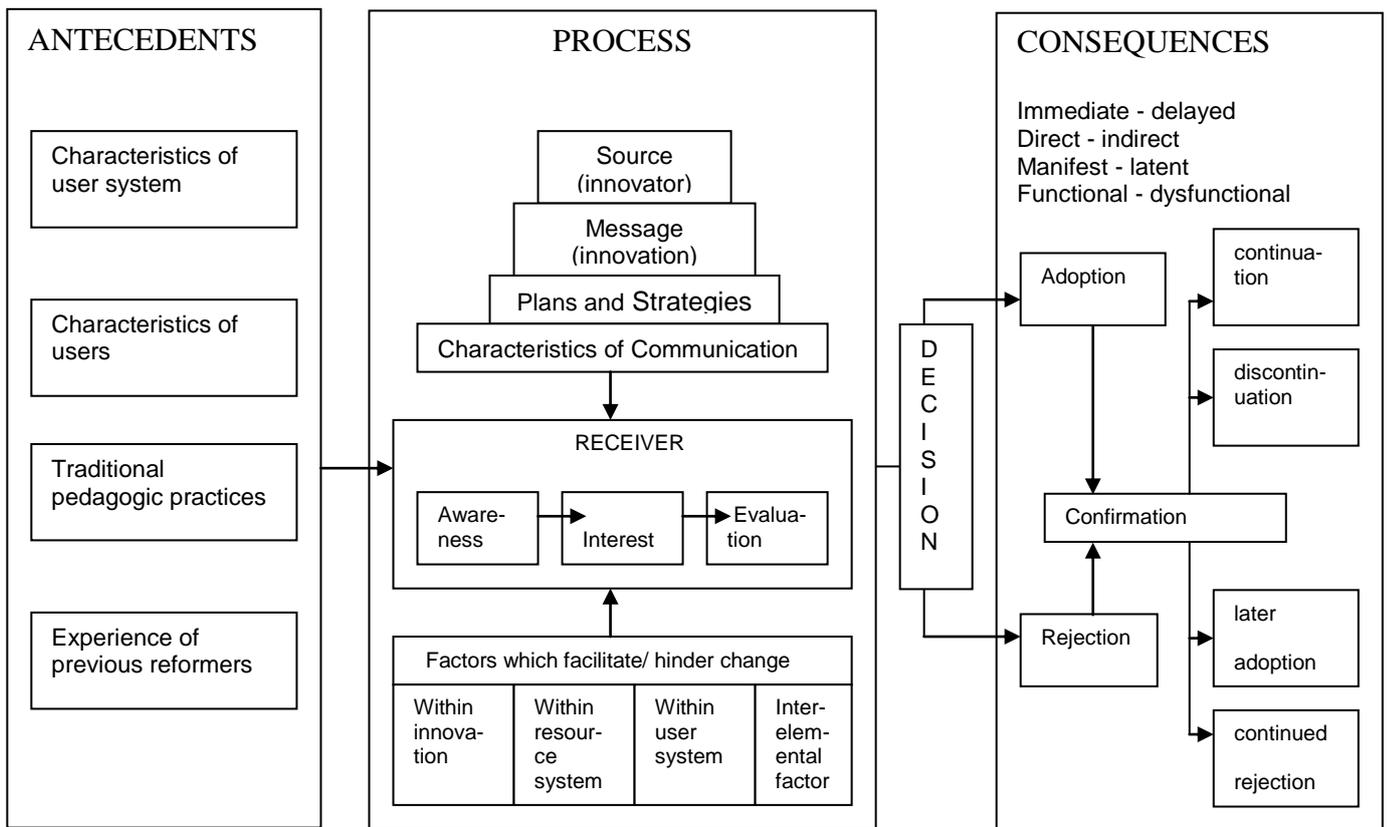
²³ The latest version of TOEFL, which is internet based.

education and highlights the importance of considering the full life of the innovation, not merely the point of its introduction.

While also drawing on various theories discussed in washback literature to think about the phenomenon (see previous chapter), the Henrichsen model seemed most appropriate since other models I had considered, such as Hughes' (1993) or Bailey's (1996) washback models (see Chapter 4), while useful in helping to conceptualise washback, seemed too simplistic to capture the complexity which I anticipated since '[N]umerous other factors than the exam are involved in determining what happens in the classroom' (Watanabe 2004: 23).

The Henrichsen model consists of three distinct interlinked phases. Firstly, the 'Antecedents' section aims at describing the situation prior to the introduction of an innovation. The 'Process' refers to the period of implementation of an innovation (e.g. the introduction of a new exams). Finally the 'Consequences' section of the model can help guide research in reporting the outcomes of the diffusion process and whether and how the innovation was successful.

Figure 4 Henrichsen hybrid model of diffusion/ implementation of innovation (1989)



This model will be referred to again throughout this chapter to explain how it was used to shape the instruments developed.

6.3 General methodological considerations

In aiming at the ‘honesty, depth, richness and scope of the data’ as mentioned in the previous chapter (Cohen et al 2000:105), the data collection will be described in as much detail as the limitations of this study allow. In this regard I have taken honesty to mean being as transparent as possible, regarding my own views towards the topics under study (see below) and being as open as possible with the informants also. There is an inherent tension in trying, on the one hand, to provide, an account which relates the real nature of a study (referred to previously as the ‘research journey’), not an idealized version, and on the other, to tell a coherent ‘story’.

The methods used to collect data in this study can be viewed in two stages: the first was an exploratory study and the second comprised the main data collection procedure on which the majority of the analysis is based. Each stage will be discussed in turn below.

6.4 Stage 1: Exploratory study

The study began with reading official documentation concerning Skills for Life, media coverage and informal interviews with ESOL teachers; some interesting themes began emerging. From this process several preliminary research questions formed about the range of assessments being used and what effect they were having on ESOL classes.

The research questions for this study were simply:

- what is the full scope of assessments of ESOL students which take place?
- how have Skills for Life assessments affected the ESOL teaching?
- how has Skills for Life affected the assessment of ESOL students?

I decided an exploratory study was needed to investigate the issues regarding the changes in assessment practices being brought about by the Strategy in order to help me design the main study and better formulate the research questions, namely address the boundaries, as the nature of this study is exploratory rather than testing any specific hypotheses. Given unlimited time and resources such boundaries may be less vital but in a study of this type some idea of the extent and direction of the investigation, without in any way prescribing the potential findings, are necessary.

I was interested in a small-scale in-depth approach and rather than merely describing current practices I wanted to try to explore why they occur. I therefore decided interviews backed up by observation were the best way to proceed.

In the exploratory study I sought information which told me more about the stakeholders and their environment, or, in the terms the model uses, about the 'Users' themselves, as well as the 'User system' they worked within i.e. the institutions and programmes of study they were involved in, as well as their then current 'Pedagogical Practices' as regards assessment. ('Experience of previous reformers' is not dealt with as this is not relevant since no comparable overhaul of assessment of ESOL students had previously taken place). Even though the new Skills for Life exams had not yet at that point been accredited, awareness of imminent changes was known so I was interested in categories suggested by the Process stage of the model, primarily Characteristics of Communication and the Receivers (of the innovation).

In washback studies the norm is to undertake a baseline study in order to better describe and understand the situation prior to the change which is the subject of the study e.g. the introduction of a new exam into an educational system. The reason for not undertaking a baseline study in this case, was that there was no centralised system of exams prior to the period of data collection in this study. The ideas suggested by the Antecedents stage of the model assisted in building a picture of the previous state of affairs, before the Skills for Life strategy has been introduced, for example, information about the teaching situation, and the nature of the ESOL students (provided in Chapter 2).

6.5 Institutional and teacher profiles

In order to situate this study, and enhance transferability of findings, key aspects of the data collection need inclusion. The type of institution and the nature of the teachers interviewed are thus outlined below.

6.5.1 Institutional profile

The exploratory study, 'A week in the life of ESOL', was undertaken at a local Adult Education College in May 2004. The college used for the exploratory study is based in a medium sized town. It at that time²⁴ offered a range of 'hobby' courses as well as 'care in the community' (offering educational opportunities for students with learning and health difficulties) and vocational and certain academic (mostly foreign language) courses. It offered a range of qualifications in various subjects up to NQF Level 2 as well as 'leisure' courses, and was relatively well resourced in terms of IT resources, classrooms and social space for students and staff. For the type of students it aimed at and for the range of courses offered it was deemed a fairly typical example of this provision.

6.5.2 Teacher profiles

The three teachers who were interviewed and the DoS had mixed teaching backgrounds and experience but had trained and worked as EFL teachers and had more recently moved into ESOL. All had worked abroad teaching EFL at some point in their careers. They were qualified to PGCE and /or MA level. They all had some experience with various EFL exam boards as item writers and/ or examiners, which I believe may have given them some privileged insight into the examination process since such experience involves examination of the exam production process and

²⁴ Since then, due to funding changes, the profile of courses available has changed drastically.

depending on the training involved some insight into testing theory, even though probably at only a basic level. All the teachers in the exploratory study (not all were interviewed) are profiled in Table 5 below to give a picture of the staff at this site.

Table 5 Teacher profiles

	Interviewed	Years of teaching experience (rounded up)	Highest EFL related qualification
T1	✓	20	MA
T2	✓	20	MA
T3	✓	20	Diploma
T4	✗	15	Diploma
T5	✗	5	Certificate
T6	✗	20	MA
T7	✗	15	Diploma

6.6 Observations

I sat in on ESOL classes to undertake the observations, locating myself outside the main working area but sometimes joining in with the lesson or helping the teacher as requested. The schedule I used was based on observation schedules I had previously designed and used in other small scale action-research projects but had adapted in light of my experience for this occasion

6.6.1 Format

A framework was needed to systematise the activity I observed. The format chosen for the exploratory study was a semi-structured observation schedule. It was not a highly structured instrument such as the COLT A/B - The Communicative Orientation of Language Teaching Observation Scheme devised by Spada and Frohlich (1995) since I did not intend a quantification of the results but rather needed a much more descriptive analysis.

It was designed to allow a description of the class, a systematic recording of certain key features (such as any mention in class of exams or assessments). The data was intended primarily to take on a support function by providing a context for the interview data. The format facilitated a mixture of event sampling, allowing specific features to be focussed on, to be searched for systematically, and also for instantaneous sampling, which offers a chronology and provides the context of the lesson (Cohen et al 2000: 308). It incorporated the categories already discussed regarding the physical, human, interactional and programme settings (Morrison 1993). The instrument was consequently composed of four parts:

Table 6 Description of the observation schedule

Section	Content	Rationale	Morrison's (1993) observation framework
Part 1	Checklist for assessment specific features	Systematicity	human
Part 2	Chronological description Codes & timings	Context Highlights incidents related specifically to assessment	interaction/ human
Part 3	Summary notes specific notes on critical incidents	Global view (sum of parts)	
Part 4	Space for class map, use of resources etc.	Aide-memoire, context	physical/programme

See Appendix 2 the observation instrument used

Part 1 consisted of a list of possible typical behaviours which may be observed. As they were noted in a particular class, these were ticked off on the list. Part 2 required more in-depth analytical input. This comprised short notes describing the main activity observed (e.g. Teacher explains task) and a timing to indicate length of each activity. The interaction patterns were also noted here (e.g. Teacher to individual student, plenary, pairs, small groups) to give some indication of the class teaching style. After the class ended a quick summary section was completed which focused on

critical incidents observed such as a teacher explanation of a linguistic feature which may occur in the exam e.g. the teacher emphasized the importance of correct spelling in an Level 2 exam. The final section (part 4) comprised a space to draw a rough map of the classroom and note relevant equipment and resources used in the lesson. The observation sheet can be found in Appendix 2.

6.6.2 Recording

As I was not intending any detailed interaction or discourse analysis of the classrooms I did not make audio recordings of the sessions. Moreover I did not want to make any of the students uncomfortable through the presence of video camera since some are in delicate situations as regards their official status in the country. The issue of trying to address researcher bias through taking full and thorough notes and using various methods of recordings had to be compromised in this case. Notes were as detailed as possible according to the framework I had devised and sometimes this included short sections of conversations for illustrative purposes.

6.7 Interviews

The interviews in the exploratory study were more heuristic and conformed less to a strictly prepared question schedule than did the subsequent interviews (for the main data collection). The aim was primarily to augment my own understanding of the topic so I needed to allow the teachers to 'tell their own stories'. I wanted a framework nevertheless to help set boundaries, as mentioned above, and this was thus produced based on the categories within the Antecedent stage of Henrichsen's model (1989).

My interview schedule was based on a core of questions relating to the categories listed above but was aiming to be flexible enough to allow for anecdotal evidence and

additional information which the teacher may see as relevant. This was very important since, as already stated, this was an exploratory study and this approach indeed proved very fruitful.

My interview questions covered areas such as the decision-making process regarding the choice of external exams and the influences upon it, how informed students, teachers and the DoS were in terms of the exams available, how exams were promoted, which exams were chosen by the DoS to be made available (if any, and if none why not), what pressures the providers were under to enter students for exams and whether, to their knowledge, this was uniform nationally or whether local conditions exerted different pressures, how internationally recognised the 'EFL' type exams were and how widely recognised the Skills for Life exams were. As regards internal assessments I was interested in the extent to which its monitoring function was perceived as offering added value or as simply a matter of 'jumping through hoops' for the sake of securing funding.

I interviewed the director of studies (DoS) prior to the study to gain approval and understanding of my research, and to gain an overview of the ESOL provision. I then undertook observations of as many classes as possible (14 in total) during one week and then interviewed three teachers in depth.

6.8 Opportunistic data

Fortunately, as well as the teacher and director of study interview data, I also managed to collect some additional unplanned data in the form of interviews with a group of students who had just taken a Pitman's ESOL oral exam. This gave me insight into their reasons for taking an exam, their preparation process and perceptions of the exam itself.

As well as the interviews and observations, I was able to sit in on a staff meeting, which was also useful in witnessing part of the process of internal assessment. The teachers were discussing students' levels of attainment in terms of the Skills for Life Curriculum and it was useful to observe how such decisions were reached. This also informed the later interview questions in provoking the inclusion of a wider view of assessment than just the external exams.

6.9 Analysis

Summaries of the interviews were made and observation notes as well as all field notes were scrutinised. Key points were then transferred to a grid (informants and themes as the axes) for better comparison of responses. Observation notes were added to this grid where my observations backed up any points the informants had made. Qualitative analysis software packages (such as Atlas –ti or NVivo) were not used to analyse the texts in detail as it was not felt necessary at this stage due to the relatively small amount of data. The aim of the study was simply to draw out major themes and these could be extracted by careful iterative readings and note taking.

6.10 Outcomes of the exploratory study

The importance of understanding the informants and outlining their profiles as far as possible is that comparison with other sites may be understood better for knowing how similar (or not) the informants are on a range of features. This is important for the transparency (as discussed in the previous chapter) in that the suggested outcomes can be put forward by the researcher but with enough information the reader can make perhaps equally valid conclusions given the information provided.

This is in the nature of qualitative research that the outcomes made can only be proffered as one possible interpretation of the observed phenomenon.

6.11 Emerging themes

The key themes from the teacher and DoS data which proved fruitful in the exploratory study were as follows. They are grouped according to categories of the Henrichsen model. These are the topics which were subsequently built on to produce the interview schedule for the main study.

Table 7 Key themes emerging from the exploratory study

Henrichsen category	Topic
Characteristics of the user	Profile data (training, background etc)
	Attitudes to assessment in general
	Views on testing for citizenship
	Awareness of current range of exams
	Evaluation of current range of assessment practices
Characteristics of the user system	College Profile
	Differences between EFL and ESOL students
	Confusion and frustration over new qualifications requirements
	Relationship between exam results and college targets
	Decision making process re. exams taken
	Appropriacy of the exams being taken
	Quality of diagnostic and placement testing
	Role and purpose of ILPs
	Problems with the way ESOL teaching is currently organised for students, for teachers, for society
	Stakes related to various assessment practices
Current practices	Perception of increased workload from cross-referencing materials and records to the curriculum
	Quality and use of diagnostic and placement testing
	Achievement testing – internal and external measures
Characteristics of Communication	Channels and flow and nature of information – Government to Teachers

	Flow of information between teachers
	Satisfaction level re. quality of communication
Receivers	Awareness of the new S4L rationale
	Awareness of the new S4L exams
	Awareness of role and purpose of new Skills for Life exams
	Evaluation of Skills for Life

The students' responses to questions concerning why they were taking the exam, and their attitudes to taking exams, were varied and not what I expected so I decided therefore to include student interviews in my main study (see Appendix 3). What became apparent was that they did not have the same reasons for taking the exam and were not necessarily taking the exams as seriously as each other. For this reason the main study investigated differential stakes, which was an aspect not prompted by the Henrichsen model.

Regarding washback, the main finding from the exploratory study showed that external exams did not impinge on the content or methodology of the ESOL classes to any great extent at that site at that time (apart from the obvious exception of the one class whose sole aim was to prepare students for the IELTS exam). What had a far greater influence were the internal assessment mechanisms and records such as the ILP (individualised learning plans), and class workplans, exemplified by the amount of class-time spent on completing various pieces of paperwork needed for monitoring and tracking purposes by the college. What prompted further study of the external exams was the DoS's and tutors' concerns about the effects which they anticipated. They suggested washback would be inevitable, which arose as a strong theme given the then current understanding of what was about to be instigated in terms of the centralised set of external exams, a new phenomenon for ESOL in the UK.

6.12 Stage 2: Main study

6.12.1 Revised research questions

As a result of the exploratory study the research questions were refined and consolidated. They were consequently framed as:

RQ 1.a) What is the range, nature and function of assessment practices in UK ESOL teaching?

RQ 1.b) How are these practices linked to the Skills for Life strategy?

RQ2.a) To what extent is there evidence of washback from the assessment practices?

RQ2.b) Is any washback related only to the assessment practices resulting from Skills for Life?

RQ3) What are the factors which may drive washback?

The diagram in Figure 5 summarises the main aspects of this research. The Research Questions (RQs) are indicated in bold, and the diagram aims to show how they relate to each other. The 'call out' boxes indicate the sources of data to be drawn on for each aspect.

the format, for example, T5-12:1425, which translates as: these words are from the teacher labeled as Teacher 5 for the purposes of this study, document (in this case an interview transcript) number 12, line 1425.

6.13.1 Teacher profiles

T1 was an experienced teacher having taught for over twenty years. Her initial training had been in teaching modern foreign languages and she later moved into teaching English as a foreign language (EFL), beginning in a volunteer teacher capacity. Then after moving to her current hometown she gradually moved into ESL work (T1: 2.28). In her current post she tended to teach the higher level classes, having only students wishing to take Cambridge main suite exams (T1: 2.1281). In this respect her work was more of the nature of EFL teaching than ESL. She also acted as exams co-ordinator for ESOL for her college (T1: 2.592) which entailed the administrative work of organising exam timetables and related logistics. Teacher 1 was generally less forthcoming than the other teachers during interview.

T2 had more than twenty five years' of teaching experience, and was thus the most experienced teacher in this study in some respects, although this appellation must be tempered by the fact that all her experience had been in the same institution. She had worked in the field of ESL nearly all of her working life, having initially trained as a fully qualified state school teacher after finishing a degree at university, but soon moved into ESL work combined with her main subject and then after a career break moved into ESL full-time (T2: 3.03, T2: 3.27). She tended to teach up to Entry 3 level students, i.e. lower level ability students (T2: 3.623).

T3 had also undergone a teacher training course (PGCE) after her university degree and had briefly taught modern languages in the English state school system before

moving abroad to teach English as a foreign language for a few years (T3: 6.20). She began ESL teaching on her return to the UK in another town (T3: 6.73). She acted as a team leader within her department. Unlike other teachers she talked of a career in ESOL, planning to move into teacher training (T3: 6.156). She taught a range of classes and had recently started teaching IELTS preparation classes (i.e. high level students needing qualifications to pursue higher education, professional training or permission to undertake professional practice) and reported that she found it quite challenging (T3: 7.434).

T4 had the most varied background of all the teachers in that he had worked in various professions before training to become an ESOL teacher only approximately two years prior to this research. He was enthusiastic about his work and his colleagues and appeared very keen to learn and try out new ideas (T4 9.702). He believes a certain amount of pressure is good for both teachers and students (T4 9.433).

T5 had trained as an EFL teacher and worked abroad for one year before returning to the UK and taking up ESL teaching (T5: 12.157). She had gone on to study for an MA and was consequently the most highly qualified teacher in this group. She had gone on to undertake work in a variety of areas within EFL, such as EAP (English for Academic Purposes) and ESP (English for Specific Purposes), within the region. She worked part-time at her current institution, holding simultaneously other posts elsewhere (T5: 12.22) so had a point of comparison between Site 3 and other institutions regarding such matters as management, organization of teaching and reactions to Skills for Life.

T6's main teaching experience was within ESL. She had trained more than five years before at the same college where she now worked having successfully completed the

initial teacher training programme offered by the same college (T6: 13.195). She held the position of team leader. However, she worked only part-time which had implications for the work-load that this role entailed (T6:12.487). T6, clearly enjoyed her work but also had a clear sense of her boundaries between work and her home life and was scathing of those teachers who allowed their ESOL work to take over their lives (T6: 13.592 and 13.617).

In the recent ESOL Effective Practices Project (Baynham et al 2007), the profiles of the forty teachers taking part in that study were not dissimilar to the teachers in this study: the number of years of experience ranged from one to thirty, the average being about ten; the average time teaching at their current college was about four, with a range from less than a year to over twenty. These teachers were also divided between those whose main experience teaching English had been gained overseas and those who had worked primarily in ESL in the UK (Baynham et al 2007).

The teachers in this study will be discussed further in Chapter 9 (see Table 18 for further points of comparison).

6.13.2 Directors of Study Profiles

The three Directors of Study varied quite interestingly in their backgrounds, which offered various perspectives on the main issues under discussion regarding assessment and testing.

DoS1's provenance was firmly from within the world of Basic Skills. She was primarily a literacy teacher and had some experience of numeracy teaching but not of ESOL (DoS1: 1.34). She, however, at the time of my research due to circumstances at her college, was managing the Basic Skills team, which included ESOL. She had had some

training by shadowing her predecessor but this had not included experience of ESOL classrooms (DoS1: 1.45).

DoS2 was shaped by her experience in state sector teaching at secondary level (i.e. from ages 12 to 16 approximately) which she had been engaged in for more than twenty years since the dominance of the assessment regime in state education arose in our interview several times. She had moved to the Further Education sector about five years previously (DoS2: 4.61) in the role of ESL teacher (DoS2: 4.53) and had only recently become DoS. She spoke positively about the role assessment plays in teaching (DoS2: 5.31), but without going into any detail or specific exemplification.

The background of DoS3 was initially in ESL, as it was termed at the time, more than thirty years previously when he began work in that. He also had many years experience of working overseas teaching English as a foreign language (EFL), as well as engaging in teacher training and trainer training in the same field. On return to the UK he had worked in teacher training and then returned to ESL (DoS3: 10.25). He reported having reflected on the role of testing and having gained from working with experts in testing in one of his previous posts (DoS3: 11.715).

All the Directors of Studies still had teaching duties in their ESOL departments, some quite substantial (DoS1 18 hours, DoS2 6 hours, and DoS3 16) which meant they kept in touch with the issues their teachers were facing on a daily basis.

6.13.3 Institutional profiles

All three sites offered ESOL classes both at the main college site and also at satellite centres for those members of the community needing not only very basic ESOL teaching, e.g. for students with no English whatsoever, but also an element of social

contact with other students newly arrived in the country. The satellite centres generally are designed to be more accessible in two ways: firstly, geographically in that they are not within the main college setting which is not necessarily local to the communities they aim to target. Secondly in affective terms they have traditionally tried to be less academic in nature and just as interested in the process of settling into life here as they are in language development. Details of the sites can be found in Appendix 4.

6.14 Interviews: content

The interviews commenced with profile questions to establish the interviewee's experience and background. In addition, the more factual questions were asked first in order to 'warm up' the interviewees before moving onto the questions requiring opinion or more in-depth consideration. Thus, what I felt to be the more challenging questions, came later on. Various questions were posed indirectly via a scenario setting technique such as 'If you had a new member of staff who is new to the world of ESOL and you were asked to give them a brief background to the recent changes in ESOL teaching, what would you concentrate on/ tell them?' These question types had worked well and proved fruitful in the exploratory study. The interview schedule consisted of 69 questions in all and the average length of interview was 2 hours (see Appendix 5).

In the following section I will describe the differences between the DOS and Tutor interview schedules, the student interviews and various key considerations for their production.

6.15 Interviews: format

The interview schedule drawn up for this study most resembles the third of four types of interview schedule listed by Cohen et al (2000: 271) namely 'standardized open-ended interviews', where the 'exact wording and sequence of questions are determined in advance' (see Appendix 1 for the interview schedule). In the standardized open-ended type of interview 'all interviewees are asked the same basic questions in the same order' and, as Cohen et al suggest, this method 'facilitates organization and analysis of the data' (2000: 271). However, the order was not observed strictly; some deviation was allowed where I deemed this likely to yield useful supporting information not elicited by the prepared questions.

The advantage of undertaking all my own interviewing, using a schedule which had come out of the previous study, was that I was absolutely clear about the aims of each question. Adhering exactly to the wording of each question was thus not necessary as it may be for a larger study involving a data collection team in which consistent wording may be necessary to avoid multiple-interpretations of the point of the questions, which may affect later data interpretation.

In addition, the other feature of such questions i.e. being open-ended, 'can [also] result in unexpected or unanticipated answers which may suggest hitherto unthought-of relationships or hypotheses' (Cohen et al 2000: 275). The teachers had freedom of response in order to open up channels of information and not to restrict the scope of their answers.

Both the Teacher and Director of Studies interview questions were grouped into themes, all of which were encapsulated by Henrichsen categories and aimed to cover the main themes of the research questions from various perspectives. I reorganised the questions after the first draft so that they sounded more natural in their topic

progression. The topic areas covered are set out in Table 8. The full interview schedule, annotated to indicate which questions dealt with which area below, and which RQs they were aimed at is available in Appendix 5.

Table 8 Topics covered by the interview schedule

Topic areas	Henrichsen categories probed by this topic
Background (teacher profile)	Characteristics of the Users Characteristics of the User System
College profile	Characteristics of the User System
Communication	Communication Characteristics of the User System Characteristics of the Innovation
Assessment practices internal assessments	Characteristics of the user system Receiver awareness Receiver evaluation Characteristics of the Innovation
Assessment practices external assessments	Characteristics of the user system Receiver awareness Receiver evaluation Characteristics of the Innovation
Attitudes to assessment	Characteristics of the user
Citizenship issues (stakes)	Receiver awareness Receiver evaluation Characteristics of the user
Effects of Skills for Life (ESOL)	Characteristics of the Innovation Receiver evaluation Characteristics of the user Characteristics of the user system

The student interview questions (see Appendix 3) were based on the main themes which had come out of the student interview in the exploratory study. They were much shorter and simpler than the Tutor/DoS interviews. These questions had a very specific focus, namely the student's perception of any washback which may have occurred in their classes (question 4-9), their reasons for taking the exam, namely by choice or coercion (question 10-12), the stakes involved (question 13 combined with 10) and the values of certification versus simply receiving a grade (question 14).

Questions 1-3 were for warm up purposes primarily and for checking which exam they had taken.

6.16 Considerations entailed in the methods employed

6.16.1 Two way benefit

The research process need not necessarily only be to the benefit of the researcher. As Cohen et al (2000: 273) say (citing Kvale 1996: 30), it can be ‘a rare and enriching experience for the interviewee, who may obtain new insights into his or her life situation’. I hoped that a rough cost-benefit analysis for both researcher and informant would result in a balanced outcome. I considered this two-way benefit model of data collection to be an ethical approach which I felt comfortable with. Most participants were enthusiastic to talk about their situation and though certainly not wanting to stage a ‘therapeutic’ type interview, I felt confident that this approach would yield richer data than ‘mining’ data in a closed- response questionnaire type interview, which allows little of the informants’ personality and voice to influence the outcome. While this can be a danger in that interview, unless well managed can easily stray from the research focus, I believe the extent to which it enriches the data outweighs such dangers. I intended the research to be a positive experience for the informants and this informed the choice of interview format.

6.16.2 Researcher practitioner issue

I would place myself half way along Gold’s (1958) cline from ‘complete participant’, to ‘complete observer’ since I sat in on ESOL classes to undertake the observations, locating myself outside the main working area, but sometimes joined in with the lesson since the teachers were aware of my EFL teaching background and also experience teaching ESOL classes, and recruited me to help with certain activities. I

felt that having some insight from my professional background but yet having the advantage of sufficient distance through not experiencing the day-to day reality of ESOL teaching within these teachers' institutions helped me observe more clearly and question better.

6.16.3 Recording

In an effort to reduce interviewer bias all interviews were recorded, with the consent of the interviewees. I needed a record of the interaction and wanted to keep note-taking during the interview to a minimum to be able to concentrate on the interviewee and employ active listening techniques and to make the situation seem as informal as possible to maximise the conviviality of the situation as I believed that the more relaxed the informant felt, the richer the data I was likely to obtain. I also felt too much interesting data would probably be lost if I relied only on concurrent notes. Digital recording was chosen for ease of storage and transfer of data. In addition, the equipment is both robust and small enough to be discrete while providing good sound quality.

6.16.4 Interview management

The questions, although planned in advance, were not delivered verbatim from the interview schedule (as already mentioned) to allow a more natural, more conversational-like discourse. The overall interview style aimed at was what Mason has called 'conversation with a purpose' (Mason 2002: 67). The purpose of this is to make the interviewee relax, on the assumption this will provoke more in-depth responses and reduce inhibitions. In order to manage such conversations and maintain as natural a flow of interaction as possible while allowing informants to 'go off at a tangent' I had to be well prepared. I wanted to pursue such 'deviations' but

then engineer the discussion to return to the schedule. I had to bear in mind any points raised by the informants and weave them into the overall picture being built up as we talked and tried to avoid asking questions concerning topics which had already been covered (as mentioned above in section 5.5).

Interviews took place in the tutors' place of work, somewhere quiet to facilitate conversation and enhance recording quality where we could not be overheard so interviewees did not feel inhibited. I made detailed notes after the interview concerning the atmosphere of the interview, my perception of the interviewee's attitude, willingness or reservation, or other characteristics, and other noteworthy details as I was aware this detail would probably be lost in the subsequent transcription of the interviews.

6.17 Observation

Since the observation schedule had worked well for the exploratory study, it was used in nearly the same format for the main study. The only minor alterations were to add the sheet to complete the students' profiles during the teacher interviews, to add guiding questions for post-observation notes in Part 4, and guiding, reminder questions in Part 4 to pose during the interview, linking specifically to the observation.

6.18 Ethics

One of the issues as regards the ethical considerations of the data collection was that all interviewees were made clear who would see the data and consent forms were signed which confirmed that they understood what the research was for and that they would remain anonymous. I assured them that they were free to say what they wished

and that no management (or other management in the case of DoSs), students or colleagues would be aware of anything they told me. I also checked the questions carefully to ensure none of them could be seen to be posing 'trick questions'. I had no intention of trying to catch them out in any way or even present them with anything distressful (Mason 2002). Thirdly, I made clear before we began that they should simply tell me if there were any topics which, as they arose, they did not wish to discuss for whatever reason. In addition, since the informants had all volunteered to participate in the study I was confident that they would feel as comfortable as was possible during the interview process, which was of value to me in gaining more reliable data (in the sense of representational) than that which may be received under duress of any sort. (See Appendix 6 for a copy of the letter of consent).

6.19 Hawthorn effect

The conditions in this study are not in any way akin to those of the original study from which the term Hawthorn Effect derives²⁵, i.e. in this case it is used with reference to the effect of the study itself on the study participants. It always needs considering, but there was no evidence of this obtaining here. It is suggested that long periods of time are needed with a group of study informants in order to reduce the reactivity effect (see Cohen et al 2000: 311). As regards the students, in this study this was not feasible however, and in addition it was not necessary since the students are used to different people being in class, this not being unusual in the teaching environment. The teacher told them who I was, introducing me with my teaching 'hat' on to allay any potential fears of being someone from 'officialdom', considering the socio-political status of some of the students, such as the asylum seekers in some of the classes. The students, and indeed the teacher may of course have been behaving

²⁵ Hawthorne Effect refers to the paradoxical phenomenon whereby it is recognized those observed may not behave as they normally do, which is the purpose of observation, due to the very fact they are being observed.

differently from usual but all the teachers suggested to me this was not the case. The students seemed at ease and not concerned with my presence. The observer's paradox cannot be solved but must be considered to ensure its minimalisation.

Regarding the teachers since I observed them on more than one occasion this gave an indication of their teaching style and whether they too might have been altering their behaviour in reaction to my presence. They too are used to being observed and also since they had all invited me to their classes, I was confident they were comfortable with me being there, and not adjusting their teaching for any reason linked to my study. They were unaware that washback was the focus of my study. In addition, all teachers were used to regular observation linked to teaching quality assurance procedures.

6.20 Convenience Sampling

Before detailing the institutions and the main informants I need to discuss the matter of sampling. The inclusion of teachers in this study was on a self-selecting basis as I was reliant on teachers wishing to co-operate with me for whatever personal reasons they may have had. With that came the disadvantages of such a sample as discussed in the methodology chapter. It, of course, means that the likelihood of more proactive, interested, informed, and critical members of staff within the three ESOL departments I approached being included is higher since those more indifferent towards their profession and the current issues under study were less likely, in fact probably totally unlikely, to want to volunteer as subjects. It is important to be open about the fact that all voices in the staff room were not represented. This does indeed bias the results to some degree but does not invalidate the results since what the teachers recounted to me was what they believed and was not altered by what other colleagues may have reported. It may be argued a balanced picture of the situation

was not reached but the value of the in-depth, small scale approach is that the goal is not an 'averaging' of opinions but the offering of 'a buffet' of points of view.

I had to rely on the goodwill and curiosity of my informants to participate. One particular strength of the self-selection mode of sampling is that the informants' candour can be relied on more than if they are coerced in any way. With the self-selection model there are clear implications regarding the representativeness of my informants which must be acknowledged as this may be viewed as affecting the validity of the research.

There is not a balance of the sexes in this sample of teachers and DoSs, but the balance there is (i.e. 1:4.5 male to female) roughly reflects the current situation in ESOL teaching judging by information from practitioners (via personal communications). This is hard to verify however since there are no accurate governmental nationwide data on either the teachers or students engaged in ESOL in the UK. Estimates have previously suggested between one and 1.5 million learners (Brooks et al 2000; Schellekens 2001, cited in Barton & Pitt 2003: 8), but the numbers of teachers remains nebulous, as does data on average qualification levels and age.

For logistical reasons the research sites were chosen for being within a certain geographical location (reachable within reasonable travelling distance), being similar in size and provision and offering similar courses for ESOL students. They of course varied in minor ways such as balance between EFL and ESL students, and dominant nationality groups.

6.21 Counteracting bias

Regarding transparency, as far as disinterestedness is concerned, it is an important feature of qualitative research in that the reader needs to be able to detect possible biases which may have led to the interpretation reached. It was therefore vital for me to examine my own views on the topic before analyzing the data. I cannot change my views but by taking a reflexive approach and being aware of my position I could attempt to counteract bias and pay particular attention to a search for negative instances for any themes which related to topics I felt ‘un-neutral ‘ about.

6.22 Personal – researcher profile

We all arrive at the start of the research study with baggage; it simply needs identifying. My position is that the social consequences of testing are under-studied and not sufficiently understood. The issue of the misuse of tests and exams, particularly in high-stakes situations, had been an area of interest for a while prior to beginning this study. I had experience of researching the impact of exams having worked closely on a 5-year longitudinal study of the impact of a new version of one of the major English language exams. That experience gave me practice of successfully taking a balanced approach to seek both potential negative and positive impacts, although my personal view was not neutral due to my personal view of the effects of high-stakes exams, which I believe from the evidence I have considered can easily result in unintended negative consequences. This is the position I began this study from. It is indeed this position which was the catalyst for the study.

I undertook this research from the position of having been involved in the profession of teaching English as a foreign language, mostly abroad, for several years. On my return to the UK I had taught some ESOL classes so, while not an expert, I had a sound understanding of the differences between EFL and ESL (see Chapter 2 for a

clarification of this difference). In addition, I am a strong advocate of making language learning provision accessible for incomers to the UK and sympathetic to the difficulties of a life in a foreign country from my years living abroad and from the fact my father came to the UK from his home country as a young man and remained here his entire life.

I have also been involved in project work in the developing world aiming to improve secondary education and experienced the influence of the matriculating qualifications on the whole education system. I have also been involved in project work in a former Soviet bloc country where issues of identity and citizenship, and the power of language exams to include and exclude, were highly political issues. All these experiences have influenced my current point of view.

6.23 Data collection

The following tables describe features of the data collected.

Table 9 Data collected – Observations

Site	Level of class	Teacher ID
Site 1	-	-
	E2	T2
	E2/E3	T2
	L1 (FCE*)	T1
	L2	T1
Site 2	E1	T1
	E2	T2
	-	-
	L1	T1
	L2	T1
	Other ⁽²⁾	T2

Site 3	E1	T1
	E2	T2
	E2/E3	T1

* Key: FCE = First Certificate in English from Cambridge ESOL

Table 10 Data collected – Interviews

DoS	T1 and T2	Students	Exam to be taken by the students
✓	✓	Ss group 5	Cambridge E1
		Ss group 4, and 6-8	Cambridge E2
		Ss groups 2 & 3	Cambridge E3
		Ss group 1	National Literacy Test L1 ⁽⁴⁾
		(5)	
✓	✓	(1)	
		Ss group 8-11	Cambridge E2
		Ss group 1 - 3	Cambridge E3
		Ss group 4 - 7	Cambridge L1
		Ss group 12	Cambridge L1
		n/a	n/a
✓	✓	Ss group 1	Cambridge E1
		Ss group 7	Cambridge E2
		Ss group 8	Cambridge E3
		Ss groups 2 -6	FCE* ⁽³⁾

Notes:

(1) After experience of interviewing E1 students at Site 1 and Site 3, further E1 data collection was abandoned due to language difficulties

(2) Students on a special employability programme were at mixed levels and operated on slightly different grounds to the regular ESOL classes at that college. A S4L exam was not the class goal.

(3) These students were interviewed to gain insights into any obvious differences which may exist between perceptions of preparation for international EFL exams and the new S4L exams

(4) These students also sat FCE, out of choice, shortly after the National Literacy Test. Their aim in attending class was to sit a Cambridge international exam.

(5) A Level 2 student interview was arranged but students did not turn up.

Students were interviewed in small groups of between 2 and 4. It is important to note that the students interviewed may not necessarily have been in the classes observed and some may not have been taught by the teachers interviewed. Although desirable to have such ‘joined up’ data this was not logistically possible. In any case the main aim of the student interviews was to get a flavour of the student experience of exam preparation, insight into their levels of awareness regarding certain features of the exams and their reasons for taking the exams, in order to follow up on related issues raised by the exploratory study.

Table 11 below outlines the range of language ability levels that were covered by the observed classes included in this study. These levels refer to the UK National Qualification Framework levels. ESOL classes cover levels from Entry 1 to Level 2. See ESOL background chapter for explanation of the levels covered in UK ESOL teaching provision. (Further details such as level of the classes observed at each site, the chronology of data collection and the interview details, such as length of interview can be found in Appendix 7, Appendix 8, Appendix 9 and Appendix 10).

Table 11 Coverage of observations of students at different NFQ levels

Site	Class/ Student levels (NFQ)				
	Entry 1	Entry 2	Entry 3	Level 1	Level 2
1		✓	✓	✓	✓
2 ^{(1) (2)}	✓	✓		✓	✓
3	✓	✓	✓		

(1) one class was classified as E2/E3 mix (i.e. high E2 low E3)

(2) see note 2) for previous table

A range of nationalities was represented, 25 in all, the most commonly represented nation being Poland and Pakistan. The male: female ratio was 1:1. Of the 28 groups interviewed 9 groups were at Entry 2 level, 6 at Entry 3, 5 at Level 1, 5 at FCE, 2 at

Entry 1, and one a combined Level 1/2 class. (See Appendix 11 for a full list of all students involved in interviews).

6.24 Data preparation

6.24.1 Transcription

In the analysis process, first of all the interviews were transcribed using a basic set of conventions (see Appendix 12). Since in-depth conversational analysis was not required I did not feel transcription typical of this type of analysis was necessary, as discussed in the Methodology chapter (see 6.24.1 Transcription). Only the DoS and teacher interviews were transcribed in full, these being the core data. The amount of transcribed data amounted to some 182, 773 words (see Appendix 10 for a breakdown).

The student data was listened to carefully and instances of key themes recorded. The student interviews were not transcribed in full as they were much simpler, shorter and less productive, as assessed on the first listening. Instead key sections were noted, with recording time markers included. Points were summarised and the time recording of notable sections were noted for easy retrieval.

6.24.2 Data analysis

As a result of the data collecting process itself, note-making on the content of each interview and then the transcription process, I was quite familiar with the data by the time the analysis per se began. I had in effect begun a process of informal tacit analysis but in order to turn the data into information I needed to undertake systematic analysis, or as Miles and Huberman put it (1994, cited in Cohen et al 2000: 283), to employ 'tactics for generating meaning'. I chose to primarily follow

Mason's organizational framework for analysis (see 5.9.2 in previous chapter) to describe the process further as this is the most detailed, while also bearing in mind prescient features of the other two descriptions.

As discussed at length in the previous chapter, although I had chosen qualitative methods for my study, I did not intend to follow a grounded theory approach. I subscribe to Dey's outlook that 'an open mind is not an empty head' (Dey 1993: 229). I had already by this point formed ideas about the current ESOL situation from my reading, from the exploratory study outcomes and also from my own personal teaching experience. I allowed these ideas to guide rather than structure both the data collection and analysis but their influence has to be acknowledged. I had become interested in the topic by being alerted to possible dissatisfaction on the part of teachers with the new assessment regime. I began the investigation alert to assessment malpractice and misuse and while not expecting it was aware of its possibility.

6.24.3 Coding

By the time the first round of coding took place I was already quite familiar with the data having undertaken the interviews myself, written post-interview notes, reviewed my notes, written notes from interview recordings and then finally having transcribed the interviews too. This meant I already had some conception of themes which were emerging from the data. Coding had to be systematic and careful to ensure thoroughness and to avoid an impressionistic approach, and to ensure I was not merely searching for data to confirm my own hunches, disregarding what else the data may be able to show.

The basic method of analysing the data was by means of content analysis (as used in general social science terms, not linguistic) using the software package Atlas-ti (Scientific Software Development 2000) to organize and retrieve chunks of data. Three types of codes were all manifested in this data set: organizational, descriptive and analytic, examples of which will be described below.

The data I wanted fell into two categories a) purely profile data, such as site and interviewee descriptions, or number of students at a college, which provide context, and b) the individual, personal responses from the interviewees regarding their behaviour, attitudes and reactions to aspects of Skills for Life. As a result of this, the codes fell into three types. The first two were those covering profile information and those based on the Henrichsen framework (which thus directly linked to specific questions in the interview schedule) and, both types were thus concept-driven codes. Lastly, the third group, of data-driven codes, consisted of those encapsulating emerging themes, namely those which I had not anticipated prior to data collection.

I drew up a list of codes based on the interview schedule (see Appendix 14) and coded the interview transcriptions accordingly but as other ideas arose I checked back to the Henrichsen model to assess whether they were accommodated by the model or not. If they were accommodated, the code was added to the code list. If they were not, they were noted (using the 'open coding' function in Atlas-ti) and if other instances, or corroborative data was subsequently found then a new code was created.

Coding took place through a process of iterative reading. Searching for possible instances of all the codes was repeated several times until the code list was stable, in other words no new codes or further instances of codes were identified. This was not a neat linear process but involved working backwards and forwards through the texts over a period of time. Bearing in mind that

‘data analysis is less a completely accurate representation (as in the numerical, positivist tradition) but more of a reflexive, reactive interaction between the researcher and the decontextualised data that are already interpretations of a social encounter’ (Cohen et al 2000: 282),

it is particularly important to review the data regularly and systematically. A list of all the codes can be found in Appendix 14.

Once themes (as defined in the previous chapter) had been isolated then the data could again be checked for contradictory evidence or ‘negative case analysis’ as it is important to ensure that the meanings inevitably shaped by the researcher are fully substantiated by the data. It is through this process that personal bias can best be counteracted.

6.24.4 Observation data

To complement the interview data, the aspects of the observation data which was in text form (i.e. the chronological description: Part 3 of the instrument - see Appendix 2) was also be coded in a similar way to the interviews. Some extra codes were created to accommodate description of the classroom, based on the activity codes used during the observation (see Appendix 2 – see Part 2, and ‘Focus’ codes in Part 3). Observations were checked against the interview data from the teacher running the class under observation in order to scan for corroborating data. In addition the observation notes in their entirety were used to help describe each site and thus highlight their differences and similarities.

6.25 Making meaning – interpretation

Once the coding was complete the instances for each code were examined, code by code, to see what picture the data were painting. This process was expanded on as more data was analysed, connections made and further theories emerged. The outcome of this process is what I will report in the following section, though for reasons of space only a few examples are included.

The means of analysis basically consisted of ‘data reduction and display’ (Miles & Huberman 1994). A process of ‘categorical indexing’ was followed (Mason 2000) which involves, reading and re-reading the coded data to search for relevant themes, and undertaking cross-sectional comparisons, aided by Atlas-ti’s ability to group data into ‘families’, in other words to filter primary documents (i.e. the data files) to view relevant extracts of certain files together. This helps refine categories and define relationships between them until a coherent narrative is achieved.

6.26 Summary

The combination of the framework of the Henrichsen (1989) model and outcomes of the exploratory study led to the development of the qualitative methodology chosen for the main study which drew on data from semi-structured interviews and observations with directors of studies, teachers and students at three separate educational institutions. The analytical approach comprised a merging of methods from several previous researchers, involving the use of both concept-driven and data-driven codes.

Having explained how the two parts of the study were undertaken, the following chapters discuss the main findings.

7 FINDINGS AND DISCUSSION: ASSESSMENT PRACTICES IN THE ESOL CLASSROOM

7.1 Chapter overview

This chapter aims to describe and evaluate the various assessment practices which ESOL students benefiting from further education provision may experience i.e. those within the context of the present study. The assessments, as described by the teachers and DoSs in their interviews, will be described, in chronological order regarding the journey of a student joining the college to completing a course and leaving the college or moving on to the next level.

Description of the assessments is divided up into two groups: internal and external. The former describe any assessments which are internally generated (and also possibly validated) within the various institutions and the external are those produced by national exam boards. First the various types of internal assessment will be discussed followed by discussion of the issues related to them. Then a discussion of the various types of external assessments, and the related issues will be presented.

To clarify use of terminology I will refer to internal, in-house assessment as tests, but refer to exams from the external exam boards. In this chapter when I refer to 'teachers' or 'the teachers' I am referring only to those teachers involved in this piece of research. It will be made clear if at any point a wider population of teachers is being referred to.

The reason for exploring the whole range of assessments which the students experience rather than simply focusing on the newly introduced Skills for Life exams

was that we cannot assume these exams were the main or only influence on classroom behaviour in terms of washback. It would be easy to make assumptions about apparent observed washback but an evidential link (Messick 1996) needs establishing to be sure the behaviour is a result of specific assessments. Exploring the whole assessment range also furnished the opportunity to understand the teachers' approach to testing and assessment which I believe is vital in understanding washback.

This chapter therefore, drawing primarily on the teacher and DoS interview data, backed up by observational data, aims to engage with the first set of research questions:

RQ 1.a) What is the range, nature and function of assessment practices in UK ESOL teaching?

RQ 1.b) How are these practices linked to the Skills for Life strategy?

7.2 A student's 'assessment journey'

To gain an overall picture of the range of assessment procedures which ESOL students may experience during their college courses, and to establish how these various procedures were interlinked, and what their respective functions were, I asked the teachers about what I termed students' 'assessment journeys'. DfES literature refers to students' 'learning journeys' (DfES no date) and I have appropriated this phrase to encompass the range of assessment experiences ESOL students are subject to. I will return to discussion of this metaphor at the end of the chapter.

The majority of the assessments were internal, that is to say, devised by the teachers and not externally standardized or accredited (see Table 12). The final assessments

taken at the end of a course tended to be the only element of external assessment i.e. produced by one of the large exam bodies such as Cambridge ESOL, City & Guilds or Trinity, and these were externally accredited by QCA (For explanation of accreditation see Section 3.4, Background to Skills for Life).

Table 12 The ESOL Students' Assessment Journey: the range of assessments ESOL students' experience and the function of each type

Chronology	Validation/ accreditation	Assessment procedure	Function
Student joins course	Internal	Placement test/ interview	To place the student in a class at the appropriate level
		Diagnostic test	To ascertain areas of strength and weakness to inform course content and ILP content
		ILP (Individual Learning Plan)	To log students' individual language learning needs and achievements
		Progress tests	To check student progress
		Mocks	To check whether students are ready for the achievement exam/ assessment
		Portfolios (where relevant) - for College Certificates	To provide evidence of achievement (needed for funding purposes)
Student finishes course	External	Achievement exams/	

T6 most clearly summarized the teachers' understanding of the 'ideal' train of assessment events of a new student:

'they would have an initial assessment and the initial assessment is like a placement test and that would tell you what level they would be working at and then you do the diagnostics and that would tell you within the level what

components they can do and from that you would write the ILP' (T6-13:1928).

Table 13 The profile of assessment practices across sites

	Site 1	Site 2	Site 3
Initial assessment. screening	✗	✗	✓
Placement test/ interview	✓	✓	✓
Diagnostic test	✓*	✗	✓*
ILP (Individual Learning Plan)	✓	✓	✓
Progress tests	✓	✓	✓
Mocks	✗	✓	✓
Portfolios (where relevant) - for College Certificates	✓	✗	✗
Achievement exams	✓	✓	✓

* Combined with placement testing - see 7.3.3 below.

It can, from Table 13, be seen that, for various reasons, not all assessment practices were undertaken at each site. The following sections will go into detail on what took place and explore the rationale for these assessments.

7.3 Internal Assessment

7.3.1 Initial Interview/ screening

A student's first contact with the college was most often in an assessment situation. Site 3 gave an initial very short interview simply to check the residency status of the student and whether they were eligible for the courses on offer (T5-12:1425; T6-

13:2194). This is termed screening. In the literature from the DfES it is stated that it can be undertaken by any college staff. Whether the other sites did not undertake this, or whether it was fulfilled by reception staff, for example, was not clear. As this appeared to be an information gathering exercise and not an assessment of student ability as such, I have not included this in the assessment journey description.

7.3.2 Placement Test

The first evaluation of the students' ability was their placement test which all three sites undertook. Whether a diagnostic test was incorporated into the placement test or not varied between sites. It was not clear however whether the teachers saw any differences between placement testing and diagnostic testing. Discussions of the two exams appeared to show the two functions had been conflated in some of the teachers' minds (DoS1-1:346 and 1440; T2-3:721; T3-7:618; T3- 7:1233). These teachers are not unusual in this respect. Alderson notes '[d]iagnostic tests are frequently confused with placement tests' (2005: 4). I will therefore discuss the two separately (see section 7.3.3).

7.3.2.1 Function

As Green and Weir (2004) state, there is relatively little research into placement testing. What does exist tends to focus on tests determining access to tertiary education (e.g. Fulcher, 1997; Brown 1989) and could thus be better termed 'access' testing as it acts as a hurdle to be overcome to gain access to a particular seat of learning and generally thus involves a binary outcome: placed or not placed.

In the case of the ESOL classes in question, as in many MFL (modern foreign language) teaching establishments around the world, the primary function of the placement test is to assess the level of ability of the students to ensure they are placed

in a class which is operating at neither too easy nor too difficult a level of language in order to maximise learning potential. Ideally the test helps place students in groups which are as homogeneous as possible regarding their language ability, so that teaching can be pitched appropriately for that group. As stated in the Background to ESOL chapter, practical considerations in the investigated ESOL context relating to availability of suitably timetabled classes for example do not always allow for this (e.g. T4-9:328, DoS3-11:935). Students, may, rather than attending class at their level, attend the closest one to their ability level which runs at suitable time, fitting in with their work and home life constraints. See later sections for further discussion of this.

The results of the tests were also used to some extent at class level for informing the class content and also at individual student level to help draw up Individual Learning Plans (ILPs) (DoS1-1:1462). They are therefore also taking on a diagnostic role as mentioned above (T1-2:1372). Another key administrative, as opposed to pedagogic, role for this placement procedure was that it placed the students into a class whose QualAim²⁶ was prescribed based on the level of that class e.g. an Entry 3 Level group would aim to take an Entry 3 level exam at the end as proof of achievement. If students in that group under-performed, or even if they over-performed (e.g. managed to get to Level, the level above) the department and college were penalised financially for not hitting targets (DoS3-10:178; DoS3-11:207). This makes clear the importance of placements but the constraints students are often under cause a difficult tension for the ESOL departments.

7.3.2.2 Tests used

Whereas the end of course achievement exams were available from a (limited) range of exam boards (see Table 2), the placement test produced for ESOL providers was

²⁶ The Qualification Aims, known as QualAims. These are the qualification target set for each student, namely which level exam, they will be sitting.

developed by one body: ABBSU (the body responsible for basic skills affairs working under the umbrella of the DfES). The test was therefore often referred to by the informants as the 'official' test (e.g. DoS1-1:1559, T1-2:1407, T1-2:1581, T1-2:1407, DoS2-5:2,3 T3-7:1213).

Site 1 used a test of their own, incorporating some aspects of the 'official' test (T1-2:1382). All four skills were tested to give as complete a picture as possible of the students' abilities (T2-3:751), and to match the coverage of the ESOL Core Curriculum. Site 2 also had devised their own test, being dissatisfied with the one offered them by ABBSU. Unlike Site 1 they claimed not to incorporate a diagnostic element but planned to do so (DoS2-5:18, T3-7:1146). This demonstrates lack of understanding that a diagnostic test and placement test need not essentially be different; it is the purpose the results are put to, not the nature of the test as discussed further in the next section.

In line with the other sites, Site 3 also did not use the 'official' placement test. Instead they used a combination of a) parts of the Oxford Quick Placement Test (a commercially available generic English language placement test, not one designed for the ESOL context) for the higher level students, and b) a 'homegrown' test for the lower ones (DoS3-11:238). The Oxford Quick Placement Test did not suit the lower level students as it was aimed at too high a level of linguistic ability for some of the students to even access the test; it was not deemed 'user friendly' (DoS3-11:906) and as T6 said:

T: Not for Entry One or Entry Two ... it's a bit off-putting ... people can't read [at that level] (T6-13:1962).

T6 therefore had resorted to writing her own placement tests for these lower levels but with trepidation since, as she says:

' - well I've only been teaching six years - diagnostic test writing and

placements and stuff - that's skills I haven't got - I just had to put it all together - just to get something together' (T6-13:1915).

A danger may be that tests such as the Oxford Quick Placement Test are more grammar-knowledge oriented and the classes which students would join are (in theory) skills-based. In addition the test is oriented to grammar knowledge which EFL students in most cases have been introduced to via explicit grammar instruction, compared to ESOL students, who may or may not previously have experienced any formal language learning. This mismatch would suggest a potential problem and the tests may offer less than insightful results for appropriate placing due to this mismatch.

The danger in using 'off-the-peg' tests as some sites were (or were considering) using is that the potential close connection between test and future classes is lost. Without an accepted universal order of difficulty regarding acquisition of language items in English (see Goldschneider 2001, Lightbown 2003, Pica 2006) the constructs tested on such an off-the-peg test are unlikely to match the curriculum of a particular institution. In this case the curriculum was the ESOL Core Curriculum which was derived from the Literacy Curriculum (see Chapter 3) and was not formed on the basis of second language acquisition (SLA) research. 'Discrepancies between SLA findings and language syllabus content have often been noted' (Green and Weir 2004). While there is little evidence of reference to SLA research in the contents and ordering of the ESOL Curriculum either, by using the 'generic' placement tests there is little chance of a principled approach to placement testing.

7.3.2.3 Procedure

The teachers reported a variety of procedures to undertake placement assessment. At Site 1 a series of three hour sessions with students took place. This time was used to test several students at once, the only session undertaken on a one-to-one basis being

the speaking/ listening component and the students were taken from their small group and tested individually for that portion of the test. Therefore the placement testing procedure was rather labour-intensive (DoS1-1:345). Speaking and listening is dealt with in an integrated fashion in the ESOL Core Curriculum and therefore is also dealt with in this way in official testing. The college was receiving funding (at that time although this has subsequently has been cut) to undertake placement testing of this detail and length for each student.

The placements were typically undertaken by a special group of teachers, rather than the whole team, and typically by the more experienced teachers (DoS1-1:1493), although as T6 above indicated, experience in teaching did not necessarily equate to expertise in testing. Site 2, which appeared proactive regarding forward planning for upcoming retirements of their most experienced teachers and unforeseen staff loss, ran a shadowing programme to train less experienced members of the ESOL team in how to undertake placement testing (DoS1-1:1493). DfES recommendations state that a teacher with at least Level 4 (NQF) qualifications should undertake the testing, but due to the nature of the staffing at that time this was unrealistic, coming from a variety of teaching backgrounds and with a variety of qualifications (T1-2:49; T2-3:27, T3-6:22; T4-8:23; T5-12:168; T6-13:207), and due to on-going problems with re-training (DoS1-1:454 and 1716; T2-3:393; DoS2-5:64; DoS3-10:658).

7.3.2.4 Problems with the ‘official’ test

The description of the placement test as internal i.e. local and not standardized may surprise ABSSU, the unit which was issuing the materials and resources for use with Skills for Life classes. An Initial Assessment tool supplied in the ‘Skills for Life package’ was on offer to colleges, but they were unfortunately dismissed explicitly at two of the sites as being of poor quality:

‘I tried with all good faith – I said right this has been published I shall give it

a try - it was absolutely appalling – everybody ended up passing everything - every single level' (T5-12:681).

Others made similar criticisms (DoS2-5:20 and 23; DoS3-10:1001 and 1022; DoS3-11:241; T5-12:676 and 725 and 1456; T6-13:1941). It must be noted most of the criticisms came from Site 3, two of the complainants from there having an MA from the same institution which has a strong language testing orientation and therefore their complaints may have been based on an increased awareness of testing issues.

The main specific complaints were, firstly, that ESOL-specific Reading and Writing components are not included. Teachers were instructed to use the initial assessments designed for literacy students i.e. adult native speakers of English (T6-13:1943). One specific outcome of this is that there is a focus on spelling in these tests which was not felt to be appropriate for lower level ESOL students (T6-13:1950). While students trying to acquire English do indeed need to address spelling as part of their writing improvement, it is only one of many language components, one of many micro-skills, they are trying to master while they are developing all four macro skills and also building their vocabulary range and accurate usage of grammar. This is to be contrasted with literacy students who are tackling a much narrower range of skills, focussing only on reading and writing. How to express themselves accurately in the written form via correct spelling takes on a proportionately greater prominence for literacy students as a result of the narrower range of skills they need to improve.

Secondly, the criticisms concerned the quality of the production. Criticisms covered an insufficient range of ability or coverage of skills to allow effective placement of students, especially at the higher levels (T5-12:1451; T6-13:1940). The test was also criticised for its technical quality and poor presentation. T5, for example, found mistakes in the answer keys which she was highly unsatisfied about. She expected, and indicated the teachers could tolerate, a few mistakes but there were too many

(T5-12:1456). Another problem was that Site 1 also faced technological problems in that they were unable to run the on-line version of the test on the college system (DoS1-1:1450) which severely limited their operational ability and efficiency.

7.3.2.5 Spiky profiles and the problem of placement

‘Spiky profiles’, as previously mentioned, refers to the circumstances whereby a student is not equally proficient in all skills, which is a perfectly normal state of affairs, and pertains to L1 ability profiles also. A student, for example, may be stronger in oral and aural skills, weaker in reading and weakest of all in writing ability. A student of roughly equal ability in all four skills would have a flat profile.

‘Learners with a spiky profile, whose literacy is far lower than their oral competence, may be referred to specialist ESOL literacy provision in addition to, or as part of, their learning programme’ (DfEE 2001: 224). This is the policy laid out in an ESOL tutors’ manual from the DfES. However, evidence from the data suggests that such practice is not frequent since the reality of students’ lives (e.g. their availability for classes only at certain times of day due to work or family commitments) (DoS3-10:194; T6-13:1344) may mean even if such provision were theoretically available they would probably not consistently be able to access it. DoS3, describing some of his students, highlights this point:

‘they’re adults - I failed my teaching observation because I didn’t challenge a learner because they were late – the FE rules say if you don’t challenge a learner who comes in late you automatically fail - now I am not going to challenge a mum who has just dropped off her couple of kids at pre-school and probably had a hell of a morning ... – I’m not going to do it and I’m not going to do it to someone who’s been working until half past three in a restaurant in the morning – I’m just not going to do it – not to someone who’s twenty five years old –they’re not sixteen years’ (DoS3-10:1342).

Some students, however, did manage to access classes at different levels for different skills. For instance, At Site 2 T4 ran a class focussing solely on developing oral accuracy and fluency, and also confidence (T4-8:64 9:25; T4-9:72 and T4-9:103). This was unusual.

Spiky profiles are also relevant to achievement assessment and will be discussed further below. The main issue regarding placement is where to place students with a spiky profile: in a class which addresses their highest level of ability, their lowest or an average? Which best would suit their needs must be considered. The option of attending different classes focussing on different modes at different levels is not usually viable, either in terms of provision on the colleges' part regarding timetabling and staffing or on the students' in terms of being able to attend a mixed programme. The more common and viable approach is a weekly class at a fixed time each week. However good the placement test is at profiling the students, practicalities will most likely confound the pedagogical ideal.

7.3.2.6 Placement testing for ESOL v. literacy students

As well as spiky profiles, another aspect of the problems of testing is the potential variety of difficulties which affect a proportion of ESOL students, namely those who are newly arrived in the UK. The stress of finding work or housing, maybe culture shock as well as, let alone possible additional traumas suffered by ESOL students who are asylum seekers (Hodge et al 2004) could preclude a sufficiently representative score being achieved to place the students appropriately (T5-12:400).

One specific example highlighted the different situation concerning the various student groups: those with ESOL, with literacy and with numeracy needs. Although all three fields were grouped under the label of 'Basic Skills' within their educational institutions their students had different profiles (DoS3-10:1128). Most were located

organisationally, if not physically, within a Basic Skills Unit, or something of similar name. DoS1 compared the placement practices within ESOL with those in literacy teaching, which is where her teaching background lay. She felt strongly that students should experience the classroom environment for several lessons amongst various student groups, and be fully part of the process of evaluating which class was most appropriate for them. As well as softening the assessment experience, it led to better placement in her view. (DoS1-1:1748). DoS1 had strong views about the way ESOL students were assessed for placement:

DoS: 'they should be assessed but not on the first meeting' (DoS1-1:16524).

This was probably based on her insights concerning her literacy students who had often previously had a very negative experience of formal education, and especially assessment. She explained, when asked how she would improve the placement of ESOL students at her college:

DoS: 'I'd just improve the process where they didn't do it when they first came into college ... I'd want them to look at a class before hand even if we're not sure of levels - why can't they have a look at Entry and go right through? - a taster session' (DoS1-1:1753).

The DoS at Site 1 did not seem to recognize however some fundamental differences between the adult literacy students (i.e. adult native speakers of English) and ESOL students. What she did not seem to have embraced was that the root of the problems for most ESOL students is not previous experience of educational disruption or learning difficulties as it often is with literacy students but the problems involved with adapting to a new lifestyle in a new country and acquiring a new language as well as the fact that English is both the medium of the class as well as the content. It cannot be assumed that ESOL students have negative associations with educational institutions. To do so would show a lack of understanding of the ESOL student constituency which is vastly more heterogeneous in terms of educational background

and qualifications (Schellekens 2007: 8), as already discussed in Chapter 2 (Background to ESOL). Alignment with the practices of the adult literacy tutors regarding assessments is not necessarily the answer for ESOL students, although her point regarding assessment comprising the first contact with a college not necessarily leading to truly meaningful results is pertinent.

7.3.2.7 Administrative v. pedagogical concerns

Certain concerns were voiced regarding the problems with accurate placement. DoS3, for example, was concerned that some students may make fast progress, faster than anticipated by their placement test results, and find themselves in a group working at an inappropriate level (DoS3-10:191). It transpired that the reason why this is problematic was not primarily on pedagogic grounds. The Qualification Aim has to be recorded for each student and it is achievement of this stated goal that is used by the college and ultimately the Learning and Skills Council (LSC), the body responsible for administering funding, as an indicator of success. It is this which secured funding at the higher rate for Basic Skills students.²⁷ Therefore, assigning the wrong QualAim can be problematic in purely administrative and logistical terms yet the pedagogic impact of such considerations was not mentioned.

7.3.2.8 Conclusions

As already stated, various criticisms were made of the 'official' placement test. All these criticisms built up to a negative view of this test and it was not used as offered by DfES/ABBSU although what tutors did not comment on was the quality of the home-grown or suitability of the 'bought-in' tests. Their concerns lay predominantly with the quality of the summative assessments (see section 7.4 later in this chapter) but, it seems, not with the formative assessments which would seem to be equally

²⁷ Namely the 1.4, rate received for these students

important in pedagogical terms at least.

While placement tests are not generally high stakes assessments it seems in this particular situation the stakes are higher than normal in that the implications of the outcomes for the college, if misplacing occurs, has financial consequences, as outlined above.

7.3.3 Diagnostic Testing

7.3.3.1 Function

The prime function of a diagnostic test is to ‘identify test takers’ strengths and weaknesses, testing what they know or do not know in language, or what skills they have or do not have’ (Davies et al 1999: 23). In other words they are not designed only to establish what a student cannot yet do. Yet, it tends to be the weaknesses which result in classroom action learning or teaching points.

As noted by Alderson, diagnostic testing has generally had less attention paid to it than other main types of assessment (2005:254), such as proficiency and achievement testing, and even placement. Known testing experts themselves seem to conflate the functions of placement and diagnosis according to Alderson (2005). It is not surprising then that the teachers seem sometimes to muddle or conflate the two initial assessment practices: placement and diagnosis (DoS1-1:347 and 1440; T2-3:721).

So what makes a diagnostic test diagnostic? Is it necessarily a different type of test to placement tests? ‘The degree to which a test is diagnostic depends not so much on the purpose of the test, but on the way in which scores are analysed’ according to Moussavi (2002, cited in Alderson 2005:7). From a practical point of view it then is

understandable that the teachers in general seemed to conflate the diagnostic with the placement tests since they were often the same test. What needs to be born in mind is that not only the results, as Moussavi suggests, are used for placement. Students achieving scores within certain bands are delegated to certain classes, but the specific individual responses can also be used on an individual basis for each student. Such use is in addition to class level diagnosis of the most commonly presenting difficulties from the group as a whole. This is the ideal however, and there is little evidence this was systematically undertaken at all sites.

At Sites 1 and 3, probably due to the time-consuming nature of the testing, the initial assessment fulfilled two purposes: placement and also diagnosis of a student's strengths and weaknesses (e.g. T1-2:1372; T2-3:787). At Site 2 however a separate diagnostic test was administered after the placement procedure, once the student was settled in their designated class (DoS2-5:29). One problem found with this was that in one case at least, the teacher claimed the students had moved beyond the level of the diagnostic test by the time it was administered and it proved to be of little value:

'this week of diagnostic testing which was a nightmare – well some tutors thought it was good in that it told them stuff about their students ... but I already knew mine quite well because I had them fourteen hours a week'
(T3-7:1556).

It suggests that these initial tests were rather limited in their scope and maybe caused the inaccurate placing mentioned above. This was particularly the case for teachers who had high contact hours with their classes, as in this case.

7.3.3.2 Procedure

The tests provided by DfES were quite complex and involved teachers making a thorough assessment of the students. For speaking, as just one example, the assessors are expected to complete the grids (depending on the level - see Table 14).

Table 14 Components of the DfES ‘official’ diagnostic tests

Level	Narratives	Explanations	Presentations	Conversation (and discussion)	(Following) instructions	(Following) directions	Instructions	Speak to communicate	(Engage in) Discussion
E1	✓	✓	-	✓	(✓)	✓	-	✓	(✓)
E2	✓	✓	✓	(✓)	-	-	-	✓	✓
E3	✓	✓	✓	(✓)	✓	-	✓	✓	✓

Source: DfES (2001)

The system recommends that a three band system of evaluation be used to judge the students on a range of micro-skills and for each, the student’s skill level is deemed to be either:

Emerging – i.e. ‘the learner show little or no evidence of having skills in this areas’

Or

Consolidating – i.e. ‘the learner has some skills in this areas, but they are not yet secure’

Or

Established – i.e. ‘the learner does not appear to have problems with skills in this area’.

Once a level is decided on, in each case the recommended course of action is to a) investigate either the level below, b) include the micro-skill in an Independent Learning Plan (see below for further discussion of these) or c) move on to investigate the level above (DfEE 2001: 124).

This system is clearly designed to pinpoint a student’s level in each of the four macro - skills but it was not established practice to judge students in these terms (DoS3-

11:706); the sites used their own methods for diagnosing. One practice in evidence, other than combining with placement testing as already discussed, was that T4 used the end-of-module-test (which had been produced centrally in his ESOL department for use with all students at Site 2 as progress tests), for diagnostic purposes instead:

'we have six modules .. what I do is half way through the module I give them the test and then that gives me then more information to finish the module because if they can already use comparatives there's not much use doing a whole lesson on comparatives' (T4-8:649).

He admitted he thought he was the only one to use the test in that way (T4-8:678).

T4 also taught a group of students who had to attend ESOL courses as part of an arranged programme of further study linked to continued receipt of employment benefit. These students had the incentive of receiving a hundred pounds if they successfully passed the achievement test for their level at the end of the course. In order to help them achieve this, T4 also gave them the progress tests in the same way as for his other classes, using it for diagnosis purposes at the beginning, to identify which language components to really focus on. In this case the stakes for the students were much more tangible and consistent amongst the group and he felt a duty to help them as best he could to reach their goals. He felt he best did this through careful diagnosis (T4-9:396).

Sometimes diagnostic teaching was advocated by the ESOL department but it did not necessarily fit the nature of the courses for all types of students. For instance, Site 2 allocated a block of time near the beginning of the course to spend on diagnostic work but, as T3 said, those teachers who taught their students for a substantial number of hours each week did not feel this really necessary, by the time then having already had enough contact to make such judgements (T3-7:1156).

7.3.3.3 Conclusion

The official process while thorough seems to be a victim of its own thoroughness. The various levels, and moving up and down between levels to find as close a description of the students as possible, using the ‘emerging’, ‘consolidating’, and ‘established’ nomenclature, would obviously be time-consuming. It did not seem realistic in the time available for the test to cover all the skills in this level of detail for each student. Since all sites said they did not use the official tests in full (although some aspects were used) it is unclear whether such precise description of the students would become standard practice in future. When, as already discussed, even with a clear profile, practicalities were such that whether a suitable class to match the student’s profile would be available was not guaranteed. This made the chances of this innovation taking hold seem poor. As already discussed, even if a clear ability profile of each student could be established via such detailed initial testing, the availability of suitable classes to match that profile is slim, making this innovation (of a systematic way to describe the learners’ profiles) unlikely to be taken up.

One other problem I find with the official test is that at the same time as offering a complex assessment procedure as outlined above, they do not offer the teachers enough guidance. For instance, the Tutors’ Manual says:

‘The diagnostic grid for the speaking task does not include all the component skills of speaking which are described for each level in the Adult Core Curriculum. Those included are ones which can most easily be assessed during the course of a semi-formal conversation with a tutor. They are also skills which are *basic and important* [my emphasis], and diagnosis in the area of these skills should provide useful starting points for learning.’ (DFEE 2001:125).

Why exactly the skills selected for inclusion in the test were chosen is not made clear. What makes them ‘basic and important’ and who decided this was not explicit either.

T4's use of what were designed as progress tests may seem unorthodox but a test is labelled according to its function more than its nature. I find it interesting that as a relatively new teacher he was being creative about the use of the assessments available and used them as he saw best to maximise students' motivation and learning by not labouring points they had already mastered, for example. With reference to what Moussavi (2002) comments on above concerning the use of a test for diagnosis rather than necessarily using a separate type of test, T4 has understood Moussavi's point, perhaps intuitively.

The quality of the tests which ESOL departments designed in place of using the official test could be cause for some concern.

'Inadequate diagnosis in the context of language education is unlikely to be life-threatening, unlike inadequate medical diagnosis. And so much less attention has been devoted to ensuring the validity and reliability of diagnostic tests in the foreign language field' (Alderson 2005: 6).

Due to the confusion of the placement and diagnostic function it is doubtful the tests were as effective as they could be. They may not carry high stakes but they are a useful tool for more effective teaching. There need not be two separate tests, but as with any assessment, they should be well prepared and administered, and based on sound principles to ensure usable, meaningful results.

7.3.4 Independent Learning Plans (ILPs)

7.3.4.1 Function

ILPs are a form of formative assessment in that students are meant to identify what areas of language they personally feel they need to work on. They previously have also been accepted as proof of student achievement for funding purposes, if the ILP was

fulfilled and relevant evidence was documented. They were, at the time of data collection, used as the sole proof of achievement for Pre-Entry Level students for whom exams were not felt to be appropriate, or even exist (DoS2-5:96).

In the interim period of adjustment to a new regime, namely Skills for Life, which my study period covered, it was not clear to everyone what the exact status of the new exams was, whether they were compulsory or not, and thus the ILPs were being used instead of exams as proof of achievement (T6-13:1054) at various levels.

7.3.4.2 Description of ILPs

ILPs are documents of individual students' learning aims, in theory negotiated between each student and their tutor, cross-referenced to the curriculum and are meant to be written in SMART²⁸ terms (an example can be seen in Appendix 14). The format of ILPs is usually a one page document where the agreed targets are listed, along with curriculum cross-referencing for each of these, as well as a means of checking off when and to what extent students have reached each target. Tutor's signatures are needed on all documents for verification of completion of each goal. Examples are provided in the Tutor Instruction Manual – ESOL (2010: 228) from the DfEE but the format used at each site varied from this in a number of ways, mostly in terms of simplifying the format suggested, since it runs to six pages.

7.3.4.3 Old tool: new use

Weir (2005) states, and informants corroborated (DoS2-4/5:007; T5-12:1137; T6-13:1047), ILPs were not a new initiative of Skills for Life but, as part of general good teaching practice, had been incorporated into individual student needs analysis for

²⁸ Specific, Measurable, Achievable, Realistic, Time related.

some time. T2 reiterated, although the ILPs or similar documents were not new with the arrival of Skills for Life (T2-3:1513), their nature had changed. T6 also reported:

'I think the ILPs always used to be very light, you know, read some more, write some more – that kind of thing' (T6-13:1177).

She later returned to this to emphasise this point (T6-13:2914).

From being an individualised, rough needs analysis which teachers undertook to help students tackle personal weaknesses producing and maintaining the ILP seemed to have become a highly regulated, formulaic procedure. For example, T2 resented that proof of achievement of an entry on the ILP needed to be provided three times e.g. a piece of written work which demonstrated correct use of a certain tense, as listed on an ILP as a goal; the teacher's word was no longer sufficient she felt:

'it's got to be signed and dated by the tutor [...] I'm not quite sure if this is national requirement or a college interpretation of the national requirements- but the college has interpreted that they need three pieces of evidence for everything that you put on the ILP' (T2-3:914).

She reported feeling undermined, not trusted to make overall judgements about her students any longer (T2-3:914).

7.3.4.4 Problems in producing the ILPs

First of all, regarding the source of ILPs, they were reported in practice to be drawn up initially on the basis of the results of the diagnostic element of the placement test not based on students' own independent evaluation of their own needs as this was not felt to be viable with most levels (e.g. DoS1-1562; T4-9:398). Further ILPs were developed in tutorials which were built into classroom contact hours. At Sites 2 and 3, the amount of hours set aside for these tutorials depended on the number of students in the class (DoS2-5:33; T3-7:706; DoS3-10:1743). At Site 3 they claimed to aim at a weekly tutorial (DoS3-10:544) but given the number of students in the class and the

time needed to elicit ideas from each student a truly student-centred ILP production was not viable either.

Secondly, the teachers suggested the language level was one of the key difficulties in trying to produce the ILPs as originally intended (i.e. a statement of the student's learning needs analysis produced in SMART terms). T6 reported that producing ILPs jointly with ESOL students was indeed somewhat problematic, and felt, as did DoS3, that their ability to be involved increased as his or her language ability increased (DoS3-10:887, T6-13:1413). For example:

' – it's unrealistic for an Entry One [student] to negotiate and fully understand their ILP- Entry Two maybe – they're beginning to do it but Entry Three should start being responsible for setting their own targets and progress and working towards it' (T6-13:1300).

Note she doubted it was even unlikely for Entry One level students to understand the ILP.

Language problems were felt to be the main 'stumbling block' in ILPs. Their production is meant to be a partnership between teacher and student, but, as reported, this was very hard, for reasons of both language ability, and also 'teacherliness' (i.e. striving to make the students feel comfortable in class and making great effort to understand them etc.) Power relations to some extent also had a role to play, as T4 explained:

'as soon as the teacher tries to break down the topic/language components under discussion they inevitably start directing the discourse' (T4-9:786).

This is not desirable according to the original intentions of the ILP since the ideas are meant to come from the students themselves to enhance involvement in the learning process, and thus in theory increase relevance and motivation.

As DoS3 said, it is a fundamentally different matter to sit down with an English speaker to discuss his or her learning needs and to do this with an ESOL student. He often felt the whole culture of reflection was alien to many students (DoS3-10:408). The students expected the teacher to know what they needed; that was the teacher's role (DoS3-10/11:423). Also they did not necessarily have the analytical tools to isolate specific needs, as T6 illustrated:

'well a lot of them say 'well I want to learn English' 'well what about in the writing- what do you want ?' – 'yes writing' – 'but what exactly in writing' – 'just more writing' (T6-13:1150).

T2 offered a similar point of view, concerning the need on the student's part to analyse what their needs are:

'It's difficult for Entry Level students to really analyse exactly what they need and if you ask them about [...] where their needs are and if you review the programme and say 'what do you think?' they will usually say 'I need help with listening, speaking, reading and writing' (T2-3:865).

7.3.4.5 Literacy model

Some of the ESOL teachers showed they believed that ILPs, in their systematised, SMART form, had been adopted from literacy teaching (T6-13:003; T6-13:901).

'I think originally ILPs were set out for people who were doing these small groups and people working individually on their own – much more the literacy model - doing individual work not group work like ESOL does' (T6-13:1177.)

The format and delivery of ILPs was one way in which it appeared from the data that the 'template' from their use in literacy classes had been imposed 'from above' without considering the implications (DoS3-10:616), without realisation of some key differences between the two types of class.

Firstly, there is a significant difference between student-teacher negotiation of learning needs where English is both the topic of the ILP and also the vehicle for negotiating it (DoS3-10:408; DoS3-11:1394). While a literacy student may be able to discuss their reading and writing needs in their mother tongue with their tutor this is much harder for an ESOL student.

In addition, ESOL classes tend to be much more group-work oriented than do literacy classes. Literacy classes are traditionally more individualistic and worksheet based, with little plenary work (Hamilton & Hillier 2006). ILPs provide a structure for this style of teaching (T6-13:1078); the ILP would be more useful in providing a learning path for each student to work to. The content of an ESOL class, in contrast, is planned on the basis of it being suitable for the average class ability, with differentiated activities being interwoven into this framework as appropriate (T2-3:813). In ESOL (and EFL) classes such a detailed individual learning plan, it can be argued, is less necessary.

The difficulty of the increased cognitive overload, in that these students are being asked to analyse their needs which is a process which may be alien to them, as well as simultaneously attempting to articulate this in a language they are trying to master, does not seem to have been recognised when it was advocated that ILPs should be adopted in the ESOL context. When ESOL teachers are coping with larger class sizes than those typically experienced by literacy students, the problems of trying to help students produce their ILPs are compounded. Literacy classes, for instance, are generally smaller making this activity perhaps easier to manage, but as DoS3 pointed out:

'Basic Skills [referring to ABBSU] actually recommends there should be only eight people in a class but any FE college will tell you that they ain't going to have eight people [in ESOL classes] - they can't do that' (DoS3-10/11:623).

This simply would not be financially viable. Along with the larger classes is an increased workload for the teacher which makes the production and monitoring of ILPs even more burdensome than they would be with the recommended smaller groups (DoS3-10:583).

Further research is needed before blanket claims can be made about it being harder for ESOL teachers to negotiate an ILP with their students than it is with a native speaker of English with literacy difficulties. Native speaker students finding themselves in literacy classes due to learning difficulties may have equally great problems articulating their needs, however. Also there is no evidence that a native speaker would be better able to articulate his or her language development needs. Nevertheless, the additional problem of lower level ESOL students possibly being unable to communicate concepts and plans, should they even have them, due to poor language skills must be acknowledged as an extra hurdle in the personalised learning plan scheme. What is more, the literacy students have probably had literacy instruction previously (in school for a number of years) and have experience of what literacy means, whereas ESOL students might be entirely new to language learning and have no idea what this involves.

7.3.4.6 Compulsory or not?

There was evidence of a certain level of confusion over ILPs, regarding whether they officially had to be completed for all students at all levels or not (e.g. see T2-3:799).

T1 for example put the confusion down to lack of clear direction from line-management (T1-2:631). T2 suspected it was a higher level decision:

'I'm not quite sure if this is a national requirement or a college interpretation of the national requirements...' (T2-3: 916).

DoS1 stated, as did others, the belief that ILPs were at that point compulsory, required by local LSCs (the funding body) (DoS1-1:1258; DoS3-10:1329; T5-12:1118;

T6-13:1115; T6-13:2535). In general the range of views highlighted how the practice varied due to a lack of direction and information concerning what was required.

T2 noted what she termed a lack of logic in completing ILPs in exactly the format required. Since in the case of most classes end-of-course exams were also being taken, she was unclear why the ILPs were submitted as additional proof of achievement (and had to therefore be set out in a specified very formulaic way). Since proof of progress was achieved via exam results, as far as she understood she saw no need for this to be repeated via ILPs (T2-3:1381). This was symptomatic of the confusion surrounding ILPs at that time.

Ironically, the ILP, a tool which purports to promote individualisation of learning, a student-centred approach encouraging independent learning, was felt by some to have ended up more like a mould which shaped students, homogenising them.

'I guess again it's the sort of the idea that erm it's good for learners to set themselves goals and take control of their learning and become more independent learners which is great but again it's that forcing everyone into the same way of doing things' (T5-12:1131).

With limited class-time for many groups, and the new additional pressures of ensuring the students were prepared for the Skills for Life exam, the ideals of the individualised learning approach could rarely be fully realised (T5-12:1352).

It seemed that in a period of change in so many aspects of ESOL, in teaching to a new curriculum, lack of clarity regarding assessment practices, as well as issues surrounding updating teaching qualifications, it is perhaps understandable if the teachers seemed a little confused at times as to what was going on and exactly what official requirements were.

7.3.4.7 Attitudes to ILPs

Apart from the workload there was the issue of full comprehension of why ILPs were necessary, which seemed to add a layer of resentment to having to complete them in the ESOL class.

'I still don't see why it is necessary to actually – why we have to write it down - as long as you do it as part of your teaching – which you do because you're explaining what you're or they're going to do that day - what they're going to learn – hopefully they learn something and then at the end ... the end ten minutes [we review the class]' (T2-3:1563)

and again later in the interview:

'[it's] just good practice right – what you're doing anyway but it's now got to be written down' (T2-3:1654).

The teachers' reactions to ILPs were very mixed. There seems to have been, as referred to already, a general confusion over the purpose of ILPs. The DoS at Site 3, said that he was not convinced as to the principles underlying the practice of needing to document everything, both what is planned and achieved in class, cross-referenced to the Curriculum; he was unsure about their real value:

'I don't necessarily see any solid evidence that they are making a huge contribution toward learner progress ... it seems to be a thing demanded by inspectors– I mean if someone can show me that evidence I'd be pleased to change my mind' (DoS3-10:571).

This questioning of the rationale for ILPs lead T1, who only taught high level classes (Levels 1 and 2), to not complete ILPs with her students. She felt the need for them was unclear and had negotiated with her college validation officer who confirmed she did not need to do them (T1-2:1301). Her understanding was that the purpose of ILPs was to track student progress and she did this by alternative means, using regular

progress tests (T1-2:1319). The message that ILPs were a formal requirement by LSCs (where funding was sourced) was, she felt, entirely due to interpretation of 'rules' by line-managers (T1-2:637). Her colleague, T2, also questioned why the students seemed to be needing to fulfil two sets of formal assessments: ILPs and also external exams in the form of the Skills for Life exams (T2-3:1389). The lack of clarity was evident in that the three sites seemed to have different policies and understandings regarding ILPs. T5's assessment of the situation was:

'it's an idea which has been thought up and imposed without very much thought I think' (T5-12:1383).

An article in Reflect (the magazine from NRDC which disseminates research and other general information about Skills for Life issues) was trying to clarify issues regarding ILPs at that time and it reports it was not only teachers in my study who were unsure about ILPs (Weir 2005).

A further negative aspect of ILPs, which T6 in particular was somewhat cynical about, was that by including them as some form of assessment there was more pressure on teachers to ensure students reached all the goals laid down in the ILP. Therefore the chance of teachers setting goals to stretch their students was slim since they would want to maximise chances of students reaching all their goals easily. Rather than gambling on a goal which may be too far for a student it was securer to play it safe. As T6 related:

'they [students] were measured against their ILP and they were only set targets on their ILP that they were going to hit [...] so our achievement levels were pretty high [laughter] - funny how that works [ironic] (T6-13:2291).

However to balance T6's cynicism, the pedagogical value of setting reachable targets was expressed by T5 (in relation to the exams students are entered for) when she said:

'I do occasionally put them in for higher exams but I do think they'd be able to do it and that knocks their confidence and they want to tell their families –

how awful is it for them when they go home and tell their kids who are at university and everything and they go 'oh I failed' ? - it's horrible' (T6-13:2357).

The pedagogical and administrative tensions were seen to be often in conflict and a teacher's personal perception of the issues may dictate which rationale they lean more towards.

However, while various informants noted some of the problems associated with ILPs, their fundamental value was recognized by some at least:

'I think if they're used correctly they can be of benefit to the students and there is benefit to teachers of being able to say ' yeah – look I can see this person is weak and needs help with this' so I think they can be beneficial'(T6-13:2529).

It was not clear that there was a consensus across sites on what 'used correctly' might mean, however.

Despite the problem with ILPs, teachers seemed resigned to them, and T6, for example, advocated to her colleagues a philosophical approach:

'let's try and make it into something good otherwise it's just frustration' (T6-13:2892).

She felt the beneficial aspects (such as time spent one-to-one with students and getting them to reflect on their learning) should be focussed on rather than the doubts about their form or even the very need for producing them. Nevertheless, T6's optimism was not in general shared by other teachers.

7.3.4.8 Conclusion

From the way the ILPs were described I was given the impression that this system was generally not seen to be ideal, above all as it took time away from class teaching.

With a curriculum now structuring the courses, where none had been available before, many teachers felt pressured into covering a great deal of material in a course and preparing students for their exams became the priority, given the pressures passed down to them of the need for proof of achievement (see Section 7.4.5 for more on this). It seemed any extras, especially with dubious purpose or use, although originally designed to have a beneficial formative role were resented.

7.3.5 Progress Tests

7.3.5.1 Function

The term 'progress tests' covers a range of tests designed so that both tutors and students may judge whether the students have mastered the course materials.

Sometimes these tests were produced in-house and pooled for the whole department to use, especially where students were covering a highly centralised syllabus as in the case of Site 2. Teachers there had ready-made tests to draw on to regularly assess their students at the end of each module, which their syllabus is divided into. These had been produced by a team of the teachers themselves tasked with this project. Other sites were not so centralised.

7.3.5.2 The nature of progress tests

At Site 2, the ESOL department had put a great deal of time and effort into producing a syllabus and a set of progress tests (DoS2-5:27; T3-7:497). One member of staff there was appointed to overview assessment matters and she had co-ordinated their production which reflects the level of effort they made and how seriously they took assessment. These tests were oriented to the department's syllabus based on the new Skills for Life ESOL Curriculum, from which topic-based modules had been devised.

The ESOL team at Site 2 was rather upset when they were rather severely criticised during their OFSTED inspection for not having moved yet to externally accredited exams for their Entry Level students (DoS2-4:344). Their argument had been that they wanted to wait until the situation regarding the full range of exam boards available to choose from before moving over entirely to external exams and that these progress tests were acting as progress tests primarily, not achievement tests, although they took on this function also while external exams were not yet being used. They felt there was no recognition of the great deal of effort they had put in up to that point in producing an assessment system to act as an interim measure (DoS-3:350).

It seemed to be the choice of the individual teacher, at Sites 2 and 3 at least, whether, and if so how, progress tests were used (e.g. DoS1-1:1476). T6 used the progress tests she had devised herself in order to have a systematic way to comment on students.

Progress tests, it seemed, were preferable to progress monitoring via portfolio which groups at lower levels at Site 1 generally produced. In T6's view portfolios caused hassle for the teachers and because of the nature of their contractual status were not willing to put in the extra time needed to make their production successful T6-13:995). She seemed to have a fair amount of freedom to decide this course of action.

T1 also seemed to want to avoid a portfolio type on-going assessment (T1-2:490). She drew on progress tests offered in the course books she used, and since she had higher ability level classes who were aiming to take Cambridge main suite exams, their material was typically exam preparation books for Cambridge main suite exams, not the Skills for Life materials.

At Site 3, T5 worked in a similar way, using similar materials as she also generally taught higher level students, who were also aiming to take Cambridge main suite

exams (T5-12:1308). The assessments were typically end of unit progress tests found in many of the general EFL English coursebooks, such as Headway (by Soars & Soars from OUP), and exam preparation type textbooks. T6, also at Site 3, classified these tests as 'grammar' tests which she said were not relevant to Skills for Life (T6-13:996). By this she meant the curriculum and hence also the new exams, were skills-based, not tests primarily of vocabulary and grammar knowledge, which seemed to be the focus of the tests influenced by these coursebooks. The progress tests therefore were judging progress towards the main-suite exams not progress towards Skills for Life exams.

In contrast to the higher level students who were taking tests, the lower levels were being tracked for progress in more informal ways (T6-13:1538), as at Site 1. The most systematised method was indeed at Site 1 where all students were meant to maintain a portfolio file. Together with informal on-going assessments throughout the course, the evidence set out in these files (examples of students' work) culminated in award of a College Certificate (see Section 7.3.6 below).

7.3.5.3 Objections to progress testing

T6 felt that what was needed were 'multi-skilled' assessments and assignments. By multi-skilled she later explained that they would be integrated e.g. a task requiring writing, speaking and listening too in order to complete it (T6-13:1022). For example, students may listen to a text, write something in response to it and then complete a task which involves talking to another student on that same topic. However, this was problematic in her eyes since class attendance was often sporadic due to the type of work (e.g. shift work, subject to frequent changes etc) which students in her classes were typically involved in (T6-13:995), and such tasks had to run over more than one lesson due to the more complex, compound nature of the tasks. She felt a centrally produced set of assignments or assessments to give a regular guide to student

progress, (as per the ones Site 2 used) would be helpful but since, as team leader, she felt it would be her responsibility to design them, did not appear to relish this prospect (T6-13:996; T6-13:1015). As noted elsewhere, she reflected that she had had no training in test production and recognised that it was no easy task (T6-13:2181).

7.3.5.4 Conclusion

Progress tests were, in general, less discussed by the informants than other types of assessment. I feel they did not take on any great importance, despite their formative role. It could be that such tests were so accepted as a normal part of teaching they were overlooked and did not provoke particular comment or, alternatively, that the most contentious assessments (e.g. ILPs and the new exams) superseded other forms of assessments in the teachers' minds as worthy of discussion.

In this study, the classroom-based regular tests appeared to be the most unregulated area of assessment in the raft of practices. This stood in contrast to almost all other forms of assessment (except perhaps mock exams – see Section 7.3.7 below) whether specifically Skills for Life related or not. Compared with the highly regulated nature of the other forms of assessment (e.g. systems which the whole department followed such as for placement and achievement) the nature of progress tests was highly individualistic at two of the three sites.

7.3.6 Portfolio leading to a College Certificate

7.3.6.1 Function

The College Certificate can be categorised as an achievement award as it related to success (or otherwise) on a specific course of study. (Achievement testing will be discussed further in the next section). Internally awarded qualifications such as a College Certificate seemed to serve first and foremost for purposes of student

motivation. There was no sense that they carried any weight in the world outside the college (T1-2:1469), although T1 seemed to want to believe this might be the case, but she had no evidence it was so:

T: 'it must have some currency, mustn't it? If you got a certificate from a college in the area presumably it should have some currency

I: But you haven't any concrete evidence?

T: No, not really, no' (T1-2:1513).

Another reason the College Certificate was used was as an interim measure until the system of using externally validated qualifications had fully settled in (T2-3:1675). In the case of Site 1, the College Certificate had been introduced as a reaction to the introduction of the ESOL Core Curriculum, as it was recognised such measurement was necessary but how to do this was not yet fully formalised.

7.3.6.2 The nature of the qualification

The College Certificate was a portfolio-based qualification (T2-3:903), which was only found at Site 1. Students maintained files which held examples of their work and records of their achievements. These could, in theory, be spot-checked any time by the director of studies to monitor student progress. The files were precious in that they held all of a students' evidence of their achievements in their language learning to date at that college. Their award depended on their file.

The files which were presented for College Certificate were kept in class and since they were a way of checking on the students it was important that they were kept in an orderly fashion. This was not necessarily easy and was yet another duty for their teacher (T2-3: 914) since this need was more pressing for the ESOL department than for the students who may not be particularly concerned whether they received a Certificate or not. In addition, T2 suggested that the signing of every piece of work by

the teacher entered into the file was an extra unwelcome administrative burden (T2-3:913). The portfolio requirements adopted were apparently roughly based on those of the OCNW portfolio based qualification the ESOL departments' students had previously worked towards (T1-2:851).

7.3.6.3 Problems

In addition to providing sufficient evidence of their achievements, students needed to fulfil other criteria to receive their awards, including punctuality and regular attendance. Failure on their part to comply with these aspects was recognised by the teacher as not necessarily being due to a matter of lack of commitment on the students' part. For instance, T2 recognised that some had problems in attending regularly due to family commitments or irregular work hours, typical of the ESOL students who without sufficient English rarely had opportunities for more than unskilled jobs (T2-3:1729). T2 showed concern that this reality was not always recognized by College administration and consequently both the students and the ESOL department were penalised as a result for a lack of flexibility in accommodating their difficulties.

Where one class of students was taught by more than one teacher (e.g. covering different skills) co-ordination of what student work was to be entered into the files also had to be managed between the various teachers (T1 -2: 1278). Since some teachers were more committed to fulfilling the requirements for the portfolio than others (T1- 2:1531) this could cause extra aggravation. Quite a high work load was involved in producing and maintaining the files and T1 gave the impression that they did not feel any great loss when decisions were made to phase them out (T2-3:928).

7.3.6.4 Conclusion

At the time of collecting the data, the college was moving from one way of working to another. Rather than being a case of ‘out with the old, in with the new’ i.e. of moving cleanly from one type of achievement testing to another, which was one possible way to embrace the change, they seemed to be in a cross-over period where the new exams, ILPs and the College Certificate were all in operation at the same time (T2-3:1493). It was not always clear what the distinction between ILPs and College Certificate work was, and did not seem clear to the teachers either what their respective roles were.

7.3.7 Mock Exams

As exams got nearer, using past papers became part of classroom practice, or in the case of the new Skills for Life exams using activities modelled on the new exams as there were no past papers yet. Only on the occasions when students tried out such materials under exam-like (i.e. timed, non-‘co-operative’) conditions has it been labelled a mock exam for the purposes of this study. Administering mock exams was common to all sites, especially with high level classes (T3-7:562; T4-8:575; DoS3-11:947; T5-12:576) but they were discussed notably more by Site 3 than the other two sites.

The papers used for the mocks were past papers which had been made available for the international English exams, such as the Cambridge main suite. For the new Skills for Life exams sample papers downloaded from the Exam boards’ websites were all that was available at that point.

7.3.7.1 Function

'Mocks' are a trial run on an exam similar in scope and format to the 'real' one. Their main aim is to allow students (and teachers) to judge the level of preparedness for an impending exam. It also gives the students a chance to experience exam-like conditions in order to be better able to use the time in the exam effectively and become familiar with the rather artificial, staged nature of the exam experience (Simpson 2006), wherein language is not used naturally in terms of normal turn instigation and return, and topic choice which is dictated by the exam.

The function of the mocks in this study was several fold; as with examinations in any context, it was to gauge students' preparedness and, secondly, to familiarise students with the exam. At Site 3, the information provided by mock exams results offered an opportunity to regroup classes of students according to their results, particularly in the case of students aiming to take international exams such as Cambridge main suite (T5-12:1279). Another use of the mocks was to allow the college to try to realign the QualAim so it is in line with the actual ability of the students, and thus avoid loss of funding through a student not reaching their stated goals (i.e. passing the exam) (T5-12:1294). This re-shuffling was not mentioned at either Site 1 or Site 2 and there was insufficient data to establish whether this was an established practice ratified by management, or a team-instigated means to maximise attainment of QualAims.

7.3.7.2 Conclusion

Mocks were not commented on greatly, and were not a line of specific questioning in the interviews (note: teachers were asked to recount the complete assessment journey of their students, not prompted as to what this may include). Since it was not a topic raised by the teachers themselves this suggests maybe the teachers did not see Mocks as a form of assessment as such, but merely part of the routine classroom activities. Those who mentioned it most (T5 and T6), were the teachers whose students were

due to sit exams shortly after the period of the data collection, in other words their students' satisfactory preparation was probably very much on their minds. In the other cases the actual exam period was due to take place much later than the data collection period. This variance in experience will be returned to in the next chapter on discussion of washback intensity.

7.4 External Assessment

External assessments have been categorised for the purpose of this study as those assessments which are externally validated by recognised exam boards and accredited by the Qualification and Curriculum Authority (QCA) for Skills for Life purposes. In this section I will discuss some of the issues related to these assessments. I will first explain the external assessments which the teachers were dealing with, in the same way in which the various types of internal assessment were described. As was seen from Table 12 at the beginning of the chapter, this comprised only exams given for the purposes of measuring end of course achievement.

7.4.1 Function

As already mentioned, achievement exams are those which are taken at the end of a course of study in order to judge how much (or little) has been learnt. Apart from the certification of this achievement the external Skills for Life exams seemed to play some other roles also. Firstly, they provided motivation for students, at least for those who reacted positively to the idea of sitting an exam, which teachers reported was most of them (e.g. T6-13:1185). Secondly, the exam result was needed by the college to claim funding from the LSCs, alongside measures of regular attendance and ILP completion.

At the time of data collection, these external measures of achievement could consist of the submission of a portfolio or an exam. As will be discussed in more detail later, the exams were more popular by far and used by the participants in my study, whereas portfolios for external review at that stage did not feature at all.

7.4.2 Assessment preference

7.4.2.1 Choice of Exam Boards

Schellekens (2001) reported that Pitmans (now part of City & Guilds) and Cambridge Certificate in Communicative Skills in English, as well as NOCN (National Open College Network) qualifications were the most popular at the time of her research in 1999/2000 (NB prior to the inception of Skills for Life). As for the exams which the colleges in my study selected for use as externally validated measures of achievement, all three had chosen the Cambridge exam board (DoS1-1:596; DoS2-5:67; DoS3-10:609). Some had tried out sample papers, at least, of other Boards, or had experience of other non-Skills for Life exams from other Boards.

Several reasons for the choice of exam board were mentioned including firstly, habit (at Site 3) (T6-13:2322; and 2063, DoS3-10:1610) i.e. the institution was already registered as an exam centre or at least had already set up the administrative mechanism to offer the exams, and so decided to remain with that arrangement rather than switching to a new board (T6 13:2062). Secondly, it was felt that different bodies had their own distinct approach and familiarisation with a new Board involved a certain amount of effort, to ‘acclimatise’ the new procedures and test types for example, as DoS1 recognised:

‘staff are more familiar with it rather than getting used to a different body’
(DoS1-1:663).

A second reason given for choosing Cambridge was its reputation:

'Cambridge has some sort of kudos' (T6-13:2066).

This view was repeated in various other places (*T1-2:613; T2-3:1135; DoS3-10:1820; T6-13:2348*). Part of the effect of this reputation was that teachers felt confident that the exam would be professionally produced due to the experience and resources at Cambridge's disposal. This is what I would label 'faith validity' namely, rather than closely investigating the details of any particular new exam in a systematic, analytical way, assumptions about the quality of its exams based on 'track record' and a Board's reputation are made. The users have faith that as other exams from the Board are known and respected then any new ones will be also and provide good quality exams which test what they claim to. For example:

'we recognise UCLES²⁹ as being a world-renowned organisation' (DoS3-10:1614).

Reputation as far as the students were concerned was also seen as important.

'Cambridge has a better sound about it – students know Cambridge University – Trinity or whatever doesn't really mean an awful lot to them' (T6-13:2064).

This worked in reverse in that Boards with which members of staff had had problems in the past, either through procedure, or the nature of the exams themselves, resulted in that Board being avoided, and not considered as Skills for Life qualifications for that site, as in the case of Site 2 (*T3-6:1232*). In other cases, mention was made of why other Boards had been rejected. For instance, Trinity was found to be too 'light' in the reading section and the writing too 'heavy' so was not felt to be balanced (*T6-13:2210*). Some other exams which T6 viewed while making the exam board selection were deemed too easy (*T5-12:566*). As another example, the City & Guilds papers

²⁹ Former name for Cambridge ESOL

were accused of not being ESOL-student-friendly regarding some of the questions and rubrics (instructions) (T4-8:309).

Cambridge exams, as previously mentioned, were perceived to be advantageous also in accommodating a student's 'spiky profile'. Thus pre-entry students could be entered for just speaking and listening (since they were not capable of managing exams in reading and writing) (T3-6:1138). At that stage the 80/20 rule, explained in Chapter 3, had not been introduced or colleges had not yet understood it. This modular approach is a feature flagged up on the Cambridge Skills For Life exams website (www.cambridgeesol.org) and seems to have impressed various of the teachers (T3-6:1137). The fact is that other Boards also offered this however, suggesting that not a great deal of research on the part of the ESOL departments had gone into choice of exam boards, maybe for reasons already proposed. At best a sub-committee or the DoS alone was responsible for the choice of exams and it was not a departmental-wide decision (T2-3109; DoS2-4:833) although consultation was a least nominally in place (T3-6:608). At Site 1, it was reported to be mostly a decision made at college level, out of the ESOL teachers' hands but T1 reported the opposite (T1-2:113). This highlighted the gulf which was not discussed explicitly but was evident between the DoS and the ESOL teachers at this site, in terms of co-operation and communication.

Another proffered advantage of Cambridge ESOL was that training had been provided by Cambridge ESOL (T2-3:1217), even though this had been for a limited number of staff only. Also the teachers were satisfied with the exams from their initial perusal of the samples offered (T2-3:1130). (This was the most widely used method of familiarisation according to the data, rather than training). As T6 said, if the students passed their Cambridge Skills for Life exam they deserved it (T6- 13:2234), meaning it was a fair, well-targeted, exam, although based on gut-reaction not research.

A final reason mentioned for choosing Cambridge was merely pragmatic. T2 mentioned that they had considered at one point using Cambridge Writing papers and ESB (English Speaking Board) for listening and speaking but on reflection decided for ease of administration to keep with one Board only (T2 3: 1220). By keeping to one Board a great deal of administrative inconvenience would be avoided (T1-2:602) and this was also, hardly surprisingly, the preference of college management (DoS1-1:645).

To be fair, at this time the amount of information about the exams which was easily available was less than now since it was a period of rapid change while further exams were becoming accredited by QCA and joining the list of those which the colleges might consider. It must be noted they had to make decisions in line with their academic year, namely that decisions concerning which exam they intended to use needed to be in place by certain dates in their academic calendar, so that finance departments were informed for budgeting purposes and for teachers to know so they could begin to prepare themselves in order to prepare their students. QCA did not appear to be working to any such timetables and information about the on-going accreditation process was simply issued as each decision regarding any particular Board was made. Even if teachers had tried to research the varying boards it might not have been an easy task.

One example of the suspicions regarding the lack of credibility of these tests was reported by T5:

[name of fellow teacher] was wanting to trial it [National Literacy test] with different groups to see how the results compared to my impressions of them because I know them so she got me to trial them with four different groups that I knew – to write my impression of the level first and then mark them

and see how it all turned out and [...] they all turned out pretty much lower than I would have estimated [...] I mean one student who's just done a CAE mock who got very good results - almost Proficiency level and she came out as just about a Level One and Level One is supposed to be First Certificate level' (T5-12:1161).

T6 substantiated this view (T6-13: 2256). The teachers may not be language testing experts but they do quickly get a sense of the level each individual student is able to work at and what they may be able to achieve on their exams.

7.4.2.2 Portfolio v. external exams

While all three sites were in a state of flux at the time of the study, plans at least were to move away from portfolio-type assessments and to focus on external exam-based qualifications from the large exam boards (as described above). This pattern was pertinent however to the lower levels only who had tended to be involved in portfolio-based assessments in previous (pre-Skills for Life) years. For example, Site 1 had worked towards OCNW (Open College of the North West) awards, then their own College Certificate which was also portfolio-based. Only now was Site 1 aiming at exam-based qualifications. The higher levels at Site 1 however had been used to exams previously, since they mainly aimed at Cambridge main suite exams. In fact they now intended to continue doing this, while also, in tandem, preparing for and sitting the new Skills for Life exams.

At Site 2 equally lower levels had worked on a portfolio-based internal qualification but external exams were not exclusive to the higher level students. However, the exams the lower level students took were not taken very seriously. The main function of the exam apparently was motivational:

'even at lower levels everybody took ESB - so that was external – but at lower levels I don't think it's as serious a qualification as at the high level because at low levels you just have to turn up to get it' (T4-9:764).

Two reasons were cited for why external exams were deemed a better option than portfolios, whether externally validated or not. Firstly, DoS2 recognised that the teachers found the process of portfolio management very time consuming and rather cumbersome, and were relieved to move to a more straightforward system of only having to teach, although incorporating exam preparation into the language course, while remaining removed from the assessment itself:

'to be quite honest for a lot of the teachers here they were quite happy to have exams - it relieved the pressure of portfolio and they knew what they were teaching towards' (T6-13:2753).

DoS2 recognised the change in work load which the teachers embraced:

'the tutors' perception is 'let's do external exams and get someone else to mark' it [laughter]' (DoS2-4:1435).

The nature of the staffing at the three sites comprised a high proportion of part-time staff; at both Site 1 and 2 half the staff were part-timers and at Site 3, at that time, all but one was. For example T4 reported:

'they can't really do it [portfolio preparation] in their class because the students are going to be losing quite a lot of their teaching hours and if they're only getting paid for two hours a week teaching they don't want to do it in their own time' (T4-9:1135).

Portfolio work was viewed as time-consuming as it entailed monitoring, helping to compile and checking student work, and was seen to take up more than the teachers' paid hours (T6- 13:1950). T6 went as far to assert that all the teachers loathed portfolio work, mainly for this reason (T6-13:1713). This attitude applied equally to

monitoring ILPs, another form of the portfolio approach, in that specific evidence of achievement of specific learning goals from work produced in class was required.

Secondly, teachers in general seemed more comfortable with the notion of themselves in the role of learning facilitators with an outside body acting as judge. It seemed they did not want the two roles to be confused (T4-9: 670). This relationship between students and teachers was also mentioned as a reason for preferring an 'external judge', by T4 in particular who for example felt some students were trying to curry favour with him in preparation for the examination situation:

'the students themselves realised that if it's external you could have been a really lovely teacher but at the end of the day you're not marking it and when I told my students that their faces sort of changed [...] yes I was [previously] getting a lot of biriani and roti' [i.e. cuisine from their own culture] (T4-9: 678).

What is more, both teachers and students seemed to welcome an external judgment of their achievements, and impartial measurement boosted their confidence in their achievements (S3-3:199, S3-3:231). (See also T4- 9:671, T6-13:2275).

In summary, despite particular queries and issues as listed above, external validation appeared to have been accepted. Exams were certainly perceived as less work for the teachers who found portfolios time-consuming.

7.4.3 Currency

As regards currency, the Cambridge main suite exams were the ones sought after by the sub-group of ESOL students who would previously have been labelled EFL students (in that their aim for coming to the UK was above all for improving their

English and /or work experience and they did not intend to settle here). This group typically wanted to take home a Cambridge qualification in English (S2-3:9).

‘what good is it going to be to them to take home to Italy and everywhere else – to have Level Two Cambridge paper – no-one’s going to know what it is’ (T5-12:597).

(See also DoS2-4:1091; T5-12:606; T6-13:2065).

This desire for the ‘international qualifications’ lead to what has been referred to by ESOL professionals as piggy-backing, in that the teachers were working towards two goals (i.e. Skills for Life exams and also Cambridge exams) with the same class of students, and trying to make the class content relevant to both goals. In other words, students took Skills for Life exams in order that their ESOL department could fulfil results targets set by the LSC, and consequently draw down the funding, while at the same time these students could benefit from free English lessons within the ESOL departments. However to satisfy the students’ own main personal goals, the teachers offered preparation also for the Cambridge exams, at whichever level was appropriate to the class (T1-2:1112; T6-13:405). Evidence from ESOL teacher groups at workshops I have attended since corroborated that this practice was not unique to the three sites I studied.

A second aspect of currency was the concern amongst teachers about the value of Skills for Life exams as regards how end-users, such as future employers, would view them and how well recognised they were. DoS1 had faith in their value (DoS1-1:721) but admitted that she did not believe employers were aware of the exams and what they represented (DoS1-1:779). DoS3 felt, from his own research which he had undertaken locally, they were not known at all (DoS3-10:1172). Some teachers’ opinions appeared to be based more on hope than evidence:

'they're valuable in that it's a certificate – it's a qualification - you know it's nationally known - it's - well I don't respect it but an employer probably would' (T3-6:1331).

T3 appeared to have no concrete proof of this and was thus working on an assumption. The main issue seemed to be that the Skills for Life exams were still 'young' and teachers as yet had little feedback about how they were received by the wider world (T2-3:1253). T4 however felt the majority of the population would have no idea what the exams represented:

'if you're living in this country and somebody said you to 'what have you got?' 'ESOL E2, E3' – 99% of people in this country would not know what that was but if you said you had an A level in English – straight away' (T4-9:83).

Mallows (2009) confirms this view in reporting that employer awareness for these exams was low, and often also did not have internal recognition within colleges outside the Skills for Life departments.

T4 correctly commented that there is usually a time-lag between a new qualification's availability and its general recognition amongst the wider public. He quoted the example of the introduction of the NVQs (National Vocational Qualifications) in England when previously City & Guilds exam board had been the main provider of vocational qualifications for many years (T4-9:51). It took a matter of years before the new qualifications were familiar to the general public, and most importantly, to potential employers, and this process of acceptance is not yet complete (KPMG 2005). T6 was confident that in time the new Skills For Life exams would be recognised, but was not clear about how long this might take (T6-13:448).

Whether the Skills for Life qualification would be recognised is one issue; another matter completely is whether the different levels (Entry One, Entry Two, Entry Three,

Level One, Level Two) would mean anything to employers (T4-8:476). The terms may seem quite abstract to anyone other than those familiar with the National Qualifications Framework. It is indeed doubtful how accurately the Skills for Life exams have been situated within this framework. T3 especially expressed her disbelief at a Level Two Skills for Life qualification being equated to the GCSE grade C-A* as it does according to National Qualifications Framework.

The equivalency of exams within the National Qualification Framework has itself been cast in doubt (Sharp 1998). Sharp suggests that the drawing up of a table initially to put vocational and academic qualifications onto one scale was largely the result of political imperatives current at the time when the government was trying to promote the new concept of GNVQs (General National Vocational Qualifications) in the late 1980's when the then current Secretary of State, Kenneth Baker, was trying to revitalise the further education sector. As part of this endeavour NCVQ (National Council for Vocational Qualifications - the precursor of QCA) gained main control of the process and 'parity of esteem' between various exams became one of the main features of this drive. There was no empirical evidence gathered however that such parity existed. This thus does not provide a convincing foundation for claims that the Skills for Life exams have 'parity of esteem' with other exams simply because of their placing in this framework.

The Cambridge main suite exams however were recognised by local UK employers also, according to one of the DoSs (DoS3-10:1172 and 500). This stands in contrast to the Skills for Life exams which, as already commented, were not yet widely recognised. Teachers expressed hope that the Skills for Life qualifications would in future help their students in finding work (T4-9:40).

In contrast, DoS2 felt the literacy and numeracy exams were beginning to be recognized, but not the ESOL ones, and since the former had emerged sooner, in 2003 and thus been in existence longer, this is not surprising and may partially also be the result of the highly publicized public information campaigns (e.g. the Gremlins campaign) aimed at increasing recruitment for adult literacy and numeracy tuition amongst the UK workforce. ESOL has to date not benefited, or needed such widespread publicity, ESOL classes generally being full to capacity in most areas of the UK.

7.4.4 What is being measured?

Students have very individual learning experiences. It is the nature of learning, and especially of learning languages; it is not regular, linear and uniform across a group of students (Shohamy 2007). What students learn in their ESOL classes is not necessarily what is actually measured, especially if the assessment system is imposed centrally. Especially in ESOL classes where far more than simply the language, but a whole cultural system is being assimilated it would seem even more suitable than for other subject areas offered at colleges to measure success by evaluating personal outcomes rather than aiming for a pre-assigned, centrally controlled set of outcomes (in this case based on the new ESOL curriculum). A real achievement such as securing a job does not 'count' for funding purposes as the student is seen to have not completed the course. The department is in effect penalised financially for the students not finishing the course. A true, valid outcome such as a student leaving to find work, thus proving their language ability is not recognised as a valid outcome. This is of course not unique to FE; the same system operates in Higher Education also.

7.4.5 Administrative conflict

Further to the points made above regarding placement testing and spiky profiles, among the features of the new Skills for Life end-of-course achievement exams was their modular nature. Students could take exams at different levels in different skills, catering for (and previously described) as students' 'spiky profiles'. This refers to the recognition that students may not be (and indeed are unlikely to be) at the same level of ability in all four skills and so the ability to take exams in different modes at different levels to accommodate this uneven profile appears outwardly to be a very student-centred approach. However this flexibility, according to what some of the departments believed, at that time was not reconciled with requirements stipulated by the college, as regulated by the local LSC, in drawing down funding on proof of student achievement. This caused much frustration on the teachers' parts (e.g. T-7:115). The administrative systems set up in colleges did not accommodate such pedagogically advantageous features of the Skills for Life exam system.

The exam the groups of students in any one class were due to sit was known as the QualAim (as already discussed). A QualAim was 'assigned' to each class, and while this could be changed, on the whole, all the students in the class would sit the same exam (pre-designated via the QualAim) at the end of the course. This does not seem to be consonant with the student-centred approach suggested by the insistence on full engagement with the ILP process, encouraging individualisation of learning (T5-12:911). This was one of many dilemmas teachers were coping with at this time.

7.4.6 Summary

In this chapter the first set of research questions was dealt with:

- RQ 1.a) What is the range, nature and function of assessment practices in UK ESOL teaching?

RQ 1.b) How are these practices linked to the Skills for Life strategy?

This chapter aimed at outlining the sum of the assessment practices experienced in ESOL teaching in the UK at the time of my study to answer RQ 1.a). Although teachers were asked about all assessment practices their students might experience, much less was reported on their own progress tests than for the initial assessments, for ILPs, and for the new Skills for Life exams (RQ1. b), which seemed to be the main causes of concern to them and therefore featured more highly. The balance of the amount of concern raised appears to weigh more heavily on the side of the 'exit' assessments (the new Skills for Life exams aligned to the new curriculum) at the time of the study, compared with the entry assessments (diagnostic and placement) and formative classroom assessments, most of which existed prior to Skills for Life. A notable exception has been the change to more formal, standardised use of ILPs from being used as progress measures for accountability purposes.

The description of these assessment practices which ESOL students may experience as a 'journey' is a valuable metaphor in that it affects not only the actuality but also the ideal of a journey. In other words a journey can be seen as simply the act of moving from place A to place B. It happens with little thought, especially if someone other than the person moving has planned that journey. However, ideally a journey is an opportunity for personal development and discovery. These assessment practices may be missing out on the chance to promote students' development the more centralised, systematised and less individually-oriented they become.

In this chapter I have outlined the types of assessment that may take place in ESOL classrooms and in what ways they are directly a product of the Skills for Life strategy or not. In the next chapter I will discuss the evidence of washback from assessment practices that emerged from the study.

8 FINDINGS AND DISCUSSION: WAS WASHBACK EVIDENT?

8.1 Chapter overview

In this chapter I wish to engage with the effects of the new exam regime, which was put in place as a result of the Skills For Life strategy, as described in the previous chapter, as well as examining the effects of other, extant, assessment practices. My research questions initially lead me to search for concrete evidence of washback, namely ‘the effect of testing on teaching and learning’ (Hughes 1989: 1) which in this study has been extended to include the effects of any formal assessment, and this chapter reports on this. This process was one of unravelling: unravelling the effects of specific assessment practices instigated by Skills for Life and the effects of assessment in general. As stated already, there was no ‘clean sweep’ whereby all new practices began and all older methods of assessment were abandoned when Skills for Life was introduced. In other words, there was no ‘watershed’ point of pre- and post-introduction of Skills for Life as regards assessment methods which would have made the study of the effects much more precise and easier in terms of attributing and evaluating effects. Given the nature of the Skills For Life assessments being closely based on the new ESOL Curriculum there was also the process of unravelling what effects resulted from assessments per se and what were a result of this new curriculum.

In a time of considerable upheaval given the move to a centralization of ESOL matters (as discussed in Chapter 3: Background to Skills for Life) what became clear from the data was that there was a messy picture regarding assessment; it was not clear for the staff involved necessarily what exams were on offer and what the requirements were. It appeared to be a time of confusion in general. As an outsider to this situation this

was both an advantage and a disadvantage. I could perhaps see a bigger picture and had points of comparison gained by collecting data from 3 sites. In essence, the core question was always: is there a change in behaviour which can be traced to assessment? Can it be related to one of the types of assessment undertaken at this particular site? Is there really an evidential link (Messick 1996) that this is so? By this is meant, can it be certainly ascertained that the observed behaviour which appears to be as a result of a certain exam, is actually caused by that exam, and not any other factors. This is how I proceeded to try to answer Research Questions 2a) and 2b):

RQ2.a) To what extent is there evidence of washback from the assessment practices?

RQ2.b) Is any washback only related to the Skills for Life related assessment practices?

Whereas the data for analyzing the Assessment Practices in the previous chapter was derived from the interview data, this chapter draws on the observations as well as the interviews.

First of all in order to clarify the washback which I aim to investigate I will briefly illustrate the parameters of washback by outlining some effects from assessment in general which I observed. In other words, other identified behaviours which have not been categorized as washback will be described to highlight and clarify what specifically has been included in the analysis of washback effects. First of all definitions of washback will be clarified.

8.2 Defining washback

8.2.1 Key parameters in discussing washback

Dimensions typically occurring in discussion of washback are whether the washback affects content or methodology of the classroom, cf Alderson and Wall's fifteen washback hypotheses, in particular hypotheses numbers 3-6:

A test will influence what teachers teach

A test will influence how teachers teach

A test will influence what learners learn

A test will influence how learners learn

as defined in their seminal (1993) paper. Glover (2006) suggests the methodology of studies investigating the washback on teaching methodology has led to evidence which 'often seems unclear or contradictory' (p40). An initial rough definition divides up what (content) is taught versus how it is taught (methodology). While initially seeming another obvious dichotomy it is maybe more useful to consider washback on a cline with content at one end and methodology at the other since on closer examination of actual instances not all effects fall neatly into one or the other group.

In Chapter 4, the nature of positive and negative washback was discussed. In this chapter I aim to explore further whether this is a clear dichotomy or whether there are cases where effects may be deemed to be both or neither.

8.3 What is washback? What it isn't?

As stated in Chapter 4, in discussing washback it is important to distinguish general behaviours which are associated with the general process of assessment from specific outcomes which result directly from a specific assessment or exam. In this section I will discuss some behaviours associated with formal assessment but because they

were not identified with any assessment in particular, I have not classified these as washback.

8.3.1 Student's test anxiety

Firstly, one type of behaviour commonly associated with assessment, as mentioned by Alderson & Wall (1993), is test anxiety. This has been described as:

‘a multifaceted condition which encompasses task-irrelevant cognitions, heightened physiological arousal and inefficient study behaviour and has a debilitating effect on academic performance’ (Kirkland & Hollandsworth 1980, cited in Edelman & Hardwick 1988: 225).

It is a widely experienced phenomenon (although to varying degree), linked to various assessment situations and it must be remembered ‘test anxiety does not always have a detrimental effect on performance’ (Galassi et al 1984, cited in Edelman & Hardwick 1988: 255). The extent to which a candidate may suffer test anxiety will depend on a range of factors such as familiarity, or lack of it, with the assessment itself and also general self-confidence, along with various other affective factors (Stober 2004).

A specific factor affecting test anxiety is previous experience of an exam situation. For instance, T1 mentioned the case of one of her current students who was adamant he did not want to take any exams at all having attempted the Cambridge ESOL FCE exam the previous year but failed (T1-2:886). T1 reported this experience had a severe adverse affect on his confidence in his language abilities and this had had a profound effect on him and his attitude to further exams. T6 also mentioned the issue of students losing confidence from a previous bad exam experience (T6-13:2312 and 13:2363).

Test anxiety however has not been categorized as washback as it is the effect of assessment in general, as widely experienced and reported on, if only anecdotally, rather than being the specific effect of any one particular exam or type of exam. There are many factors which could have caused the anxiety, independent of the nature of the exam, such as poor exam preparation including sitting the exam before the candidate is ready, ill health or other distractions on the day of the exam for example.

8.3.2 Teachers' anxiety

This test anxiety may be felt by teachers as well as the exam candidates since the latter's performance is naturally viewed as linked to the teachers' ability, both specifically in teaching and also in preparing the candidates for the assessment (see Smith 1991 for further discussion of this). For example, as DoS1 says:

'at a certain stage obviously you can see there's a lot of panic around because they [teachers] obviously want their students to get the results and I suppose you start to think 'Have I delivered enough to them for them to get through?'
(DoS1-1:988).

T3 admitted feeling similarly:

'You are judged by your [=your students'] exam results no matter what anyone says (T3-6:676).

What this shows is that the teachers can be very aware of the effect of their involvement in the students' exams and feel the responsibility of their input and how this reflects in the students' ability to show their skills. Nevertheless, this too is not being categorized as washback in this study since teacher anxiety is a general phenomenon and not specific enough to link to a certain assessment. Such anxiety

may be viewed rather as an inevitable feature of the caring, reflective teacher practitioner and, again depending on level of severity, may not necessarily be purely negative.

8.3.3 Test-wiseness

Another behaviour which, according to the definition used so far, would not be classified as washback is the teaching of the type of knowledge and techniques often imparted to candidates as part of exam preparation to help them maximize their performance, which it is felt students need, in addition to their language ability. It is known as developing test-wiseness. Being generic and not applicable to any particular exam, test-wiseness was not classed as washback.

Examples of such techniques are raising awareness of the nature of the assessment e.g. types of items included (multiple choice, short answer, essay etc), the number of sections in the exam and the time allowed, and

‘we’ve got to get them used to format, the timing, the amount of writing that’s got to be done (DoS3-10:1704).

Test-wiseness encompasses behaviours and skills which are needed only for the exam situation and do not have relevance to real-world activity. Some teachers did not seem to understand this key difference:

‘well it’s reading the instructions, watching timing, checking the work if they’ve got time - which we’re encouraging – self checking at the end, planning’ (T2-3:1430).

These are practices which would be useful in working under timed conditions, or useful for producing good written work in various circumstances and this is not what is being referred to as test-wiseness.

Test-wiseness is in itself not necessarily negative. Teaching test-wiseness, it can be argued, thus ensures students do not lose marks unnecessarily by completing tasks incorrectly or wasting time during the exam trying to work out what to do. Students are able because of such preparation to display their ability without disorientation brought about by unfamiliarity with exam procedures (Simpson 2006). The variety of students' educational backgrounds is one of the challenges of the ESOL classroom. While some students are very familiar with exams and have already sat many during their education to date, others may have never been in such an environment. Cultural differences mean that expectations of what is deemed beneficial exam behaviour is not universal. Many countries are not as assessment oriented as the UK is often seen to be and students may have little experience of sitting exams.

Sometimes even 'simple' concepts such as recognising where to write one's responses may need class time devoting to them, as explained above. For example, as T6 mentioned, one of the things not all her students were familiar with, and which needed explicit teaching, was

'this idea of writing on a dotted line and ticking boxes...' (T6-13:2891)

T6 recognised that nothing can be taken for granted in ensuring the students are familiar with what they will be confronted with in their exam, and that they know how to maximize their scoring by knowing appropriate ways to present their responses.

Similarly T3 agreed:

'what I've taken on board is they've got to learn about the instructions – that is a lesson within itself' (T4-8:326).

The teachers expressed a variety of attitudes towards teaching test-wiseness. In general there was resignation about its necessity but also some resentment about the amount of time it took. What some teachers, as T2 above did not seem to recognize was that useful study skills such as planning and checking work were being imparted

which it can be argued are useful skills which are transferable beyond the study environment and exam. At a more basic level, knowing to write on the line (not below it as with certain scripts such as Bangla and Hindi) has other transferability into everyday life in for example successfully filling in forms. Certain practices taught under the umbrella of exam technique/ test-wiseness, such as writing to a strict time limit, are less obviously transferable for most people however.

8.4 Evidence of washback

Having discussed some behaviours associated with formal assessment which are not classified as washback for the purposes of this study, I will now move to reporting where there was evidence of washback. First there will be an overview of the possible washback noted from the observations. Then I will examine the possible washback from the internal measures and then the external ones.

Table 15 clarifies which questions in particular aimed at eliciting information about washback and related effects. (See Appendix 1 for full interview schedule). The relevant sections of the observation schedule are included in a later section.

Table 15 Interview questions specifically focusing on washback effects

Question	Q# Teacher schedule	Q# DoS Schedule *
What effect does external assessment have on your ESOL classes? (which in particular?)	36	40
Does it affect how you teach? How? Why?	37	41
Does it affect what you teach? How? Why?	38	42
Does it affect what students want to learn?	39	43
Does it affect the teachers' workload?	40	44
What effect does internal assessment have on your ESOL classes?	52	56
Does it affect how you teach? How? Why?	53	57
Does it affect what you teach? How? Why?	54	58
How has S4L influenced your teaching- if at all? Is it still influencing it now?	63	n/a
If so how? And why?	64	n/a

(* wording was changed to: what effect does X have on your teachers..?)

The final section of the observation sheet elicited a rough assessment of the level of ‘visibility’ of exams in the class (see Appendix 2). Table 16 **Level of exam visibility** summarises the results. Points 4) and 5) most directly relate to obvious washback. Points 1) to 3) aimed to elucidate how visible assessment was in the class and what affect this may be having on attitudes in order to relate to any washback I had observed in the class.

Table 16 Level of exam visibility

1) Did Ss show awareness of assessment?
2) Did students show any negative /positive attitudes to exams (including anxiety)?
3) Did the teacher show any negative /positive attitudes to exams (including anxiety)?
4) Was the influence of any assessment measures/ exams / tests noticed?
5) If so, did it influence content or methodology or both? How?

	T1:1	T1:2	T2:1	T2:2	T3:1	T3:2	T4:1	T4:2	T5:1	T6:1	T6:2
1	✓	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓
2	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
3	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
4	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	✓
5	C	C	-	C	C	C	C	C	C	-	C

Key: T1:1 = 1st class by T1 to be observed, T1:2 = 2nd class by T1 to be observed etc.

C= Content/ M= Methodology

Note: a tick denotes that at least some activity in the lesson was judged as manifesting washback.

Table 16 shows that washback was not seen in all the classes observed. In T2’s first class for example, none was seen. On the other hand in T6’s second class washback was seen in various ways, according to all the given parameters (1-5). In all but two

cases the students showed an awareness of the assessments and in all but two (though not the same two as previously) washback in some form was noted.

Table 17 below summarises the examples noted as a result of question 5) of this section of the observation schedule. Level of class and the QualAim of the class have been appended in order to elucidate the context in which the washback was noted.

Table 17 Observed examples of washback

Evidence of any influence of the assessments on class activity from Observations:

Class	Washback observed	Nature of activity where washback was observed	Level of class	Related to which assessment
T1:1	✓	an exam practice class	L2	S4L + FCE (E)
T1:2	✓	choice of materials – mirrored exam tasks	L1	S4L + FCE (E)
T2:1	✗	-	E2	-
T2:2	✓	speaking task practice	E2/E3	S4L exam (E)
T3:1	✓	checking previous test	E1	Mid-term review (I)
T3:2	✓	rationale for focus eg writing fast but legibly given exam focus – any real world application - exercise type flagged up as being possible question type	L1	S4L exam (E)
T4:1	✓	Explanations given in terms of exam	Mixed check	S4L exam (E)
T4:2	✓	(end of module test taken)	E2	Internal progress tests (I)
T5:1	✓	Exam practice - text types practised – flagged up as being on exam. Did practice test question for writing	E1/E2	S4L exam (E)
T5:2				
T6:1	✗	-	E1	-
T6:2	✓	Activities mirrored the exam type questions	E2/E3	S4L exam (E)

(E) = External assessment

(I) = Internal assessment

Note: The classes covered Skills for Life material and preparation for other external EFL exams

8.5 Effects of internal assessment

8.5.1 Independent Learning Plans (ILPs)

As discussed in the previous chapter, the rationale for having to complete ILPs was not fully understood except in the case of the students whose language ability level was too low (namely Pre-Entry Level students) to be measured by the available external assessments. Because of the confusion over their requirement and use, they seemed to generally be resented. This is an example of a practice which had it not been thought to be compulsory would probably not have featured in the teachers' classes in the form prescribed at least. T2 stated, teachers tend to track individuals as part of good teaching practice in any case (T2-3:811) but did not feel the need to undertake this tracking in the level of detail or in the format proscribed.

Because of a focus on differentiation (the principle of providing learning activities to address students' individual learning needs in the classroom), as well as the administrative element of maintaining ILPs, they could be time-consuming and difficult to manage. Due to class size, as DoS3 referred to above, the preparation and logistics of providing what students required in order to develop their individual linguistic needs proved challenging. T3 for instance reported:

T: *'ideally once every few weeks we should have one lesson which is workshop type lesson where they're all working on individual things and to do that I've got to plan twelve different lessons I've got to work X planning these to work on 'l' and 'r' and how can you do that on a student working alone it's finding an exercised that'll do it and the cassette and headphones get it set up and explain you know and what do I do with the rest of them while I'm doing that? Then Farhad wants to work on sentence structures so I get him something exciting*

to do that and then someone says 'oh can I have that?' 'no - you're doing 'p' and 'b'

I: and you don't have any extra floating volunteers or assistants or anyone who comes in?

T: no no – so I have before I have done it in the past when they've all had very similar targets but this time they've all got such different ones] - I can't face it - and I just feel again I could just give them all tons of worksheets to do but I won't feel right about doing that' (T3-7:662).

In addition, the lack of understanding from the College management was reported by DoS3:

'I went to talk to the Head of Basic Skills and [said] how we actually found it quite difficult to do ILPs with near beginners and do a proper needs analysis I said because of the lack of language and he said 'oh you don't need language you just need to look into their eyes look into their eyes – I can tell what they need' (DoS3-11:1394).

While we would hope this was not a typical response from management, it is symptomatic of a level of remove and lack of support reported, but which in this case only enhanced the ESOL team's sense of burden of the new regulations, as they perceived them, which was not particular to this site only, as already stated in the previous chapter.

T2 also mentioned the inherent tension in finding time to undertake the individual one-to-one time with students to help them complete their ILPs within the overall structure of the whole class approach. As T2 put it, due to:

'the nature of ESOL teaching the student likes you to be there within the class group so to actually find time to actually work individually with a student is quite difficult and so quite often you're having to give extra time over and on

top of [it] - otherwise you've got to set them tasks and if they're set a task which is not under exam conditions they want support or they want to consult you or ask you things - it is difficult – they're here to be taught as a group' (T2-3:1532).

The need for collecting 'evidence' of achievements for individuals to include in their ILPs was a constant feature of all classes, causing its own strains, again according to T2:

'what I'm saying it is hard because the students don't necessarily see the importance of the ILP because they want - they generally want to work as a group to go - move through as a group and when you're teaching you are all the time in ESOL teaching assessing all the time and you're expecting different outcomes from the students whether it's in question and answers or in written production or evidence from reading evidence for listening and speaking – the whole time' (T2-3:812).

Tensions were apparent regarding group versus individual needs and the maintenance of students' expectations.

8.5.2 Coping mechanisms in use of class time

Given the pressure some of the teachers generally felt under to cover the Curriculum (T3-7:356; T4-9:419) because the exams could sample any of the relevant sections of the Curriculum, time in class was particularly precious. The teachers had various ways of coping with the perceived pressure to complete the new style ILPs without eating extensively into this valuable time. For instance, T1, as mentioned in the previous chapter avoided ILPs altogether, while T3 for example mentioned her solution with her high level classes was not to fulfil her obligation to hold one-to-one tutorials as was recommended to the teachers since she explained

'I'm not going to do them because we've lost so much time already' [her students had lost class time sitting other non-English language exams at the college but they would still need to take Skills for Life accredited exams at the end of the ESOL course] (T3-7:641)

and feeling the pressure of the looming exam, she chose this course of action. She later in the interview backed up her decision with the confidence in her DoS's understanding of the situation and felt the DoS would understand the pressure she felt under and would condone her decision (T3-7:737), although this course of action had not yet been ratified. It seemed this relationship with her DoS was an important element in enabling her to find a solution to her problem and reduce the pressure she felt under.

Another solution to juggle completion of ILPs with maximising class time for learning which I observed was used by T3 with her low-level groups in a bid to streamline the process of producing and maintaining ILPs and to make full use of class time, in order to help students produce their personal targets for the forthcoming module. The students were given a handout of the learning outcomes for the coming module and students cut out from this list the outcomes which matched their personal ILP targets and glued these into their ILPs in the appropriate section (T3-18:59), which was concocted of a file of papers, one paper at least for each class.

Likewise, for her students to record the learning outcomes achieved in each lesson, T3 tried to elicit them from the group, wrote them on the whiteboard and students then copied them into their ILPs (T3-7:596; T5-19:56). T2 similarly wrote on the board the learning outcomes of the lesson for the student to copy down into their ILPs (T2-16:59; T2-17:70). In both cases this still took time but less time than trying to help students discern whether and how any individual targets had been reached in the lesson and how to express this in SMART terms in their ILPs in comprehensible

English. The element of individuality was lost thus but all the students had well-maintained ILPs. This activity could have been beneficial, in theory, in acting as a class review each lesson. However, again given the time constraints in class, the review of learning outcomes did not come entirely, if at all, from the students themselves at least in the classes observed. This would have been a much more valuable exercise in pedagogic terms and time taken for producing a written record instead could have been used in a more meaningful purely verbal review.

8.5.3 Summary of washback from internal assessment

Washback from the other internal assessments such as the placement/diagnostic tests, progress tests, or portfolios leading to College Certificates at Site 1, were not reported by the informants or noted in the observations. ILPs were the only internal measures which recurred in the interviews in relation to their effects on teaching and learning. It must be remembered that ILPs were in place before the instigation of Skills For Life but their format, purpose and formality changed. In summary, the washback of the ILPs was that class time was taken up with a measurement tool which the teachers seemed to have little confidence in. Pressure from the new external exams made time in class precious to cover the amount of necessary material. Drawing on mechanisms to reduce the time taken up by ILP activity seems to have led to the loss or at least distortion of individualization which was the main rationale for ILPs.

8.6 Effects of external assessments

In this section I will discuss effects related to the external exams which students were taking as set out in the previous chapter. It must be remembered that not every

single student had to sit a new Skills for Life exam due to the 80/20 rule. The alternative measures (such as ILPs) were generally used for the lowest ability students for which there were no curriculum or exams available since they began for Entry 1 level students and some students began classes as total beginners, i.e. at pre-Entry level. Throughout this section where the need for an exam as proof of progress is mentioned, it is taken as read that not all students needed to sit an exam, as described above. However in the case of my data from three sites none of the classes were at pre-entry level and all had a Skills for Life exam recorded as a QualAim.

References to what was analysed are only tentatively termed washback as will be explained at the end of the chapter. In the interim the term is used for effects noted which are clearly related to specific exams not to exams in general. This section will first discuss the washback which seems to affect the content of ESOL classes, then that which affects the methodology, then that which it is hard to categorise as either content or methodology specifically since elements of both are identifiable. Next I deal with those areas of washback which were neither a matter of content or methodology but seem to be more of an affective nature. Finally I will discuss washback which reaches beyond the classroom, and so may better be classified as impact.

8.6.4 Content

8.6.4.1 One class: two goals

As mentioned in the previous chapter, all the sites involved in this study entered some students not only for the new Skills for Life exams but also entered students for the Cambridge main suite exams. This was because of the combination of two factors. First was the nature of the funding system in place at the time of data collection,

which meant a Skills for Life exam result was needed as part of the accountability system by which funding was secured, as already discussed. Secondly, the profile of students attending ESOL classes was not necessarily that of the typical ESL student (see 2.3, Background to ESOL); it must be remembered, various students attending ESOL classes did not intend to settle in the UK. Some of the students in these classes attended English classes primarily with the aim of achieving an internationally recognized qualification in English, typically a certificate from the Cambridge main suite of exams³⁰. As T6 pointed out:

'the thing with FCE, you can get a good job' (T6-13:400).

Since the ESOL classes ran as a result of funding secured through adequate proof of student progress as decreed by the LSC, all students needed to sit Skills for Life exams. In fact it was advantageous for colleges to allow all types of students, both EFL-ers and ESL-ers onto courses, as long as they sat the accredited Skills for Life exams and thus drew down funding for the ESOL department.

Students in one of T1's classes were mainly the type of students who were of intermediate level and above who had come to ESOL classes principally to gain an internationally recognized English qualification for future job prospects. T1's students wished to sit FCE or CAE and were keen to reach the required level of language ability to do so. T1 was fully aware of their wishes, and was also as aware that it was expected of her group that they would sit the National Test, being Level 2 students (see section 3.3.6, Background to Skills for Life). To accommodate the students' wishes and the requirements of her college, she had decided to engineer her classes such that the students were being prepared for both sets of exams: both Cambridge and Skills for Life exams (T1-2:1120). I labelled this practice of teaching

³⁰ These exams are produced by what is now known as Cambridge ESOL, an arm of Cambridge Assessment. This name will not be used to avoid confusion with the Cambridge Skills for Life exams, aimed at ESOL students, which are also produced by Cambridge ESOL.

in effect two curricula at once as ‘double accounting’: having students on two sets of ‘books’ at once, referred to before as piggy backing.

The material T1 was using in class was clearly geared principally towards CAE preparation and was akin to EFL classes one may see anywhere around the world in an English medium language school rather than the style of ESOL class based upon the Skills For Life curriculum which I had observed elsewhere (T1-14:0005). She admitted to not having been influenced by the Skills for Life material or goals to date (T1-2:709); Skills For Life Level 2 materials were not being used at all, and classes were not cross-referenced to the curriculum (T1- 2:0748) as was happening not only in other classes in the same institution but also in the other institutions I visited, and which other teachers I was in contact with at the time of data collection were doing. T1 chose not to be affected by Skills for Life on the whole and found mechanisms to avoid the changes other teachers were making.

To some extent the same ‘double accounting’ was happening at the other three sites. Students wanting to take an internationally recognized English exam (typically Cambridge FCE, CAE or IELTS) were having to learn English in a class where the teachers would need to enter the students for Skills For Life exams as well as entering the students for the international English exams they wished to take. While CAE and FCE were desired typically by ‘short-stay’ students, IELTS was often required by those asylum seekers who were wanting to resume former professions, and who definitely had little choice about staying in the UK. At Site 3, for instance, some of the students studying for IELTS wished to take up their professional identities in the UK in medicine (DoS3-10:1254). The college system was not facilitating this process however.

Site 2 was in a similar situation. The DoS reported that while, at the time of the interviews, they were able to run a class focussed on IELTS preparation, the following year the next cohort of students for that class would have to take the Level Two Literacy exam and somehow find their own funding to pay for the IELTS since this was not to be recognized as an accredited qualification to draw down funding (DoS2-4:950) and therefore could not be offered by the college. DoS2 was planning to offer facilities for students to be able to prepare for IELTS as it was so important to some, if not all, of the high level students. She at that point did not have a clear plan as to how to achieve this, but was clear that in any case all students needed to sit the Skills for Life exams (DoS2-5:70). T6, at Site 3 succinctly outlined similar plans:

'we're talking about doing some sort of self-funding so they would pay up front for it and we would just do exam practice technique for those who want to do it but their funding would come though Skills for Life' (T6-13:405) (see also T6-13:421 and DoS3-10:1244).

Again, students would have to pay the exam fees themselves. T1 reported a similar proposed system for the following year, offering CAE preparation, but students would likewise have to pay the fee for this exam themselves and continue meanwhile preparing also for Skills for Life exams (T1-2:1112). Meanwhile students were in effect experiencing a somewhat incoherent syllabus which aimed to both cover the new curriculum as required, to a greater or lesser extent, depending on the teacher, and also to prepare them for international language exams. While the data did not provide hard evidence that this arrangement affected the students' learning, it clearly affected the teaching in that the tutors were juggling various needs to achieve specific ends in a specific time frame (the length of the course).

T5's view was:

'personally I think it would be good to carry on offering both as long as it – the funding - isn't affected - why shouldn't we offer both?' (T5- 12:619).

However, clearly funding was indeed affected, which was why such practices were taking shape in the ESOL departments. Once more, a pedagogical aim, i.e. students defining their own learning goals, seems at odds with an administrative function directed by the funding imperative. The irony of this situation is that the dominant discourse of the Skills for Life documentation is that of English for social inclusion and employability purposes (DfEE 2001; Papen 2005). Many students at this higher level, typically those needing IELTS, were needing and wanting to join the UK labour market as a professional, and by joining a profession social inclusion would surely be facilitated, but they could not do so without this qualification and the Strategy funding requirements did not facilitate this.

Students may not necessarily have wished to prepare for or sit the Skills for Life exams but in order to receive free tuition, this was the 'deal'. Sometimes this was made explicit to students who seemed reluctant to fulfil their part of the 'bargain' as T6 reported:

'and this year I've become really strict and said 'you get it free because there's an exam' because a lot of them were saying 'I can't come to the exam because I've got a job to go to' – 'you've had free all year - I ask one thing of you - I ask one thing - you come to the exam - you will be there' (T6-13:1762).

T6 clearly had strong views that the students needed to understand how important the exams were. However T6's focus was on the importance of the exam results for the department to draw down funding, rather than their importance as a qualification per se. Similar views were echoed by T3 (T3-6:657). As T6 said, she was not always in tune with some of her colleagues in this regard, who advocated students only sitting the exams they wished to take:

'they say people come and they want these qualifications and I say you can't look at it from that point of view – why would the government give money to

come to this country to pay someone to learn English to go back to their country – what’s in it for the government?’ (T6-13:437).

Thus as a result of these circumstances, the dissonance between pedagogical and administrative requirements meant some students were learning Skills for Life material which was not appropriate for them.

To summarise, the washback of the exams in some cases (but not all) took the form of a change in the class content. The students in such classes may not have elected to study according the ESOL curriculum, had they not been required to do so, in order to sit the Skills for Life exams, which they did in order to receive the free classes. There is also washback on the teachers to some extent in altering how they managed their classes to accommodate preparation for two sets of exams.

8.6.4.2 Focus on accuracy

One effect of the new exam regime which one of the teachers discussed explicitly is the issue of a change in class focus, especially regarding the levels of accuracy in language use demanded of students. It resulted, according to T4, from the external verification of ability:

T: ‘one of the things I was thinking about is I won’t be able to make allowances - you’re marking a person’s piece of work and they left out little things but you knew that they were OK with that - you might give them the benefit of the doubt but because you never see the examiner or the person who marks it you know everything’s got to be as perfect as it possibly can so when you’re marking things in class you’re not making allowances - you’re saying ‘well - no get this right you know’ – when they miss out an apostrophe

I: There’s a bit more attention to detail the whole time?

T: Yeah - and handwriting - we'll improve this - and spelling - you've got to get these basic words correct whereas before you'd give them the benefit of the doubt according to how much you knew about the student - now you normally can't do that - no you've got to get everything right' (T4-9:635).

This example suggests there is evidence of washback on classroom assessment and feedback to students, not just on teaching in general. What they choose to make students aware of seems to have changed in light of the marking criteria for the forthcoming exams.

Another way in which the exams had an effect on teaching was in the way in certain cases they altered the relationship between teacher and students. T4 for instance made clear he saw the external exams as of benefit for maintaining standards because of teachers' subjectivity regarding their students' abilities; his role as his students' assessor dictated his own responsibility towards them, and his evaluation of his skills to do so (T4-9:731). T2, in a similar vein, also claimed the need for an outside objective evaluation of the students' abilities because as she said:

'you're with them [the students] in a classroom - in a way you can tune in to the student and you - not make excuses but allowances and in a way assessment is objective and it should fulfil those criteria whatever the criteria - validity and objectivity' (T2-3:2267).

As did T4, she recognized that the teacher sees the whole picture of students' ability over a period of time whereas the exam takes a snapshot of that ability taking only certain explicit, demonstrated ability into consideration whereas the teachers will probably consider the whole person plus circumstances which may affect their performance on any specific occasion when their ability is being assessed. The exam judges the product (displayed ability) whereas the tutor may judge the process (the

learning journey) as well as the product. In other words teachers may feel they understand a student's competence whereas an exam can only judge performance. The exam allowed her to focus on teaching, without having to be concerned about her role as assessor and having to reconcile both roles with her students, and thus affected her classroom behaviour. The matter of alterations to student – teacher relationships will be returned to later.

A focus on accuracy in the classroom is in itself neither negative nor positive but depends entirely on the overall balance between accuracy and fluency (Skehan 1989). The effect of changing a class to increase work on accuracy may indeed be viewed as positive washback, but would depend on the nature of the classes beforehand. If the change made the class more balanced then indeed it could be deemed positive. A shift in balance from one to the other could be detrimental however and the change viewed as negative washback. Given this, it only holds if the class goals are general language development and do not have some specific aims which may occasion a skew towards either fluency or accuracy depending on what that aim may be e.g. fluency may be less valued as a skill if the ultimate aim of the language learning is to translate into L1 or, for ESOL students, a focus on accuracy in their written work may not be so necessary if their use of English for work and socialising is mostly of a verbal nature.

Observation of T4's classes did not suggest an approach dominated by a focus on accuracy since the larger proportion of his class was taken up with students speaking in small groups to fulfil a task (in two separate activities) where the aim was to find out information from each other, not to demonstrate accurate use of English (T4-20:18 to 60; T4-21:19 to 43). In this case we only have the teacher's impression as to the nature of the changes in his class in general. This highlights the need for baseline

information to accurately evaluate washback however, as we do not have data on the nature of his previous classes, prior to the new regime.

8.6.4.3 Effect intensity

The Skills for Life exams are available in modular form, there being 3 modes: Speaking and Listening; Writing; and Reading, as per the divisions of skills in the ESOL Curriculum National Standards and Descriptors (DfES 2001). These modules can be taken at different times in the academic year and some sites spread out the scheduling of exams. T6 discussed how the focus in her classes changed depending on which module of the exam was imminent, for example she reported, since her students had taken their speaking and listening exam already at the point of my interview with her, the class had moved onto concentrating primarily on reading and had recently done very little explicit speaking practice (T6-13:2414). She admitted that for a couple of weeks prior to the speaking exam they had concentrated solely on this skill (T6-13:2461) and classroom work had taken the form of activities which closely mirrored the exam format (T6-13:2409).

'when we got like the speaking exam – so you would do it where you talk to each other about something or interview each other ... but really since we've come back from half-term the speaking's gone because we haven't needed to do the speaking' (T6-13:2412).

Again, due to specific goals, namely the Skills for Life exams, teachers and DoSs reported feeling under pressure to prepare students for these exams in the period as the exams approached.

DoS: *'because of the pressure we're under by the college our final term which - is eight weeks long - our final term is dominated by exam practice*

I: *how do the teachers feel about that?*

DoS: they feel it's what they want to do, they feel it's what the learners want to do, everyone wants success' (DoS3-10:1693).

This was echoed by T1, and verified by T5 who spoke about how the year was divided up, with the first term and a half spent on learning English and the final half term spent on exam preparation (T5-12:979) namely, focusing on the familiarization with format and item types on past and sample papers rather than continued language development. (Further language development may happen incidentally while working on such material, but this is not in a planned, informed manner as it would be in non-exam focused coursework, hopefully). Other common practices such as 'cramming' were mentioned, in T4's case in the form of increased homework loads prior to the exam (T4-9:716), although T6 admitted she had doubts about the effectiveness of cramming (T6-13:2417).

This reinforces the notion of skills being developed for the exam, not for language development per se, where the exam is simply one component of this learning process which gives a measurement of ability at a certain time; the exam has become the end in itself rather than merely part of the means to an end. Pedagogical principles seem to become skewed in the shadow of looming exams, what Cheng refers to as 'intensification' (2004).

8.6.4.4 Materials used in class

Washback on the content of classes was manifested also in the form of choice of teaching materials as teachers mentioned how the exams dictated their choice of materials for use in class. For example, T6 was finding it challenging to produce materials to prepare students for the new exams:

'those materials [ones used in a class I observed] were FCE materials that I rewrote and chopped stuff out and rewrote in a slightly different way for

them – that's the time consuming thing - I've found having to rewrite material - creating materials - has been a nightmare' (T6-13:2426).

These did not then resemble at all the Skills for Life materials which accompanied the new Curriculum when it was released.

T5 claimed the choice of textbook for her classes was based on what students needed to fit the pattern of the course in that they began with a focus on English language development and only concentrated on exam preparation in the latter half of the course and would choose appropriate coursebooks for each half (T5-12:1235). This was a class taking Cambridge main suite exams and therefore the teacher had a choice of books available, unlike the Skills for Life exam which being so new did not have commercially prepared textbooks at that point, only a few sample papers made available by the exam boards. DoS2 was looking forward to when such materials might be produced by Cambridge to complement their Skills for Life exams (DoS2-4:1065).

Teachers mentioned their resentment when papers were changed (T6-13:1404 and 13:2883). Most typically it is item type and exact task specifications (e.g. length of time allowed for a paper or number of questions to respond to) which is altered, rather than the underlying constructs. This highlights one of the issues raised by the teachers' use of material in that, teachers usually resort to maximising students' access to parallel exam forms, or what are believed to be parallel forms, in trying to prepare students for the exam in question. Admittedly often, if not usually, it is hard for teachers to gain access to explanation of the underlying constructs of the exams as these are rarely made publicly available (Hamp-Lyons 1998). Since only a certain number of non-live exams are made available for practice purposes by Exam Boards, when these are exhausted, examples from alternative sources are used, but these cannot be guaranteed to be truly parallel as the means to produce such practice

exams (such as specifications) are not generally made available by Exam Boards (see Wall & Horak 2011 for further discussion of this point). Materials do not necessarily do other than familiarise students with exam item types and cannot thus be deemed to be causing positive washback, fulfilling no other function than to achieve better exam results, rather than developing language skills further.

8.6.5 Methodology

While many of the reported effects of the new exam regime concern course content, a second group of effects were related to the methodology of ESOL classes. In the washback literature, it is generally felt content is more commonly affected by the influence of assessment than methodology. This was proposed by for example Alderson & Wall (1993), and Alderson & Hamp-Lyons (1996). Studies since have suggested methodology can indeed be affected, for example Glover (2006) who suggests that, based on his findings from his study examining differences in teacher talk, teachers tended to take a more inductive rather than a deductive approach in exam preparation classes compared to those which are not exam focussed. In my study there were some, though limited, examples of how methodology was affected by the new exam regime.

8.6.5.1 Interaction patterns

Some teachers reported they felt changes to how they taught were taking place. T3, for example, felt she now taught in a more teacher-led, lockstep way than before (T3-7:466) since she, as with T1 (T1-2:747), felt she had more work to get through due to goals (the Skills for Life exams) which were much more specific and exacting than previously had been the case. She felt uncomfortable deviating from her lesson plan to address students' arising queries or apparent linguistic needs:

*'if it wasn't for the exam I wouldn't mind at all but in the back of my mind
'Oh we need to get this exam done'* (T3-7:479)

The two teachers, it must be remembered, worked at different sites so any effect of college or department ethos is not pertinent here and is more likely to be due to the exams.

T6 described how she felt the style of her ESOL teaching was now becoming much more akin to literacy classes (T6-13:1067), which have tended to operate on a more individualistic basis, with students working their way through worksheets matched to their own particular need, as mentioned earlier. The teacher acts as guide, monitor and helper but does not tend to orchestrate interaction and language practice for the whole group, trying to make classes student-centred, in contrast to what in general more EFL-style classes typically aim to do (T3-7:123). Nevertheless, having observed T5's classes (T5-22:19-35) it was unclear why she felt this, as this was not the impression I gained of her classes. For instance the activities in the first observation consisted of students undertaking first discussion in pairs on peer correction of some of their written work, class level discussion, and then a further piece of writing. No worksheets were in evidence and students worked on the same tasks at the same time.

However not all teachers by any means agreed that this move to a literacy style was happening. T2 reported a typical class as:

'some of them do like to work occasionally by themselves but on the whole they enjoy working as a group with group activities and when planning I tend to plan group activities to begin with and then it would break down into individual work or extension activities – so we're starting from the group and then moving out so they're getting the sort of the individual work or extra work on a particular language item' (T2-3:822).

8.6.5.2 Exam practice

While test-wiseness, as discussed above, refers to general familiarization with beneficial exam-taking habits and behaviour to propagate exam success, exam practice refers to the students experiencing parallel forms of the exam to be taken, and also sometimes as exams approach in exam-like conditions such as working according to prescribed timing, and with no conferring between candidates. Most commonly the aim is to familiarise students with the item types found in the target exam, and to experience what they may know in theory about the style and format of an exam and put their test-wiseness into practice.

Exam practice was an often recurring theme in the data and there were mixed views as to the role of exam practice, and mostly in regard to how to teach it and when. For example, although without labelling it as such to students, T6 claimed exam practice was introduced far earlier into the course than had been done before exams became the standard as an end of course achievement measure (T6-13:2464). This had the overall effect of making the course more 'exam-streamlined', in other words the exam took on an importance which shaped the course content rather than simply acting as the culmination of the course, as appeared to be the case with exams before the Skills for Life.

In contrast, T5, at that period early on in the change towards the new exam regime, saw the exam merely as a final achievement measure which deserved no great emphasis during the course (T5-12:1745) which suggests, since they were based at the same site, other factors were at play to influence their differing outlook. These may have included the level of students these teachers taught, T6 generally teaching the lower ability classes, and T5 the higher classes, which typically had a higher proportion of students used to doing exams. Another factor may be their own personal attitudes towards exams (which will be discussed further in the next

chapter) or even their role in the department. In summary the exams did not exert the same effects.

It must be considered how far it was the effect of the new system which influenced behaviours. T1 said that if they had a stock of past papers, exam preparation would take place throughout the year and not be focused into the period immediately prior to the exam (T1-3:1471). T6 said practice would also be introduced throughout the year and not be crammed at the end ideally (T6-13:2464). A longitudinal study could track whether such practices persisted or indeed exam practice became integrated, more subtly perhaps, into courses in time once materials and familiarity with, and experience of, the exams had been accumulated.

While one theme of discussion in the data was concerned with when exam practice took place, another was how much of the course was taken up with exam practice.

DoS3 reported:

'because of the pressure we're under by the College our final term which is eight week long our final term is dominated by exam practice' (DoS3-10:1693).

and he reiterated this later:

'the profile of the college is we want to be a grade one college and therefore we have to have good results and from about this time of year we will increasingly be looking at exam preparation techniques' (DoS3-11:550).

Similarly, DoS2 reflected on the effect which a previous external exam, the CELS (Certificate in English Language Skills from Cambridge ESOL), prior to the Skills for Life regime had had. She projected what she expected to happen once the Skills for Life exams were in place. (At that point they had been chosen but none had yet been taken at her institution, Site 2). She said:

DoS: 'when we introduced the CELS exams at Entry Three they had quite an impact on the latter half of the year in terms of washback from the exam – into teaching – some of which is positive and some of which might not be - so it's – I'm not saying it's all bad but certainly they had an effect and I would anticipate that the Skills for Life will.

I: And do you think that's more in the content of what the teachers are teaching – the actual subject matter?

DoS: Yeah. I think it concentrates people's minds on what they're going to have to do in the exam and to lots more exam prep – exam practice and so on' (DoS2-4:1331).

The degree to which exam practice features in a course will depend on the evaluation of the teacher (and DoS also perhaps depending on their level of involvement in class planning) as to the importance and need for such practice. Some groups of students however may adopt more agency, pushing for more exam-practice material in class than the teacher would wish to include. Typically such groups are those who come from exam-oriented academic cultures and who have joined the college specifically to achieve an internationally recognized English language qualification. (This was described above under 'double accounting'). This was found to be the case for T5:

'I suppose towards this time of year I get more into just the photocopying and doling out test papers with answers - the students just like the answers so my role had changed from teacher to photocopier' (T5-12:1246).

Her colleague, T6, also reported that students requested extra exam practice as the exam approached (T6-13:2419). This behaviour was also reported by their DoS:

'I mean there's a very prevalent attitude amongst most - some of our learners from certain parts of the world who believe that taking the test again and again and again will actually help you do the test - I don't think there's many of the teachers who actually believe that but quite a lot of the

students do believe that so if they feel they're not getting exam practice then we're letting them down' (DoS3-10:1699) (and DoS3-11:544).

Another way in which exam practice changed the nature of the classes in the period leading up to exams, T3 suggested, related specifically to the dominant interaction patterns:

'it was much more them [the students] on their own working and you wouldn't have that in a normal class - you would have much more - a much greater interaction' (T6- 13:2443)

She used the term 'normal class' to differentiate them from those classes devoted to exam practice. In contrast, rather than a change to 'normal routine' it seems some classes, due to exam practice, would be focussing on and intensifying techniques usually included in 'normal' classes:

'we will do more pair interview work because it's on the exam we - will definitely get people talking to each other looking at each other reacting to what each other is saying those sorts of things' (DoS3-11:576).

Site 3 was due to take the Cambridge Skills for Life exams and the format of the oral is a paired candidate interview. Once again, without concrete baseline data it is hard to verify that this behaviour was indeed previously the norm however.

Exam practice is worthy of investigation because in general it has become a normalised classroom practice. None of the teachers raised any issues in relation to including exam practice in class. There was no indication it was seen as unethical or detrimental (see Hamp-Lyons (1998) and Mehrens & Kaminski (1989)). Exam practice should be viewed with caution and its role and prominence in a course investigated. It must be remembered that testing and teaching have different purposes: 'while the primary purpose of other components (of an instructional program) is to promote learning, the primary purpose of a test is to measure'

(Bachman & Palmer 1996:18) and thus the more exam practice encroaches on a learning programme, the more language development is likely to be reduced. The aim of the practice is to familiarise students with the exam format and its aims are entirely related to passing the exam, since the skills required to pass a test are not necessarily or comprehensively the skills required in a target language use domain (Bachman & Palmer 1996). In this respect exam practice may be viewed as having negative washback, yet it can fulfil the function of revision so cannot be entirely negative. In addition, depending on the timeframe, teachers may take the opportunity to fill in gaps in students' language skill and knowledge, but this will be in line to the expected content of the exam rather than according to an informed well thought out programme of language development, as already mentioned.

The classroom behaviour regarding exam practice has been viewed as washback for the purposes of this study because although assessments had been undertaken previously, there was a reported change in practice in relation to the new regime. This is labelled only tentatively however. As mentioned before, without concrete baseline data and many more examples of specific classroom behaviour it can only tentatively be classified as washback.

8.6.6 Content or methodology, or both?

8.6.6.1 Narrowing of the learning experience

One situation described by T4 raised an interesting issue highlighting the complexity of cause and effect relationships in the learning environment. T4 reported changes which affected both what he taught and how he taught it. He no longer exploited out of class activities such as taking students to visit local amenities, as he had done previously:

'I'd love to take mine [my students] out into [town name] ... - well some of that can go by the wayside - you think oh well I can't afford time on that ...' (T4-9:419).

He clearly felt that this was a loss to these students and it therefore had a negative outcome for their learning opportunities. This situation has parallels with the issues of narrowing of the curriculum much cited as a common result of high takes testing, as discussed in Section 4.10 (e.g. James, 2000).

Whether the restrictions are merely perceived or real, teachers may feel under pressure to focus on work which has a direct relationship to curriculum content on which students will be examined and to exclude any 'optional extras'. The problem T4 felt he was faced with, as he understood it at least, was that of trying to adequately cover the new core ESOL Curriculum material. The teachers interviewed indicated there was confusion in general over how closely teachers needed to cross-reference to the curriculum i.e. to show how the curriculum was being covered, everything they did in class (T1-2:747; T3-6:556 and T3-7:1104) but T4's interpretation was that there was no room (i.e. no spare time) for 'extras' such as educational trips into town to visit local amenities for example, which would have a strong pedagogical justification for such students, enabling them to learn about facilities available to them in their new homes, in the case of those who had come to settle in the UK. T2 backed up DoS2's view that time could feel restricted in class especially in the evening classes, which were generally shorter than day-time classes (T2-3:2039).

The DoS at this institution however counteracted T4's claims of less time being available for such activities. She said

'people said so initially certainly that some of the things they'd done in the past in terms of in particular going to the market - doing that sort of thing - it was too rigid and it wasn't allowed - in actual fact there's plenty of time -

it's a little bit more difficult if you're on an evening class where there's less [sic] hours but there's still plenty of time to get through the syllabus and do other things as well' (DoS2-4:1360).

Rather than the actual requirements, what is important is T4's understanding of what he was meant to do and it is that which affected his teaching. At this site (Site 1) from field notes I would report a very supportive teaching environment (for example observed interaction between teachers, their reference to mutual support and friendliness in discussions with me). T4 may have felt under more pressure than more experienced teachers; he was relatively new to the profession and still developing his professional skills as he made clear in his interview (T4-9:751). In addition, a factor influencing T4's behaviour may have been his own character, being, as became obvious from his interview, a conscientious teacher (T4-9:700; T49:507) willing to teach to the best of his ability and do what was asked of him by management. T4 stood in contrast to T1 for example, who gave the impression of much less cohesion with her ESOL teaching team, she had many more years of experience as a classroom teacher, and she did not seem to feel duty bound to follow to procedures (T1-2:772).

While it could be argued this reduced scope of T4's classes was a case of washback, affecting both the content and methodology of his lessons, from the curriculum rather than the washback from the exams, I propose that the pressure of exam success at the end of the course, and exam content being based closely on the curriculum meant he felt his classes could not afford time away from the process of covering the curriculum since omissions may lead to exam failure, exams potentially sampling any aspects of the curriculum at the appropriate level for this group. However, as there was an alternative view from his DoS, in this case there was no clear evidential link satisfactorily explaining why T4 felt under this pressure, which changed the nature of his classes. His relative lack of experience, his attitudes towards

the assessments, or communication of the expectations of ESOL Skills for Life classes from management could all have accounted for T4's actions.

8.6.6.2 Atomistic approach

Another effect of the nature of the exam content influenced not only content but a methodological approach to the class content for certain teachers, in terms of an atomistic approach rather than a holistic approach to language. The ESOL Skills For Life Curriculum breaks language down into specific language elements which should be covered at each of the five levels (from Entry 1 to Level 2), with a framework of language development at word level, sentence level and text level, as already discussed in Chapter 3 (Background to Skills for Life – section 3.3.1). Teachers reported effects of their teaching from this type of linguistic description. T2 actually appreciated this change in working since what she saw as a return to a focus on grammar suited her approach to language teaching. She reported that having been involved in ESL teaching for over 25 years she had seen the trends change:

'there was a swing from what you'd call topic and functional based language and it swung to actual looking at language as it stands rather than structure and grammar so we've moved really from function I think to the grammar' (T2-3:460).

She clearly did not approve of the topic and functional approach, feeling it did not provide students with a systematic enough approach to language learning (T2-3:474).

While T2 saw the way language was broken down into components in the new ESOL Curriculum as positive, T3 on the other hand, expressed the disadvantage of what she seemed to see as a 'checklist' approach to language teaching. She explained:

'The E1's [= Entry One i.e. the lowest level group] were having a discussion giving opinions - agreeing and disagreeing fantastically – better than an E3 [a higher level] could do but if I had to prove that I would have to write down

in a little box: I can use 'I agree or I disagree' but they weren't doing that – they were doing it all wrong – the grammar wasn't there but they were communicating – but according to the ticky boxes they can't do this' (T3-7:992).

While going on to acknowledge the students needed to work on improving their accuracy, she valued their ability to communicate which is after all the aim of learning a language for most people in most circumstances. Her frustration at having to evaluate students according to their mastery of a list of language elements was palpable in her use of the phrase: *'according to the ticky boxes they can't do this'*. Her frustration lay in a potential mismatch between what the assessments aimed to measure and what really needed measuring in her view, namely their communicative ability in a more holistic manner.

Equally, vice versa, simply because a specific aspect of language had been covered in class because it was in the Curriculum (and thus may appear in the exam) did not mean students would understand the need to demonstrate what they had learnt in class if the task did not explicitly require it. She recalled:

'what was interesting for me is that we've just been doing a lot about like plus ING – we've done it do death - we've done it in pairs - they've asked other people - done it several days - done it in writing - done listening – reading - done it to death - we did the test on Monday - none of them used it so what does that tell me? I don't know' (T3-7:1009).

Allwright (1984) reminds us that learners do not necessarily learn what teachers teach but that does not mean they are not learning in class, and as professionals we need to understand the nature of the language learning process, especially in relation to how we assess what we feel they should have learnt. As Ivanič and Tsung (2005: 8) report, 'learning is rarely a simple sum of what has been taught in class'.

8.6.6.3 Change in attitude towards ESOL classes

A third change, not directly related to either content or methodology of the ESOL classes related to affective aspects of ESOL classes. One of the washback effects T4 noted was what he saw as a change in attitude towards the teacher on the part of the students, as a result of the presence of an external exam at the end of the course. This relates to Alderson & Wall's (1993) hypothesis #11: 'a test will influence attitudes to the content, method etc. of teaching' but this it must be noted relates to the 'etc.'

An example of this change is that once the students realised that T4 was not responsible for their final exam grade, but that an external examiner was, they changed their behaviour toward him. As he already explained in the previous chapter, they had been bringing him small gifts in the form of home prepared food which they knew he liked in order to build a good relationship with him (T4-9:671), since as he later went on to say:

'I think they [his students] think they can - friendship can buy marks - I don't mean in any way that they're bad people by it but they do think that friendship can buy marks' (T4-9:731).

T2 suggested a similar clarity of definition of the teacher's role regarding formal assessments improved the classroom working atmosphere (T2-3:1096).

The external assessment helped to clarify this situation and made T4's relationship with his students easier to manage. He felt they took the class, and him, more seriously (T4-9:731). He described how he tried to convey the ethos in class that they were there, as a group, to work hard towards a goal and he was willing to work as hard as they were to achieve that goal. He went on to say:

'all my classes are going to be working very very hard because we do know that it's an external exam and that I can't give them any help so I think that's

the difference – that it's probably making us work harder because there's no lee way' (T4-9:718).

In this way T4 felt the presence of external exams had a positive effect on the class in general.

The data suggested the nature of courses had changed in another, related, way. The social aspect of an ESOL class, helping students to settle into UK society by making new friends in their class, was no longer such an important feature of ESOL classes (Rosenberg 2009). The tenor of classes at the three sites seemed to have been formalised. For example, the attitude reported by two teachers was that if the students were not prepared to work, attend regularly and to take exams they should not come to class.

I think it [class atmosphere] used to be all too much all like the Out Centres [classes held in the community away from the main sites] - all a bit of social thing whereas now it's a serious thing - no you can't come one day a week - if you can't manage four days a week then sorry no (T3-6:839)

and

'it's not a social club' (T6-13:1701).

However, there is insufficient evidence in the data that this change in approach was brought about only due to the introduction of formal assessment in the form of external exams. Various other factors such as a change in student group profile, may have had such an influence. Once more the evidential link that this effect was a direct result of the new exams is lacking.

8.6.7 Summary of washback found from external assessments

Content was seen to be affected in terms of ‘double accounting’, in increased focus on accuracy, increased intensity of exam related work as exams drew near, and choice of materials. Methodological washback was observed in interactional pattern change and exam practice activities. Narrowing of the curriculum was cited by certain teachers, an atomistic approach to the language to be covered which affected who it was taught, and change in attitude towards ESOL classes could be classed as neither specifically relating to content or methodology.

8.7 Impact

As discussed earlier, exam effects which are perceived to operate beyond the realm of the classroom itself are better referred to as impact rather than washback, to distinguish the sphere of influence of the exams. Two examples were noted from the data and these are outlined below.

8.7.1 Raised profile of ESOL

Another reported effect of the new regime of external examinations was that the external exams were raising the profile of the ESOL departments with the colleges (T3-7:136). T3 felt the college did not really understand the work of the ESOL department and value its contribution to the college, and this was a sentiment expressed at all three sites (T1-2:1670; T2-3:1761; T3-7:114; T3-6:868; T4-9:439; T4-9:618; DoS3-10:111, T5-12:673; T6- 13:1254; T6-13:2858) and explicitly expressed by T6:

‘I think you might be pushed to find anyone in college [beyond the ESOL department] who understood anything about ESOL’ (T6–13:2863).

Amongst the reasons given for this was the ESOL students were not in most cases taking already well recognized exams, such GCSEs or NVQs i.e. exams from the suite of those taken by UK students. Not taking the known exams seems to have been equated with being unimportant.

This was a strong recurring theme and a phenomenon which sheds light on the general perception of ESOL, and offers another reminder of the power of exams, not only for the candidates, but to the wider college community, also pointed out by Bailey (1996). Nevertheless once again there is insufficient evidential link that this perceived change in attitude was purely due to the external exams taken by the ESOL students. It may equally have been due to the increased amounts of funding being brought into the college which afforded greater visibility, due to the then generous budget of the Skills for Life strategy, which, while related, is not exactly washback.

Once more this is an effect which relates not to content or methodology but works at a higher affective level, of raising the status of the ESOL department and potentially boosting morale. It also works at a level outside the classroom and according to various definitions in the washback literature would be more correctly described as impact, working at the macro level where washback works at the (relatively) micro levels (Bachman & Palmer 1996, McNamara 1996 and 2000, Green 2007).

8.7.2 Increased opportunity

Another perceived effect did not directly alter classroom practices but, rather, affected the students' opportunities, and it concerned access to qualifications. It was mentioned that it was not uncommon for ESOL teachers particularly those coming from an ESL background compared to those from an EFL background, to have negative attitudes towards even formal assessment, let alone exams (T3-6:651). This

was not the case for most of the teachers in my study however (elicited by interview questions 10, 56, 57, 67, 68) (T1 -2:1542; T2-3:1661; T3-7:979; T4-9:501; T5-12:1471; T6-13:2562). A level of understanding was expressed for those wary of formal assessment nevertheless, due to the fact, as already discussed above, that such assessment puts pressure on teachers in potentially being judged by their students' results (T3-6:675).

The reported wariness towards formal assessment (including external exams or, for lower levels, formalised portfolios), could no longer prevent students accessing qualifications, which some informants reported had been the case previously (DoS2-4:581; T3-6:651 and T3-6:691). DoS2 and T3 felt teachers adverse to formal assessment had deprived their students of opportunities, and imposing their own outlook on examination was not fair. The required Skills for Life exam results now precluded this.

Increased opportunity in general is unarguably a positive effect. Blanket imposition of formal assessment, on the other hand, even on those students not interested or in need of qualifications, may not be as positive and does not conform to the individualization of learning agenda espoused by the ILPs.

8.8 Issues

8.8.1 Differential effects

It appeared that the framework regarding both the scope of learning and the assessment of outcomes which Skills For Life had imposed had not had a uniform effect across all the participants, as suggested by the last of Alderson & Walls' (1993) 15 hypotheses: 'tests will have washback effects for some learners and some teachers but not for others' (p121) which was adapted in Alderson and Hamp-Lyons to: 'tests

will have different amounts and types of washback on some teachers and learners than on other teachers and learners' (1996: 296). In this section I will suggest some possible reasons for this.

Firstly, I will address the teachers. There was not a homogenous reaction to the requirements of the Skills for Life regime from all teachers. For example, the stress caused by ILPs was much discussed, but as already mentioned different teachers found different ways of coping. How and when exam practice was introduced into classes differed. Some worked closely with the curriculum and Skills for Life exam preparation, and other did not. T1 suggested a reason for the variance in teachers' reactions to the stipulations of the new regime:

T: 'I think it's [Skills For Life] affected some more than others – or who've chosen to be affected by it.'

I: so - how they interpret what's needed

T: well yes - I think so' (T1-2:1531).

She had chosen not to comply with the suggestions on how classes were to be run, or on the paperwork required. This could have been for various reasons: she was established at the college whereas many staff in ESOL departments are temporary or on fixed term contracts (T6-13:1134 and T613:965; DoS3-10:596; T1-2:176; T2-3:2233; DoS2-4:173) in which case they may feel the need to 'tow the line' for job security reasons (T4-8:622). T1 had considerable years of experience giving her confidence in her teaching; she had seen different educational and organisational trends come and go in the past. Maybe personal traits also came into play, or maybe her attitude was affected by leadership from someone outside her field (her current DoS was not an ESOL teacher) (DoS1-1:34) and there may be yet more reasons.

The cause is not important but the fact she felt she had a choice when others did not is the key and this leads me to examine the role of communication, and the various

teacher related factors which may have been instrumental, to try to understand and analyse the range of responses to the Skills For Life regime. I will explore this in the following chapter.

It was not only the teachers who reacted differently, but students too. It would be easy to assume all students would take the exams seriously, and wish to work harder for them in the period prior to the exam. However the data did not hold this to be the case uniformly. T6 in discussing her group, who were at that point just about to sit their Skills for Life exams, offered a very different picture in that she reported attendance had been erratic, and certainly no better than at other times in the year (T6-13:2388), which suggests either the students did not feel the need to prepare, were not interested in their success in the exam, or other factors in their lives were preventing them from attending, all proving more important than the exams.

The teachers may stress exam preparation but if the students do not feel the pressure, or conversely feel too much pressure which they resent, or do not have the motivation to succeed, or see little value in the qualification they are sitting, then the very presence of exams will likely have little or no effect on them. Because a group of students are sitting the same exam, does not necessarily mean they will all have the same stakes in their success. This theme of why washback was not the same amongst participants will also be explored in the next chapter.

8.8.2 Washback from the Skills For Life exams or from the curriculum?

From the examples of possible washback reported above, apart from exam practice itself, in general it is hard to argue that the behaviours noted result from the exams being in place as distinct from the ESOL Curriculum being followed.

The effect of encouraging an atomistic approach to language, seeing it as a series of building blocks which simply need assembling, and not taking into account equally important aspects of language such as pragmatics, raised the issue of how far some of the effects mentioned are the result of the exam or are indeed the result of the Curriculum, and how far the two can be teased apart.

As mentioned already, often in washback studies, washback is noted as being caused by lack of alignment within the educational situation in that curriculum content may not be covered if the curriculum does not match exams, and material which matches, and prepares for the exam takes over as course content. In the case of Skills for Life the teachers reported a high level of alignment:

'I think they've [the exams] been quite representative of what they've had to do and they've [the exams] been reasonably easy to teach towards because they do represent actually what they [students] have to do from the curriculum' (T6 -13.2231).

T2 corroborated this view, saying the exam content is what they would cover anyway from the curriculum, as covered by the Skills for Life materials (T2-3:1486).

However this was not a universal view. I wondered how far teachers may be for whatever reasons saying what they thought should be the case rather than what was actually the case. T1 had admitted to not using the materials and not taking Skills for

Life as the 'skeleton' for the courses (T1-2:709), for instance. Additionally, T6 appeared to be struggling with matching what she would usually teach with the new prescribed content.

"you're working to serve the curriculum rather than the curriculum working to support you and it really is a matter of a slave situation and the curriculum is the master (T6-13:2715).

Her use of this strong analogy suggested Skills for Life was not being easily adapted to the classroom reality. The reason it took a main role in her planning was fundamental though:

'the curriculum gives you the money' (T6-13:2713).

She clearly saw the direct relationship between what had been imposed and what was expected, the funding being the life-blood of the existence of ESOL classes. The role of funding will be further explored in the next chapter as regards stakes.

8.9 Reflection on the washback identified

No evidence was found that the exams were put in place specifically to act as drivers to pull proposed revised course content into line, as advocated by the exam driven reform described as Measurement Driven Instruction (Popham 1987) (see Background to Washback chapter). In the current age of accountability the exams simply act, in principle, as a means to measure student's achievements and to ensure value for money. Whether they are the right measures, measuring the right achievements is debatable, but the introduction of the Strategy which offered aligned curriculum, materials and exams is laudable, in principle. However, given the reality of such a variety of ESOL students, a single curriculum, although at five levels, is unlikely to consistently match student need. The exams have come to take on prominence because they are the visible manifestation of compliance with the Strategy, not because they dictate the nature of the Strategy and associated learning.

Reasons for the weak washback can be summarised. The time at which the study was undertaken in relation to the introduction of the new exams may have been of importance, for instance. If the study had been undertaken once the exams had run through several seasons of sittings then the results would very probably have been different. Another reason may have been the close alignment between the curriculum and the exams. Classroom teaching and learning behaviour was not recognized by the participants as being heavily influenced by the exam as it was influenced by the curriculum and thus already accepted as the norm and not noteworthy, having been in place since 2001.

This study has highlighted in particular how washback can take the form of a change in student-teacher relationships. Another area which this study has shed light on is that washback can be differential; we cannot assume that a single exam will have the same consequences for all stake holders (some may have no or very low stakes) and thus the washback may well vary. It cannot be assumed groups are heterogeneous, but this, as a consequence, poses a problem for how to effectively study washback.

8.10 Summary

In conclusion, in answer to the second set of research questions:

RQ2.a) To what extent is there evidence of washback from the assessment practices?

RQ2.b) Is any washback related only to the assessment practices resulting from Skills for Life?

there is evidence of some washback, but it is not strong and not consistent. The washback is described as not strong because there are poor evidential links between

cause and the effects (Chapman & Snyder 2000). The evidential link is vital to report with any degree of certainty why certain behaviours have resulted from exams.

As regards whether washback was identified only from Skills for Life exams, there was some evidence from both internal and external assessments, from Skills for Life exams and from non-Skills for Life assessments too. The strongest examples of possible washback are instigated by Skills for Life assessments, although one was an external measure and the other internal, namely ILPs and exam practice for the new Skills for Life exams.

As regards whether the washback suggested Skills for Life had had a positive influence, a positive effect was noted in terms of students and teachers working harder and classes being taken seriously. However some negative effects were that teachers were feeling constrained and forced to introduce aspects of assessments they did not find beneficial (namely ILPs). Deciding whether washback is positive or negative is not necessarily clear cut. Washback, as said many times before (see Tsagari 2009; Alderson 2004), is complex.

9 FINDINGS AND DISCUSSION: FACTORS AFFECTING WASHBACK: THE ROLE OF STAKES, COMMUNICATION, RECIPIENTS AND OTHER FACTORS

9.1 Chapter overview

The previous chapter dealt with the washback which was detected. This search highlighted the complexity of ensuring an evidential link, and being able to confidently categorise behaviours as washback, that is to say those caused by specific assessments, rather than merely behaviours associated with assessment. It was also noted that the occurrence of washback was very irregular, there was what is termed differential washback, i.e. not all stakeholders were subject to the same effects. In this chapter I am therefore concerned with the areas which were suggested by the result of the exploratory study as being of interest in providing possible causal factors as to why washback may or may not happen, and why it may be differential. This chapter thus aims to address the research question:

RQ3) What are the factors which may drive washback?

In analysing the data, several themes proved useful in particular in elucidating the situation. Firstly, Alderson & Wall's (1993) hypotheses #12 and #13 propose:

‘tests which have important consequences will have washback’

and conversely

‘tests that do not have important consequences will have no washback’

(1993:120). The previous chapter showed that although students in any one group were sitting the same exam they clearly did not perceive the stakes in equal terms, according to their teachers. The teachers and DoSs equally did not seem to be affected by the exams in a uniform way, and they mentioned different influencing factors. This differentiation of the stakes will therefore be explored in this chapter.

In addition, the Henrichsen model on which the initial research outline was framed proposes key components for successful adoption of an innovation. The Henrichsen model is divided into three main sections: the Antecedents, the Process and the Consequences. The Process section relates to the period when an innovation is being implemented i.e. the period described in my study. Central to the Process section are the Receivers, as they are labeled, and this refers to those whose practices should change as a result of the innovation. In the case of my study this would be above all the teachers and the DoSs. The model suggests they will be influenced by features of the innovation itself on the one hand, and various factors which may hinder or facilitate change on the other. This suggests that by examining the nature of the Receivers, the nature of the washback may be elucidated.

Moreover, within the Process section of the Henrichsen model other aspects which predict successful innovation are set out. As already mentioned, these include features of the innovation itself. The findings from the Exploratory Study suggested that one of the features which came to the fore as a recurring theme was communication and thus this became a theme in the interview schedule. The data arising from this therefore was examined further as potentially fruitful to explain the different approaches to the exams. Finally there is the field of factors which may hinder change, another field within the Henrichsen model, which I investigated. This was emergent, not having been specifically elicited by interview questions but themes became apparent in the analysis and were accounted for by this area of the model.

So, in short, the themes I focussed on in order to investigate further the presence or absence of washback were the stakes involved for the various stakeholders, specific features of the teachers and DoSs (the Receivers), the communication concerning

delivery of Skills for Life in a wider sense and the exams in particular and emerging hindering factors. To do so I explored the interview data.

9.2 Stakes

First of all, what is meant by stakes needs clarifying. Gipps draws on Madaus, who 'defines a high-stakes test as one whose results are seen, rightly or wrongly, by students, teachers, administrators, parents or the general public as being used to make important decisions. A low-stakes test, by contrast, is one that is perceived as not having important rewards, or sanctions tied directly to test performance' (1994: 34).

It is the notion of incentives (rewards) and sanctions which will be returned to when discussing the stakeholders of the Strategy. The phrase 'rightly or wrongly' highlights that perception of a situation is as powerful as the reality in causing classroom behaviours relating to exams.

In the literature, stakes are often defined in relation to an exam e.g. IELTS or TOEFL. What must be noted is that in relation to stakes in particular there is a difference between what are often termed proficiency exams and achievement exams.

Proficiency exams are generally described as those judging the language ability of a candidate at a specific moment in time. An achievement exam on the other hand is one associated with a certain course of study, and the exam evaluates (in principle at least) what has been learnt on that course. In the case of proficiency exams candidates have chosen to sit that exam for a particular purpose, which is most often career or study prospects, joining a specific profession or for immigration purposes. A 'fuzziness' in these definitions arises from the provision of short courses to prepare candidates for proficiency exams, as are widely available for IELTS and TOEFL. The

sole aim of such courses is to prepare candidates for the exam and often focus on test-wiseness. Their aim is not language development as such, but to ensure the language covered by the exam has been mastered.

In relation to stakes, the question to be asked is: is the student's main goal a programme of education (which may have an exam at the end, or may not) or is their goal a specific exam which they need for specific purposes (and for which they may take a series of associated preparation classes, or may not)? In most school situations (i.e. in the UK at least up to 16 years old) the GCSE exams taken are thus achievement exams, the culmination of a course of study, or more usually courses of study in a variety of subjects.

Conversely, students who sit IELTS and TOEFL for gaining entry to an English speaking university, for example, simply need the exam result and are taking the exam because they have (hopefully) already gained a certain level of ability over the course of time elsewhere and need proof of this. The stakes are likely to be similar for all candidates in the case of proficiency exams since they are all taking the exam for a specific purpose. When it is asked 'what do you have to lose or what will the consequences be if you fail this exam?' it can be imagined that similar responses from each cohort of students would occur.

The majority of the washback studies to date relate to either of the two categories of exam outlined above. They discuss either the situation arising on courses preparing students for high stakes proficiency exams, such as IELTS (Green 2007; Hayes & Read 2004) or TOEFL (Alderson 1996; Wall & Horak 2006; Wall & Horak 2007; Wall & Horak 2008; Wall & Horak 2011) or they discuss the situation where a new achievement exam is introduced, typically, into a school system (e.g. Smith, 1991).

The latter is sometimes introduced with the specific aim of changing classroom

teaching practice to bring it into alignment with a new curriculum (Cheng 1998). The students in that case have not elected to sit the exam in question; it is simply the culmination of that particular education process. This is not to say they do not value these exams or recognize their value. When we ask however ‘what do you have to lose or what will the consequences be if you fail this exam?’ the responses may be quite varied as students will have engaged in different ways with the teaching programme, and have varying motivations and plans for their lives post compulsory education.

The Skills for Life exams are complex in regard to how to define their nature and possibly defy the definitions proposed. For some the Skills for Life can be viewed as proficiency exam, for example providing proof of ability for purposes of applying for citizenship. The certification of ability is the primary goal for these candidates and the focus is thus the exam result. For others the course of study is the main concern and the sole reason for attending a course. The exam result is merely one part of this learning process and in no way the goal. The focus is thus very different in these two cases, and of course these lie at the extremes. For other students their situation may begin resembling one of these examples, but later change into the other. This is one reason potentially why the washback picture was not clear since the stakes are varied due to the range of approaches taken to the exams by students, and may vary over the course of study.

In this section I will consider the stakes of the externally imposed regulated assessment, namely the exams and the ILPs, since these instigated more discussion amongst the teachers and DoSs than the other less regulated assessments as outlined in Chapter 7 on the assessment practices, and the instances of washback noted in the previous chapter were predominantly related to these assessment measures. These will be viewed through the lens of the students, the ESOL teachers, and the DoSs,

ESOL departments and the colleges in order to investigate whether the nature of the stakes was related to why little concrete washback was discovered.

9.3 Chain of pressure: the funding imperative

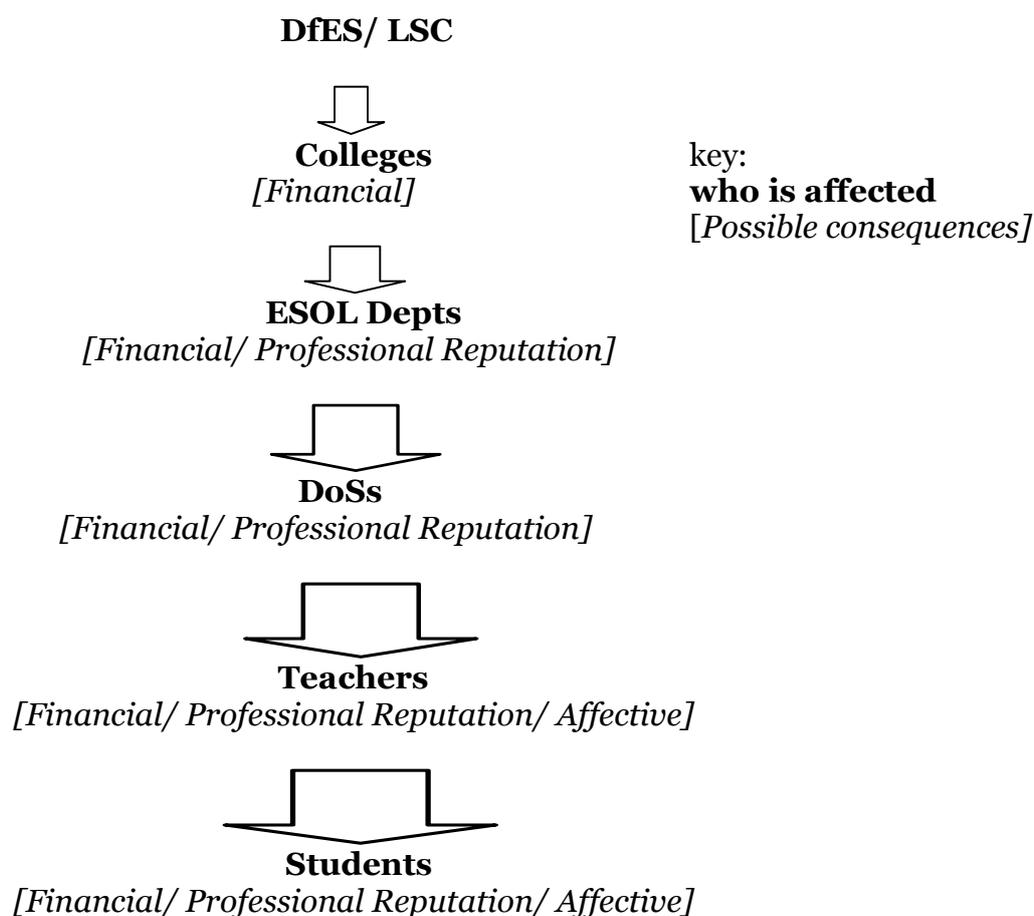
Having viewed stakes in terms of the incentives and sanctions associated with students' exam results, the picture which emerged from the data was that of a chain of pressure affecting a range of stakeholders. These pressures, in one way or another, all related to funding issues. What was of interest for this study was not the actual procedures in place at each site regarding the direct relationship between securing funding and the number of student passes gained on the Skills for Life exams but the study participants' perceptions of what may happen, which caused them stress, and to act in certain ways. As Madaus posits:

‘The power of tests and exams to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon. If students, teachers or administrators believe that the results of an examination are important, it matters very little whether this is really true. The effect is produced by what individuals perceive to be the case’ (1988: 8).

Because the staff had different perceptions of the consequences of the exam results, or ‘important decisions’ made, to use Madaus’ term, the exams thus took on different stakes for them personally. This will be discussed further below.

This chain of pressures originated from the funding providers, i.e. DfES but it was administered via the LSC (Learning and Skills Council), and this is shown in Fig. 6 below.

Figure 6 Chain of pressure



A few examples of the ways in which the increase in exam stakes were manifested at each level will be outlined, and how these were differentiated at teacher and DoS level. For these levels in the chain first-hand reporting of personal stakes took place, unlike the higher levels of the chain, which were reported by the teachers and DoSs.

9.4 Stakes for the institution

The local LSC sets the student achievement targets by which the success of the college is judged and funding is allocated accordingly. The targets are dictated by Public Service Agreement targets (KPMG 2005). After years of neglect of ESOL, the then current generous funding allowance of a 1.4 rating for ESOL, literacy and numeracy

students was welcome. This funding was received on the basis of departments proving success through hitting targets for exam results as well as retaining students recruited onto courses and also proving acceptable levels of attendance at class. This was not a matter purely in the remit of the administrative departments; the DoSs and teachers in this study were keenly aware of this link with students and funding (T3-7:334, T5-12:1000). Below are illustrations of the way they displayed this awareness:

'everything is funding lead er can't get away from that - this isn't a charity it's [a] business which provides a service - now that service has to be funded and funding comes from whichever body provides you and so every student has to have a QualAim and they only get funded if they receive that - they only get full funding if they get that QualAim' (T4-9:363)

'it's not just the achievement but the auditors come in – the attendance and retention yeah that's through LSC because if the retention or attendance is bad you lose money so if they come and they look and they say so and so's been away you haven't taught them and they won't just take money away for that person they'll do a percentage up and they'll take it for all the students' (T6-13:2084)

'I don't know quite how it works but basically for everyone who enrolls - everyone who finishes the course and then if they pass their exams you get chunks of money for each' (T3-7:158).

This was reinforced by DoS2:

DoS: the pressure on me comes from our Performance Management Unit who see to it that if we don't get our students through achievement then a certain amount of money will be clawed back

I: yeah – so there’s some department in the college that is really seeing it as a formula – you know students in students out equals X amount of money?

DoS: exactly right

I: so they’ll hassle you to make the formulas balance right?

DoS: exactly and the LSC put pressure on them’ (DoS2-4:712).

Another of the DoSs concurred:

‘your funding is constantly being cut back so obviously the more results you get the more funding you get’ (DoS1-1:930).

The LSC therefore held a powerful role in the network of ‘actors’ within adult education provision.

The data revealed that systems put in place by colleges to deliver the required information to LSCs regarding such achievement were not always compatible with the specific vagaries of ESOL students who tend to be more mobile and varied than other adult students. Many students have to face the pressures of adapting to a new life in the UK which can affect attendance in class and result in students who are registered for exams not being able to sit them and thus they do not achieve their QualAim. Asylum seekers (i.e. refugees recently arrived in the country who are applying for leave to remain) can be moved to another part of the country at very short notice and so do not complete their course and achieve their QualAim. Equally, without evidence of a good level of English many ESOL students take on unskilled work and the concomitant lack of security and inconvenience (such as shift work) which are often associated with such posts. This is then viewed as a failure in terms of student achievement figures. The college thus pays the price financially for what is neither their fault nor the students’ fault. As one of the teachers said:

‘the main problem is if people leave before they do an exam – then the college doesn’t get funding - complete funding for them because they are ‘non-achievers’ (T1-2:453)

which was supported by T6:

‘if they don’t go in for exams - don’t pass exams - you lose money because you get money for entering and passing exams’ (T6-13:2086).

The predominance of the funding imperative was pervasive throughout the data.

Nevertheless, it must be noted the dominance of targets and achievement measures is found throughout the English education system and is not unique to ESOL. It is in the interests of institutions to try to retain students within their institution long enough to gain formal measurable outcomes for their studies, even if it is in the students’ interests to move on to alternative studies or leave the formal education setting.

9.5 Stakes for ESOL departments

First of all, the Directors of Study, in their role of leading the ESOL departments, were often under pressure from higher management at their institutions to ensure the department hit achievement targets, as set out by the local LSC. The pressure to maintain provision through continued funding cuts fell above all to the DoSs.

However it affected the whole department to some extent in that it was the teachers who had to prepare students for exams to try to maximise success, as defined by the local LSC’s targets. Discussion of the ESOL department here is in terms of the group identity which the teachers and their DoSs take on as a unit.

One of the ways the students’ results affected the ESOL departments was how it affected their relative stance within the college. For example T2 reported on the monitoring of the relative success of each department:

- I: 'how important are the students' exam results for the ESOL department? what affects does it have ? ...*
- T: well for the funding yeah*
- I: how would it affect the department if you weren't -?*
- T: I've forgotten what they - there's something but I can't remember the terminology – it's just gone out from me but apparently each group is monitored for the achievement*
- I: what? you're on a kind of league table?*
- T: errr it's a hidden agenda - put it that way erm ... when all the result have been analysed ... – it is fed back into the system*
- I: there is an element of saying this group has done well this group hasn't?*
- T: so it is monitored put it that way – they know -*
- I: this is a successful department and that is not so successful?*
- T: yes - so it is monitored' (T2-3:1291).*

The term 'hidden agenda' suggests this was not a transparent practice, and may not even have been verifiable. Again, what procedures were actually in place does not matter. It is a matter of the perceived conditions being as important as the actual ones, a point referred to earlier. If the departments felt under pressure to succeed because they felt they were judged, it does not matter whether in fact they were or not as this belief will have affected their actions. The fact that there may not have been such formal monitoring in place was verified by the fact that T1 at Site 1 reported she was unsure what happened if an ESOL student did in fact fail their exam; she showed no awareness of what consequences may ensue (T1-2:1152).

On a similar theme, at Site 2, it was reported that the 'ranking' of the department was known within college because, as T4 explained, when his department underwent an

OFSTED inspection, all the results were available college-wide on their Intranet (T4-9:602). The ESOL department's evaluation was clear for all to see and weigh up against other departments' grades. Specific consequences above and beyond reputational effect or esteem within the college were not reported. T4's DoS reported that

DoS: our self-assessment report at the end of the year which has to go into the LSC - that's [students' exam results] one of the things that's on it

I: right – so numbers of who you put in - who succeeded?

DoS: that's right

I: which affects funding?

DoS: well it affects our [OFSTED] grade

I: ok

DoS: each area of the college gets a grade according to all the criteria and achievement is one of the criteria that affects your grade and so you're your grade will affect future funding

I: ok so it affects you

DoS: yes it does

I: and what effect does it have say if within the college you get a slightly lower grade than another department? does it mean within the college that department will get more money than your department having a lower grade?

DoS: not in – no – not as blatantly as that [laughter]' (DoS2-4:1196).

Site 3 also reported a similar feeling of being weighed up and judged:

'if the section as such begins to seriously underachieve we will be investigated - someone will come along and probe around because we will be seen as at risk and therefore when the great inspection comes along if we don't get our grade - ... we will drag down the College' (DoS3-10:1651).

While members of staff were aware of the role of targets in their relationship to funding for the college, and so also for the department, T5 for one was not exactly clear about what the consequences may be should the targets not be reached:

I: how important are the exam results to your ESOL department?

T: fairly important because they have the targets and I guess they're supposed to show they're meeting the targets but as to what happens if they don't hit the targets I don't think anything

I: does [name of DoS] get it in the neck if suddenly there are terrible results?

T: I don't think anything drastic would happen if we didn't meet the targets

I: you wouldn't be the pariahs of the college [laughter]

T: I hope not' (T5-12:1208 and see also 12:1000).

This may suggest either that the stakes were not particularly high or that management shielded the teachers 'at the chalk face' from unnecessary, and potentially detrimental worry. There is insufficient evidence to indicate either.

In summary, as a department the teams seemed to feel pressure as much from the financial imperatives of student success, as potential consequences of damage to their professional reputation but all reactions were built on foundations of unconfirmed facts.

9.6 Stakes for directors of studies

The stakes for DoSs are potentially two-fold in their mid-way position in the staffing structure; they sit in the position of management within their departments and therefore have to liaise with senior management and take on the responsibilities

which that entails, but they are also members of the teaching team and thus identify with the issues faced by the other teachers. In this section I have concentrated on reporting effects which seem to align with the former role, and deal with the latter in the following section.

Not only were the future funding levels for the ESOL departments dependent on departments reaching targets set by LSC but failure to hit targets could have affected the DoSs personally in financial terms, they seemed to suggest. In theory at least, they could suffer a pay cut if the ESOL department was not deemed to be successful, as reported by DoS3:

'well I don't get any increments unless I hit targets ...' (DoS3-10:1651).

Another way in which it was reported that DoSs were in a role of responsibility, accountable if targets were not reached, was suggested by one of the teachers, T5, who when asked about how important the exam results were to the college, said:

'I would say pretty important so if it did develop that our little department was suddenly getting really appalling exam results – which we might do then – yeah that might – it wouldn't affect me - it might affect [name of DoS] ... and having to justify what was going on' (T5-12:1223).

T5's DoS substantiated this by reporting:

'the whole of it - funding and all the things for accountability and attendance and achievement that go along with it when it goes wrong my major headache is not my learner - it is not my teachers - it is my Funding Managers -my Achievement Managers who come along and they won't report to me - they report to my Head of School and they say ESOL is – it's got 52% and it should be at 75 % and then the Head of School will come along to me and say well this is what they [the managers] say' (DoS3-11:1305).

As already reported above regarding pressure from the Performance Management Unit of her institution DoS2 also felt this pressure from more senior management levels, though it was rather less animatedly expressed (DoS2 -4:712).

The staff at the other site (Site 1) did not mention such specific pressures on the DoS, in terms of the pay related effects or the accountability regarding reaching targets. It cannot thus be assumed this was a universally felt pressure, but may have been a localised effect due to interpretation of the LSC requirements at Site 3 or the effect of that specific management style which transferred stress down to the DoS, whereas at other sites this was managed in alternative ways not causing such stress. Without specific follow-up data it is hard to verify.

To summarise, the three DoSs differed in how they expressed the stakes involved for them regarding students' results. All three worked in the same system under the same funding regime so this difference might be explained by their own role in management, their previous teaching experience, how much they had invested in the system (e.g. how close to retirement or leaving they were). They gave different impressions and so factors specific to them must have influenced their perceptions of the pressures involved. This will be pursued later, in the section of this chapter which deals with 'Receivers'.

9.7 Stakes for teachers

The teachers were quite aware of the role and importance of funding to the operation of all they did, and were also aware of the reason for management involvement in their field at that time of change for ESOL departments. As one said:

'ESOL brings in a lot of Government money for the College' (T3-7:157).

However, the financial aspect was not the only pressure that teachers felt under. In this section the main areas causing the teachers stress in relation to external student formal assessment will be outlined. These points have been explored by asking once more, what the consequences of these assessments for the teachers are.

9.7.1 Vicarious success

As reported in the previous chapter's discussion on washback, one of the main concerns for teachers as regards the consequences of their students' exam results was the teachers' desire for the students to be successful.

'how does it affect me personally? – well it's - obviously you get a sense of achievement that your students have achieved (T2-3:1270)

and T6 reinforced this:

'personally it's a bit of a matter of pride that they get it [good exam result] because you've taught them and also it's rewarding because when they get it it means a lot to them' (T6:13-2275)

and added

'it doesn't affect what they know but it is upsetting when you know what they can do and they fail it because of nerves or something like that' (T6:13-2282).

The immediate consequence is nothing more than a sense of pride and not wanting to see their students fail (T4-9:497). The stakes in this case may be termed localized and internalised in that they are noted by, and relate to no-one other than, the individual teachers and are of an affective nature. This type of stake relates to the teachers' sense of professionalism.

While this may be viewed as an empathetic but self-referential effect of student success (or otherwise), it was noted that how the teachers themselves were judged in

relation to student performance was another factor mentioned. At Site 1, for example:

'each tutor is looked at from an achievement point of view – if you've got a class of twelve students how many of those students are achieving?' (DoS1-1:936).

T3, expressed this quite openly.

I: how important are the students' exam results to you?

T: mm – they're very important ... you know - all teachers feel they are judged on their exam result so there is a pressure' (T3-7:9)

and later went on to say:

T: 'yes if students in my class failed or did badly then people would think 'oh she hasn't done very well' so it's pride' (T3-7:30).

From the context, and subsequent comments she made clear she did not mean people in general but her colleagues and management. Once more, professional respect is at stake.

9.7.2 'Payment by results'

Apart from these affective and professional effects, as with the DoSs, the teachers perceived that student exam success rates were linked to their pay, and while this did not appear to be their chief concern, it was expressed several times when they described their reactions to the new exam system. For example:

'I mean the only way that it does affect me is erm because we have performance – performance related erm I guess it's performance related pay - an annual increment which is related to the performance review' (T5-12:1025).

Her hesitancy may suggest a lack of certainty or clarity on this point but later she returned to this:

- I: how important are the students' exam result to you personally?*
- T: erm well important for the performance review for a start [laughter]
kind of fairly important' (T5-12:1199).*

At the same site, her colleague T6, referred to what may be referred to as a payment by results system:

'I do get paid according to their results so if my results are good it goes forward to – not solely on that – but it is part ... all the teachers who want to get raises every year their results are part of it' (T6:13-2284)

and she also returned to this theme later (T6-13:2319). Their DoS backed up their views:

'for [sic] our own point of view our achievement record is based on whether people pass or don't pass – our wages depend on it' (DoS3-11:544).

However, T5 admitted she doubted that there would, in reality, be any effect on pay:

'I don't think you would fail a performance review or not get your performance increment if if you - you're - the rate wasn't up to target' (T5-12:1046).

T3 also expressed similar doubts about any particularly grave consequences.

I: 'is there any way in which it would actually have an effect on you as a teacher - I mean it would actually change the responsibilities you're given or -

T: er I don't think so unless it was consistently all the students failed' (T3-7:34).

Yet again there is a lack of a clear message on this issue and teachers are basing their opinions on hearsay and rumour. These were potentially quite serious issues for the teachers and their state of awareness relates to the communication issues discussed

later in this chapter.

It is noteworthy that the relationship between pay and results was not a theme raised by the participants from Site 1. As noted previously, funding issues were generally a topic raised less at that Site. In addition this effect on pay was not discussed by any of the participants with certainty, neither teachers nor DoSs, and it may be expected the latter would be more knowledgeable about such employment practices. Although a rather nebulous potential threat (and, it could be argued, not valid enough to be taken seriously) it was part of the teachers' exam related stress factors and thus it was categorised as one of the stakes, which may affect classroom behaviour.

9.7.3 Targets

At the time of the data collection, the talk of targets (which is not confined to Further Education in the UK but a dominant discourse in all levels education and indeed the rest of the public sector) was not concrete in terms of teachers discussing the specific terms of what targets they were personally expected to reach. However they were aware that these targets existed in the colleges and that they feared they probably would be affected more directly, more personally, by them in future.

DoS: 'I think that will be something that increases as well to the point where I think teachers are going to have targets

I: right – individual - they're going to know exactly

DoS: yeah - it hasn't come in yet but I can see the signs of that happening – it happened in schools' (DoS2-4:1271).

T4, at the same site concurred:

'you do know at the end of the day that this is the way everything's going - everything's going to be funding lead because the way the college is at the

moment anyway they keep throwing figures at us you know - you know that you've got to be successful' (T4-9:389).

At Site 3 teachers also seemed to be in agreement with T4 about the way targets were imposed, and that the mechanisms by which they were arrived at, or their rationale, were not usually clear and in fact sometimes perceived as unfair:

'every year you're supposed to set yourself targets and erm obviously the college want you to meet the achievement targets for the college as [a] whole and for your school but it's completely impossible to erm use those figures in any useful way because as an individual I'm not wholly responsible for a student's or a group of student's erm pass mark ... I don't solely teach the - those groups - there are other teachers on my group ... so it's kind of a bit of a farce really (T5-12:1044).

T3 summed up the feeling of others being trapped between the two causes of stress: the pressure to help their students and the pressure to hit targets:

'so I'm feeling a bit pressured so that's really had a major impact on my teaching and ... - how I feel about that class - I feel a bit worried because - ... if they don't do well it's going to have an effect on me and affect their future' (T3-7:242).

It is interesting that this issue of targets, and predicting their increasing influence, was most prominent at Site 2. This may have been influenced by the fact that they appeared to be a close team, with good communication, and that their DoS had extensive experience of secondary education (which she alluded to above) and predicted a similar dominant discourse of the so-called 'target culture' which was prevalent there reaching ESOL departments which had previously been fairly autonomous in their operations.

Merrifield (2006) reports that Skills for Life teachers were experiencing a tension, being pulled between two types of professionalism: a 'responsive professionalism' and 'new professionalism' (p7). By this is meant the ability to respond to their students to 'fine-tune their teaching to make it relevant to people's lives' (Ivanič et al 2006:36) but also adapting to the requirements of the Skills for Life Strategy in meeting targets, delivering the curriculum, complying with paper work requirements etc. Merrifield was discussing Skills for Life tutors in general, but the findings of this study back up her findings.

The three main ways in which teachers were affected by students' exam results were firstly that their sense of professionalism was potentially in question. They also feared that a proportion of their wages may be linked to their students' results but this was an unconfirmed fear. Lastly stress was being caused from an undefined, nebulous concern which existed that targets set by their colleges may shortly play a much more direct effect on them.

9.8 Stakes for the students

As McNamara & Roever note, the 'ESL population is the one whose group membership is most in flux' (2006: 239) and this continual change, with its consequent alteration of goals, motivations and pressures will I believe affect the students' reactions to exams, and affect the students' perceptions of the stakes for different students. Andrews et al (2002: 207) concluded from their research that 'the precise nature of washback seems to vary from student to student'. In this section I will outline the variation between students detected in this study and the various stakes noted.

Firstly there is the matter of the ESOL students' motivations for learning English. As stated already, the students come to classes for various reasons. As T2, said:

'I think they [exams] are important because there's a sense of - it is a sense of achievement if you get - if you pass an exam so I do think it is important but again it depends what their initial reasons were for coming to a class and for them if it was to improve confidence or socially they're meeting people and communicating then - it's very individual isn't it - very individual' (T2-3:1284).

Students themselves reported a range of motivations, mirroring what the teachers had reported. According to the teachers, some needed qualifications for securing work, rejoining a previous profession, or applying for citizenship (T4-8:114; DoS3-10:357; T5-12:258; Obs T4-21:67). However, not all students had specific goals. Some were taking classes for general self-improvement. In order to exemplify the variety within the student groups three of the groups at one of the sites will be described.

At Site 3, for example, in the Group 1 student interview, one student reported he wanted to improve English to be able to travel, another to improve his skills and get qualifications for future work prospects but with no specific profession in mind, and the last one had a very specific goal of being a manager in the hotel industry. He needed to do well because as he said:

'I want because er my dream is have very very big family - more children - I need the money' (S3-1:25).

At the same site, Group 2 students wanted to take the exams for either improving job prospects or to embark on another course at the College, in one case to top up qualifications gained in hairdressing in her home country to be able to find work in the UK in her profession (S3-2:40). Two of the group admitted they had no career or professional goals and also wanted to attend ESOL classes just to be able to

communicate better in the community where they live (S3:2- 63). This contrasted with Group 3 above, none of whom intended to stay in the UK.

In Group 3, the members of the group also contrasted highly with Group 2, in that one needed the exam results for citizenship application³¹, while the other two were taking English just for general improvement purposes, and were interested in finding out what their level of ability was from the exam results, and no more than that (S3-3: 17; S3-3: 48; S3-3: 71).

The teachers were asked for their perceptions of the students' view of how important the exams were for them, with a view to gaining insights into the stakes involved. T5, for instance, suggested most of the students were highly motivated regarding exams:

'the majority do really really want to pass so yeah I guess they see it as pretty important' (T5-12:1190) (and similar views were repeated at T5-12:1195).

Her colleague at the same site, Site 3, however explained how the students, while all reasonably motivated, differed markedly in their motivation. The higher level students, who above all wanted Cambridge main suite qualifications, had classes at one of the college's sites, and the lower ability groups in general met at another of the sites. She suggested a difference between the groups as follows:

'it's more rewarding for these [the lower ability students] than it is for the others because for the others you feel they've probably done a lot of exams and it's just another thing – you know and it doesn't really have the same [significance] - but down here it's such a confidence booster - or it would be on the odd occasion when they've passed [laughter] – and so it's quite

³¹ At that point, the regulations required that students supply an Entry 3 level qualification or supply proof of progress across from one level to the next, as well as passing the 'Life in the UK' test, to apply for citizenship in the UK.

important – and as I said to them at the end it's not the be all and end all'
(T6:13-2277).

She suggested that students of the lower groups often gained more than just a qualification; they gained self-confidence and self-esteem often referred to as the 'soft skills'. T6 was not the only teacher to note such a role, of a more affective than instrumental nature, for the exams in the lives of her lower level ESOL students. T3 reported:

'I think for the lower levels they perceive it [taking exams] as important in that I don't know - for their self-esteem - to show concrete progress – to show they have finished that level (T3-7:53).

T6, at Site 3, concurred in believing in the role of the exams for boosting self-esteem. Conversely, the consequences of failure were loss of face, and self-respect, as T6 suggested:

I: how important are the students' exam results to the students?

T: I think quite important – not for any other reason than just confidence self-esteem - that's the word - and I think it's desperately important ... they want to tell their families – how awful is it for them when they go home and tell their kids who are university and everything and they go 'oh I failed' - it's horrible ... no matter how well we say 'you're fantastic - you did all the course - well done' ... it is very important and I think the reason why some of them didn't want to do the exam was because they didn't want to fail' (T6:13-2354).

These illustrations remind us a student group comprises individuals each with their own circumstances and motivations. The consequences cannot be placed on a cline or scale of severity, with some outcomes deemed more important than others, as each will take on their own gravity for each individual which is not valid for comparison.

Another way in which one sub-group of the ESOL student population reacted differently to others was, it was noted, that the exams were taken very seriously by the asylum seeker sub-group of ESOL students. As DoS2 suggested:

DoS: they've got their own motivations but erm exams have become a lot more important recently and I think the asylum seekers have had a positive effect on that actually erm there's a quite large number of the asylum seeker cohort who value education before they come so they've gone through some form of education and ... they expect proper lessons and proper outcomes whereas a lot of the local community Asian population come from regions where there is very little education and

I: less literacy within their communities?

DoS: yes and the important thing for them has been to understand enough to get by

I: sure – and maybe not to go beyond that

DoS: whereas the asylum seekers are aiming high – they want to get to university they want to –

I: or continue with professions they had beforehand

DoS: that's it – yeah – exactly – so I think they've made a difference' (DoS2-4:1242).

The stakes for the asylum seekers were clearly much higher than for other groups of students in that they had little choice but to plan seriously for the future. In general, they had neither the community support that some ESOL student groups had, such as the local Asian population mentioned above, nor the option to return home at some point as did what here are referred to as the short-term residents (Hodge et al 2004).

It was not only the asylum seekers of course who wanted to gain specific qualifications with a view to joining the UK labour market. Which of the students needed qualifications, would depend on what types of work the students were aiming for, DoS1 suggested:

DoS: ‘.. and it’s very much ‘have you got the qualification’?

I: so it is quite important they have something

DoS: I’d have thought it depended on what sort of job it was rather – if it was a person to person job maybe more on the sort of skills that they’ve got regarding can you talk to people rather than what you’ve got on paper’ (DoS1-1:969).

She seemed to suggest that language ability certification would be less necessary for work requiring extensive and effective oral communication skills. This is hard to understand unless she is saying she did not have faith in the ability of the Skills for Life exams to accurately assess communicative ability. Equally she may have meant the employers did not recognize this ability in the (admittedly as yet relatively unknown) new exams. In either case it is clear she felt the Skills for Life qualifications were not necessarily suitable for all job-seekers.

In some specific cases, the consequences of the exam results had a very concrete financial implication. One specific group was part of a programme facilitating the unemployed in getting back into work. The college had various goals for the group and various incentives were offered, as T4 explained:

‘I’ll give you an example - over at [name of the employability programme] if they go up one level they get - the college gets the funding for the course - if they go up two levels the student gets £100 bonus’ (T4-9:368).

This was evidently an unusual example, but the nature of the process of commercialisation of education means that learning for learning sake is not necessarily any longer the sole driver for gaining an education in the adult education

sector. Funding for courses is sourced through various means, and so the motivations and incentives for students are likely to be equally varied. In this case the target-driven culture of financial incentives for learning may have overruled intrinsic desires for self-improvement. It must be noted the students did not choose to come on the course as attendance was part of their conditions for continued access to unemployment benefit so expectation of such motivation may well have been low in any case. Field notes recorded that the class atmosphere was strained and not particularly co-operative and stood in contrast to this teachers' other ESOL class in this respect.

In short, the reasons why students chose to attend ESOL classes can be reduced to two main categories, namely for Personal Development, to gain in confidence, and in pride of their abilities, and secondly Career, to enhance their potential or reach specific work goals. However the ESOL students as a whole have a highly complex profile so their individual motivations and the consequences of their studies cannot be predicted.

9.9 Differential washback

Differential washback refers to the notion that washback of a single specific exam is not uniformly experienced. It may affect some stakeholders and not others. This chapter is aiming to see why washback may or may not have occurred and to investigate the nature of that washback and who is affected to try to explain this. At each site any one class at a certain level was aiming to sit the same exam, but different perceptions of the effects on them and the consequences of those exams were identified.

It was notable that not all the themes concerning stakes were raised by all the teachers; they had varying concerns as regards student exam results. Overall, the teachers did not appear to feel the same stakes applied to them all. This would explain why washback was not consistent as far as they were concerned. The nature of the students varied enormously at that point when funding for ESOL students was so generous; boundaries between ESL and EFL students were very fuzzy, all potentially coming under the 'umbrella' of ESOL. This could account for why the stakes differed for students, since they themselves differed so widely.

However the teachers, from what was suggested regarding LSC requirements, for example, were under similar pressures, as the stakes were similar. This does not explain why they should react differently. No-one at Site 1 for example mentioned the link between teacher pay and exam results. I propose that the role of communication of information about the new Skills for Life Strategy, combined with individual teacher related factors, account for the range in reactions. The profile of the six teachers and three DoSs (who were also teachers) varied, in terms of work background and length of experience, attitude to exams, all of which could affect their reactions to the changes they faced and how they perceived the consequences of the exam results.

The homogeneity of the stakes of a certain candidate group could provide insight into the likelihood of washback. Conversely assumptions cannot be made as to the likely washback when the stakes are varied within the candidate population.

The chain of pressure outlined above does not fully explain why there was a differential effect regarding washback. To investigate further, other aspects of the situation were probed. First, matters relating to the Receivers, i.e. the teachers and

DoSs, to see which factors may have influenced why they behaved differently, and then communication of the innovation will be discussed.

9.10 The Receivers: the teachers, their experience and attitudes.

This study relied on the evidence provided by nine participants and the rich data of their exact words, not summaries, allowed a sense of the individuals to come through. It proved hard to distil this vivid sense of the individuals into words due to the word limit of this study. This however fed into the overall picture of nine individual professionals and their approach to the Strategy and, in particular, the assessments.

In Henrichsen's work on the diffusion of innovations (1989) he refers to the actors in the chain of events who implement the innovation as 'Receivers'. Key features of the Receivers which Henrichsen suggests will affect the adoption (or otherwise) of an innovation are the Receivers' awareness of the innovation, their interest in it and their evaluation of it. Equally, various other researchers have recognized the pivotal role teachers play in implementing educational change. Burrows (1998), for example, categorised teachers into three groups characterized by their reaction to the introduction of educational innovation: adopters, adapters and resisters and this is pertinent in investigating how teachers may decide, or not, to engage with the five strands of Skills for Life. The teachers in essence did not have the choice to reject the innovation i.e. implementation of Skills for Life, but they had the choice of how enthusiastically to embrace it. Having examined the stakes, I hoped investigating how far they engaged with the exams in particular would shed light on why washback, did or did not happen.

One important reason to examine the Receivers as Henrichsen suggests, rather than making assumptions about what outcomes will occur when changes are made to any system, is because, as Becker suggests:

‘taking everything into consideration, people do whatever they have to or whatever seems good to them at the time, and that, since situations change, there’s no reason to expect that they’ll act in consistent ways’ (1998: 45).

Thus in order to understand what has happened as a result of changes, it is prudent to examine in each case what has happened as people are unpredictable; in each case a different set of circumstances (variables) will be at play upon them, influencing their behaviour.

The reason these variations between the Receivers were interesting to examine was that having investigated the stakes involved, it was noticeable that although the teachers were in the same circumstances regarding what they had to deliver, they did not all appear to react to a similar degree to the same incentives and sanctions of the Strategy. For example, some seemed less concerned than others by the theoretical sanctions of performance related pay linked to students’ exam results and the need for close monitoring of their adherence to the curriculum through the paperwork they were expected to submit (e.g. workplans including cross referencing to the curriculum). This then affected their subsequent teaching-related behaviours in different ways.

Even if the message about how the Strategy was to be rolled out had been clearly transmitted (see next section), in a consistent way, how stakeholders subsequently reacted to this would probably depend on a range of factors. The ones I have identified are the teachers’ professional experience, attitudes towards Skills for Life and assessment in general, and their evaluation of Skills for Life. It must be stressed that the focus was on the teachers’ reaction to the strategy and their attitude towards

the resulting assessments. What was of interest was whether any patterns arose concerning how they had arrived at their personal evaluations of the exams and the attitudes expressed towards assessment and how this translated to their classrooms.

9.11 Experience

In order to elucidate why the teachers and DoSs may have reacted differently to the Skills for Life exams and been influenced differently by them, in this section I examine the teacher and DoS profiles. In the following table various parameters are set out to sketch out their profiles. The terms EFL and ESL are used, as described in Chapter 2, to refer to the traditional differences in teacher training, methods and materials in teaching English to two types of students: those normally resident in an English speaking environment and those who are not.

Table 18 Teacher profiles

	T1	T2	T3	T4	T5	T6
Years of teaching experience*	20+	25+	10+	1+	10+	5+
Years at current workplace	10+	25+	1+	1+	1	5+
EFL or ESL background	EFL	ESL	ESL	EFL	EFL	ESL
Teaching background	MFL EFL	ESL	MFL EFL	ESOL only	EFL	ESL
Non state-sector experience	✓	✗	✓	✗	✓	✗
Experience as examiner/ exam writer for Exam Boards	GCSE & CAE marker	OCNW assessor	GCSE & Pitmans interlocutor	No	IGCSE marker	No
Contractual status	F-T	F-T	F-T	F-T	P-T	F-T

Table 19 DoS Profiles

	DoS1	DoS2	DoS3
Years teaching experience*	5+	25+	30+
Years in current position	1+	5+	1+
EFL or ESL background	neither	ESL	EFL + ESL
Teaching background	Adult Literacy No ESOL	Adult + Secondary level non-English subject + special needs ESOL	Adult + EAL (school level) EFL ESOL
Non state-sector experience	✘	✘	✓
Experience as examiner / exam writer for Exam Boards	✘	✘	✘
Contractual status	F-T	F-T	F-T
Current teaching load + DoS duties?	✓	✓	✓

Key:

*years rounded to multiples of 5 (i.e. 5+ = between 5 and 9 years.)

MFL = modern foreign language

9.12 Parameters

The main parameters (as given in Table 18 and Table 19) which can inform the staff profiles are as follows:

Years of experience. Despite the often repeated aphorism in the world of teacher training that it is possible that a teacher may have twenty years' experience, or they may equally only have twenty times one year of experience, the number of years of teaching can indeed give an indication at least of their potential professional sophistication. For example it has been noted that less experienced teachers can be overly reliant on a curriculum (Baynham et al 2007) and understanding such

attributes can help understand why an experienced teacher may react differently to an inexperienced one.

Length of time in current workplace. Length of time working within their particular current teaching team informs us about possibility for the growth of a group identity, and establishment of any institutionalised practices. While this is not guaranteed it is an indicator of the possibility of this.

EFL or ESL background. While much current literature suggests a narrowing of the gap between EFL and ESL teaching (see Chapter 2 for discussion of this), this gap has traditionally existed and depending on when a teacher trained, this difference may or may not appear prominent and influence their teaching of, and attitudes towards, current practices. As well as training, where they have gained their teaching experience in the methodologies, materials and learner needs pertinent to these two different student groups may also lead to variation in the skills set teachers develop.

Teaching background. As with the influence of training methodologies mentioned above, experience of other areas of teaching also may influence a teacher's perspective of ESOL. Comparisons of adult teaching with children and youths or modern foreign language teaching with ESOL may inform practice.

State sector v. private sector. Experience of differences between the two in terms of operational methods and funding may influence practices. Private sector operations will have to run on entrepreneurial principles whereas the state sector relies on government funding primarily, or at least has done so traditionally, and management, organisation and operational decision-making will differ as a result. Teachers' work practices and expectations may change accordingly.

Exam experience. Involvement in producing exams furnishes an insight into exam principles and design which may have afforded a perspective which other teachers did not have access to. Their knowledge in the field of testing would be higher than those without, which may influence understanding and evaluation of exams. Due to a general lack of training, as De Vincenzi has said, most teachers do not know a great deal about test development so conclusions drawn from what they say about standardized testing is potentially flawed (1995). Such exam writing or examining experience may counter such general lack of insight.

Contractual status. Contractual status informs us about the time they can spend within a particular teaching team, and possible opportunities for cross-fertilisation of ideas and practice from other posts elsewhere within other institutions. It may also inform the level of commitment such staff can offer to the institution.

In describing the profiles of staff, when referring to teachers only those participating in the study are referred to, not the staff at each site in general. In addition, the terms EFL and ESL are used to distinguish two distinct teaching situations as outlined in Chapter 2, where the usual everyday language environment of the student differs.

9.13 Tutor and DoS descriptions

First of all it must be reiterated that all the DoSs had teaching hours on top of their DoS duties and so understood the classroom situation in addition to the management perspective. The profiles show that at Site 1 the DoS was relatively inexperienced, both in number of years of teaching, and regarding ESOL. The teachers participating in this study however had considerable experience; one had come from a modern foreign languages route into EFL and then into ESOL, while the other had specialised

in ESOL her whole career. At Site 2, the DoS had wide teaching experience, all within the state sector. The teachers were less experienced; one having changed careers and the other, like T1, moving into ESOL via EFL and, previously, MFL teaching. DoS3 was also very experienced, and had non-state sector teaching experience in EFL as well as a solid background in ESL and EAL teaching, prior to Skills for Life, so had good points of contrast. The teachers at Site 3, also differed in the routes they had taken into ESOL, one having trained in EFL and having experience working abroad and the other training in ESL. At each site one of the two teachers was in the position of their entire teaching experience to date being in their current workplace. They thus have not had the opportunity to gain first-hand experience of alternative work practices and institutional habits and ethos. All study participants had minimal experience in exam production, though some had acted in assessor role which affords a certain level of insight. All but one had full-time contracts.

The teachers and DoSs in this study group thus represented a range of experiences and this may start to explain their different approaches to Skills for Life, how far they tried to comply with administrative requirements, and how they conducted their classes, and how far assessments were an explicit part of the classes and how preparation for exams was addressed.

9.14 Teachers' attitudes towards assessment

One key feature which may explain variation in classroom behaviour and attitudes regarding assessment preparation practice for the Skills for Life exams is the teachers' attitudes to testing and assessment in general and to Skills for Life assessments in particular. The next section explores this dimension.

When asked about their attitudes towards assessment, all the teachers reported positive examples of how assessment aids the learning process (T1-2:1542; T2-3:1661; T3-7:761; T5-12:1330). As T6 said:

‘if you say you can’t assess them that’s like saying they can’t learn and I think they can learn and they can be assessed and they get something out of assessment and teachers get something out of assessment’ (T6-13:2565).

T1 in particular related her current views of assessment back to her own experience of sitting exams as a student herself (eg.T1-2:1555), which had been a routine but positive experience. However, even T4 who had not come from an academic background talked animatedly about the benefits of assessment, primarily for judging student progress as well as for developing his skills as a teacher by examining the students’ results (T4-9:509).

9.15 Advantages of the new external exams

Compared with the ‘international exams’ (in most cases referring to the Cambridge main suite exams) which acted as the only comparable benchmark that far, since other qualifications up to that point had been portfolio-based and /or internal college certificates, the new exams were deemed to be more relevant and appropriate than these (T2-3:1155). For example,

‘the context does tend to be a lot more real for ESOL students’ (DoS2-4:1042).

Another teacher commented on the appropriacy not just of the content but also of the format of the exam:

‘the speaking and listening has worried me because the listening seems so light with the speaking but in actual fact when you actually sat down with the speaking exam it was a much better representation of their listening abilities than it is to have one of those listening tapes where they go through and tick the boxes and fill in the words – it was much more representative of

the listening to instruction – answering –replying [..] I quite liked that’ (T6-13:2236).

Another advantage according to T6 was that she felt the exams were quite fair (T6-13:2231) and T3 described them as more ‘ESOL-friendly’ (T3-6:1266) than the exams they had taken previously such as the Cambridge main suite. DoS2 also said some improved exams, tailored more towards the ESOL population, had been needed for a while (DoS2-4:1057) and she felt they were better than what had been used previously:

‘compared with EFL tests which is what have been available in the past the context does tend to be a lot more real for ESOL students’ (DoS2-4:1042).

A further advantage of the new Skills for Life exams system was that they were taken more seriously according to T1 (T1-2: 1250). As many (if not all) students valued them more, the teachers did also, affecting overall motivation.

In summary, the exams were generally seen as suitable for the students, to be fair and that they affected motivation positively.

9.16 Disadvantages of the new external exams

While the teachers had voiced various positive comments, they had plenty of criticisms also, which is natural in a period of adjustment to a new regime. While they were generally satisfied with the exams they had chosen, once they had experience of them, and also had received feedback from the students, several issues did become apparent. For instance, it was reported by T2 that the writing seemed set at too high a level on average and most students would be sitting a lower level exam than the other

modes they were taking. This had been decided after the experience of the mock exams (T2-3:1203).

Several teachers made comments regarding the fact that the exams from different Boards did not seem to be at comparable levels (DoS2-4:503; T5-12:562 and 574), (for example, as noted already above, Trinity exams were generally deemed easier than Cambridge). Different formats, item types, and approaches to testing are what distinguish the Boards but differences in the levels, since they were all meant to be aligned to the ESOL curriculum, is hard to explain.

In a similar vein, complaints about certain boards were made in terms of their technical production. Some were felt to be slacker than previously, for example exam marking schemes were reported as not being accurate, and descriptors and advice less helpful than one teacher who had more experience with the Cambridge main suite exams, was used to (DoS3 - 10:971). As teachers only had access to practice versions of the exams, there is no evidence that the 'real' exams were also poorly produced; however, the impression gained from their first interaction with the exam has affected their opinions regarding quality.

As the exams were still so new, some teachers seemed to feel rather at a loss and unguided. They had no past papers to refer to in order to anchor their sense of the different levels and the requirements of each. Spratt (2005) has discussed this issue in terms of how teachers often rely heavily on commercially produced exam-preparation materials in the wake of a new exam until they feel familiar and gain confidence in their exam preparation. She terms the resulting effect of the reliance on the coursebooks as the 'fruit of uncertainty' (2005:23). Unfortunately due to the nature of the (general lack of) ESOL purchasing power and small market size, few ESOL specific coursebooks suitable for the UK market were available then, let alone

books specifically designed for Skills for Life preparation. The teacher thus had very little support other than practice papers and some limited teacher training.

In addition, there were concerns about whether the exams had been trialled on appropriate groups. This was relevant since the teacher in question knew that the majority of the Cambridge ESOL suite of language exams were designed for the international EFL market. As T6 said:

'they say 'yes we have trialled them' but who did you trial them with [hesitation] people who have fossilized problems or did you trial them with people who might have learning difficulties – or were you trialling them with the people who were already doing KET and PET – who are Europeans'
(T6-13:1423).

The exams had been introduced over a very short timeframe compared to a normal exam development schedule (at least two years typically for a major high stakes exam) and it is plausible that the usual procedure may not have been followed due to time pressures to have a suite of exams ready for the market as soon as possible. T6's concerns are credible in terms of validation concerns in that as Davies et al state:

'If the test is to be used for a different population or purpose from that for which it was originally developed [...] it may be appropriate for a further 'local' validation study to be carried out' (1999:220).

A disadvantage which some touched on was that they felt de-professionalised, in that their judgements were no longer found to be valid enough indicators of student ability (T2-3:936; T3-6:566). However, this is an example of differences in individual approach since others had discussed their role as assessor and how they felt uncomfortable about it, while simultaneously also being a support and guide for students (e.g. T2-3:2267). It is a reminder that it is unlikely any system would be

acceptable to an entire teacher group as it comprises individuals and consequently their own personal belief systems regarding effective pedagogy will also differ.

DoS3 expressed doubts about the role the exam had taken on:

'yes I think it's helpful I think to have a syllabus - to have a curriculum - to have materials - to have training opportunities - all those sorts of things yes definitely that's definitely yes but you know is it the tail wagging the dog or the dog wagging the tail' (DoS3-11:1226).

At various times during our interview DoS3 raised this issue of the role of the assessments and how they had come to dominate classroom practice. T3 also raised concerns about how the exams were being executed though:

'so it is good for our students because they've got a lot more money coming to ESOL so that's great but at the same time they're not really tested - they [Admin dept / Management] just want to be able to tick the boxes and everything and like with education can it be ticked in a box and can you measure it ? - I don't know' (T3-7:1066).

To sum up an overall feeling about the assessments, teachers and DoSs recognized their value in general, and were generally positive about Skills for Life exams (unlike the teachers in Alderson & Hamp-Lyons' (1996) study about TOEFL washback), but were concerned about their implementation. The frustrations were summed up by T5 however who asked for the relevant bodies, the exam commissioners, and the Exam Boards, *'to take the ass out of assessment'* (T4-9:1255).

9.17 Evaluation of Skills for Life

The reason for being concerned about the teachers' evaluation of Skills for Life is based on the premise that a member of staff who is positive about the strategy will be

more likely to be positively disposed towards the changes they were expected to make. When the teachers were asked in the interviews for their evaluation of the Skills for Life Strategy as a whole, a range of views were expressed, generally reflecting the verdict of a neutral stance or a positive reaction (e.g. DoS3-11:1242) to the change:

'it was just at the right time for us – we were wanting to break away from the old style ESOL teaching and move on' (DoS2-4:785).

Yet there were caveats; for example, DoS2 seemed to sum up the views expressed in various other ways by the other participants:

'the move towards it [Skills for Life] is very positive and I think everybody's seen it that way actually but it's the erm one size fits all mentality you know the lack of flexibility within it' (DoS2-4:764).

T5 expressed the most negativity about Skills for Life. With reference to other classes which she taught outside Site 3 (as she only worked there part-time). She said:

I tried using some of the materials that'd been published but there was just no depth to them and it wasn't what the students wanted – they didn't want to learn about kind of the employment market in Britain and British culture and they wanted to learn about English and what tenses are and why we use them (T5-12:716).

These ESOL students were not of the typical ESL profile, but were mostly short-term residents in the UK wanting to improve their English before returning home. This situation has, it was noted, continually had an effect on the evaluation of the Strategy in that many aspects of its aims and methods may suit migrants who intend to stay in the UK but various groups, now under the ESOL umbrella (see Chapter 2) were expecting a course more typical of an EFL situation where topics are of a universal nature and /or concentrate on the mechanics of English rather than 'English for survival and development' purposes, as ESL classes tend to. Criticism was thus not of

the Strategy itself but of its operationalisation which had caused changes in the target student group, which it had not been originally aimed at and thus did not fully suit them.

Based on these parameters, the teachers and DoSs were identified as adopters, adapters or resisters, using Burrows' (1998) terms. Henrichsen's model also describes, in the Consequences, stage of the model, potential levels of adoption of an innovation. His analysis suggests two outcomes: adoption or rejection, and variations within these parameters. A follow up study to consider the impact of Skills for Life on ESOL once the innovation of the centralized assessments had been in operation for a considerable amount of time could usefully utilize these descriptors, but since the research for this study was considering the situation shortly after the introduction of these assessments the Burrows typology was found more appropriate. This categorisation is set out in Figure 7.

Figure 7 Adopters, adapters or resisters

	Adopter	Adapter	Resister
DoS1	✓		
T1			✓
T2	✓		
DoS2		✓	
T3		✓	
T4	✓		
DoS3			✓
T5		✓	
T6			✓

The distance between attitude to assessments in general and attitude towards the specific assessments, both internal and external required by Skills for Life accountability regulations, were the main criteria taken into consideration to reach this categorisation. Adopters were the most 'compliant' and appeared willing to implement procedures as recommended by the Strategy; Adapters, were not unco-

operative but adapted the procedures to collude more with previous practices; Resisters seemed to be less willing to make changes and made minimal changes to their practices.

The overview gained from this is that the Strategy would have been better received if there had been more foresight about possible resisters and a more effective and extensive training programmes and an additional information dissemination strand been included in Skills for Life.

All this having been said, having looked at the Receivers themselves, and the stakes, this in itself did not explain the patchy washback. The various stakeholders' engagement with communication of the Strategy merited investigation to see if this factor could further help explain this.

9.18 Communication

The role of communication in a project is prominent in much of the literature on change management (see for example, Kennedy 1987, Markee 1983, Rogers & Shoemaker, 1971). Henrichsen's model (1989) likewise includes communication as one of the key factors in determining success or otherwise of an innovation. As he states:

‘...although change in a desired direction is possible it seldom happens by itself. Innovation is seldom sufficient on its own. Neither is merely communicating the news of an innovation to the appropriate audience enough to bring about change’ (Henrichsen 1989: 4).

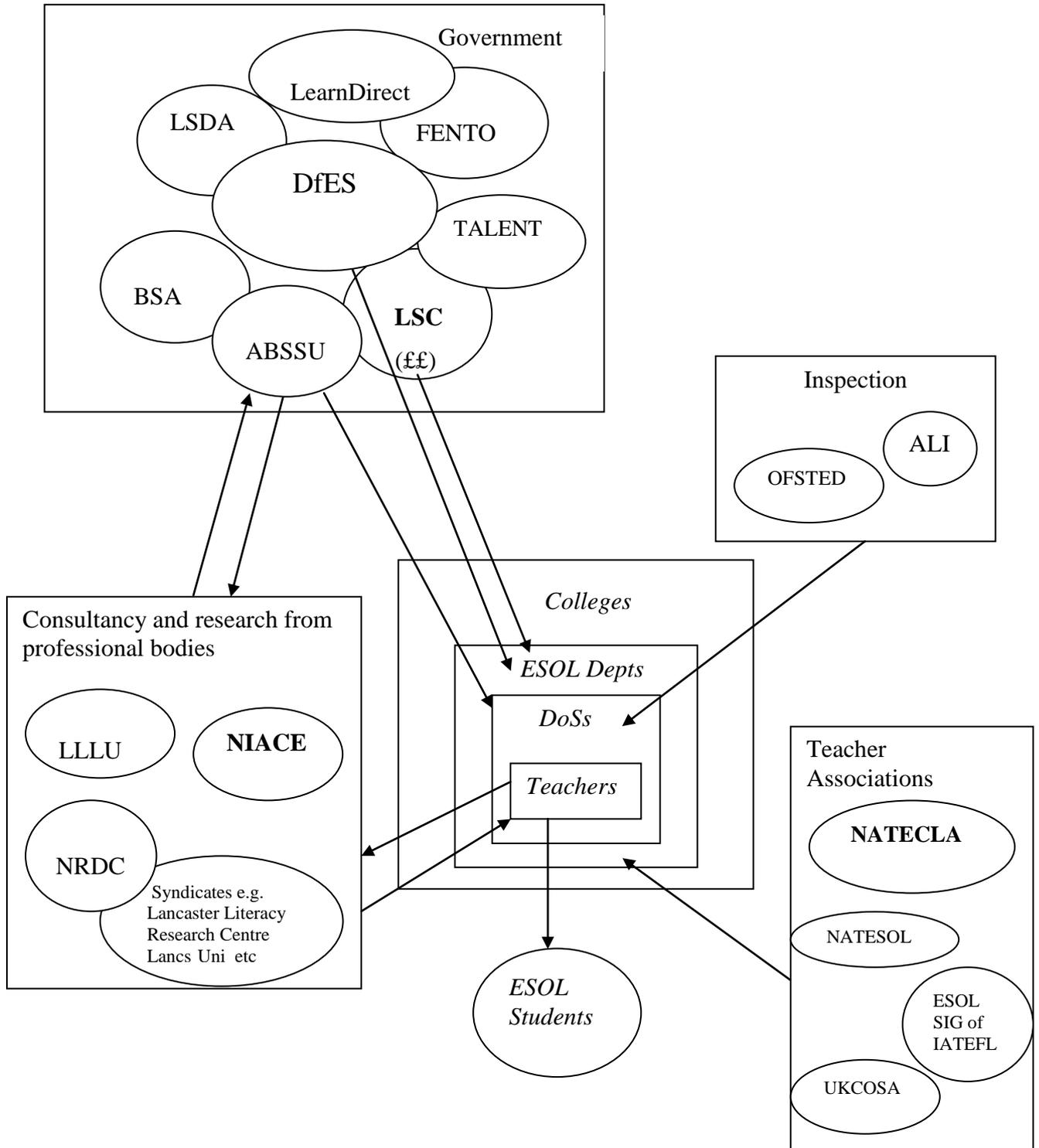
In other words, the nature of the information and how it is disseminated is paramount.

I wanted to establish what the ESOL staff understood about the Strategy and whether they connected any potential beneficial washback (or even unbeneficial washback) with the aims of the Strategy. As established already, the evaluation of stakes can affect behaviour; whether the consequences are real or perceived, e.g. the belief that pay cuts may follow poor student exam results, is not important. It is the belief of what may happen which drives behaviour (Donaghue 2003). The teachers' beliefs would be based on the ideas they have garnered about the Strategy from various sources, both official and unofficial. Thus, establishing what they knew about the Strategy and how they found out about it could elucidate subsequent behaviour. Monitoring how messages about an innovation are communicated and received is crucial to understand the success or otherwise of a projected change. As discussed in Section 6.14 in Chapter 6, questions focussing on communication of matters relating to Skills for Life were included in the interviews to get a sense of how the messages about Skills for Life were being related and also were being related to.

From an examination of the reporting bodies through documentary sources such as reports, professional newsletters, websites, and traffic on professional discussion lists it became clear that there were various sources of information, from the top down (from government level) and from the bottom up (from user level). From the top, there were the 'official' governmental bodies such as ABBSU and LearnDirect who were providing information on Skills for Life. There were also professional bodies with a direct keen interest in the issues, such as NATECLA who aimed at keeping the ESOL workforce abreast of developments, and had been consulted on the draft curriculum (DfES 2001). Other bodies were of a semi-official nature such LLLU who also had been involved in the development of the new ESOL curriculum (DfEE 2001). Relationships between the various bodies involved in the Skills For Life programme is of interest in seeing how information was disseminated; Figure 8 outlines the various sources a teacher may have encountered and highlights why messages may not always

have been most effectively transmitted, there being so many potential links in the chain of communication.

Figure 8 The communication network - sources of information drawn on in this study



In the following sections I will report on the main themes emerging from the data in relation to discussion of communication in an attempt to identify the main areas of concern from the point of view of the teachers and DoSs.

9.18.1 Official sources

One of the main sources of communication was from Government bodies, and as Figure 8 shows, various bodies were involved in some way or other with Skills for Life. However ABSSU (Adult Basic Skills Strategy Unit), created in 2001 (Hamilton & Hillier 2009) was the main one mentioned by the study participants. The information from Government bodies was found to often be confusing however (T3-6:950 and T3-6:922). DoS3 criticised one sort of Government level information for not being effective:

‘the LSC isn’t connected to the people who it administers unto and it would seem to me there’s an ESOL industry somewhere you know that is really not communicating very well to people to – especially the people who are more on the periphery of ESOL’ (DoS3-10:1401).

ABBSU was most often seen as having a controlling role but not as disseminating useful information. Three of the participants had only heard of it and had little idea of its real function (DoS2-4:229; T3-6:434; T4-8:515); T5-12:437) and even DoS3, who may have had more contact directly in the DoS’s co-ordinator role, was rather unclear about its purpose (DoS3-10:1448). Yet its aim was to be the hub for all information and advice relating to the Strategy, according to official documentation and websites of the time.

The LSC (Learning Skills Council), which has subsequently been succeeded by various other organisations, was mentioned but only by the DoSs. None of the teachers listed them as a source of information, but showed they knew of its existence by mentioning

it in relation to the pressure they felt under to comply with targets set by the LSC (see previous chapter). The DoS in their management role would have been aware in discussion with college management about targets set by LSC and thus it is natural they were aware of them and their role. That no information came from LSC which the teachers recognized as such, suggest their contact was in practice, primarily if not totally, administrative.

TALENT (Training Adult Literacy, ESOL and Numeracy Teachers) which is a body concerned with professional development was also mentioned briefly (DoS3-10:1464; DoS3-11:272). It was found to provide useful information and materials. Site 2 found materials which helped them design their placement material, for example (T3-7:622). This is a website acting as a repository for information rather than a professional body.

9.18.2 Professional bodies

Among other professional bodies mentioned was UKCOSA (The Council for International Education) which supports the needs and interests of foreign students in the UK, particularly those in institute of higher education, and UK students studying abroad. It is notable that the site which mentioned the latter, was one with as many students of an EFL profile rather than ESL, since this organization has EFL matters as their main concern so it was interesting they were mentioned as part of the communication chain of Skills for Life issues.

NATESOL (Northern Association of TESOL), in existence since the mid-1980's is another professional association, with a regional focus. It was recognized as offering professional support and information, but as regards Skills for Life, the former rather

than the latter (T3-6:581). Staff at Site 2 were the only ones who appeared to engage with this body however.

The professional body most referred to by the teachers and DoSs was NATECLA which is an active interest group for teachers in ESOL, founded in 1978, which also provides input into initiatives involving ESOL, such as the Skills for Life review by NIACE which took place in 2006 (Hamilton - & Hillier 2009). NATECLA was reported by teachers to be a better source of information than ABSSU in that the information was targeted to their needs and was simpler to understand (T3-6:915 and T3-6:950) and was spoken about by most of the study participants (DoS1-1:208; T1-2:234; T2-3:179; DoS2-4:253; T3-6:442; T4-8:302; DoS3-10:1448; T6-13:642). Other sources included professional bodies including the exam boards who produced the Skills for Life exams.

NIACE (National Institute of Adult Continuing Education) whose remit is to encourage all adults to engage in continued learning³² was almost as well known by the participants as NATECLA.

9.18.3 Management

Generally, it was reported that 'official' information was filtered down from Management via the DoSs (DoS1-1:300; T2-3:566; T3-6:892; DoS2-11:285, DoS3-11:285). Teachers did not receive this information (such as matters concerning funding issues) directly from official sources, e.g. from the local LSC or ABSSU, as alluded to already. Rather, the DoSs sifted out and passed on what they felt was necessary for teachers to know. They did this as they felt there was a great deal of

³² See <http://www.niace.org.uk/about-us> for further details.

information available, in fact too much for the teachers to process (DoS2-11:285) and some of the DoSs struggled to manage the information load themselves:

'I do get all the information you could possibly want - whether I believe a lot of it or understand' (DoS2-4:736).

The teachers reported they found out more about Skills For Life through various other means such as signing up to e-mail discussion lists than through management (T2-3:570, T3-6:894). They received this information directly, but these were not generally from 'official', sources. DoS1 in fact reported she thought maybe her teachers were more informed on ESOL matters relating to Skills for Life than she was and admitted her team may have had sources of information about developments in ESOL she did not know about (DoS1-1:485), but it must be remembered she was a Literacy, not an ESOL, specialist.

Thus the role of managers (including DoSs) as a conduit was vital for the transmission of the Strategy aims and procedures, but they were not the only conduit.

9.18.4 Peers

Another route for information to reach teachers was via their peers. T5 suggested how important the role of the staffroom was for teachers for sharing information about Skills for Life (T5-12:832). Two of the sites (Sites 1 and 2) reported having staff meetings, of various formats and levels of formality, and they appeared to be important in spreading and sharing information about developments regarding Skills for Life (DoS1-3:301; T1-2:435; T2-3:350; T3-6:609; T4-8:575). At Site 3 the topic of meetings did not arise directly except for T6 mentioning that there were no formal staff meetings, mostly due to pressures on part-time staff (T6-13:957). Having said that most of the teachers were involved in staff meetings, these involved mostly the

full-time teachers. It was hard for teaching teams to hold staff meetings including part-time staff as their time for this was not paid and they also often needed to work hours elsewhere (T2-3:2232;T6-13: 796).

There seemed to be a rather divisive ‘them and us’ mentality expressed on occasion between the full-time and part-time staff, arising over such issues as keeping informed of developments. Firstly, in rather critical terms, T3 suggested the part-time staff could do more to keep involved:

‘it tends to be the part-time tutors who don’t come in that much don’t tend to pick up on what’s going on - and the part time tutors they don’t check their e-mails they don’t check the pigeon holes you know so it’s ‘get with it - wake up a bit’ (T3-6:910).

They were not generally keeping ‘in the loop’, again due to the nature of their time being split between various jobs and not being at any one site for the length of time as expected of full-time staff (T3-6:909). Not being around amongst colleagues between classes also reduced the opportunity for sharing information, both formally and informally.

There was some evidence of a lack of sympathy for the constraints and pressures the part-time staff were under. However, this was not the case across the board. T5, at that time having a part-time contract herself offered a reason for part-time staff’s behaviour:

‘I feel that I kind of take control over how much I get involved with because otherwise being a part-timer you could just be overloaded too much with too much information’ (T5-12:833).

Contractual status is of importance in this matter as so many ESOL departments consist of a high proportion of part-time staff. At Site 1 for example only half the staff

had full-time contracts (T1-2:177) and this kind of profile was found in other studies (KPMG 2005; Baynham, et al 2007).

9.18.5 Exam boards

An indication of how far the exams, even though just one strand of the whole strategy, were synonymous with Skills for Life in some teachers' minds was made clear by T5's comment, when asked about how she kept abreast of matters concerning Skills for Life in general, not specifically assessments:

'I guess my main source of information would be the various exam boards - that's my first source of information for whatever exam I'm teaching' (T5-12:464).

Other teachers also hinted that Skills for Life for them meant the exams, not the new curriculum and materials, or other strands of the strategy (T1-2:709; T2-3:502; T3-6:820). With this in mind, the value to them of information received directly from exam boards can be understood. There are two reasons for this being problematic nevertheless. Firstly, while the exams were closely based on the curriculum, this being a requirement for gaining accreditation, they were produced by a variety of Boards. There were several accredited Boards and therefore each would have their own understandings and interpretation of Skills for Life. Secondly, the information was in a state of flux as further exams were being developed and gaining accreditation and support material was being made available. This would make for a rather unclear picture, which was continually changing. Simply due to the amount of information available on exams its importance seems to have outweighed information on the curriculum, for example, which had already been published and distributed for a while by the point of my data collection.

The capacity for an exam to come to stand for the whole of an educational programme has been noted. Shohamy, for instance, suggests that

‘[i]n situations when pedagogical knowledge is minimal, the test becomes the substitute for other ways of communication such as curriculum, in-service training etc’ (2001: 68).

This may be too harsh in this case to suggest the teachers’ limited teaching expertise was why the exams took on such a prominent role but its prominence may indeed flag up insufficient training especially as regards the concepts underpinning Skills for Life (See below).

9.18.6 Training

Training has a key role in information dissemination. There was little mention however of training as a source of information about Skills for Life for teachers. DoS3 reported it was hard for teachers to access training due to financial and time constraints; many staff were part-time and not paid to attend training or found it hard to get to training having several different jobs, and the colleges were not providing financial support for training (DoS3-10:65).

At Site 2 some internal training was held (T4-8:256) but for the message to be transmitted from the Strategy ‘headquarters’ a centralised approach to dissemination appeared lacking. A programme of training for all staff may have ensured aims were been more clearly transmitted and provided understanding of exams and counteracted the quite dominant position of information from Exam Boards.

9.18.7 Issues with the dissemination of information

As well as there being several routes via which information was being received, there were several problematic themes which also arose. These are the subject of the next section.

9.18.7.1 Uneven dissemination pattern

The teachers gave the impression that how information was disseminated to teachers was above all a matter of individual teacher initiative in signing up for e-mail discussion lists, e-mail shots, or newsletters etc. from various bodies (T3-6:894). T3 suggested this was also why part-timers were much less likely to be well-informed, not feeling they had the spare time to attend to this.

Teachers discussed various strategies for keeping informed. T3 suggested that the role of continual professional development in the form of joining professional bodies, and attending conferences and training events was quite important in keeping informed:

'if you've been to an event like a NATECLA event and you've registered at the beginning – and quite often Oxford University Press or Cambridge will be sponsoring the event and then they've got your details then and then you'll be on their mailing list but if you don't go or you're not a member of any association you won't necessarily find out' (T3-6:906).

T3's colleague, T4, suggested keeping well-informed was a matter of actively seeking out information; there was no room for passively waiting for information simply to arrive. Keeping one's eyes open and asking questions of colleagues was the approach he advocated to become adequately informed (T4-8:247 and 292).

The communication patterns which were noted were probably not poor due to the Skills For Life initiative. It has to be noted that the teachers talked about their role in college and how they felt misunderstood in that the rest of the college did not appear to understand their role or their work in teaching English to non-native speakers of English (T1-2:1669; T3-6: 867; T4-9:439). Another example was that at Site 3, T6 was not sure if they had staff meetings at the other site, this being a split campus institution, and teachers tended to be based at one or the other (T6-13:965). This was in indication in itself of the levels of intra-institutional communication. Poor organization-level communication in general is probably the culprit. It is notoriously hard to manage well. As George Bernard Shaw is said to have expounded: ‘The single biggest problem with communication is the illusion that it has taken place’.

9.18.7.2 Inconsistent messages

There were also complaints about who received what information, and how, as well as the quality of the information. DoS3 expressed his frustration at, and the effect of, lack of clarity about the information the ESOL teams were receiving:

‘I would like to get hold of the Learning and Skills Council and our Funding Managers and tell them to go away and get real – because we spend countless hours worrying about is it funded isn’t it funded have we done the right number of guided learning hours is it going to count towards achievement – get those people ... and take them into a room and not let them out until they’ve come out with a solution’ (DoS3-11:1284).

A similar view was expressed by one of his teachers. She was obviously frustrated by the sense that ideas had not been fully thought through before implementation:

‘I feel I do get information but I feel that when they give the information they don’t know all – all aspects I’ve never been anywhere where I’ve asked

question and I've always got answers to my questions - I've been to so so many conferences and they go 'that's a good question - write it on the bottom of your questionnaire' - I'd have thought -... they can't be things nobody else has thought of (T6-13:1892).

T4, was of a similar opinion:

'I don't think you should start to tell people that things are going to change until you know the direction they're going to go in' (T4-4:441).

In the same vein, T6's frustrations concerned lack of information regarding the practicalities of trying to plan imminent courses:

'it's up and down up and down and you don't know where you are - you get used to it and it changes again - and to be fair to the college the college doesn't know either but erm I don't know about the funding hours I don't know about the mode - the level - who can take it - I don't know about how many times they can fail and retake it - I don't know about any of that and I finish in three weeks and I'm not back until the beginning of September' (T6-13:2817).

It was unclear at what level of management her frustration was targeted, and it appeared that she herself was not clear either. This resulted in an almost tangible sense of powerlessness as she spoke. Lack of information which affected the day to day work of getting classes advertised and running obviously would affect those having to concern themselves more with this end of college activity, as opposed to the general policy end. Better communication between the target setters and the implementers would have helped avoid such frustrations. There seemed to be too much change. This was also noted by Davies (2005) in the interim evaluation of the Skills for Life programme.

However, in regard to the communication of any 'grand plan' it must be noted the teachers' annoyance and frustration was voiced above all about the training and

qualifications required of ESOL teachers rather than anything regarding the students' exams (e.g. T3-6:928; T4-8:504). The system of specific qualifications required to be deemed an appropriately qualified ESOL teacher was still being developed at that point and as courses cost teachers both time and money this clearly was on their minds (T2-3:204; T4-8:431; T6-13:846). It was claimed they were faced with a string of mixed messages (DoS2-4:694; T1-2:452) and this led to an atmosphere of uncertainty, which appeared to have a spillover effect onto their view of communication of the Strategy matters in general.

T1 may have summed up the true reason for much of the manifest frustration which was evident in my interview with teachers. She stated clearly that the amount and manner of delivery of information from 'above' i.e. college management and from Government level, was fine in her opinion; the problem was simply that the systems which the information was informing them about were not what the teachers wished for (T1-2:805); expectations were not being reached. The nature of change management is that expectations need to be managed (Henrichsen 1989) but it does not seem to have happened in this case.

It can be argued that the initial stages of Skills for Life at least may have been more effective and efficient if a research, development and diffusion model had been adopted (Markee 1997). However, as Markee says, 'This leadership style is based on the change agent's status as an expert' (1997:65). It is doubtful this role could have been established in the eyes of the ESOL community. Who exactly would have been recognized as the change agent? DfES were too far removed from the day to day work of ESOL practitioners and ABBSU who were more closely responsible for the information sent out also seemed distant from the ESOL professionals judging by the results from my study. Another contracted body may have done better. Longer was probably needed to research, to plan dissemination and involve practitioners. While

TALENT was quoted as useful by some (T3-7:622) it did not seem to have the recognition as a disseminator, being, as stated, a repository for useful information for ESOL practitioners. Pathfinder Project, a pilot scheme to highlight ways the Strategy could be implemented by teaching teams around the country, appeared to be an attempt to showcase Skills for Life in action, but it was not mentioned by any of the participants at all. Despite this DoS3 commented:

‘what I fundamentally feel is that there were a lot of things that went on which didn’t have much consultation bottom up and then a policy was formulated which got imposed’ (DoS3-10:648).

Markee concludes from extensive studies of educational innovations that:

‘[c]hange agents must develop formal communication networks among participants: Indeed they cannot afford to leave developments of such communication networks to chance’ [original emphasis] (1997:174).

Markee (1997:174) further says the spreading of the message should not be left to ‘a single channel’ but in this case too many channels may have been a significant cause for the Strategy to have not been clearly represented.

9.18.1 Awareness of the rationale for Skills for Life

I felt it was necessary to investigate what the teachers understood about Skills for Life and its very rationale as I believed this may have had a bearing on their classroom behaviour and general acceptance of the changes. As T6 put it:

‘it’s really important the teachers are on board because if they’re not it’s going to be a mess’ (T6-13:1170).

On the whole, it transpired the rationale was not generally well understood. Teachers had heard of the Moser Report (see Chapter 3 for discussion of its role) but not a great deal more (DoS1-1:1703). T1 admitted openly she had no idea what the

rationale was (T1-2:1630), while her colleague, T2, said it was clear but did not actually manage to articulate what her understanding of it was (T2-3:1959). T6 also said the rationale was unclear to her (T6- 13:2713). Both T3 and T5 said they thought the Strategy had been introduced to address immigration issues (T3-7:1057; T5 - 12:1589), and T3 went on to elaborate:

'I suppose it's a result of politics immigration asylum refugees coming into the country long term residents who are costing money by not working Asian housewives who aren't putting anything back into the system from the government point of view so it is all influenced by politics so there's the result of that – and that's the same with the ABE the native English speakers again to try to get them off the unemployment role isn't it – in its more cynical sense' (T3-7:1051).

This seems indicative of a lack of clear communication of the true aims of the Strategy.

The emerging pattern was one of irregular communication of the intentions and methods of the Skills for Life strategy. The KPMG report on Skills For Life (2005) concluded there was no shared understanding of issues and priorities in ESOL and this thus backs up my own findings. For an innovation to be successful the message concerning its intentions cannot be left to chance.

To sum up the communications issues, overall the teachers felt informed, but it appeared they were informed via a rather haphazard variety of routes. DoSs were important for keeping their teams informed, and filtered out vast amounts of data to make the amount of detail to be processed more manageable for their teachers. This made their role in the transmission vital but not sufficient. Part-time staff, who comprise a large part of the ESOL workforce, were seen to be at a disadvantage as regards keeping informed, and whose fault this was, was somewhat contentious.

All in all it is hard to say whether the intended messages about the Strategy were being disseminated as through perusal of Strategy documentation I found no proof as to whether a clear dissemination plan had existed. There was discontent over aspects of how the Strategy was being handled up to that point of my data collection, particularly regarding teacher qualifications. There also seemed to be a lack of teacher agency in the communication process. Information came from the top down and was often deemed insufficiently complete to enable teachers and DoSs to plan effectively.

9.19 Factors which may hinder effective change

The Henrichsen model suggests that as well as considering Communication and the Receivers in trying to understand the progress of an innovation several other parameters of the situation may be considered and these are labelled: the factors which facilitate/hinder change. The model suggests investigation of the innovation itself (Skills for Life), of the resource system (those instigating the innovation, in this case DfES/LSC), of the user system (those who implement it, namely tutors and DoSs and ESOL departments) and inter-elemental factors amongst others. While the detailed breakdown of the factors Henrichsen suggests was not fruitful for analysing the data for this study, its consideration did lead to investigation of aspects of the Skills for Life programme, specifically regarding the assessments, which proved to be illuminating.

In using these terms to look again at the data, various tensions and clashes between functions and aims of aspects of the Strategy became apparent. This took the study beyond exploring purely washback, but since primary investigation in this area proved of interest it became apparent that it was worth moving beyond the initial

study parameters . What stood out as of potential importance was that the financial imperative pervaded the discussions of Skills for Life with the teachers and DoSs and were predominant when the justification for courses of action related to assessments, were the focus. This was found to be a strong theme. The teachers felt a keen pressure to deliver good student results (see previous sections regarding stakes). Apart from the allusion to effects relating to pay and professionalism, which had no concrete basis, they discussed results in terms of hitting targets. None of them however included any conception of the purpose of these targets. In considering this, by returning to the data, it became clear there were a number of classroom practices which, while not a manifestation of washback (i.e. classroom behaviour caused directly by a certain assessment) were a result of the system the assessments operated within. These findings suggested the situation could be of wider significance, to not only this particular set of exams, or even just this specific educational setting, but potentially to policy in a variety of areas, in terms of the consideration of accountability measurement methods.

While exploring the classrooms in the search for evidence of washback, some classroom practices stood out as not being pedagogically defensible and in order to understand this, the functions of the assessments involved was examined. The unifying factor where dubious practice occurred appeared to be the monitoring role that the exam results (and ILPs) took on. Exam results acted as a means of reporting for accountability purposes.

9.19.1 Inappropriate candidacy

Having reviewed the data some tensions became apparent. Teachers reported sound practices which aided language development via the curriculum, culminating in an assessment which reflected students' competence. However, some of the assessment practices were distorting this development process. They seemed to be a result of the

needs of the wider system, as is common to all areas of the public sector, to gather data to provide a measurement of success for accountability purposes. There were three main areas where such distortion was observed. All these examples have already been discussed in previous chapters outlining the findings.

Firstly, one of the ways in which practices were distorted by the need to fulfil accountability measure was that students were not necessarily sitting exams at the appropriate level, i.e. matched to their competence at the time of the exam, because of the system of ESOL departments being required to assign QualAims to students at the beginning of their courses. Success was judged in terms of targets reached, i.e. in number of exam passes, rather than allowing other measure which were a direct reflection of success such as a student leaving a course because they had found work (see Section 9.7.3). Also because of fear of not hitting targets it was mentioned that students were entered for exams it was felt they would comfortably pass, rather than one which might reflect their current competence 'ceiling'.

Secondly, students who were not described by the profile of those the Strategy aimed to support and develop, namely people who had come to make the UK their home, were included in ESOL classes. In some cases the exams were not appropriate exams for students who had no need of Skills for Life qualifications in their daily lives, as they were fully intending to return home shortly after their studies. The practice of 'piggy-backing' or 'double accounting' to cater for both the requirements of such student and also the requirements of the accountability system, highlight the lengths tutors and ESOL departments were going to in order to accommodate both sets of needs (see Section 8.6.4.1).

The piggy backing is not a result of a poor exam but a candidate population being directed to the wrong exams which do not match their target language use aims. The

boundaries of who is more suitable for which exams had become 'fuzzy' due to the very accountability requirements which caused them to be sitting these exams, rather than the ones they may have chosen to take themselves, such as from the Cambridge main suite. By offering a higher rate of funding for ESOL students without careful definition of eligibility, this led to an expansion of the potential ESOL student population who were able to benefit from Skills for Life, and allowed in students who would traditionally be labelled EFL students. Piggy backing would appear to be an unnecessary doubling of efforts on the part of the teachers and the students simply to satisfy a system which had caused the situation in the first place.

Equally, other students who did fit the Skills for Life ESOL student profile, also needed exams other than the relatively unknown, new Skills for Life exams to further their job prospects (e.g. to practise as doctors – see Section 8.8.1). However they also had to study for and sit Skills for Life exams to satisfy the accountability measures as all students needed to gain a Skills for Life qualification for the college to draw down funding for those students. A more meaningful outcome may have ensued had these specific students, needing specific English language qualifications (such as IELTS) been able to prepare and sit these exams alone. This success, being noted as a successful completion of a college ESOL course if not a Skills for Life course, could satisfactorily provide a suitable measure of success for a system, which aimed to enhance employability of students it must be remembered. A one-size fits all approach could potentially impede or at least delay certain students from gaining employment in their chosen field.

Another way in which exams appeared to be taken by a candidate constituency which was not necessarily suited to those exams was in the case of the higher level Skills for Life exams (namely Level 1 and Level 2). It must be remembered that it was deemed by the system of the Strategy that by the time ESOL students reached Level 1 they

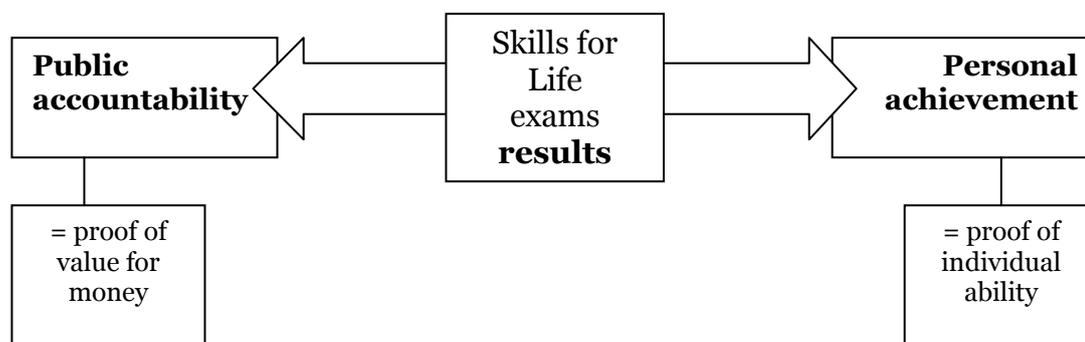
should be judged by the same means as native speakers of English so ESOL students sat the National Literacy exams at Level 1 and Level 2, not ESOL specific exams (as stated in Chapter 3). The ESOL students, as did literacy students, covered four skills in class as set out in their curriculum, yet, as per Level 1 and Level 2 exam format, were only tested in reading and indirectly tested in writing (with an emphasis on spelling and punctuation). Again, results gained from such exams do not represent the range of a Level 1 or Level 2 ESOL students' English competence. Such a result gives no information about ESOL students' abilities in speaking and listening or ability to produce text. Once more the system seems to have imposed requirements which did not result in useful information for students or future employers. While it can, maybe erroneously, be assumed a basic level of ability of speaking and listening for a native –speaker of English this cannot be assumed for someone with English as L2. Equally their ability to produce written text was not reflected in the Level 1 and Level 2 qualifications (at that time), which at least is on par with the situation for native speakers, neither being able to provide an adequate account of this skill from those exams. It is beyond the remit of this study to examine the reasons for the format for the exams at these levels but they did not appear to reflect the candidates' skills adequately or fully.

The issue at stake is not that the Skills for Life exams are of poor quality, as already stated. To be accredited, the exams boards had to prove the exams (the Entry Level exam at least) were closely aligned with the ESOL Core Curriculum and many of the exams (at least those then being used) were produced by some of the largest, most well-known and most respected exam producers in the UK. Their quality, or lack of it, is not of concern in this study. The issue is how the exams were being used, who was sitting them and when.

9.19.2 *Dual functionality*

As stated above, the study moved from a focus on washback to other phenomena which became apparent and this arose especially through turning attention to assessment functions. As a result, I believe a negative consequence of Skills for Life was an increased focus on external assessment for accountability purposes, which distorted the meaning of the scores gained from achievement testing since the scores had a second function beyond that of measuring a student's language competence. This second function of the results was their utilization for accountability purposes to prove funding was being well spent. Proof was required that a student had undertaken a course of learning appropriate to their ability level, had taken advantage of the facilities available through regular attendance and been taught adequately. The exam results did not necessarily provide this information however, as outlined already. In the case of Skills for Life the very measures put in place to supposedly ensure that teaching was effective and students were learning English as effectively as possible, and thus that public money was being well used, encouraged assessment practices which did not necessarily provide such information. The problem lies in the dual functionality of the exam results, which Figure 9 aims to represent.

Figure 9 The dual functionality of exam results and its inherent tensions



The issue I believe is what I would like to refer to as misguided accountability. Any taxpayer would be keen for measures to be in place to ensure public money is being wisely dispersed and used effectively for state education, as alluded to already.

Accountability is a necessary procedure in a democratic society to ensure money is spent to achieve best value for financial input. However the way this is engineered is the crux of the matter since a procedure which cannot ensure such an assurance and warps the process it is trying to measure is not an effective method. Goodhart's Law sums this up in stating that 'when a measure becomes a target it ceases to be a good measure' (Amrein & Berliner 2002). In a similar vein, a social sciences version of Heisenberger's Uncertainty Principle³³ can be drawn on to explain this phenomenon. This version of the Uncertainty Principle states that the more important that any quantitative social indicator becomes in social decision-making, the more likely it will be to distort and corrupt the social process it is intended to monitor (Amrein & Berliner 2002; Linn 2000).

In the case of this study the social indicator is the monitoring of public funds being spent on the Strategy and it is this which has caused the distortion in classroom practice, as mentioned. The cause of the distortion lies in the fact that at least one of these two functions is consistently of a high-stakes nature i.e. using the exam results from external awarding bodies for monitoring, which secured continued funding. The other function is of high-stakes on a differential basis, in that some students required exam passes for applying for citizenship while for others there were no concomitant stakes at all. The internal assessments which the ESOL students experienced had low-stakes attached to them, and thus such tension did not arise (see Section 7.3 above).

³³We believe we have gained *anschaulich [often translated as perceptible or physical]* understanding of a physical theory, if in all simple cases, we can grasp the experimental consequences qualitatively and see that the theory does not lead to any contradictions'. (Heisenberg, 1927, p. 172 cited in Hilgevoord, Jan and Uffink, Jos, "The Uncertainty Principle", *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition)*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2011/entries/qt-uncertainty/>>.)

In the same way that Haladyna et al (1991) talked about 'test score pollution' meaning that a score cannot be reliably understood to represent the true ability of a candidate when there has been excessive exam practice leading up to the exam, I suggest this term may also be applied when an exam has been taken which is at the wrong level for a candidate, or the wrong type of candidate has taken the exam. If a student is not taking an exam which indicates the upper reach of their ability then the result does not supply useful information. It tells the end user (e.g. an employer or educational institution) a certain level of English which the candidate can manage, and it will probably be assumed this is the top range of their ability. Likewise, if an exam designed for a certain type of student is taken by a different student group then this does not provide useful data. This does not do justice to the student and neither does it convey accurate information to the end user.

One of the results of this misguided accountability has been a range of tensions, beyond that pointed out, i.e. neither of the functions of the exam results being satisfactorily fulfilled. For example, what became evident from the data was that teachers were, on the one hand, experiencing tensions between what they felt they would need to teach to fulfil student need and, on the other hand, the requirements of the system they worked within which required the use of ILPs and as high a level of exam passes as possible to secure continued funding. This is backed up by the results of other studies, for example, which found

'a constant tension between the teachers' understandings of their learners on the one hand, and their perceptions of the policy demands and audit culture of FE and ESOL on the other' (Baynham et al 2007:35).

Nevertheless, such tensions are not unique to ESOL. McNess, Broadfoot and Osborn (2003:243) found from the PACE³⁴ study that primary teachers in the UK were affected by similar concerns arising from the tension of fulfilling government requirements and dealing with the children's learning requirements.

Other studies have come to a similar conclusion that dual functionality for a single exam is detrimental to the chief aims for examining. See for example (Qi 2004) who looked at the NMET (National Matriculation English Test) in China. In that case one function consisted of selection for university entrance and the other was to instigate developments in the teaching and learning of English. One key difference between that study and this one was that a clear intended function of the new test consisted of improved teaching and learning via washback whereas with Skills for Life apparently no specific intentions (other than increasing the skills of the workforce) were documented and exams were not designed to be taking a primary role in bringing about change via washback, acting as 'lever for change' (Pearson 1988).

9.19.3 Dual functionality and validity

Gipps has questioned the validity of exams which are used for varying purposes. In her example she contrasts the formative use of exams for pupils, with the use of exam results for accountability, not just at school level (and maybe also at class level) but also for informing policy (1994). As she says:

'Since the uses are so clearly different at the different levels it seems highly unlikely that the same test can be considered equally valid at all levels, which is the same as saying that a test cannot be valid for all purposes' (Gipps 1994: 64).

³⁴ The Primary Assessment, Curriculum and Experience Project – University of Bristol Graduate School of Education.

She does not clarify sufficiently why an exam might not be valid since it will produce information about a specific candidate's ability and if that information may be put to various purposes, e.g. to diagnose depending on the type of information supplied by the exam board, or to certify competence (proficiency), that need not detract from the exam's validity.

The onus, Gipps claims, is on the exam boards to ensure appropriate use of exams.

As she says:

'Most test users cannot carry out validity studies, and so it is the test developer's role to articulate the uses to which a particular test may be put. For this we can pose the question: for what use, and of which construct, is this a valid indicator? Test developers must address this question as a priority in the design of tests and their evaluation of construct validity. This implies an opening up of the test's constructs to users and for the test developers to commit themselves to appropriate test use' (1994:170).

However, ultimately an exam producer cannot anticipate all eventualities of exam use. In the same way a manufacturer of a hammer cannot be held responsible if it is used by a murderer to kill someone, an exam producer cannot be accountable for the use of an exam beyond its designed usage, so long as this is clearly articulated. As Gipps suggests, clear information regarding the intended use and candidate population must be provided but whether end users then adhere to that recommendation for the exam's usage is beyond the jurisdiction of the Exam Board. While an Exam Board can monitor exam quality closely through standardisation procedures and ensure scores are meaningful through extensive research programmes and adhering to good exam production practice, they could not conceivably reliably monitor the usage of the scores. Dual functionality may cause neither of the functions to be adequately fulfilled, and cause practices associated with the exam to be distorted, and this will thus affect the validity of the exams, since

validity lies in the use to which the results are put, not in the nature of the exam itself according to the Messick-ian view of validity (1996) (see also Davies 2003).

This final section of the findings aimed to describe the effects of aspects of the system within which Skills for Life operated. It highlighted how certain practices which the teachers and DoSs had to operationalise, such as setting easily attainable QualAims to ensure accountability, themselves caused what they were aiming to measure to become distorted. Being within an educational setting the paramount aim is effective learning and anything which mars this must be seen as problematic. While accountability is an issue of public concern, systems put in place to return accountability reassurances must be managed adequately to ensure they do not interfere with the learning process or hinder effective change to an educational programme.

9.20 Summary

The washback identified was characterised by its differential nature and this chapter aimed to explain why this may be. One main reason was that the stakes were not found to be uniform for the various stakeholders. The nature of the teachers themselves was also found to be of significance in the way they responded to the Strategy, having different attitudes towards it and towards assessment, as well as differing amounts of professional experience which may have affected how they reacted to challenges they faced in adapting to new ways of assessing their students. Equally, the nature and quality of the communication of the Strategy in general, and in particular about the new exams, was not consistent and centrally managed and the innovation literature highlights the importance of a clear message for an innovation to be realised effectively.

Finally, effects beyond washback were also uncovered. Other factors within the system (the new Strategy) suggested that the two clear functions of the exams to provide proof of progress for accountability purposes, as well as providing a measure of achievement for individual students were at odds, due to the perceived pressure to reach targets. Distorted classroom practices occurred as a result.

10 CONCLUSION

‘Measure what is measurable and make measurable what is not’. (Galileo)

10.1 Chapter overview

In this chapter I will first review the aims of the research and then summarise the findings from this study. I will next set out some implications from this study, the value and role of washback studies and also potential solutions to the dilemma of one of the findings, namely that dual functionality has distorted the meanings of the exam results.

10.2 Review of the aims

The aim of this study was to explore the washback from assessment practices resulting from the Skills for Life Strategy which was introduced in 2001. The Strategy aimed to increase the skills base of the UK workforce and focussed on people who had not yet gained Level 2 (of the National Qualifications Framework) qualifications in literacy or numeracy. Large sums of money were injected into Adult Literacy, Numeracy and ESOL classes. Assessment was one of five strands to the Strategy and the new Skills for Life external examinations were introduced in 2004 after the establishment of the new ESOL Curriculum (in 2001), which they were closely aligned to. With the Strategy came a much more centralised approach to the teaching of ESOL, where previously education providers had worked relatively autonomously, in terms of course content and assessment.

The research questions, which were grouped into three sets, for this research were:

RQ 1.a) What is the range and nature of assessment practices in UK ESOL teaching?

RQ 1.b) How are these practices linked to the Skills for Life strategy?

RQ2.a) To what extent is there evidence of washback from the assessment practices?

RQ2.b) Is any washback related only to the assessment practices resulting from Skills for Life?

RQ3) What are the factors which may drive washback?

Having collected data at three colleges from interviews and observations, which were analysed entirely using qualitative methods, the following main findings resulted.

10.3 Review of the findings

10.3.1 Assessment practices

The findings of this study showed that a variety of assessments existed in the ESOL classroom, from students' first contact to exit from a course of study. The assessments ranged from low-stakes assessments, which were undertaken for purely pedagogical ends, such as diagnostic testing, to those which were of higher-stakes, with purposes beyond the classroom, for example to provide data for accountability and for citizenship purposes. Some of the assessments had come to be of some importance. For example, the external exams and ILPs took prime position regarding their impact

due to the stakes involved in acting as measures of success from which funding for the colleges was secured.

10.3.2 Washback

The washback, although not strong, was in evidence but primarily caused by the new skills for Life external exams and from the ILPs. The ILP-related washback concerned the classroom activities taken up with completing these documents, which were handled in a variety of ways in the various classes to try to ensure they remained useful but did not take up excessive class time. Class time was felt to be pressured because of the exams which the students would sit, and the pressures on departments to maximise good results to reach college targets (as discussed in relation to stakes).

The washback concerned the timing of when more exam-oriented practices occurred (washback intensity) and also an increased focus on accuracy over fluency as this was perceived to be needed for exam success. As regards the materials used, some washback was noted in the use of practice materials and efforts to produce materials which emulated the exams, and these did not resemble the Skills for Life materials made widely available to accompany the ESOL curriculum.

As regards methodology, some washback was noted in that interaction patterns changed according to which exam was to be taken next. For example, a higher intensity of practice of talking in pairs occurred prior to the speaking exam. What is more, exam practice took on a more prominent role once the Skills for Life exams had been introduced, but to varying degrees in different classes. A rather atomistic approach to the way the language was taught, in accordance with the approach taken by the curriculum, was noted, but it is hard to distinguish whether this was the effect of the exam or the curriculum.

Some examples of what was labelled affective washback occurred in that the students' attitude towards the teacher changed as a result of the judgment of their ability being undertaken by an external source. This freed the teachers to be support and guide rather than also assessor.

Reasons for the weak washback may have been associated with the timing of the data collection or the close alignment of the exams with the curriculum. The most notable aspect of the washback was that it was variable from site to site and from class to class. The examples cited above were not found consistently across all sites. This led to further investigation of the factors which may influence whether washback takes place or not.

10.3.3 Stakes

The role of stakes played a key part in determining how the assessments were perceived by the study participants, and also in the strength of washback, and these stakes varied, not just between stake holder groups but also within the groups. As with validity, the stakes are not a feature of the exam per se but of the use it is put to by individual users; equally an exam is not valid in itself but is valid for a specific use. The research likewise reinforces that the stakes are not a feature of the exam itself but of the individual in terms of the effect it will have on that individual. Not all exams will therefore have the same stakes, and any one exam will not have specific 'fixed' related stakes as individuals will have their own reasons for taking the exam. The washback is therefore likely to be differential as a result of the strength of the effect of the consequences in individual cases. It must be noted again that whether these consequences are real or perceived makes no difference; they equally will effect different behaviour.

10.3.4 Receiver factors

The first implication for considering key players in an innovation is to avoid considering them as a homogeneous entity. Assumptions should not be made about how the Receivers will deal with the introduction of an innovation without investigating them in terms of various parameters, as demonstrated here. The richer the investigation, the clearer a picture of the influences on their reactions and behaviours will be.

The teachers and DoSs can be observed from the perspective of influences on them, both internal, i.e. collegial and institutional practices, as well as external, such as previous teaching experience in EFL and MFL, or in other professions. The length of their teaching experience is to be considered as well as experience with examination boards. All will have influenced their beliefs and evaluation, and determined whether they resembled adopters, adapters and resisters of innovation.

10.3.5 Communication

Through considering the communication routes whereby the participants learnt about Skills for Life, and in particular the exams, it became clear that a centralised, standardised method of communication from the Strategy led to teachers did not exist. They relied on information firstly, from government level and management filtered through the DoSs and secondly, from any professional bodies they chose to gain information through signing up to discussion lists, or e-mail drops, amongst other methods. The professional body, NATECLA was the most frequently cited as providing useful, accessible information for the teachers.

Another widely used source was peers and staff meetings were important for dissemination which had implications for part-time staff for whom it was harder to attend such meetings. The Exam Boards themselves were the other main source of information.

The overall perception was that the dissemination of information was irregular in method and quality. As an indication of the effectiveness of dissemination of information about the Strategy, the participants were asked about the rationale for the Strategy and this proved to not be clear to them, which indicated a lack of effective means for diffusion of information.

10.3.6 Factors hindering effective change

The final area to be examined in order to better understand the occurrence of washback in the situation being studied was the combination of factors already investigated. Through considering such combination, the research moved beyond purely a study of washback, in that the scope of some of the effects moved beyond the classroom. Some tensions became apparent which were found to be probably due to a combination of the stakes involved, communication issues regarding the Strategy as a whole and also teachers' own experience levels and their perceptions of the Strategy, which affected how far change was embraced.

The greatest influence on assessment practice however proved to be the requirements of the Skills for Life administrative systems in providing measures of success for accountability purposes. The need for accountability measures affected which assessments took place, and who took them, a distortion which confounded the functions of the assessments: to prove students had made appropriate progress for

accounting purposes and for students' own purposes. This outcome is of significance for policy in this area and also elsewhere in educational settings where accountability measures are put in place, in the form of exam results. Where targets are used as a means to enhance efficiency, effectiveness or productivity there is the ever present danger of unintended consequences of such schemes. The stakes involved, and the higher they are the more of an effect they are likely to exert, are likely to provoke behaviours which subvert the original aims of the targets. In the case of the Skills for Life exams it appears one such outcome was some less than ideal pedagogical practices and evidence of some distortion of the validity and meaning of some exam results. The issue of dual functionality was thus found to be as notable, if not more so, than the washback found.

10.3.7 Summary

As regards this washback study, the variables, namely, the range of stakes, the effectiveness of communication of the new Strategy, the influences upon the various Receivers detailed in this study, as well as others not focussed on here, begin to account for the variety in washback noted, weak though it was. Other reasons for the weak washback were probably the time at which the study was undertaken in relation to the introduction of the new exams, and the close alignment between exams and an established curriculum.

Yet, more importantly, the original aims of the research to study possible washback became just one aspect of this work, by the time of its conclusion. In uncovering effects beyond the classroom the study moved into exploration of impact, with implications beyond this field. Other findings of perhaps wider significance were the effect of the dual functionality of the exams in that this phenomenon may be of

relevance beyond this specific setting and this will be expanded further in the section below.

10.4 Implications of the study

10.4.1 *Theoretical contribution*

This research has contributed in particular to the field of washback by highlighting the fact that washback can be differential, which emphasises that washback is complex and to capture it meaningfully, to be beneficial for monitoring exam use and additionally informing exam quality, is no light undertaking. Washback studies need to avoid being simplistic and missing key factors which illuminate contextual detail. The nature of washback can easily be masked by superficial investigation.

This study has also contributed by exploring effects which go beyond the content or methodology of classes, but touch on the affective aspect in that evidence was found of a change in student's attitude towards classes due to the exams. In addition, and linked to this, the student-teacher relationship was found to change in some circumstances because of the presence of external judgment of the students' abilities.

Positive washback is manifested as the desirable classroom activities which lead to improved learning, brought about by the introduction of a certain type of assessment. It is that assessment which is the lever for change. Washback is not always planned and intentional however. Washback studies aim to chart this change in classrooms either to ensure it has happened as predicted and desired, or to chart why it has not happened, but also, as washback is probably as often incidental as intentional, to chart and therefore learn from various teaching and learning situations, from which others may in turn learn. Many studies chart the attempts to effect change through exams. This research charts what may be termed 'incidental washback'.

A baseline study can clarify a situation prior to a new exam's introduction and classroom behaviour can in this way more easily be attributed to the new exam. Yet, there is a call for washback studies where an exam has already been in place for a while, or as in the case in this research, where there was not a 'clean-sweep' with a single exam superseding a prior assessment method but introduction overlapped with previous methods. The complicating factor of various new exams being introduced also added a layer of complexity. In such cases a baseline is not always possible. This research has showed how washback studies can still be of use nevertheless in such situations.

10.4.2 The value of washback studies

It should be considered whether, since washback studies pose various problems, they are worth undertaking. Securing the evidential link (Chapman & Snyder 2000) is a particularly taxing aspect of washback studies and too often behaviour is attributed to an exam without thorough investigation of other possible reasons for it. Making assumptions about data derived through observations needs triangulating with study participants' verification, but this is very hard to secure convincingly. Declarations of identified washback should be treated with caution. In addition, making predictions about washback should be made with caution as such a variety of factors affects whether washback will happen or not. So it may be asked what the value of studying washback is if in fact it is so demanding and imprecise.

Washback studies aim to elucidate the effect of the exam on classroom teaching and learning. The aim of the studies is to inform the exam format to highlight whether it promotes beneficial washback (i.e. promoting educational activity and classroom behaviour which enhances learning, not simply enhances the ability to pass an exam) and any detrimental effects are reduced or ideally eliminated. Exam Boards need to

account for the complexity in assessment practices and to ensure that as far as possible they promote effective learning, and do not encourage detrimental classroom practices. To do so they need to collect appropriate data in the form of washback studies, despite the difficulties involved.

Results from washback studies should therefore routinely feed back into on-going test development or project management. Understanding the stakes for all involved is important to whether, and how, washback is manifested and the stakes cannot be assumed to be homogeneous. The complexity of the stakeholders (Receivers) also should ideally be explored to understand the range of responses to the exams which may pertain. If the Receivers and the stakes are not understood, the data from washback studies will be of only limited value. How well the message about a new educational programme, or simply a new exam, is conveyed to all responsible for delivery may affect its operation and monitoring this facet of the process also is beneficial and why help explain if plans for roll-out do not transpire as planned. Various other factors have the potential to add insight (such as other areas suggested in the Henrichsen model) but the practical considerations such as the amount of resources and time available for a washback study will dictate how much detail a washback study can go into. The nature of washback can easily be masked by superficial investigation.

10.4.3 Dual functionality : washback to impact

Another important implication of this study is the consideration of the tensions between, on the one hand, inevitable administrative systems (their management and execution being so much more complex when working on a national scale) and, on the other, pedagogical prerequisites for ensuring effective classrooms. The two sides need reconciling to provide usable data for each to ensure effective practice is taking

place and to promote further development of teaching and learning practices in ESOL, while meaningful monitoring is also undertaken. A one-size-fits-all approach while easing administration does not cater for the realities of ESOL students' lives, for example. Also careful examination of the stakes involved, in practice not simply in theory, needs to be taken into consideration and monitored to ensure teaching and learning practices do not become warped.

As regards, paying attention to accountability, indisputably, sound measures are needed to ensure ESOL courses are effective, that public money is well spent, but there is a need for measures that in the process do not cripple the possibilities for provision which ESOL departments wish to offer. For example, they need to be able to offer classes specifically to help student needing IELTS scores so they can rejoin professions held previously in their home countries if this is what local job-market demand dictates. In brief, alternative ways to assess effectiveness are needed, both in terms, firstly, of meeting the demands of the local job market and of the individual student's goals in life, and secondly also of accounting for the expenditure of public money, but in ways which encourages students taking appropriate qualification.

It is a waste of time and finances to collect poor data for such purposes i.e. exam scores of doubtful interpretation. Solutions may lie in considering the following: divorce student achievement from target-hitting mechanisms. This may be attempted by various methods. Dorn reminds us that we do not have to rely on statistics as the primary means to report on success and problems since a variety of means have been used in the past (Dorn 2007:18). Linn has also suggested that '[m]ultiple measures are needed for monitoring, and accountability systems' (2000: 9). However, in theory, multiple measures already existed in the Skills for Life system in that 'success' was being judged on exam results, but also factored in were satisfactory fulfilment and recording of ILPs and student retention and attendance. The problem is not

necessarily multiple measures, but having the right measures, ones which measure what needs measuring; scales are no good for measuring cloth and a ruler is equally no good for sugar.

One means to gain a measure of students' progress and ability would be to draw on continuous assessment performed by the teachers themselves. The situation reported in Canada and Australia (see Chapter 2) of assessment of ESOL students, using such methods caused difficulty regarding rigour which made them unsuitable for accountability purposes (Burrows 2001, Pierce & Stewart 1997). The data also suggests that this may not be the best method. The teachers in my study reported dissatisfaction with internal measures of student progress and ability used for high-stakes purposes. This related to how taking on the role of judge affects the teachers' role as guide and support to students. Also they expressed doubts in their own expertise in testing and assessment, affecting their ability to assume the role sufficiently well, and, as the consequences are of import, teachers were left feeling discomfort in that position.

Equally the problem of outcomes potentially being distorted by accountability pressures could arise again. It is feasible that pressure on teachers from departmental or management level to take whatever means necessary to achieve good success rates (because funding is secured on that basis) would place teachers in a difficult position. This may simply shift the way in which pressure was exerted but not remove it.

Precedents set by the compulsory education sector seem to often be followed in ESOL, literacy and numeracy, such as the introduction of a core curriculum and associated exams for checking progress. If this is the case, then we may expect the Skills for Life exams to be short-lived as SATs are being abolished Key Stage by Key

Stage (Marshall 2008). In their place at primary level are APPs (Assessment of Pupil Progress) which is more teacher-led than the centralised SATs. This move is one which various teachers in the current study would seem not to welcome, as already discussed.

One way whereby the assessment could have been externally derived and so internal pressures on teachers eliminated, while still maintaining national monitoring, would have been, for example, to undertake spot testing. In this method, only a percentage of students nationwide are examined and an overview extrapolated from their results.

Spot testing could have been combined with an improved inspection programme undertaken only by teams with considerable experience in ESOL. An inspection system would shift the emphasis onto the process of teaching and learning, away from a product (exam results) as the means to judge success. ILPs are supposed to measure student progress but it has been seen that they do not fulfil this function due to the problems encountered, as already discussed. As stated already, the teachers in my study felt they generally were not supported by inspection as it stood, did not generally benefit from professional development as a result of the inspections and felt they were being unfairly judged when the inspectors were not necessarily ESOL practitioners (current or former). An improved system which the teachers had faith in could help feed into development of all the other strands of the ESOL strategy.

In relation to this point, in one of the teacher interviews (T6) the teacher recounted an interaction in one of her classes. She referred to an incident witnessed in the class as part of the observed class data collection. Some well-known fairy tales were used as class material. T5 recounted:

*well you remember the Princess and the frog story [the material used in class
– where the princess kisses frogs in the search for her prince who is in the*

guise of a frog because he is under a curse] - well we did a bit more on the Wednesday and we extended it and we were talking about that and they all had the story but May who's Chinese said 'but we don't kiss the frog - we throw it against the wall and then he turns into the prince' - I said 'it's a bit bad - what if he wasn't a prince?' - 'well you have a dead frog' (T6-13:1261).

I see this tale as an analogy for two cultures which can take two very different approaches to the same situation. It is understandable that the administrative/management culture of institutions prioritise the accountability measures given our target-driven audit culture. Conversely teaching staff naturally prioritise learning and teaching. With increased research and improved communication between the two 'cultures' perhaps a means to satisfy all needs can be found, and thus leave more frogs intact.

10.5 Close

To close this thesis, I return to the quote from Galileo, which I set up as a central theme because I took it not as advice from Galileo but as a warning about 'forcing square pegs into round holes' in the pursuit of neat statistics to satisfy target demands which claim to prove success on specific measures. This fallacy goes beyond the world of Skills for Life and touches our target driven culture in the UK in a wider sense. More effort should be made in finding the right measure for the right task or the ensuing data will be, if not meaningless, then less useful than it might be.

11 LIMITATIONS AND POSSIBLE FUTURE RESEARCH

11.1 Limitations

In this section I will outline various areas of the study which on reflection may have been undertaken differently and thus proffered more rigorous results. These limitations concern consideration of the time frame of the study, data collection techniques used, as well as the role of the researcher.

11.1.1 Time frame

The data would have provided a stronger picture if the data collection had been more iterative in nature. Such an approach was part of the original plan but circumstances beyond the researcher's control intervened and there was a delay between data collection and analysis, and data collection cycles were reduced from the original plan. A further round of interviews to fill in gaps which came to light during the first rounds of analysis was not ultimately possible. Further clarification of points and expansion of interesting ideas would have enhanced the data greatly, but this was unfortunately not possible.

11.1.2 Diachronic approach

The research as undertaken in this study could be described as having taken a synchronic approach, a single snapshot of a particular situation. To better explore the concept of Burrows' (1998) concept of Adopters, Adapters and Resisters a diachronic approach would have been preferable. Such an approach would also have allowed a move into the Consequences aspect of the Henrichsen (1989) model.

11.1.3 Elicitation techniques

One consideration which arose during the interpretation process was a concern that the data received was too much the result of the prompts presented to participants in the form of the questions posed. While they were flexible, and were developed ‘on the hoof’ to exploit emergent information streams, the framework of the questions still shaped the information which the participants offered. Other topics which may have been equally enlightening might have been lost by taking this approach.

Alternative methods which may have proved more fruitful, in reflecting more accurately the participants’ own views and priorities may have been elicited by for example using more general topic prompt cards which the participants could have talked to or rejected as per their own interests and knowledge.

11.1.4 Student data

Another area which could have been better approached with an alternative methodology would have been alternative student data. The students initially interviewed in the exploratory study were not representative of students in the subsequent study, being on average of a much higher level of language ability. The student data from the main study could have been better exploited if it had been of better quality. This extra source of data would have reinforced triangulation and adding rigour to the study.

The difficulty faced was mainly due to the low level of language ability and the lack of resources in accessing data other than through the medium of English. Using interpreters would secure more in depth data but is costly and was beyond the scope of this study, but has been used successfully elsewhere (e.g. see Baynham et al 2007).

An alternative would be a questionnaire produced in the students' L1s, which would however also require extensive resources but would have been more plausible. Nevertheless, assuming literacy in L1 for the lowest levels in particular, is not advisable. Also the data would not have been like for like as questionnaires had not been used elsewhere in the study. Consideration of this at the earliest stages may have necessitated a revision of the data collection techniques to be used.

11.1.5 Sole researcher

It is the nature of PhD work that the researcher is the sole researcher on the study to prove one's own ability as an independent researcher. Having considered the process undertaken in this study I believe however that the analysis would have been vastly improved by the perspective(s) of a team or at least a couple of researchers. However far one tries to abide by a rigorous methodology, the nature of such analysis is that it is prone to subjectivity and bias.

Coding can be subjected to inter-rater reliability checks but in my experience this is rarely truly satisfactory. The depth of knowledge to fully understand the research theatre adequately to understand the categorisation of instances which is needed in order to second code effectively, is unlikely.

Another way in which possible researcher bias may have been manifested is, as already discussed above, that the set of questions inevitably framed the areas of information included in the data set. The study could have become broad and unmanageable if participants were allowed simply to talk on whatever topics related to my study theme, they wished but a method to focus the participants' attitudes and

opinions into a manageable framework could have countered possible researcher bias.

For a study bound by any restrictions, such as PhD, which needs to be undertaken by a sole researcher, a mixed methods approach may be preferable to counteract potential bias and add rigour which may be missing due to the lack of the insights gained from joint data coding and analysis by two or more researchers.

The Henrichsen model, while providing a useful initial framework for considering the complexity of a situation where an innovation, such as a new exam, has been introduced, also posed several problems. Its very complexity can lead to a labyrinthine intricacy, and insufficient data on all points to make adequate conclusions. For this reasons, combined with the time frame within which the study took place, which prevented the three phases of the model to be fully explored, only certain aspects of the model were utilised in the study.

11.2 Possible future research

The concept of differential stakes would benefit from more in-depth study. This would require more data, in greater depth, from the students themselves. As outlined above, this would require careful consideration of data collection techniques.

As mentioned above, as a possible limitation, a long range study to cover the Consequences section of Henrichsen model would allow more insight into the nature of washback stability. This could build on Cheng's concept of washback intensity (2004) and add the dimension of other factors which may affect how washback may change later over time, for example, as teacher familiarity with a new exam alters (see

Spratt and the fruits of uncertainty). Availability of further exam preparation materials may also affect a subsequent change in washback.

Another profitable area of washback study could be a comparison of a situation where, firstly, an exam has been introduced in order to effect change in the relevant educational system, such as provoking an increased focus on oral skills by introducing a speaking exam, and secondly, where an exam has been introduced with no clear intention of causing change. A detailed study of whether washback fundamentally differs in these two situations, of intended and unintended washback, may allow insights into how far, and in what ways, such interventions may be effective. Clearly each washback study is heavily context-bound by the individual nature of each exam, yet some patterns may emerge.

Madaus stated among the seven principles underlying test impact on teaching that ‘a high-stakes test transfers control over the curriculum to the agency which sets control of the exam’ (1988:97). Given the situation within which Skills for Life has operated, whereby the curriculum has remained central to the programme and the exams are offered by a range of providers, it would be of interest to follow up how far Madaus’ claim may apply.

The environment within which ESOL classes as described in this study now operate has altered owing to drastic changes in funding structure. Therefore a follow-up study to examine and compare the effects of the examinations currently used in the world of ESOL in the UK should be undertaken for all the reasons outlined above as to why washback studies should be undertaken.

REFERENCES

- Ade-Ojo, G. (2004). The New National ESOL Curriculum: Is it Determined by Convention? *RaPAL Journal*, 53, 21-27.
- Alderson, J. C. (1986). Innovation In Language Testing? In M. Portal (Ed.), *Innovation in Language Testing* (93-105), Windsor NFER/Nelson.
- Alderson, J. C. (2004) *Forward* In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (ix-xii). New Jersey: Lawrence Erlbaum Associates Inc.
- Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: the Interface Between Learning and Assessment*, London: Continuum.
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A Study of Washback. *Language Testing*, 13(3), 280 -297.
- Alderson, J. C., & Wall, D. (1993). Does Washback Exist? *Applied Linguistics*, 14(2), 115-129.
- Allwright, D. (1984). Why Don't Learners Learn What Teachers Teach? The Interaction Hypothesis. In D. M. Singleton & D. G. Little (Eds.), *Language Learning in Formal and Informal Contexts* (3-18). Dublin, Ireland: IRAAL.
- Amrein, A., & Berliner, D.C. (2002) High-Stakes Testing, Uncertainty And Student Learning, *Education Policy Analysis Archives* 10 (18)
- Andrews, S. (1994a). The Washback Effect of Examinations: Its Impact upon Curriculum Innovation in English Language Teaching. *Curriculum Forum*, 1, 44-58.
- Andrews, S. (1994b). Washback or Washout? The Relationship Between Examination Reform and Curriculum Innovation. In D. Nunan, R. Berry & V. Berry (Eds.), *Bringing about change in language education: proceedings of the International Language in Education Conference 1994*, (67-81). Hong Kong: University of Hong Kong.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting Washback: A Case Study. *System*, 30, 207-223.
- Bachman, L., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bagley, C. A. (1992). Back to the Future: Section 11 of the Government Act 1966, Local Authorities and Multicultural / Antiracist Education. London: NFER.
- Bailey, K. (1996). Working for Washback: A Review of the Washback Concept in Language Testing. *LanguageTesting*, 13(3), 241-256.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse Society*, 19, 273-306.

- Balliro, L., & Bickerton, B. (1999). The Workforce Investment Act and new multiyear RFP. *Bright Ideas*, 8(4), 20-21.
- Barton, D., Ivanič, R., & Hamilton, M. (2000). *Situated Literacies: Reading and Writing in Context*. London and New York: Routledge.
- Barton, D., & Pitt, K. (2003). Adult ESOL pedagogy: A Review of Research, an Annotated Bibliography and Recommendations for Future Research. Lancaster: Lancaster University for NRDC.
- Barton, D., & Papen, U. (2005). *Linking Literacy and Numeracy Programmes in Developing Countries*. Lancaster: NRDC.
- Barton, D., Appleby, Y., Hodge, R., Tusting, K., & Ivanič, R. (2006). Relating Adults' Lives and Learning: Participation and Engagement in Different Settings: National Research and Development Centre for Adult Literacy and Numeracy.
- Baynham, M. (1995). *Literacy Practices*. London and New York: Longman.
- Baynham, M., Roberts, C., Cooke, M., Simpson, J., Ananiadou, K., Callaghan, J., & Wallace, C. (2007). *Effective Teaching and Learning ESOL*. London: National Research and Development Centre for Adult Literacy and Numeracy (NRDC).
- Becker, H. (1998). *Tricks Of The Trade. How To Think About Your Research While You're Doing It*. London: University of Chicago Press
- Black, P. (1998). Learning, League Tables and National Assessment, Opportunity Lost or Hope deferred? *Oxford Review of Education*, 24(1), 57-68.
- Blewchamp, P. (1994). *Washback in the TOEFL Classrooms: An Exploratory Investigation into the Influence of the TOEFL Test on Teaching Content and Methodology*. Unpublished dissertation, University of Lancaster, Lancaster.
- Block, D. (2007). *Second Language Identities*. London: Continuum.
- Breen, M. (2001). The Social Context for Language Learning: a Neglected Situation? . In C. Candlin & N. Mercer (Eds.), *English Language Teaching in its Social Context: A Reader*. London: Routledge.
- Brindley, G. (1998). Outcomes-based Assessment and Reporting in Language Learning Programmes: A Review of the Issues. *Language Testing*, 15(1), 45-85.
- Brindley, G., & Hood, S. (1994). Curriculum Innovation in Adult ESL *The Second Language Curriculum in Action* (232-248).
- Brooks, G., Giles, K., Harman, J., Kendall, F., & Whittaker, S. (2000). *Assembling The Fragments: A Review of Research on Adult Basic Skills*, London: National Foundation for Educational Research/ DfES.
- Brown, J. D. (1989). Improving ESL Placement Tests Using Two Perspectives. *TESOL Quarterly*, 23(1), 65-83.
- Burgess, R. G. (Ed.). (1984). *The Research Process in Educational Settings: Ten Case Studies*. London: The Falmer Press.

Burns, A. (2009). Adult ESL in Australia: Policies, Programs and Practices. *Language Issues*, 20(1), 31-39.

Burrows, C. (1998). *Searching for Washback: An Investigation into the Impact on Teachers of the Implementation into the Adult Migrant English Program of the Certificate of Spoken and Written English*. Macquarie University, Sydney.

Burrows, C. (2001). Searching for Washback: The Impact of Assessment in the Certificate of Spoken and Written English. G. Brindley & C. Burrows (Eds.), *Studies in immigrant English language assessment* (Vol. 2). Sydney: National Centre for English Teaching and Research.

Burrows, C. (2004). Washback in Classroom-Based Assessment: A Study of the Washback Effect in the Australian Adult Migration Program. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (113-128). New Jersey: Lawrence Erlbaum Associates Inc.

Centre for Canadian Language Benchmark (1981) Canadian Language Benchmarks

Centre for Canadian Language Benchmark (2000) Canadian Language Benchmarks

Carlsen, C. (2007). Language Testing - a Matter of Ethics. In C. Carlsen & E. Moe (Eds.), *A Human Touch to Language Testing - A Collection of Essays in Honour of Reidun Oanaes Andersen on the Occasion of her Retirement June 2007*. (97-107) Oslo: Novus Press.

Chapman, D. W., & Snyder, C. W. (2000). Can High-stakes National Testing Improve Instruction: Reexamining Conventional Wisdom. *International Journal of Educational Development*, 20, 457-474.

Cheah, Y. M. (1998). The Examination Culture and its Impact on Literacy Innovations: The Case of Singapore. *Language and Education*, 12(3), 192- 209.

Chen, M.-L. (2002). *Taiwanese Junior High School English Teachers' Perceptions of the Washback Effect of the Basic Competence Test in English*. Ohio State University,, Columbus, Ohio. Retrieved 09/ 04/04, from:
<http://rave.ohiolink.edu/etdc/view?acc%5Fnum=osu1039116364>

Cheng, L. (1997). How Does Washback Influence Teaching? Implications for Hong Kong. *Language Education*, 11(1), 38-54.

Cheng, L. (1998). Impact of a Public English Examination Change on Students' Perception and Attitudes Toward Their English Learning. *Studies in Educational Evaluation*, 24(3), 279-301.

Cheng, L. (2004). The Washback Effect of a Public Examination Change on Teachers' Perceptions Toward Their Classroom Teaching. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (147 -190). New Jersey: Lawrence Erlbaum Associates Inc.

Cheng, L. (2005). *Changing Language Teaching Through Language Testing* (Vol. 21). Cambridge: Cambridge University Press.

Clavell, J. (1975). *Shogun*. London: Hodder & Stoughton.

- Cohen, A. D. (1994). *Assessing Language Ability in the Classroom* (2nd ed.). New York: Heinle and Heinle.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research Methods in Education* (5th ed.). London: Routledge Falmer.
- Cook, V. (1999). Going Beyond the Native Speaker in Language Teaching. *TESOL Quarterly*, 33(2), 185-195.
- Cooke, M., & Simpson, J. (2008). *ESOL: A Critical Guide*. Oxford: Oxford University Press.
- DfEE. (1999). *A Fresh Start: Improving Literacy and Numeracy*. London: Department for Education and Employment.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *A Dictionary Of Language Testing*. Cambridge: Cambridge University Press
- Davies, A. (2003). Three Heresies of Language Testing Research. *Language Testing*, 20 (4) 355-368
- Davies, P. (2005). Study of Impact of the Skills for Life Learning Infrastructure on Learners - Interim Report on the Qualitative Strand: NRDC.
- DCSF (2007). *The Languages Ladder - Steps to Success*. Nottingham: DCSF Publishing
- de Lotbiniere, M. (2002, Thursday 12 December). Concern over 'Britishness' Test Grows, *Guardian Weekly*.
- Denzin, N. K. (1970). *Sociological Methods: A Source Book*. Chicago IL: Aldine.
- Denzin, N. K. (1978). *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: McGraw Hill.
- Denzin, N. K., & Lincoln, Y. S. (2003). Introduction - The Discipline and Practice of Qualitative Research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Landscape of Qualitative Research - Theories and Issues*. (1-45) Thousand Oaks: Sage.
- Derrick, J., Ecclestone, K., & Merrifield, J. (2007). A Balancing Act? The English and Welsh Model of Assessment in Adult Basic Education. In P. Campbel (Ed.), *Measures of Success - Assessment and Accountability in Adult Basic Education* (287-326). Canada: Grass Roots Press.
- De Vincenzi, F. (1995) Language Tests And ESL Teaching. Examining Standardized Test Content: Some Advice For Teachers, *TESOL Quarterly*, 29 (1), 180-187
- Dey, I. (1993). *Qualitative Data Analysis: A User-friendly Guide for Social Scientists*. London: Routledge.
- Dey, I. (2004). Grounded Theory. In C. Seale, G. Gobo, J. Gubrium & D. Silverman (Eds.), *Qualitative Research Practice* (80-94). Thousand Oaks Sage.
- DfEE. (2000). *Breaking the Language Barriers: The Report of the Working Group on ESOL*.

- DfEE. (2001). *Adult ESOL Core Curriculum*. London: Department for Education and Skills.
- Donague, H. (2003). An Instrument to Elicit Teacher's Beliefs and Assumptions. *ELT Journal*, 57(4), 344-351.
- Dorn, S. (2007). *Accountability Frankenstein - Understanding and Taming the Monster*, Charlotte, NC: Information Age Publishing
- Edelmann, R., & Hardwick, S. (1986). Test Anxiety, Past Performance and Coping Strategies *Personal Individual Differences*, 7 (2), 255-257.
- Engle, R., W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, 11(1, February), 19-23.
- Ferman, I. (2004). The Washback of an EFL National Oral Matriculation Test to Teaching and Learning. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (191 -210). New Jersey: Lawrence Erlbaum Associates Inc.
- Flick, U. (1998). *An Introduction to Qualitative research : Theory Method and Application*. London: Sage.
- Fowler, Z. (2003). *Politically constructing adult literacy: A case study of the Skills for Life strategy for improving adult literacy in England 1997 - 2002*. PhD, University of London.
- Frame, P. (2001). *Report of Strand E ESOL Qualifications*. London: Institute for Education
- Frederiksen, N. (1984). The Real Test Bias: Influences on Testing and Teaching. *American Psychologist*, 39, 193-202.
- Frederiksen, J., & Collins, A. (1989). A System Approach to Educational Testing. *Educational Researcher*, 18, 27-32.
- Fulcher, G. (1997). An English Language Placement Test: Issues in Reliability and Validity. *Language Testing*, 14(2), 113-139.
- Fullan, M. (1982). *The Meaning of Educational Change* (2nd ed.). New York: Teacher's College Press.
- Galassi, J. P., Frierson, H. T., & Siegel, R. G. (1984). Cognitions, test anxiety, and test performance: a closer look. *Journal of Consultative Clinical Psychology*, 52, 319-370.
- Garner, B. (1999). Nationwide Accountability: the National Reporting System. *Focus on Basics*, 3 (B).
- Geertz, C. (1973). Thick Description: Toward an Interpretive Theory of Culture. In C. Geertz (Ed.), *The Interpretation of Cultures: Selected Essays* (3-30). New York: Basic Books.
- Gipps, C. (1994). *Beyond Testing – Towards a Theory of Educational Assessment*. London: The Falmer Press.

- Glaser, B., & Strauss, A. L. (1967). *The Discovery of Grounded Theory*. Chicago Aldane.
- Glover, P. (2006). *Examination Influence on How Teachers Teach: A Study of Teacher Talk*. PhD Thesis, Lancaster University, Lancaster.
- Gobo, G. (2002). Sampling, Representativeness and Generalizability. In C. Seale, G. Gobo, J. Gubrium & D. Silverman (Eds.), *Qualitative Research Practice* (405-426). Thousand Oaks Sage.
- Gold, R. L. (1958). Roles in Sociological Field Observations. *Social Forces*, 36, 217-223.
- Goldschneider, J.M., & DeKeyser. R.M. (2001) Explaining the "Natural Order" of L2 Morpheme Acquisition in English: a Meta-Analysis of Multiple Determinants *Language Learning* 51 (1) 1-50.
- Gorard, S., & Taylor, C. (2004). What is 'Triangulation'? *Building Research Capacity*(7).
- Gordon, S. P., & Reese, M. (1997). High -Stakes Testing: Worth the Price? *Journal Of School Leadership*, 7, 345 -368.
- Gosa, C. M. C. (2004). *Investigating Washback: A Case Study Using Student Diaries*. PhD Thesis, Lancaster University, Lancaster.
- Green, A. (2003). *Test Impact and English for Academic Purposes: A Comparative Study in Backwash between IELTS Preparation and University Pre-sessional Courses*. PhD Thesis, University of Surrey, Roehampton.
- Green, A. (2007). *IELTS Washback in Context - Preparation for Academic Writing in Higher Education* (Vol. 25). Cambridge: Cambridge Univeristy Press.
- Green, A., & Weir, C. (2004). Can Placement Tests Inform Instructional Decisions? . *Language Testing*, October(21), 467-494.
- Greenslade, R. (2005). *Seeking Scapegoats. The Coverage of Asylum Seekers in the UK*. London: Institute for Public Policy Research
- Haladyna, T. A., Nolen, S. B., & Haas, N. S. (1991). Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution. *Educational Researcher*, 20(5), 20-25.
- Hamilton, M., & Hillier, Y. (2009). ESOL Policy and Change. *Language Issues*, 20(1), 4-15.
- Hamilton, M., & Merrifield, J. (2000). Adult Learning and Literacy in the United Kingdom. In Comings., J, Garner. B., & Smith. C., (Eds.), *NCSALL Annual Review of Adult Learning and Literacy*. San Francisco: Jossey-Bass.
- Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles and Practice*. London: Routledge.
- Hamp-Lyons, L. (1998). Ethical Test Preparation Practice: The Case of the TOEFL. *TESOL Quarterly*, 32(2), 329-337.

- Hawkey, R. (2006). *Impact Theory and Practice - Studies of the IELTS test and Progetto Lingue 2000*. Studies in Language Testing 24. Cambridge: Cambridge University Press.
- Hayes, B., & Read, J. (2004). IELTS Test Preparation in New Zealand: Preparing Students for the IELTS Academic Module. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (97-112). New Jersey: Lawrence Erlbaum Associates Inc.
- Henrichsen, L. E. (1989). *Diffusion of Innovations in English Language Teaching: The ELEC Effort in Japan, 1956- 1968*. New York: Greenwood Press.
- Herman, J. L., & Golan, S. (1993). The Effects of Standardized Testing on Teaching and Schools. *Educational Measurement: Issues and Practices*, 12, 20-25.
- Hodge, R., Pitt, K., & Barton, D. (2004). *"This is Not Enough for One's Life": Perceptions of Living and Learning English in Blackburn by Students Seeking Asylum and Refugee Status*. Literacy Research Centre, Lancaster University. Lancaster.
- Hughes, A. (1988). Introducing a Needs-based Test of English Language Proficiency into an English Medium University in Turkey. In A. Hughes (Ed.), *Testing English for University Study: ELT Documents Vol. 127* (134-153). London: Modern English Publication in association with the British Council.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (1993). *Backwash and TOEFL 2000*. Unpublished report for Educational Testing Service.
- Hughes, D. (2004). ESOL Certificates Worth Building On, *The Guardian*, Feb 19th.
- Huhta, A., Kalaja, P., & Pitkänen-Huhta, A. (2006). Discursive Construction of a High-Stakes Test: The Many Faces of a Test-Taker. *Language Testing*, 23(3), 326-350.
- Ivanič, R. (1998). *Writing and Identity: The Discoursal Construction of Identity in Academic Writing*. Amsterdam: John Benjamins.
- Ivanič, R., & Tseng, M.L. (2005). Understanding the Relationships Between Learning and Teaching: an Analysis of the Contribution of Applied Linguistics: National Research and Development Centre for Adult Literacy and Numeracy.
- Ivanič, R., Appleby, Y., Hodge, R., Tusting, K., & Barton, D. (2006) *Linking Learning and Everyday Life: a Social Perspective on Adult Language, Literacy and Numeracy Classes*. London: NRDC
- James, M. (2000). Measured Lives: The Rise of Assessment as the Engine of Change in English Schools. *Curriculum Journal*, 11, 343-364.
- Jones, J. (2004). Opinion, *The Guardian*, 2nd March.
- Kellaghan, T., Madaus, G., & Airasian, P. (1982). *The Effects of Standardized Testing*. London: Kluwer Nijhoff Publishing.

- Kellaghan, T. & Greaney, V. (1992) *Using Examinations to Improve Education - A Study in Fourteen African Countries*. Washington, DC: The World Bank
- Kelle, U. (2004). Computer-Assisted Qualitative Data Analysis. In C. Seale, G. Gobo, J. Gubrium & D. Silverman (Eds.), *Qualitative Research Practice* (443-460). Thousand Oaks: Sage.
- Kennedy, C. (1987) Innovation for a Change: Teacher Development and Innovation, *ELT Journal* 41 (3) 163-170.
- Khan, R. (2002). Impact Assessment Of An In-Service Teacher Training Programme In Bangladesh: A Research Report. *Journal of NELTA*, 7(1-2).
- Khanna, A. L., Verma, M. K., Agnihotri, R. K., & Sinha, S. K. (1998). *Adult ESOL Learners in Britain: a Cross Cultural Study*. Clevedon: Multilingual Matters.
- King, N., & Horrocks, C. (2010). *Interviews in Qualitative Research*. London: Sage.
- Kirkland, K., & Hollandworth, J. G. (1980). Effective test taking: skills-acquisition versus anxiety-reduction techniques. *Journal of Consultative Clinical Psychology*, 48, 431-434.
- Koretz, D., Linn, R., Dunbar, S., & Shepard, L. (1991). *The Effects of High Stakes Testing on Achievement: Preliminary Findings About Generalization Across Tests*. Paper presented at the AERA/ NCME, Chicago.
- KPMG. (2005). KPMG Review of English for Speakers of Other Languages (ESOL), Retrieved 03/01/09, from: www.lsc.gov.uk
- Kunnan, A. J. (2009). Testing for Citizenship: The US Naturalization Test. *Language Testing Quarterly*, 6(1), 89-97.
- Kvale, S., & Brinkmann, S. (2009). *InterViews - Learning the Craft of Qualitative Research Interviewing* (2nd ed.). Thousand Oaks: Sage.
- Lakatos, I. (1982). The Methodology of Scientific Research Programmes *Philosophical Papers 1*. Cambridge: CUP.
- Lam, H. P. (1994). Methodology Washback - An Insider's View. In D. Nunan, R. Berry & V. Berry (Eds.), *Bringing About Change in Language Education: Proceedings of the International Language in Education Conference 1994* (83-102). Hong Kong: University of Hong Kong.
- Latham, H. (1877). *On the Action of Examinations Considered as a Means of Selection*. Cambridge: Deighton Bell & Company.
- Lavender, P., Derrick, J., & Brooks, B. (2004). *Testing, Testing, 123 - Assessment in Adult Literacy, Language and Numeracy - A NIACE Policy Discussion Paper*. Leicester: NIACE.
- Lavender, P. (2007). *Preface*. In Ward, J. (2007). *ESOL: The Context for the UK Today*. (v-ix). Leicester: NIACE.
- Life in the United Kingdom Advisory group. (2003). *The New and the Old*. Croydon: Home Office.

- Lightbown, P. (2003) SLA research in the classroom/SLA research for the classroom, *Language Learning Journal*, 28 (Winter), 4-13.
- Lii-shih, Y.-h. E. (1991). Teaching to the Test: A Case Study of Taiwan. *Studies in English Literature and linguistics*, May, 51-63.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic Enquiry*. Beverley Hills, CA: Sage.
- Linn, R.L. (2000). Assessments and Accountability, *Educational Researcher*, 29 (2), 4-16.
- Locastro, V. (1990). The English in Japanese university entrance examinations: a sociocultural analysis. *World Englishes*, 9, 343-354.
- London, N. (1997). A National Strategy for System-wide Curriculum Improvement in Trinidad and Tobago. In Chapman., D.W., Mahlick, L.O., Smulders, A. (Eds.) *From Planning to Action: Government Initiatives for Improving School Level Practice*. Paris: International Institute for Educational Planning.
- Loofbourrow, P. T. (1992). *Composition in the Context of the CAP: A Case Study of the Interplay Between Assessment and School Life*: Center for the Study of Writing: University of California Berkley, Carnegie Mellon University.
- LSC (Producer). (2006). Policy Requirements For Planning: Managing the Balance and Mix of Provision (2006/7). Retrieved 07/07/08, from: <http://readingroom.lsc.gov.uk/lsc/2006/funding/streams/nat-policyrequirementsforplanningmanagingbalanceandmixofprovision-ps-2006.pdf>
- Madaus, G. (1988). The Influence of Testing on the Curriculum. In L. Tanner (Ed.), *The politics of reforming school administration* (83-121). London: The Falmer Press.
- Mallows, D. (2009). New Research into ESOL Skills for Life Qualifications. *NATECLA News*, 89(Summer), 13.
- Markee, N. (1993). The Diffusion of Innovation in Language Teaching. *Annual Review of Applied Linguistics*, 13, 229-243.
- Markee, N. (1997). *Managing Curricular Innovation*. Cambridge: Cambridge University Press
- Marshall, B. (2008) The End Of Testing And Future Possibilities: An Examination Of The Demise Of Sats In England And Possible Alternative Assessments. *English Teaching: Practice and Critique*. Volume 7, Number 3 (December). 65-73.
- Marshall, C., & Rossman, G. B. (2006). *Designing Qualitative Research*. Thousand Oaks: Sage.
- Mason, J. (2002). *Qualitative Researching*. London: Sage.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- McNamara, T. (2001). Language Assessment as Social Practice: Challenges for Research. *Language Testing*, 18(4), 333-349.

- McNamara, T., & Roever, C. (2006). *Language Testing: The Social Dimension*. Malden, MA: Blackwell Publishing.
- McNess, E., Broadfoot, P., & Osborn, M. (2003) Is the Effective Compromising the Affective? *British Educational Research Journal*, 29(2), 243-257.
- Mehrens, W. A., & Kaminski, J. (1989). Methods for Improving Standardized Test Scores: Fruitful, Fruitless or Fraudulent? *Educational Measurement: issues and practices*, 8(1), 14-22.
- Merrifield, J. (2006). From the Tyranny Of Targets To a System That Learns. *Language Issues*, 18(2), 2 -11.
- Messick, S. (1989). Validity. In L. R (Ed.), *Educational Measurement* (3rd ed.). Washington: Macmillan.
- 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(3), 241-256.
- Miles, M., & Huberman, M. (1994). *Qualitative Data Analysis* (2nd ed.). Beverly Hills: Sage.
- Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation can improve American Schools*. New York: The Free Press.
- Morrison, K. R. B. (1993). *Planning and Accomplishing School- centred Evaluation*. Norfolk: Peter Francis Publishers.
- Morrow, K. (1986). The Evaluation of Tests of Communicative Performance. In M. Portal (Ed.), *Innovations in Language Testing*. London: NFER/Nelson.
- Moser, C. (1999). *Improving Numeracy and Literacy: A Fresh Start*. London: DfEE.
- Moser, C. (2000). *Breaking the Language Barrier*. London: DfEE.
- New Zealand Immigration Service. (1995). *New Zealand's 'Targeted' Immigration Policies: Summary of October 1995 Policy Changes*. Wellington: New Zealand Immigration Service.
- NIACE. (2004). *Citizenship Materials for ESOL Learners*. London: DfES and NIACE.
- NIACE. (2006). *More Than a Language...*, Leceister: NIACE
- Nicholls, A. (1983). *Managing Educational Innovations*. Hemel Hempstead: George,Allen & Unwin Ltd.
- Norton, B. (2000). *Identity and Language Learning*. London: Longman.
- Papen, U. (2005). *Adult Literacy As Social Practice - More Than Skills*. Oxon: Routledge.
- Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A Developmental Perspective on Standardized Testing. *Educational Researcher*, June-July, 12-19.

- Pearson, A. (2005). ALI inspection and the CFI, IATEFL Conference Presentation, Cardiff.
- Pearson, I. (1988). Tests as Levers of Change (or 'Putting First Things First'). In D. Chamberlain & R. Baumgartner (Eds.), *ESP in the classroom: practice and evaluation* (98-107). London: Modern English Publications in association with the British Council.
- Pica, T. (2006). Adult Acquisition Of English As A Second Language Under Different Conditions Of Exposure. *Language Learning*, 33(4), 465-497.
- Pierce, B. N., & Stewart, G. (1997). The Development of the Canadian Language Benchmarks Assessment. *TESL Canada*, 14(2), 17-31.
- Pitt, K. (2005). *Debates in ESOL Teaching and Learning*. London: Routledge.
- Poland, B. D. (2003). Transcription Quality. In J. A. Holstein & J. Bubrium (Eds.), *Inside Interviewing: New Lenses, New Concepts*. Thousand Oaks, CA: Sage.
- Popham, W. J. (1987). Two-plus Decades of Educational Objectives. *International Journal of Educational Research*, 11(1), 31-41.
- Popham, W. J., Cruse, K. L., Rankin, S. C., Sadifer, P. D., & Williams, P. L. (1985). Measurement-driven Instruction: We're on the Road. *Phi Delta Kappan*, May, 628-634.
- Qi, L. (2004). Has a High-Stakes Test Produced the Intended Changes? In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (171-190). New Jersey: Lawrence Erlbaum Associates Inc.
- Qi, L. (2005). Stakeholders' Conflicting Aims Undermine the Washback Function of a High-Stakes Test. *Language Testing*, 22(2), 142-173.
- Rees, S., & Sunderland, H. (Eds.). (1995). *Working with the Criteria for the assessment of English Language Skills - examples of current practice*. London: LLLU.
- Richards, J., Platt, J., & Platt, H. (1992). *Dictionary of Language Teaching & Applied Linguistics* (2nd ed.). Harlow: Longman.
- Robb, T. N., & Jay, E. (1999). A Study of the Effect of Direct Test Preparation on the TOEIC Scores of Japanese University Students. *TESL-EJ*, 2(4), 1-22.
- Roberts, C., Baynham, M., Shrubshall, P., Barton, D., Chopra, P., Cooke, M., & Whitfield, S. (2004). English for Speakers of Other Languages (ESOL) - Case Studies of Provision, Learner' Needs and Resources: National Research and Development Centre for Adult Literacy and Numeracy.
- Roberts, M. (2002). TOEFL Preparation: What are our Korean Students Doing and Why? *The Korea TESOL Journal*, 5(1), 81-106.
- Rogers, E.M. (1995). *The Diffusion of Innovations* (4th ed.). New York: The Free Press.

- Rogers, E.M. & Shoemaker, F. (1971) *Communication of Innovations: A Cross Cultural Approach 2*, New York: Free Press
- Rosenberg, S. (2007). *A Critical History of ESOL in the UK, 1870- 2006*. Leicester: NIACE.
- Rosenberg, S. (2009). 'Why Don't They Speak English?', ' Why Can't They Speak English?', 'Why Won't They Speak English?' and 'Whose English is it Anyway?'. *Language Issues*, 20(1), 40-52.
- Rumsfeld, D. (2002). There Are Known Knowns; There Are Things We Know We Know.
- Salmi, J., & Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy*, 19(2), 31-68.
- Saville, N. (2010). Developing a Model for Investigating the Impact of Language Assessment. *Cambridge ESOL: Research Notes*(42, November), 2-8.
- Say, R. (1997). *Stress in Teaching: the Prevalence and Causes of Stress Amongst ESOL Teachers in a College of Further Education in the UK*. M.Ed, Sheffield.
- Schellekens, P. (2001). English Language As A Barrier to Employment, Training and Education. Norwich: DfEE.
- Schellekens, P. (2004). *EFL and E(S)OL: Common Learning Needs, Common Teaching Goals?* Paper presented at the IATEFL Conference, Brighton.
- Schellekens, P. (2007). *The Oxford ESOL Handbook*. Oxford: Oxford University Press.
- Schellekens, P. (2009). Cause and Effect: the Impact of the Skills for Life Strategy on Language Assessment. In L. Taylor & C. Weir (Eds.), *Language Testing Matters: Investigating the Wider Social and Educational Impacts of Assessment - Proceedings of the ALTE Cambridge Conference April 2008*, Studies in Language Testing 31. Cambridge: Cambridge University Press.
- Schellekens, P. (2011). *Teaching and Testing the Language Skills of First and Second Language Speakers*, A report produced for Cambridge ESOL
- Schwarz, N., & Sudman, S. (1996). *Answering Questions* (1st ed.). San Francisco: Jossey-Bass Inc.
- Scientific Software Development. (2000). Atlas-ti (Version 4.2): Scientific Software Development.
- Scott, C. (2005). *Washback in the UK Primary Context with EAL Learners: Exploratory Case Studies*. PhD, University of Bristol.
- Seale, C. (1999). *The Quality of Qualitative Research*. London: Sage.
- Seale, C., Gobo, G., Gubrium, J. F., & Silverman, D. (Eds.). (2004). *Qualitative Research Practice*. Thousand Oaks: Sage.
- Sharp, P. (1998). The Beginning of GNVQs: an Analysis of Key Determining Events and Factors *Juornal of Education and Work*, 11(3), 293-311.

- Shih, C.-M. (2007). A New Washback Model of Students' Learning. *The Canadian Modern Language Review/ La Revue Canadienne des Langues Vivantes*, 64(1), 135-162.
- Shohamy, E. (1992). Beyond Proficiency Testing: A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning. *The Modern Language Journal*, 76(4), 513-521.
- Shohamy, E. (2001). *The Power of Tests. A Critical Perspective on the Uses of Language Tests*. Harlow: Pearson Education Ltd.
- Shohamy, E. (2007). Language Tests as Language Policy Tools. *Assessment in Education*, 14 (1), 117-130.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test Impact Revisited: Washback Effect over Time? *Language Testing*, 13(3), 298-317.
- Silverman, D. (1993). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*. London: Sage.
- Simmons, C. (1988). Contrasting Attitudes to Education England and Japan. *Educational Research*, 30(2), 146 -152.
- Simpson, J. (2006). Differing Expectations in the Assessment of the Speaking Skills of ESOL Learners. *Linguistics and Education*, 17 (1), 40-55.
- Skehan, P. (1989). *Individual Differences in Second Language Learning*. London: Edward Arnold.
- Smith, M. (1991). Put to the Test: the Effects of External Resting on Teachers. *Educational Researcher*, 20(5), 8-11.
- Somerset, A. (1983). *Examination Reform : The Kenya Experience*. Washington DC: World Bank.
- Spada, N., & Frohlich, M. (1995). *The Communicative Orientation of Language Teaching Observation Scheme (COLT)*. Sydney: The National Centre for English Language Teaching and Research (NCELTR). Macquarie University, Australia.
- Spolsky, B. (1994). The Examination of Classroom Backwash Cycle: Some Historical Cases. In D. Nunan, V. Berry & R. Berry (Eds.), *Bringing About Change in Language Education in Hong Kong (55-66)*: University of Hong Kong Department of Curriculum studies.
- Spolsky, B. (1995). *Measured Words*. Oxford: OUP.
- Spratt, M. (2005), Washback And The Classroom: The Implications For Teaching And Learning Of Studies Of Washback From Exams, *Language Teaching Research*, 9(1), 5-19
- Stecher, B., Chun, T., & Barron, S. (2004). The Effects of Assessment-Driven Reform on the Teaching of Writing in Washington State. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods (53-72)*. New Jersey: Lawrence Erlbaum Associates Inc.

- Stites, R. (1999). A User's Guide to Standards-based Educational Reform: from Theory to Practice. *Focus on Basics*, 3 (C).
- Stobart, G. (2003). Editorial. The Impact of Assessment: Intended and Unintended Consequences. *Assessment in Education*, 10(2), i-ii.
- Stober, J. (2004). Dimensions Of Test Anxiety: Relations To Ways Of Coping With Pre-Exam Anxiety And Uncertainty. *Anxiety, Stress, and Coping*, 17(3), 213-226.
- Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research*, London: Sage
- Street, B. V. (1993). *Cross Cultural Approaches to Literacy*. Cambridge: Cambridge University Press.
- Swain, M. (1985). Large-scale Communicative Testing. In Y. P. Lee, C. Y. Y. Fok, R. Lord & G. Low (Eds.), *New Directions in Language Testing*. Hong Kong: Pergamon Press.
- The Basic Skills Agency. (2005). *Police ESOL: The Basic Skills Agency*.
- Trlin, A., North, N., Pernice, R., & Henderson, A. (1998). *The New Settlers Programme: Encounters, Responses, Policies: An Introduction To A Research Project*. Paper presented at the 3rd National Conference, 1997, New Zealand Federation of Ethnic Councils, New Zealand Federation of Ethnic Councils (inc), Christchurch
- Tsagari, C. (2006). *Investigating the Washback Effects of First Certificate in English Exams in the Greek Context*. PhD thesis, University of Lancaster.
- Tsagari, D. (2009). *The Complexity of Test Washback*. Frankfurt, Germany: Peter Lang.
- Tusting, K. (2005). Language & Power in Communities of Practice. In D. Barton & K. Tusting (Eds.), *Beyond Communities of Practice – Language Power & Social Context*. Cambridge: Cambridge University Press.
- Van Duzer, C. (2002). Q&A : Issues in Accountability and Assessment for Adult ESL Instruction Retrieved 04/03/04, from:
<http://www.cal.org/ncl/digests/accountQA.htm>
- Van Duzer, C., & Berdan, R. (2000). Perspectives on Assessment in Adult ESOL Instruction. In Comings., J, Garner. B., & Smith. C., (Eds.), *NCSALL The Annual review of Adult Learning and Literacy*. San Francisco: Jossey-Bass.
- Wall, D. (1996). Introducing new tests into traditional systems: insights from general education and from innovation theory. *Language Testing*, 13(3), 334 -354.
- Wall, D. (1999). *The impact of high-stakes examination on classroom teaching – a case study using insights from testing and innovation theory*. PhD, Lancaster University.
- Wall, D. (2005). *The impact of high-stakes examination on classroom teaching – a case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.

- Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan Impact Study. *Language Testing*, 10(1).
- Wall, D., & Horak, T. (2006). The Impact Of Changes On The TOEFL Examination On Teaching And Learning In Central And Eastern Europe - Phase 1: The Baseline Study. *ETS Monograph Series*. Princeton, NJ: ETS.
- Wall, D., & Horak, T. (2007). The Role of Baseline Studies. *Assessment in Education*, 36(March), 56-64.
- Wall, D., & Horak, T. (2008). Coping With Change: Phase 2. *ETS iBT-Research Series*. Princeton: NJ, ETS.
- Wall, D., & Horak, T. (2011). The Impact of Changes in the TOEFL Examination on Teaching in a Sample of Countries in Europe - Phase 3, The Role of the Coursebook, and Phase 4, Describing Change. *ETS iBT-Research Series*. Princeton: NJ, ETS.
- Ward, J. (2007). *ESOL: The Context for the UK Today*. Leicester: NIACE.
- Warschauer, S.F. (1995). Exam-Based Reform Stifles Student Writing in the UK. *Educational Leadership*. 52 -(6) (March).26-29
- Waters, A. (2001). *In-Service Teacher: Learning in ELT Projects and Programme: An Integrated Approach*. PhD, Lancaster University.
- Watanabe, Y. (1996). Does Grammar Translation Come from Entrance Examination? Preliminary Findings from Classroom-based Research. *Language Testing*, 13(3), 318-333.
- Watanabe, Y. (2004). Methodology in Washback. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing - Research Contexts and Methods* (19-36). New Jersey: Lawrence Erlbaum Associates Inc.
- Watts, N. (2001, 16 -18 November 2001). *Pathways to Learning in ESOL: Learner and Provider views*. Paper presented at the AMEP Conference, Adelaide.
- Weir, C., & Milanovic, M. (2003). *Continuity and Innovation : revising the Cambridge Proficiency in English Examination 1913- 2002* (Vol. 15). Cambridge: Cambridge University Press.
- Weir, M. (2005). ILPs: a Trivialisation of Teaching and Learning? *Reflect*(2), 29-30.
- Wells, A. (1996). ESOL - About Time for a Fresh Start? *Adults Learning*, 8(1), 8-9.
- Wesdorp, H. (1983). Backwash Effects of Multiple-choice Language Tests: Myth or Reality? In J. van Weeren (Ed.), *Practice and Problems in Language Testing* 5. Arnhem.
- Williams, D. (2004). When the Classroom is a Haven: Teaching Asylum Seekers Has Proved a Sharp Learning Curve for FE Tutors, *The Guardian*, Feb 3rd.
- Woods, P. (1988). Pulling out of a Project: Twelve Tips for Project Planners. *ELT Journal*, 42 (3), 196-201.
- Wrigley, H. S. (1992). Q&A: Learner Assessment in Adult ESL Literacy.

Zeng, K., & Le Tendre, G. (1988). Adolescent Suicide and Academic Competition in East Asia. *Comparative Education Review*, 42(4), 513-528.

APPENDICES

Appendix 1: Alderson & Wall's (1993: 120-121) 15 washback hypotheses

1. A test will influence teaching
2. A test will influence learning
3. A test will influence what teachers teach
4. A test will influence how teachers teach
5. A test will influence what learners learn
6. A test will influence how learners learn
7. A test will influence the order and sequence of teaching
8. A test will influence the rate and sequence of learning
9. A test will influence the degree and depth of teaching
10. A test will influence the degree and depth of learning
11. A test will influence attitudes to the content, method etc. of teaching and learning
12. Tests that have important consequences will have washback
13. Tests that do not have important consequences will have no washback
14. Tests will have washback on all learners and teachers
15. Tests will have washback effects for some learners, but not for others teachers

Appendix 2: Observation Sheet

OBSERVATION SHEET – Class Profile (pt 1)

Date:

Teacher:

Class name:

Ability Level:

Time start:

Time end:

Materials used:

Student Profile

	M/F	EFL/ESL	nationality	notes
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

15				
16				
17				
18				
19				
20				

OBSERVATION SHEET - observable influence of assessment on class (Pt 2)

Key: I = internal assessment E=-external assessment

(Codes to use: M= mentioned only, U = used / practised in class)

Mention of	✓	examples
internal student 'tracking' (e.g ILPs)		
internally set tests /exams: which?		
external exams – which?		

Possible features of assessments used/ mentioned in class	✓	I / E	Example (to show how teaching or learning is being influenced)
references to assessment(s) (T or S)			
test/exam taking strategies			
features/type of grammar tested			
features/type of voc tested			
which skills are tested			
sub-skills tested			
topics to be expected			
text types to be expected			
item types to be expected			
exercises/tasks which mirror exam/test			
sample papers /past papers (from Boards)			
practice /mock exams			
exam-oriented coursebook/materials			
feedback to Ss on their progress			
class time spent on discussing exams available			

learning goals set with assessment requirements in mind			
discussion of connections between employment/citizenship and assessment(s)			
Other:			

OBSERVATION SHEET - Chronological description (Pt 3)

Time	Focus	Activity	Notes

OBSERVATION SHEET – Overall impressions of the class – notes (Pt 4)

- Did Ss show awareness of assessment?
- Did students show any negative /positive attitudes to exams (including anxiety)?
- Did the teacher show any negative /positive attitudes to exams (including anxiety)?
- Was the influence of any assessment measures/ exams / tests noticed?
- If so, did it influence content or methodology or both? How?

(From interview)

Number of students registered for this class?

What types of assessment has this class experienced so far?

Placement/ diagnostic/ progress/ achievement/ proficiency
All students? If not, why not?

Is this explicitly an exam preparation class?

Appendix 3: Student Interview Schedule

Date:

Location:

	Question	
1.	What exam did you sit today?	
2.	What did you think of it?	
3.	How confident do you feel about it? Do you think you did OK?	
4.	Did you know what to expect before you went into the exam? / Did you feel prepared before you went into the exam?	
5.	Was there anything in the exam which surprised you?	
6.	Did your teacher do preparation with you in class before your exam?	
7.	If so: What did he/ she do with you?	
8.	Did you find that preparation helpful today in the exam?	
9.	If so: In what way?	
10.	Why did you do the exam?	
11.	Did you have a choice about taking the exam or not?	
12.	Did you have a choice about which exam to take?	
13.	How important is it to you to get a good mark? Why?	
14.	Is it important for you to get a certificate? Or is the mark enough?	

Student names:

Appendix 4: Site Profiles

	Site 1:	Site 2:	Site 3:
Type	FE College	FE /HE and Sixth Form College	FE College
Number of ESOL students (approx)	120-150	800	150
Number of FT Teachers	6	11	1
Number of PT Teachers (= hourly paid)	6	10	10*
ESOL provision at more than one site?	✓	✓	✓
Informant labels:			
Director of Studies	DOS1	DOS2	DOS3
Teacher	T1 T2	T3 T4	T5 T6

* includes 2 members of staff on half-time

Source: interview data from DoS at each site

Appendix 5: Interview Schedule

TEACHER INTERVIEW SCHEDULE

<p>How the question relate to the framework.</p> <p>Key:</p> <p>1) Background info: Profile</p> <p>2) RQ= Res. Question :- What were practices? What was their link to S4L? AP Any WB? Linked to S4L? WB What factors was driving the WB? (Factors within the system - cf Henrichsen- : Stakes , Communication, The Receivers i.e. Teachers,) FAC -C, FAC-R, FAC-S</p> <p>* only relevant if interviewee was teaching prior to S4L</p>
--

	Questions	
	Background – Teacher Profile	
1	How many years of teaching experience do you have?	Profile
2	Has this been completely within a TEFL/ TESL context?	Profile
3	How long have you worked at this college?	Profile
4	What are your teaching qualifications?	Profile
5	Have you had training specifically in teaching ESOL? (omit if necessary- see q. above)	Profile
6	Are you an examiner with any exam boards? Have you ever been?	Profile
	College Profile	
7	What type of students are at this college? (e.g FLers/ ESLers: Refugees, Migrants)	Profile
8	Do most teachers come from an ESL or EFL background? (might need clarifying)	Profile
	Communication	
9	How far do these bodies influence your ESOL teaching or attitude to ESOL: NATECLA, FENTO, NATESOL, OFSTED/ALI , ABBSU, NRDC ? (which effect does each have?)	FAC –C
10	Do you get the chance to discuss assessment issues in staff meetings? Or informally in the staff room?	FAC –C
11	What are the main issues which come up? (adapt for DOS)	FAC –C
12	Have these changed from 6 months ago?	FAC –C
13	From 12 months ago?	FAC –C
14	Are you aware of any regional differences in assessment practices?	FAC –C
15	Please outline the key features of recent changes in ESOL? e.g. If you had to explain the situation to new teacher what would you say	FAC-S
16	How well informed do you feel about changes in ESOL?	FAC –C
17	How do you find out about these developments?	FAC –C
18	Are you satisfied with this level of information and how you get the information?	FAC –C
19	If not, what would you prefer to see happening?	FAC –C

	Assessment Practices - External (i.e. externally moderated)	
20	Which external exams do students at this college sit? Any?	AP
21	Who decides whether a student will sit an exam?	FAC-S
22	Who chooses which exam?	FAC-S
23	What is this decision based on?	FAC-S
24	How do exam results relate to college targets – as far as you understand?	FAC-S
25	As far as you are aware, what is the range of exams that an ESOL student might take (not necessarily just those available at this college)? (NB if new to ESOL think about reducing face threatening situation)	AP
26	What do you think of them?	FAC-R
27	What are their strengths and weaknesses?	FAC-R
28	Are you aware of the new ESOL exams due are now becoming available?	AP / FAC-C
29	What can you tell me about them (e.g. which Boards are offering new exams, how they are different)?	AP
30	What do you think about them?	FAC-R
31	What is the value of the Literacy Exams (L1 & L2) for ESOL students?	FAC-R
32	How important are the students' exam results - to you personally? Why?	FAC-S
33	How important are the students' exam results - to the student? Why?	FAC-S
34	How important are the students' exam results - to the ESOL dept. as a whole? Why?	FAC-S
35	How important are the students' exam results - to the College? Why?	FAC-S
36	What effect does external assessment have on your ESOL classes? (which in particular?)	WB
37	Does it affect how you teach? How? Why?	WB
38	Does it affect what you teach? How? Why?	WB
39	Does it affect what students want to learn?	WB
40	Does it affect the teachers' workload?	WB
	Assessment Practices: Internal assessment (college moderated procedures)	
41	Could you explain to me any mechanisms for on-going (continuous) , classroom-based assessment of ESOL students? - i.e. undertaken by the teachers at the College	AP
42	Do you do ILPs? Do you consider ILPs as a form of assessment?	AP
43	Do you do mid-term reviews? Do you consider them as a form of assessment?	AP
44	How do the students react to ILPs, mid-term reviews etc?	FAC-R
45	How do the teachers react to ILPs, mid-term reviews etc?	FAC-R
46	Does internal assessment – if any happens – affect teacher's workload?	AP
47	Do you do all the following at this college ? Placement, diagnostic, progress, achievement testing or assessment? Other?	AP
48	Who does this? Always class the teacher?	AP
49	What weight do these internal assessments carry? (why are they done?)	FAC-S

50	How does the current internal assessment situation differ to the situation before S4L?	FAC-C
51	And has it changed again since?	FAC-C
52	What effect does internal assessment have on your ESOL classes?	WB
53	Does it affect how you teach? How? Why?	WB
54	Does it affect what you teach? How? Why?	WB
55	What is your attitude towards these procedures/ this system? (e.g. is it helpful / formative or more of a hindrance/ a necessary evil)	FAC-R
	Views on Assessment	
56	Do you think students' learning can be assessed? Should it be assessed?	FAC-R
57	Has your attitude to this changed since S4L reforms?	FAC-R
	Citizenship issues	
58	What is your understanding of the current position on citizenship and language qualifications?	FAC-S
59	What is your opinion regarding the inclusion of English language assessment among the criteria for qualification for citizenship?	FAC-S
	Effect of S4L	
60	Is the rationale behind S4L clear? What is your understanding of it?	FAC-R
61	Are the changes in assessment practices resulting from S4L compatible with its aims?	FAC-R
62	Were changes to the assessment system easily manageable for teachers?	FAC-R
63	How has S4L influenced your teaching- if at all? Is it still influencing it now?	WB
64	If so how? And why?	WB
66	What would you do, if it were within your powers, to improve the current ESOL assessment practices – internal or external - for the ESOL students? / the ESOL tutors? / the College?	FAC-R
67	What are the negative features of the current ESOL assessment practices? (if not covered before)	FAC-R
68	What are the positive features of the current ESOL assessment practices?	FAC-R
69	Is there anything else concerning assessment issues which you think might be of relevance or interest that I haven't raised already?	

Appendix 6: Letter of Consent

Tania Horak's PhD Study (Lancaster University): Assessment issues in ESOL in the UK – the effect of the Skills for Life Strategy on assessment practices

Consent form

I understand that

- the interview data will only be used for the above PhD study and related academic articles
- I will be made anonymous in data viewed by anyone beyond the researcher and her supervisor
- a summary of the interview will be made available for checking before being used in this study

I agree that data from my interview can be used in the above study.

Name:

Signed:

Date:

Position:

Institution:

Appendix 7: The Quantity of Data

Informants		Interview (length – in words)
Site 1	T1	13,009
	T2	18,358
	DoS1	15,851
Site 2	T3	23,915
	T4	22,683
	DoS2	9,216 *
Site 3	T5	16,120
	T6	29,858
	DoS3	28,384
	TOTAL	182,773

Appendix 8: Chronology of Data Collection

Date	Site	Observation which teacher: student level	Interviews			
			DoS	T1	T2	Ss
6/12/04 (& 15/12/04)*	Site 2		✓			
6/12/04	Site 2	T3: E1				
6/12/04	Site 2	T3: L2				
12/1/05	Site 2	T3: L1				
19/01/05	Site 2			✓		
9/2/05 (& 23/2/05)*	Site 3		✓			
15/2/05	Site 2	T4: Mixed**				
15/2/05 (& 8/3/05)*	Site 2				✓	
23/2/05	Site 2	T4: E2				
2/3/05	Site 1	T1: L2				
7/3/05	Site 1	T2: E2				
23/3/05	Site 1			✓		
22/4/06	Site 1	T1: L1 (FCE)				
26/4/05	Site 1	T2: E2/E3				
29/4/05	Site 1		✓			
3/5/05	Site 1				✓	
13/5/05	Site 3	T5: E2				
14/6/05	Site 3	T6: E1				
16/6/05	Site 3	T6: E3/E2				
16/6/05	Site 3				✓	
17/5/05	Site 3			✓		
26/5/05	Site 3					✓ (x1)
27/5/05	Site 1					✓ (x1)
8/6/05	Site 3					✓ (x5)
13/6/05	Site 1					✓ (x4)
15/6/05	Site 1					✓ (x3)
20/6/05	Site 3					✓ (x2)
7/7/05	Site 2					✓ (x4)
11/7/05	Site 2					✓ (x8)

* These interviews were conducted in 2 parts due to time constraints on the part of the informants.

** Employability programme students – mixed levels

Appendix 9: Observation Details

Informants	Teacher	Level of class observed	Date
Site 1	T1	L1 (FCE)	22/4/06
Site 1	T1	L2	2/3/05
Site 1	T2	E2	7/3/05
Site 1	T2	E2/E3	26/4/05
Site 2	T3	E1	6/12/04
Site 2	T3	L2	6/12/05
Site 2	T3	L1	12/1/05
Site 2	T4	Other*	15/2/05
Site 2	T4	E2	23/2/05
Site 3	T5	E3/E2	16/6/05
Site 3	T5	E1	14/6/05
Site 3	T6	E2	13/5/05

* ESOL for employability programme class – various levels

Appendix 10: Interview details

Informants	# student(s) in group	Interview (length in mins)	Date
Site 1			
T1	-	1 hr 28 m	23/3/05
T2	-	2 hr 05 m	3/5/05
DOS 1	-	1 hr 31 m	29/4/05
group 1	2	21 m	27/5/05
group 2	3	21 m	13/6/05
group 3	2	10 m	13/6/05
group 4	4	11 m	13/6/05
group 5	1	14 m	13/6/05
group 6	1	11 m	15/6/05
group 7	5	12 m	15/6/05
group 8	4	14m	15/6/05
Site 2			
T3*	-	2 hr 37 m	19/01/05
T4*	-	2 hr 29 m	15/2/05 & 8/3/05
DOS 2	-	1 hr 20 m	6/12/05 & 15/12/05
Ss group 1	3	19 m	7/7/05
Ss group 2	2	14 m	7/7/05
Ss group 3	2	16 m	7/7/05
Ss group 4	2	14 m	7/7/05
Ss group 5	3	14 m	11/7/05
Ss group 6	2	10 m	11/7/05
Ss group 7	2	11 m	11/7/05
Ss group 8	2	17 m	11/7/05
Ss group 9	2	15 m	11/7/05
Ss group 10	2	12 m	11/7/05
Ss group 11	2	12 m	11/7/05
Ss group 12	2	11 m	11/7/05
Site 3			
T5	-	1 hr 34 m	17/5/05
T6	-	2 hr 39 m	16/6/05
DOS3*	-	3 hr 23 m	9/2/05 & 23/2/05
Ss group 1	4	16 m	26/5/05
Ss group 2	2	9 m	8/6/05
Ss group 3	2	22 m	8/6/05
Ss group 4	2	15 m	8/6/05
Ss group 5	2	11 m	8/6/05
Ss group 6	2	10 m	8/6/05
Ss group 7	3	19 m	20/6/05
Ss group 8	3	15 m	20/6/05

* interviews were conducted in two sessions

Appendix 11: Nationalities, gender and language ability level of student informants

Informants	Student nationality	male	female	level
Site 1				
Ss group 1.1	Turkish, Czech	0	2	L1
Ss group 1.2	Polish, Lithuanian, Chinese	3	0	E3
Ss group 1.3	Italian, Romanian	0	2	E3
Ss group 1.4	Polish	2	2	E2
Ss group 1.5	Afghani	1	0	E1
Ss group 1.6	Polish	1	0	E2
Ss group 1.7	Ethiopian, Pakistani, Indian	0	5	E2
Ss group 1.8	Pakistani, Iraqi, Indian	5	0	E2
Site 2				
Ss group 2.1	Afghani	1	0	E3
Ss group 2.2	Pakistani, Iranian	0	2	E3
Ss group 2.3	Kenyan, Sri Lankan	0	2	E3
Ss group 2.4	Pakistani	1	2	L1
Ss group 2.5	Pakistani	1	1	L1
Ss group 2.6	Lebanese, Pakistani	2	0	L1
Ss group 2.7	Polish, Pakistani	1	1	L1
Ss group 2.8	Pakistani	0	3	E2
Ss group 2.9	Pakistani	2	0	E2
Ss group 2.10	Somali, Iranian	1	1	E2
Ss group 2.11	Lebanese, Pakistani	1	1	E2
Ss group 2.12	Equatorial New Guinea, Pakistani	0	2	L2/L1
Site 3				
Ss group 3.1	Italian, Polish	4	0	E1
Ss group 3.2	Swiss, Chinese	2	0	FCE*
Ss group 3.3	Polish, Slovak	1	1	FCE
Ss group 3.4	Polish	1	1	FCE
Ss group 3.5	Italian, Polish	2	0	FCE
Ss group 3.6	Brazilian, Slovak	0	2	FCE
Ss group 3.7	Thai, Afghani, Brazilian	1	2	E2
Ss group 3.8	Singaporean, Malaysian, Spanish	1	2	E3

* at this site of the college (which has 2 sites) students were grouped according to the Cambridge exam they were aiming to sit, not by Skills For Life level.

Appendix 12: Transcription Conventions

capital letter for names only – not for beginning sentences

question marks for genuine questions only (not questions tags)

only punctuation used are question mark and dash

- for boundary of meaning units

-- for a pause

[extra info – usually nonverbal action] e.g [laughter] [further chat re admin issues]

contractions used as close to as that heard as possible (can't, don't, they've etc)

ID

T: Teacher

D: DOS

I: Interviewer

(?) unclear / best guess

erm - transcribed to show hesitancy

mmm – agreement noise

BUT - not indicated:

Overlaps

Emphasis

Length of pause

Appendix 13: Sample Coding (using Atlas –ti)

The screenshot displays the S4L ESOL software interface. The main window shows a transcript of a conversation between two individuals, I and T, with line numbers from 0653 to 0688. The transcript text is as follows:

0653
0654 I: so there you're talking about end of module tests that you've given maybe a bit early
0655 to see if you need to keep doing it longer
0656
0657 T: because we give the whole level E2 test at the start of the year and then give
0658 it their ILPs - their
0659
0660 I: that shapes the other bits of assessment
0661
0662 T: yeah but then obviously you've got on-going assessment issues because everything
0663 looks great on paper but when you start to access that and then you get suddenly one of
0664 these students who was totally useless at comparatives the penny suddenly dropped and
0665 it's no longer an issue and that can you know the next time you come to look at your
0666 paperwork there it's no longer an issue it wasn't really a problem it was a hidden
0667 problem they hadn't quite grasped it - you've got that sort of on-going issue as well that
0668 you're forever changing your paperwork - I did all mine when I did the diagnostic because
0669 we get an extra week now at the start of the year for the diagnostic because the college is
0670 funding it and I did all mine in colours so I could see any fault lines and then a few
0671 weeks later those fault lines had disappeared and new ones had appeared
0672
0673 I: well that's the thing yes it's very different to the teaching of some other subjects -
0674 it is something that's constantly moving - but you found ways of coping with that?
0675
0676 T: well yeah because the thing is as well I like to have formula for everything and er
0677 even though that formula it's the same formula the results are changing all the time I
0678 don't think the others do it but I use my modular tests as a diagnostic tool so all the time
0679 instead of having to think what to put in their ILP you've got this student sitting in front
0680 of you and you're thinking what is he weak at what is he strong but I don't need to - I
0681 mean this happened to the whole class were weak on comparatives so I said right ILP
0682 - it doesn't need to be in the ILP it's a class target so we had a whole lesson re doing
0683 comparatives until they got it and then we could move on so I'm very well -I'm [?] I'm
0684 schizophrenic - I'm very big on having this diagnostic in place because that tells
0685 me but also you've got to have it all in here as well
0686
0687 I: having it all in your head -juggling which
0688

On the right side of the transcript, there are coding annotations in red text:

- internal assessment~ (bracketed next to lines 0662-0671)
- diagnostic test~ (bracketed next to lines 0667-0671)
- diagnostic test~
ILPs~ (bracketed next to lines 0678-0685)

The software interface includes a menu bar (File, Documents, Quotations, Codes, Memos, Networks, Views, Extras, Help), a toolbar, and a status bar at the bottom showing the current file path and system time (12:19).

Appendix 14: Complete List of Codes

Concept driven: P = Profile data
S = relating to the interview/ observation schedule
Data driven: E – emergent codes

Code	type	Code	type
'double accounting'	E	exam admin	E
'institutional validity'	E	exam preparation	E
ABBSU	E	exam procedures	E
accountability	E	exam technique	E
accreditation	E	exam tips	E
achievement asst	E	exams 'contract'	E
adapting to S4L	E	external assessment	S
admin tension	E	FCE	E
assessment journey	S	FE environment	E
asylum seeker issues	E	feedback	E
awareness of S4L exams	S	fees	E
CAE	E	FENTO	E
Cambridge ESOL	E	funding	E
CELS	E	grammar	E
change management	E	green files	E
changes to Literacy t'ing	E	habit	E
choice of board	S	HW	E
citizenship test attitudes	S	IELTS	E
citizenship testing	S	ILPs	S
City & Guilds	E	info from exam boards	E
college certificate	E	innovation	E
college profile	P	inspection	E
college resources	E	internal assessment	S
communication	P	lack of guidance	E
consequences of S4L	S	lack of standardisation	E
currency	S	lack of understanding	E
curriculum	S	learning hours	E
dept issues	E	info from exam boards	E
diagnostic test	E	lesson summary	S (obs)
DoS attitude to assessment	S	lesson topic	S (obs)
DoS aware's of ESOL t'ing	S	listening	E
DoS eval'n of S4L exams	S	literacy mould	E
DoS profile	P	LSC	E
Edexcel	E	main v. satellite	E
editing	E	materials	S
ESB	E	methodology	S
ESOL dept. structure	E	mid-term review	E
ESOL dept. profile	P	lesson summary	S (obs)
ESOL misfit	E	lesson topic	S (obs)
ESOL provision	S	mobility	E
ESOL teaching skills	E	mock exams	E
ESOL v. Literacy	E	NATECLA	E
ESOL/ESL/EFL	S	NATESOL	E
evaluation of S4L	S	National Test	E
evaluation of S4L exams	S	NIACE	E

NRDC	E	trust	E
OCNW	E	Ts adapting to change	E
OFSTED	E	tutorials	E
paperwork	E	uncertainty	E
past paper	E	vocab	E
pedagogy v finances	E	washback	S
placement tests	E	workload	E
plans	E		
portfolio	E		
positive effects	E		
pre-Entry Ss	E		
prediction - changes in ESOL	S		
pressures on teachers	E		
problems of PT Staff	E		
professionalism	E		
progress tests	E		
pron	E		
prov'g what Ss learnt	E		
Q type	E		
QualAim	E		
raised standards	E		
reading	E		
regionality	E		
relationship with Man't	S		
restricted provision	E		
role of ESOL classes	E		
scoring	E		
soundbites	E		
speaking	E		
spelling	E		
Ss analy'g lang needs	E		
stakes	S		
T att to asst.	S		
T att to ESOL exams	S		
T att to ESOL t'g	S		
T attitude to exams	S		
T exam expertise	S		
T profile	P		
T qualifications	P		
T role	E		
T training	S		
T view of Sfl	S		
targets	E		
time limits	E		
Trinity	E		

Appendix 15: Example ILP (from DfES ESOL Tutor Manual 2001)

Individual learning plan

Achievements (look back at your targets and short-term goals)

How do you feel about your progress? Use this space to record other achievements and/or ways in which you have used new skills outside the course.

What do you need to work on next (include results of further diagnostic assessment)?

Targets	By date	Specific skills curriculum reference(s)	Achieved

Signatures

Learner: _____ Tutor: _____

Start date: _____ Date of next review: _____

Individual learning plan

Target 1

Date	Activity	Resources	Complete

Target 2

Date	Activity	Resources	Complete

Target 3

Date	Activity	Resources	Complete

Target 4

Date	Activity	Resources	Complete

End of course review

To be completed at the end of the course

What do you feel you have gained from this course?
Accreditation:
Achievement of targets and goals: Evidence:
Next steps:

Signed

Learner: _____ Tutor: _____

Date: _____

Appendix 16: Information on the exam boards offering Skills for Life qualifications

- Cambridge ESOL – formerly known as UCLES – - operates internationally
- City and Guilds – taken over Pitman’s exams - – one of oldest exam boards: founded in 1878 – S4L all levels all 3 modes – externally set and internally marked (source: www.cityandguilds.com) operates internationally - See British Council Website
- Edexcel – operates in internationally - largest awarding body in UK - owned by Pearson company- former ‘incarnations’ London Examinations and BTEC - (source: www.edexcel.org.uk/subjects/a-z/sfl)
- ESB – English Speaking Board – focuses on oral assessments – founded 1953 - operates in internationally - (source: www.esbuk.org/esb_intro.htm)
- OCNW - Open College of the North West– part of Open College check - offers– established 1975 partnership between universities (who provide validation) and colleges of FE and Sixth Form colleges (who offer courses) aiming to provide access to HE. Nationwide although originating in North West. (source: www.ocnw.com/OCNW/tabid/68/Default.aspx)
- Trinity – offers exams in XX - operates internationally

Appendix 17: Acronyms used in this study

Acronyms	Full terms/ names
ABE	Adult Basic Education (UK)
ABSSU	Adult Basic Skills Strategy Unit (UK) – previously SfLSU
ALI	Adult Learning Inspection (UK)
ALLN	Adult Literacy, Language and Numeracy
ASLPR	Australian Second Language Proficiency ratings
BETS	Basic Employability Training Scheme (UK)
AMEP	Adult Migrant Education (now renamed English) Programme (Australia)
CAE	Certificate of Advanced English (Cambridge ESOL) (UK)
CELS	Certificate in English Language Skills (UK)
CSWE	Certificate in Spoken and Written English (Australia)
CCLB	Center for Canadian Language Benchmarks
CLB	Canadian Language Benchmarks
CBLA	Canadian Language Benchmarks Assessments
DfES	Department for Education and Science (UK) – now known as DfEE
DoS	Director of studies
EALTA	European Association of Language Testing and Assessment
ED	Department of Education (USA)
EFL	English as a Foreign Language
ESL	English as a Second Language
ESOL	English for Speakers of Other Languages
ETS	Educational Testing Service
EU	European Union
FE	Further Education
FENTO	Further Education National Training Organisation (UK)
FCE	First Certificate of English (Cambridge ESOL) (UK)
FE	Further Education (UK)
GCSE	General Certificate of secondary Education (UK)
HE	Higher Education (UK)
IELTS	International English Language Testing System
ILP	Independent Learning Plan
KET	Key English Test (Cambridge ESOL) (UK)
LSC	Learning and Skills Council (UK)
NALDIC	National Association for Language Development in the Curriculum (UK)
NATECLA	National Association of Teachers for English and Community Languages for Adults (UK)
NATESOL	National Association of Teachers of English to Speaker Other Languages
NEC	Neighbourhood English Classes (UK)
NIACE	National Institute of Adult Continuing Education
NRDC	National Research and Development Council (UK)
NRS	National Reporting System (USA)
NVQ	National Vocational Qualification (UK)
OCNW	Open College of the North West (UK) now known as Ascentis
OFSTED	Office for Standards in Education (UK)
PET	Primary English Test (Cambridge ESOL) (UK)
PGCE	Post Graduate Certificate of Education (UK teacher qualification)
QCA	Qualifications and Curriculum Authority
SAT	Standard Assessment Tests (UK)
TOEFL	Test of English as Foreign Language (ETS) (USA)
TOEIC	Test of English for International Communication (ETS) (USA)
UCLES	University of Cambridge Language Exams Syndicate (UK) now Cambridge ESOL