



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Lin, Shwu-Wen

Title:

The requirement of English language proficiency for graduation in Taiwanese universities : its impact on non-English majors and their English curriculum

General rights

The copyright of this thesis rests with the author, unless otherwise identified in the body of the thesis, and no quotation from it or information derived from it may be published without proper acknowledgement. It is permitted to use and duplicate this work only for personal and non-commercial research, study or criticism/review. You must obtain prior written consent from the author for any other use. It is not permitted to supply the whole or part of this thesis to any other person or to post the same on any website or other online location without the prior written consent of the author.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to it having been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you believe is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact: open-access@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline of the nature of the complaint

On receipt of your message the Open Access team will immediately investigate your claim, make an initial judgement of the validity of the claim, and withdraw the item in question from public view.



**The requirement of English language proficiency for
graduation in Taiwanese universities: Its impact on
non-English majors and their English curriculum**

Shwu-Wen Lin

**A dissertation submitted to the University of Bristol in accordance
with the requirements of the degree of Doctor of Philosophy in
Applied Linguistics in the Faculty of Social Sciences and Law**

October 2010

ABSTRACT

In response to government's policy to improve English language proficiency of university students, some Taiwanese universities now require their students to reach a certain level of proficiency, as evidenced through scores obtained from formal language tests, before they are allowed to graduate. Various English language proficiency tests are dictated in the requirements of different universities. This study examined the impact of such requirements on the English for Academic Purposes curriculum for non-English majors, on the English classroom teaching and learning and on the students themselves. Data were collected from two universities, one with the graduation requirement, and the other without, through classroom observations of sixteen English lessons and interviews with seven teachers and nineteen students. In addition, the learning power of a selective sample of 454 students (including the interviewees) from these two universities was assessed, using the Effective Lifelong Learning Inventory.

The research findings indicate that the washback of the locally-developed English proficiency test, the General English Proficiency Test (GEPT), on teaching and learning was evident, although limited. The influence of other language tests was minimal. Furthermore, the GEPT washback seemed to have resulted less from the implementation of the requirement *per se*, than the importance of the test as viewed by the general public. The implementation of the requirement seemed to have reinforced the influence of the GEPT in universities. The students' learning power can offer some insights into understanding their varied perceptions of the graduation requirement and its impact. Students with stronger learning power, and in particular those with a higher level of resilience to challenges and difficulties were more likely to prioritise their English learning and test taking over simply fulfilling the requirement. Those with weaker learning power and lower resilience experienced a higher level of anxiety in taking English language tests to meet the graduation requirement.

This present study is significant in two ways. Conceptually, it took into consideration the social agenda of a language test, a crucial factor in understanding the impact of the graduation requirement and the test. Methodologically, the exploration of students' learning power offered opportunities to better understand their varied perceptions of test impact.

DEDICATION

This dissertation is dedicated to those who always have faith in me even when I have lost faith in myself.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and the help of the following individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

I would first like to express my deepest gratitude to my supervisor, Dr. Guoxing Yu, for guiding me through the whole process of my PhD research and helping me overcome each and every hurdle that came along the way. I am especially thankful for his patience, attentiveness, the encouragement he has given me, and the large amount of time he has spent on my dissertation despite his hectic schedules. I look up to him as his enthusiasm in language testing research, his pursuit of perfection and his continuous efforts are what make him a successful non-native researcher in the field.

I would also like to thank Professor Pauline Rea-Dickins for her guidance through the first three years of my PhD. I really admire her sharp wits, her encyclopedia-like knowledge of related research and her ability to pinpoint the weakness of my chapters. I am especially thankful that she has constantly reminded me the value of my research and the doubts I had throughout the process were driven away.

The conduct of this research would not have been possible without the generous help from my friend, Dr. Chen, for helping me gain access to a case university, and from my former colleagues and all the teachers who voluntarily participated in my study. In addition, the dissertation would not have shaped up so well without the help of my proofreader, Dr. Sharpling.

Thanks to CREOLE staff and my fellow schoolmates of GSoE for the inspiring discussions we had of our research, our frustrations and our achievements. My PhD years would have been such a lonely path without them. Special thanks goes to Dr. Chia-Yen Huang, who has always been there when I need him. I am extremely grateful for his complete understanding of my swaying moods and my selfishness from time to time. Life would not have been the same without him.

Most important of all, I would like to thank my dearest parents, who love me unconditionally, who have always believed I am the smart one and as long as I work hard enough, I can always achieve my goals.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the regulations of the University of Bristol. The work is original, except where indicated by specific reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

SIGNED:

DATE:

TABLE OF CONTENTS

ABSTRACT.....	i
DEDICATION	iii
ACKNOWLEDGEMENTS.....	v
AUTHOR'S DECLARATION.....	vii
TABLE OF CONTENTS	ix
LIST OF TABLES	xv
LIST OF FIGURES.....	xvii
CHAPTER 1 Introduction	1
1.1 Overview	1
1.2 Background	1
1.3 Research context.....	5
1.3.1 The implementation of the graduation requirement for English proficiency.....	6
1.3.2 The General English Proficiency Test	10
1.4 Personal experiences.....	14
1.5 Overall research aims	16
1.6 Structure of the dissertation	17
CHAPTER 2 Literature Review	20
2.1 Introduction	20
2.2 Theoretical framework	20
2.2.1 Definition of key terms: washback and impact	20
2.2.2 Language tests and related policies	23
2.2.3 Direction of Washback: Washback as positive, negative or neutral.....	28
2.2.4 Washback Variability	31
2.2.5 Strength of washback: washback intensity.....	34

2.2.6 Mechanisms of washback and its conceptualisations.....	38
2.3 Empirical studies of washback and impact	47
2.3.1 Washback and impact on teachers and teaching	48
2.3.1.1 Teaching Contents.....	48
2.3.1.2 Teaching materials	51
2.3.1.3 Teaching Methods	56
2.3.1.4 Implications for this study	60
2.3.2 Washback and impact on learners and learning	61
2.3.2.1 Empirical studies on washback on learners and learning	66
2.3.2.2 Implications for this study	72
2.4 Studies related to the educational and research context in this study	74
2.5 Approaches to Learning and Assessment: The Use of Effective Lifelong Learning Inventory (ELLI)	78
2.5.1 Approaches to Learning and Assessment.....	78
2.5.2. Learning how to learn and learning power.....	80
2.5.3 The seven dimensions of learning power	82
2.5.4 ELLI and impact of language test	87
2.6 Summary	89
CHAPTER 3 Methodology	92
3.1 Introduction	92
3.2 Philosophical position.....	92
3.2.1 Postmodernist influences in this study	93
3.2.1.1 The concept of washback.....	94
3.2.1.2 The nature of washback	95
3.2.1.3 Power relations.....	96
3.2.2 The positivist elements	97
3.2.3 Philosophical underpinnings as manifested in the study	98
3.3 Research questions	98
3.4 Pilot study.....	99

3.5 Research context and research approach.....	102
3.5.1 Ethnographic case study approach.....	102
3.5.2 The two case universities.....	105
3.6 Data collection.....	107
3.6.1 Identification of teacher participants.....	108
3.6.2 Classroom observation.....	112
3.6.3. Interviews	119
3.6.4 ELLI questionnaire for students.....	123
3.7 Methods of analysis.....	124
3.7.1 Transcription.....	126
3.7.2 Translation.....	127
3.7.3 Data analysis for observation and teacher interview data	128
3.7.4 Analysis of interview data	132
3.7.5 Statistical analysis of the ELLI questionnaire data	133
3.7.6 Integrated analysis for ELLI and student interview data.....	135
3.8 Trustworthiness of this research	135
3.8.1 Respondent validation	136
3.8.2 Triangulation and Thick description	137
3.9 Ethical issues	138
3.9.1 Informed consent	138
3.9.2 Confidentiality and anonymity	139
3.9.3 Participants' rights.....	140
3.9.4 Researcher's responsibilities.....	140
3.10 Summary	141
CHAPTER 4 Impact on the EAP Curriculum for Non-English Majors.....	143
4.1 Introduction	143

4.2 Presentation of classroom observations in timeline	145
4.3 Impact on teaching	155
4.3.1 Teaching materials.....	155
4.3.2 Teaching content	172
4.3.3 Explicit reference to the test	182
4.3.4 Testing and assessment.....	184
4.3.5 Summary	190
4.4 Teachers' perceptions of the impact of the graduation requirement.....	191
4.5 Factors that shape the impact and washback on EAP curriculum for non-English majors	195
4.6 Summary	198
CHAPTER 5 Impact on the Non-English Majors.....	201
5.1 Introduction	201
5.2 The student interviewees	204
5.3 Learner's knowledge of the graduation requirement.....	204
5.4 Washback of GEPT and other tests	206
5.5 Perceived impact of the graduation requirement on non-English majors.....	212
5.5.1 Teachers' Perspective	212
5.5.2 The learner's perspective: attitudes towards the graduation requirement.....	217
5.5.3 Learning power: the comparison between the two cases.....	224
5.5.4 The learning profiles of the interviewees	229
5.5.5 Learner's perspective.....	230
5.5.5.1 Students who perceived little impact	231
5.5.5.2 Students who perceived stress and anxiety towards the requirement	240
5.6.6.3 Summary	254
5.6 Summary of the chapter.....	255
CHAPTER 6 Discussion of the Findings.....	259

6.1 Introduction	259
6.2 Summary of major findings	259
6.2.1 What are the effects of the English requirement for graduation on the on the EAP curriculum for non-English majors in Taiwanese universities? (RQ1)	260
6.2.2 What are the effects of the English requirement for graduation on the non-English majors in Taiwanese universities? (RQ2)	262
6.2.3 To what extent students' learning power explain their perceptions of the washback on learners and their learning? (RQ3).....	265
6.3 The Power of the GEPT.....	267
6.3.1 The social impact of the GEPT	268
6.3.2 The reinforcement of the graduation requirement.....	272
6.3.3 GEPT and the language education policy	274
6.3.4 Summary	276
6.4 Factors that mediate washback on learners and learning	278
6.4.1 The role of publishers	279
6.4.2 Intrinsic factors: Personal perceptions of test, test stakes and test difficulty	280
6.4.3 Intrinsic factors: Learning power.....	283
6.4.4 Washback on learner motivation.....	286
6.4.5 Summary	289
6.5 Summary of the chapter.....	289
CHAPTER 7 Conclusion and Implications	291
7.1 Overview	291
7.2 Strengths and limitations of the study.....	292
7.2.1 Research strengths	292
7.2.2 Research limitations.....	294
7.3 Directions for further research.....	297

7.3.1 Expanding the length of the study and involving more teacher participants for classroom observation	297
7.3.2 Expanding the size of the study in including other related stakeholders and more case universities	298
7.3.3 Tracking learning in longitudinal studies and expanding ELLI along with extended qualitative accounts of learners.....	299
7.4 Implications for stakeholders.....	301
7.4.1 Implications for policy makers	301
7.4.2 Implications for stakeholders in charge of determining the graduation requirement and related curriculum change.....	303
7.4.3 Implications for teachers and learners.....	305
7.5 Reflection and concluding remarks	306
REFERENCES.....	309
Appendix A GEPT Test Construct	322
Appendix B Effective Lifelong Learning Inventory: validated English version and ELLI items corresponding to dimensions of learning power	325
Appendix C Field note for Amy’s EIST II class, lesson on 07.05.2008	330
Appendix D ELLI traditional Chinese version	333
Appendix E Research ethics form for ethical considerations in this study	337
Appendix F Consent form for teachers.....	340
Appendix G Consent form for the students	342
Appendix H Sample mid-term and final test papers that modelled on the item types of the GEPT elementary level.....	343

LIST OF TABLES

Table 2.1 Overview of studies exploring washback and impact on the learners and their learning.....	63
Table 3.1 Planned and achieved observation data collected	117
Table 3.2 Overview of Research Design and Methods of Analysis	125
Table 3.3 ELLI items corresponding to the seven dimensions	134
Table 4.1 University A: teaching materials	156
Table 4.2 University B: teaching materials.....	157
Table 4.3 Comparison between University A test papers and GEPT elementary level.....	187
Table 5.1 Differences in student mean scores on ELLI dimensions	228
Table 6.1 Comparison between University A test item and the GEPT elementary level	270

LIST OF FIGURES

Figure 2.1 A basic model of washback direction	30
Figure 2.2 A model of washback intensity	37
Figure 2.3.1 Traditional washback theory: A stimulus-response model.	39
Figure 2.3.2 1990s view of washback: A “black box” model	41
Figure 2.3.3 Proposed view of washback: A curriculum innovation model.....	41
Figure 2.4 Bailey’s model of washback.....	42
Figure 2.5 Green’s model of washback	44
Figure 2.6 Shih’s washback model of students’ learning.....	46
Figure 2.7 Complex and embedded description of learning power and learning dispositions	82
Figure 2.8 A map of some of the variables integral to motivation for learning and learning power.....	83
Figure 3.1 Transana clip screenshot	127
Figure 3.2 Holliday’s (2002) process of analysing qualitative data	130
Figure 4.1 Timeline of Adam’s lesson for English Skills Integrated Training 2008.05.21 ...	148
Figure 4.2 Timeline of Alice’s lesson for English Skills Integrated Training 2008.05.09	149
Figure 4.3 Timeline of Amy’s lesson for English Skills Integrated Training 2008.05.14.....	150
Figure 4.4 Timeline of Anna’s lesson for English Skills Integrated Training 2008.04.28	151
Figure 4.5 Timeline of Becca’s lesson for English Reading and Writing 2008.05.27	152
Figure 4.6 Timeline of Ben’s lesson for English II 2008.06.03	153
Figure 4.7 Timeline of Betty’s lesson for English II 2008.06.03	154
Figure 4.8 Monthly-issued English learning magazine: reference to GEPT preparation	165
Figure 4.9 University A: Transana series keyword percentage graph.....	173
Figure 5.1 Mean Learning Profile for University A students (N=216)	225
Figure 5.2 Mean Learning Profile for University B students (N=238)	226

Figure 5.3 Sample spider diagram	230
Figure 5.4 Brad's Learning Profile	231
Figure 5.5 Bianca's Learning Profile.....	232
Figure 5.6 Blair's Learning Profile.....	232
Figure 5.7 Bridget's Learning Profile.....	232
Figure 5.8 Brenda's Learning Profile	233
Figure 5.9 Archer's Learning Profile	236
Figure 5.10 April's Learning Profile	237
Figure 5.11 Alvin's Learning Profile.....	239
Figure 5.12 Bess' Learning Profile.....	241
Figure 5.13 Abel's Learning Profile	241
Figure 5.14 Anson's Learning Profile	242
Figure 5.15 Aiden's Learning Profile	242
Figure 5.16 Billy's Learning Profile	246
Figure 5.17 Bonnie's Learning Profile	246
Figure 5.18 Andrew's Learning Profile	247
Figure 5.19 Bryan's Learning Profile	250
Figure 5.20 Aaron's Learning Profile.....	250
Figure 5.21 Alex's Learning Profile	250

CHAPTER 1 Introduction

1.1 Overview

This study is an exploration of how the implementation of the requirement for English proficiency for graduation affects the university students and the English curriculum. The majority of the students in Taiwanese universities, the non-English majors, are required to provide formal evidence of English proficiency from one of a number of English proficiency tests in order to receive their degrees. The effects of the requirement on the students and their English for Academic Purposes (EAP) curriculum are explored through the washback and the impact of those tests.

In this chapter, I will describe the research context of this study, and explain what motivated me to conduct this study. An overview of how this study is situated in the washback literature is then presented. At the end of the chapter, the structure of this dissertation is outlined.

1.2 Background

It has been more than a decade that language testing researchers acknowledged the lack of empirical evidence to this long-asserted phenomenon, washback (Alderson and Wall, 1993), test influence on teaching and learning. Since then, empirical washback studies in different educational contexts incorporating a variety of research methods have flourished, ranging from quantitative survey studies (e.g. Shohamy *et al.* 1996; Stecher *et al.*, 2004), qualitative and ethnographic classroom observations (e.g. Wall and Alderson, 1993; Burrows, 2004; Read and Hayes, 2004; Watanabe, 2004) to studies

that employ both quantitative and qualitative research methods (e.g. Cheng, 2005; Green, 2006a,b, 2007a,b). These studies have revealed that the nature of washback is complex and multi-faceted: it is not as simple as the assertion that a test will influence teaching and learning. There can be washback from exams on various areas of teaching and learning but the washback effects may vary in 'form' and 'intensity' (Cheng, 1997; 2005) in probably unpredictable manners. Manifestation of washback effects of a test varies between individual stakeholders within the same stakeholder group, and between different groups of stakeholders (Alderson & Hamp-Lyons, 1993; Green, 2006a,b, 2007a,b). It has been a general trend as seen in the previous studies on washback of language tests that the research focus tended to be more on teachers than learners – the central stakeholder group in education. As Broadfoot (2005) argues that learners and learning should be given higher priority in washback research because the most important purpose of tests and assessment practices is to enhance learning. Yet, not enough evidence has been collected on the impact of language testing on learners. The current study thus seeks to explore the less explored, washback to the learners.

As mentioned above, the majority of previous washback studies explored the washback processes in relation to teaching; with learners as 'peripheral' (Green, 2006a, 2007b). A few recent studies that had a focus on learners contributed substantially to our understanding of washback process and also presented some important conceptual development in washback studies (e.g., Green, 2006a,b, 2007a,b; Gosa, 2004; Scott, 2005; Shih, 2006, 2007; Tsagari, 2006). Firstly, learners are treated as equally important in the washback process as teachers because they play a major role in the presence or absence of washback to learners (Gosa, 2004; Shih, 2006, 2007; Tsagari, 2006). Secondly, it is acknowledged that learners, like teachers, may demonstrate

individual differences in their experiences of exam influences (Gosa, 2004; Scott, 2005; Shih, 2006, 2007; Tsagari, 2006) but their expectations of a course and their perceptions of washback may differ greatly from those of teachers (Green, 2006a, 2007b). Lastly, the relationship between different stakeholders involved in the washback process is dynamic and interactive (Tsagari, 2006) and thus, to understand washback on learning, the role of schools, teachers, courses and parents should also be considered as they may be influential in shaping washback to learners (Green, 2006a, 2007b; Scott, 2005; Shih, 2007; Tsagari, 2006). The limitations of these studies, however, lie in the methods they have used to elicit students' perceptions and attitudes. The student diaries in Gosa (2004) and Tsagari's (2006) studies were indeed useful in revealing individual differences among students. However, they had the drawbacks of having very limited number of participants, uncontrollability of diary contents and difficulty in differentiating washback from influences of other factors (e.g. other tests not in the scope of the study). Likewise, the interviews as the only tool in eliciting students' views in Shih's (2007) study also revealed individual differences but nonetheless, it was not easy to develop a consensus among the diverse perceptions and attitudes. Green's survey study allowed for a large number of participants across several schools. Yet, to probe the relationship between the teachers, programmes and the learners further, he indicated the need to use more sensitive instruments in conjunction with qualitative methods for in-depth investigation. These methodological limitations of the previous studies call for mixed-method approach in order to capture a more elaborated picture of washback on learners and their learning while allowing depth and individual variety.

One aspect of test influences on learners yet to be explored is the role of learner's

individuality in determining the different extent of washback they perceive and experience. Thus, this study aims to investigate washback to learners by proposing a quantitative measure of students' learning power along with interviews that elicit their perceptions towards the requirement and the tests included in the requirement (see 2.5, 3.6.4).

The educational context of this study (to be described in section 1.3) presents two interesting topics that have received little attention in washback studies so far. The requirement accepts scores of not just one English proficiency test, rather, students can provide evidence of their English language proficiency from any of the English proficiency tests listed in the requirement (1.3.1). Most previous studies centre on the influences of one particular high-stakes test or assessment system, which is closely related to the curriculum (Cheng 2005; Watanabe 1996; Wall & Alderson 1993; Green 2007; Wall 1996; Alderson & Hamp-Lyons 1996). In the very few studies that have probed into effects of more than one test (Shohamy, 1993; Shohamy et al., 1996; Watanabe, 1996, 1997, 2001, 2004), the contents of those tests are still aligned with the prescribed curriculum. However, none of the tests stated in the graduation requirement in this research context (See 1.3) are developed according to Taiwanese university English curriculum. This presents an interesting and rare opportunity to explore which test among the list of English proficiency tests has the strongest degree of washback and why.

Secondly, the implementation of the English requirement for graduation is one of the recent developments in English curriculum in Taiwanese universities (Shih, 2007) and has received little formal research. Recent washback studies in Taiwan have

investigated more in the context of high schools than in that of universities (Chen, 2002; Huang, 2004; Wu and Chin, 2006). Both Chen (2002) and Huang (2004) studied washback of the Basic Competence Test in English (that is linked to the junior high school curriculum) on teaching in junior high schools while Wu and Chin (2006) explored the potential washback of the General English Proficiency Test (GEPT) on senior high school English curriculum. The latest study by Shih (2006, 2007, 2009, 2010), which related closely to the present study, also investigated the GEPT washback on learning in the context of higher education. The context of his study was similar to mine but he chose to investigate solely on washback from the GEPT. His study revealed that the GEPT only brought about limited degree of washback on learning. One possible explanation for this was that the participants in Shih's studies were all English majors. How the GEPT may influence the majority of university students, the non-English majors, is left unexplored. Another difference worth pointing out here is that he studied GEPT impact on students from universities of technology in Taiwanese higher education system. It will be interesting to examine the potential difference in the manifestation of washback of the graduation requirement on students from comprehensive universities. (For further review, see Chapter 2).

1.3 Research context

Alderson and Wall (1993), in their seminal washback article, argue that the educational context in which the test is used should be looked at because there may be forces other than the test that might affect the nature of washback. The educational context this study entails is a highly complex one. The implementation of graduation requirement for English proficiency relates to several aspects, including language educational policies, university assessment, university autonomy and a number of English

proficiency tests, which will be described in the next sections of this chapter.

1.3.1 The implementation of the graduation requirement for English proficiency

The implementation of the graduation requirement for English proficiency can be considered as a reactive action towards several language educational policies in Taiwan. In the past decade, the Ministry of Education (MOE) announced a series of policies concerning the English proficiency of university students. These policies can be presented in three stages: 1) The initial stage: e-Generation Man Power Cultivation Plan, 2) The development stage: Adoption of the Common European Framework of Reference, 3) The assessment stage: Projection 2005-2008.

The concept of the graduation requirement for English proficiency can be traced back to the 'e-Generation Man Power Cultivation Plan', a sub-plan in the national development plan, 'Challenge 2008', which started from 2002 (see www.moe.gov.edu.tw). In order to promote international competitiveness, at the level of universities, the plan suggested the establishment of a common index of English proficiency for university students. Since 2002, a number of prestigious universities began to require their students, both English majors and non-English majors, to reach certain level of English language proficiency in order to obtain their degrees. The implementation of such regulations was controversial and seriously debated because it was not the MOE's intention to align English proficiency tests which were external to the university English curriculum with students' graduation. As a result, instead of pushing every university to establish the requirement, the MOE adopted a less controversial approach by asking the universities to encourage their students to reach a certain level of English proficiency instead of enforcing a requirement nationwide.

In 2005, the MOE adopted the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR) (Council of Europe, 2001) as its source to establish the target levels of English proficiency for the English learners in Taiwan (See also Wu and Wu, 2007). How the scores of the available English proficiency tests in Taiwan were mapped against each level of the CEFR was made public. At around the same time, in the projection 2005-2008, the MOE started to materialise the promotion of English proficiency among university students by setting a target percentage, each year, of students passing English proficiency tests at the level right for them. It was expected that by the end of the year 2008, 55% of university graduates would meet the threshold of English proficiency equivalent to the CEFR-B1 level by proof of English proficiency test scores. In order to see whether the target had been met, each university was required to report to the MOE the percentage of their students in taking an English proficiency test and the percentage of them reaching the targeted standard. In addition, how each university 'performed' on this dimension would be taken into account in the evaluation of a university. As a result, there was an increase in the number of universities implementing the graduation requirement so that it would help push their students' performance in English proficiency tests to achieve the target set by the MOE.

However, the complex nature of higher education complicates the seemingly simple goal of establishing a common index of English proficiency for university students. The complexity lies in the fact that universities enjoy a larger degree of autonomy than schools that provide compulsory education (primary school to senior high school). There are some universities that reject the idea of implementing English graduation requirement, arguing that universities are not cram schools and that the English

curriculum should not be exam-oriented. They prefer to require their students to attend more English courses instead. Furthermore, there are currently 165 universities, colleges, institutes or universities of technology providing equivalent degrees that accept students with diversified levels of English proficiency. It becomes a huge challenge for universities to ask all their students to reach the same standard of proficiency before they graduate. In some universities, it is obligatory only for students from language-related departments to meet the requirement; some universities allow students who have already met the requirement to waive credits on compulsory English courses while others create additional courses in their curriculum to assist students at lower levels of proficiency. In short, the universities differ in their approaches to attending the MOE policies. The English graduation requirement implemented by different universities may also be different in the detailed regulations (e.g. curriculum change, compensation plans) and in the targeted student population (e.g. English majors or non-English majors).

Despite the possible differences as stated above, the main message of the graduation requirements is similar: it is about which English proficiency tests are accepted and what test scores are set as threshold. As an example, undergraduates in the National Taiwan University (National Taiwan University·Guidelines for Advanced English Study, 2002, amended in 2008, 2009) have to meet one of the standards stated below before being awarded the Bachelor's degree:

- High-Intermediate Level of General English Proficiency Test Stage 1
- TOEFL 550 and above
- Computer-Based TOEFL (CBT) 213 and above or TOEFL iBT 79 and above
- IELTS 6.0 and above

- Foreign Language Proficiency Test (FLPT) each written test 70 and above
- Cambridge ESOL FCE Level B and above
- Other English proficiency tests that are approved by the Advanced English Study planning group and the centre for general education
- A degree received from universities in English speaking countries that are recognised by the Ministry of Education

Other universities may have different standards for their students, but most of the tests included in the requirement are indeed English proficiency tests that are not tied to any university English curriculum.

Among the tests clearly stated in the above requirement, the General English Proficiency Test (GEPT; Shih, 2006, 2007, 2008, 2010) seems to be one of the most important tests because of its popularity. Although the FLPT is the first English proficiency test developed in Taiwan (since 1965), it is not as popular as the GEPT. According to statistics of the Language Training and Testing Centre, which develop both tests, there were altogether 348,378 GEPT test takers of all five levels in 2008, (over 81,000 college and university students) (GEPT Elementary Level, Intermediate Level, High-Intermediate Level Scores Statistical Report, 2008, LTTC Annual Report, 2008; see 1.3.2 for further descriptions of GEPT) while the total number of test takers for the FLPT was 4694 (LTTC Annual Report, 2008). There are quite a number of university graduates taking the TOEFL/CBT/iBT or the IELTS to study abroad. Yet according to the statistics provided by Bureau of International Cultural Educational Relations, 23,665 people applied student visa for studies the United States, the United Kingdom and Australia in 2009. This number was likely to include the number of test

takers for both TOEFL and IELTS and also those who were not university students. .

My search of the graduation requirement implemented by different universities on the internet revealed two other English proficiency tests that have been recognised by Taiwanese universities: College Student English Proficiency Test (CSEPT) and TOEIC. The CSEPT is a test the LTTC has developed particularly for Taiwanese college and university students. The official statistical report of the CSEPT showed that in 2008, the number of test takers was 43,638 (CSEPT Statistical Report, 2008), including students from general universities, universities and institutes of technology. The test that also, receives high popularity among university students is TOEIC, with a total of 186,649 test takers (over 73,000 college and university students) in 2008 (http://www.toEIC.com.tw/toEIC_news_02.jsp, TOEIC Scores Statistical Report, 2008). The numbers stated above have shown that among those tests recognised by the English graduation requirement, the GEPT and the TOEIC could be more influential to Taiwanese undergraduates than other English proficiency tests.

As TOEIC has been discussed more widely in the testing literature (Robb and Ercanbrack, 1999; Newfields, 2005; see more studies on ETS TOEIC research), the following section will introduce the GEPT, the locally-developed test that has a record of over 3.8 million test takers in a decade.

1.3.2 The General English Proficiency Test

The GEPT is an English proficiency test developed by the Language Training and Testing Centre in 1999, commissioned by the Ministry of Education with the goal of promoting life-long learning and encouraging English study. According to the annual

report provided by the Language Training and Testing Centre (2004), the test aims at providing authenticity in its items and thus, the test contents are able to reflect the living experiences within the local Taiwanese culture. It is a criterion-referenced test with five levels: Elementary, Intermediate, High-Intermediate, Advanced and Superior. Each level is differentiated by criterion levels of English proficiency, which is made clear by a general description of the English proficiency expected at that level and a detailed skill-area description specifically for the four components of the test, listening, reading, writing and speaking. The Superior level, which has the criterion of native proficiency in dealing with advanced academic language, is provided on demand by institutions, rather than individuals. For the other four levels, there are two stages of the test, and test takers have to pass the first stage in order to be advanced to the second stage. The first stage consists of the listening and reading components while the second stage consists of the writing and speaking components.

As what Roever and Pan (2008) have reviewed, the five levels can be divided into two groups according to the similarity of task types. For the three lower levels (Elementary, Intermediate, High-Intermediate), all the listening and reading components consist exclusively of multiple choices. The task types for the listening component include the selection of a correct description to a picture (Elementary and Intermediate), the appropriate response to a question, and the comprehension of a short conversation. Except for the High-Intermediate level, the picture description is replaced by the comprehension of a short talk. In the reading component, the task types are vocabulary, cloze and short passage reading comprehension for all three levels. The speaking component is 'tape-based' (Roever and Pan, *ibid.*), not interview-based. Test takers are asked to produce responses in the form of repetition (Elementary), read-aloud

(sentences for Elementary and short passages for Intermediate), short answers to pre-recorded questions (all three lower levels), picture description (Intermediate and High-Intermediate) and topic discussion (High-Intermediate). In the Elementary level writing component, test takers have to write sentences for tasks including rearranging scrambled sentences, combining sentences and rephrasing. They also have to compose a short paragraph to describe a picture. Differently, the Intermediate and High-Intermediate level consist of a Chinese to English, paragraph translation task and an extended paragraph writing task for an indicated topic.

The Advanced and Superior level are different as they are more academically oriented and consist of tasks other than multiple choice items. For the Advanced level, the listening component remains multiple-choice with longer conversations and talks. The reading component comprises other task types including matching, short answer, fill-in-the-blank and summary, which require reading for gist or for specific details. For the speaking component, test takers face an interviewer, either engaging in monologic self-introduction, short talk on a topic, or in dialogic information exchange task or topic discussion. For the writing component, there is no translation task but two guided writing tasks, which require summary of texts or interpretation of visually presented information as charts or graphs before further discussions of the issues presented.

Different from other levels, the Superior level consists only of two sections, which assess integrated skills. The first section requires reading a long article and listening to a long talk before summarising the main ideas and writing an essay. Based on the ideas of the listening and reading input, in the second section, test takers make an oral presentation and answer questions raised by the interviewer concerning what they have presented.

The immense popularity of the GEPT among the Taiwanese citizens including the university students indicates the likelihood of strong GEPT influences on the students' English learning. The reason why the GEPT receives greater attention than other available English proficiency tests may be because of its good publicity from the media and also the support from the government and the MOE. It is not only recognised by the government agencies as a criterion for promotion, but also used by the MOE as a criterion to evaluate the English proficiency of the applicants to its scholarship program. In addition, its popularity is also fuelled by its status as an English proficiency test that targets at all English learners, with little restriction on age (except on children under 12), profession, or education backgrounds. The GEPT can be considered as one of the most well-known English proficiency tests in Taiwan. Therefore, by exploring the impact of the graduation requirement for English proficiency on university students and their English curriculum, this study aims to investigate whether the GEPT has brought about strong extent of washback as reflected from its popularity. However, judging from the number of university students taking other tests, such as TOEIC or CSEPT, the GEPT should not be the only test that can have washback on the students and their learning. This study will thus also explore whether there is washback and impact of other tests in the universities.

As stated in the general description of the High-Intermediate level (see Appendix A for the construct of the GEPT High-Intermediate level), test takers who pass this level have the proficiency equivalent to that of a non-English major undergraduate (see also Vongpumivitch, 2010). A search of the requirement regulations in Taiwanese universities also shows that universities with a ranking above average mostly set up the High-Intermediate level of the GEPT as the standard in their requirement. Universities

with lower ranking may accept a pass at the GEPT Intermediate level, which is equivalent to a high school graduate's English proficiency or even the Elementary level, which is equivalent to that of a junior high school graduate, as the threshold.

What facets of GEPT washback will be expected in relation to the test constructs and design characteristics of the Intermediate and High-Intermediate level will be discussed in 3.6.2.

1.4 Personal experiences

My motivation to do this research derived from my previous experience as a full-time instructor for the English department in a Taiwanese university from 2002 to 2006. It was during that period when the GEPT was just introduced and started to receive attention from the media, society and educational institutions. Around the same time, policies concerning the common standard of English proficiency among Taiwanese university students were also announced. Over the course of my four years dealing with both academic and administrative affairs in the university, I have been involved in several discussions with colleagues over the feasibility of establishing the English graduation requirement for students. I have also attended a panel discussion, as a representative of the department, and discussed the requirement with representatives of English departments from other universities in Southern Taiwan. Two important issues that emerged from those discussions attracted my attention: the dilemmas and the consequences.

First, I have found that there was far more complexity than I expected in putting educational policies into real practice in universities, and the introduction to such requirements could face serious dilemmas and challenges. On the one hand, I realised

that the establishment of the requirement was essential for the universities and the English departments to comply with the policies and to receive good evaluation in the university assessment. As a teaching member, I also felt the pressure to boost the number of our students to pass an English proficiency test and knew that the requirement would be helpful to some extent. On the other hand, I acknowledged the huge difficulty for students with their relatively low levels of proficiency to pass English proficiency tests. I could also understand the challenges the universities would face if the majority of their students failed to meet the requirement in their fourth year, the last year of university. It was thus a huge challenge for the English departments and the university authorities to determine whether or not to implement the graduation requirement, and what regulations should be included, considering what would be the best for the universities, the departments and the students.

Secondly, I have realised that despite numerous debates and discussions being made on the topic, insufficient attention has been paid to the consequences that the implementation of such requirement could bring to the university English curriculum, the teachers and especially the students. Although the discussions of the dilemmas and challenges included some considerations of possible consequences, the consequences referred to were more related to the consequences that universities would confront in the face of university assessment, reputations or student performances. There have not been many concerns about how the requirement might eventually influence actual English teaching and learning in the universities, which I regard as more important, or at least, equally important. I believe that as an English instructor who can be involved in making important decisions for all the students in the university, we have the responsibility to understand what happens after a decision has been made. I hope that

this study may provide a glimpse of the ‘after’ scene and stimulate much more discussions among teachers and researchers to move beyond the implementation, to focus more on the consequences the implementation of the requirement bring. I am particularly concerned about what the consequences are for the university students. A review of literature (Chapter 2) has also identified the need for more studies to shed light on washback to the learners and thus, a key focus of this study is to understand to what extent the learners have been affected by being compelled to take and pass an external English proficiency test for graduation.

1.5 Overall research aims

The central aim of the present study is to explore the impact of the implementation of graduation requirement for English proficiency in Taiwan on the university students and the English curriculum for them. University students for the purpose of this research refer to only non-English majors. English learning for them becomes more than merely attending and passing the basic required courses. Since one important focus of this study is on the learners, another aim of the study is to explore the relationship between learning power (Deakin Crick et al., 2004; Deakin Crick, 2007; Deakin Crick and Yu, 2008) and washback to the learners. Therefore, this study will attempt to explore the following:

- 1) To what extent and in what ways has the English graduation requirement influenced the English for Academic Purposes (EAP) curriculum for non-English majors?
- 2) To what extent and in what ways has the English graduation requirement influenced non-English majors?
- 3) In what ways can a learner’s learning power inform an understanding of the process of washback to learners?

1.6 Structure of the dissertation

The structure of the dissertation is as follows

Chapter 2 – Literature Review: This chapter reviews the literature on the main theme of the study: washback/impact and learning power. The review of washback/impact studies from both general education and language testing fields lay out the theoretical framework of this study and highlight the gap that this study attempts to fill. The review of approaches to learning and learning power provides a theoretical link between learning power of students and the washback of tests on them.

Chapter 3 – Methodology: This chapter first presents the philosophical position this study takes on, describing the post-modernist influences on the epistemological and methodological considerations underlying the study with the complementary positivist element, especially on the systematic presentation of learner characteristics, as defined by learning power. It delineates the research approach of this study; presents the research questions and different methods of data collection and the methods of analysis. It also includes a discussion of the attempts to achieve trustworthiness and of the ethical issues that impinge on the conduct of this study.

Chapter 4 – Impact of the graduation requirement on English curriculum for non-English majors: This chapter provides an analysis of the impact of the requirement on teaching and reveals GEPT washback mainly on teaching materials and testing materials. The findings also highlight the role of stakeholders other than teachers and learners in the washback mechanism.

Chapter 5 – Impact of the graduation requirement of non-English majors: This chapter focuses on an analysis of learners' perceptions of and attitudes towards the requirement. The findings reveal students' individual difference in perceiving the impact of the requirement and the differences that may result from the threshold level, their learner characteristics as defined by learning power, and influence from other stakeholders.

Chapter 6 – Discussions: This chapter presents discussions of two major issues emerging from the findings in Chapter 4 and 5. The first concerns the social impact of GEPT, which has been reinforced by the graduation requirement in the university settings. The second presents a further operationalisation of washback to the learners and washback on learning by relating the findings of this study to the previous conceptualisation of washback.

Chapter 7 - Conclusion: This chapter provides a summary of the main findings in accordance with each research question, a critique of the strengths and limitations of the study, directions for further research, and lastly, implications for the different stakeholders related to the graduation requirement, ranging from policy makers to teachers and learners in the classrooms.

CHAPTER 2 Literature Review

2.1 Introduction

In this chapter, I review the literature relevant to this research in four sub-sections (Section 2.2~2.5). Section 2.2 defines the key terms (washback and impact) used in this study (2.2.1) and discusses the impacts of educational and language policies in particular, policies related to language tests, in different contexts (2.2.2). It further reviews the key issues of washback, e.g. the direction of washback (2.2.3), washback variability (2.2.4) and washback intensity (2.2.5), washback mechanism and different conceptualisations of washback in the literature (2.2.6). Section 2.3 presents a critical analysis of empirical studies on washback on teaching (2.3.1) and learning (2.3.2), and highlights the research gaps this study aims to fill. In Section 2.4, I review a limited number of washback conducted in the same educational and research context of this study. The last section (2.5) situates the ELLI in the literature of assessment and approaches to learning and describes the theoretical underpinnings and the seven dimensions of the Effective Lifelong Learning Inventory (ELLI) which is used to assess the student participants' learning powers. This section provides a rationale for the use of this tool in this washback study.

2.2 Theoretical framework

2.2.1 Definition of key terms: washback and impact

In the fields of general education and language testing, there are a number of terms with similar definitions that all refer to the influences of a test (Cheng, 2005). The term that has gained wide currency in applied linguistics is 'washback'. Other preferred terms in applied linguistics include 'consequential validity' (Messick, 1989, 1996) and test

impact (Bachman and Palmer, 1996; Shohamy, 2001). Related terms in the general education measurement literature date back to Popham's (1987) notion of 'measurement-driven instruction' in referring to the role of tests in driving teaching and learning. Another term that is similar to Popham's is the alignment of the content and format of the curriculum to those of the test, which Shepard refers to as 'curriculum alignment' (1993). While Madaus (1988) argues against 'teaching to the test', as this will result in detrimental effects that may outweigh short-term benefits, Frederiksen and Collins (1989) have a more positive view of the curricular and instructional changes, according to what a test is designed to measure, and propose the concept of 'systemic validity' of a test for such deliberate positive alignment. Following Green (2007) on acknowledging the implicit differences of approach in the terminology, I believe it is also important to differentiate the terms used in this study from other terms in the literature.

There is a diverse view in applied linguistics of what the term 'washback' should encompass. The various definitions given to the term reveal differences in 'scope, actor and intentionality' (Spratt, 2005). Washback, generally defined, is the influence of testing on teaching and learning (Alderson and Wall, 1993; Bailey, 1996; Hughes, 1988, 2003). Messick (1996) further refines the definition to "the extent to which the test influences language teachers and learners to do things they would not otherwise necessarily do" (p.241). Cheng (1997) adds another dimension to the definition of washback as she uses the term as "an active direction and function of intended curriculum change by means of the change of public examinations" (p.38), arguing that not only "accidental side-effects of examinations" (Spolsky, 1994, cited in Cheng, 1997) can be considered as washback. However, she (*ibid.*) also points out that in a complex

situation such as curriculum development, intended curriculum change would also bring about unintended and accidental side effects.

Some researchers consider washback to lie within the scope of a larger phenomenon, test impact. Bachman and Palmer (1996) do not use the word 'washback', but refer to the effect of test on individual students and teachers as the micro level of test impact, which comes under the macro level of test use and its social impact. Hamp-Lyons (1997) also argues against the limitation of focusing test effect on teaching and learning and refers washback as one dimension of test impact, which "pervades every aspect of our instruments and scoring procedures" (p.299). Wall (1997) makes a distinction between impact and washback, referring the former to "any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as whole" and the latter to "the effects of tests on teaching and learning" (p.291). Some researchers do not take on the distinction between the two terms, and refer to both test effects at the micro and macro level as washback (Andrews et. al, 2002, Scott, 2005).

Regardless of the different conceptualisations of washback and impact, many language testing researchers now locate both concepts within Messick's theoretical notion of 'consequential validity', in which washback is an instance of the consequential aspect of a test's construct validity. Messick further explains that the consequential aspect of test validity includes:

evidence and rationales for evaluating the intended and unintended consequences of score interpretation and use in both the short- and long-term, especially those associated with bias in scoring and interpretation, with unfairness in test use, and with positive or negative

washback effects on teaching and learning. (1996, p.251)

He also emphasises that neither washback as only one form of social consequence of testing, nor test consequences, can be viewed alone as a separate aspect of test validity. Other researchers have also associated washback with test validity such as the concept of 'washback validity' by Morrow (1986) and 'systemic validity' proposed by Fredericksen and Collins (1989), both asserting the importance to evaluate test validity according to its effect on teaching and learning. However, in echoing Alderson and Wall's (1993) serious doubt on the direct link between washback to test validity, Messick (1996) emphasises that washback is a test consequence 'that bears on validity only if it can be evidentially shown to be an effect of the test and not of other forces operative on the educational scene' (p.242). Thus, he argues about the need to 'seek validity by design as a likely basis for washback' instead of considering washback as a sign of test validity (p.252).

In line with McNamara (2000) and Shohamy (2001), both 'washback' and 'impact' are used in this study, adopting Wall's (1997) distinction between the two concepts with small adaptation to this particular educational and research context. 'Washback' is narrowly defined as referring to the effects of any test stated in the graduation requirement on teaching and learning. The term 'impact' encompasses two major aspects: (1) the influences a test may bring to the stakeholders involved within the classrooms, the universities, the educational system and the society, (2) the influences of the graduation requirement in the universities.

2.2.2 Language tests and related policies

Language tests are not merely tools that measure language proficiency, but are

instruments embedded in educational, social and political contexts. Shohamy (1993, 1996, 1998, 2001) calls for attention to the complex agendas that language tests can entail:

Critical language testing assumes that the act of testing is not neutral. Rather, it is both a product and an agent of cultural, social, political, educational and ideological agendas that shape the lives of individual participant, teachers and learners. (Shohamy, 1998, p. 332)

She further argues as to how tests could be used for power and control in creating intended washback and impact such as manipulating certain language knowledge, behaviours or pedagogy when high-stakes decisions are made based on the test results (1996, 2001, 2007).

In this decade, there are a number of empirical studies which have examined how tests have been used to bring intended effects by the authorities concerned. Some focused on the role of tests in curriculum innovation (Chapman and Synder, 2000; Andrews et al., 2002; Burrows, 2004; Ferman, 2004; Stecher et al., 2004; Cheng, 2005; Qi, 2004, 2007) while others were more concerned about the political agendas of tests (Shohamy, 2004; Evans and Hornburger, 2005; McNamara, 2005; McNamara and Roever, 2006; Menken, 2006, 2009; Kunan, 2009 a, b).

Shohamy's (2004, 2006, 2007) studies examine how centrally-controlled educational agencies in multilingual and multicultural nations use tests to influence language policies which appear to reflect democratic pluralism in becoming *de facto* policies that promote homogeneity. Her studies have revealed that tests are powerful tools that can determine the status, the hierarchy of languages and also suppress the diversity in languages.

The political functions of language tests can also include tests being used as instruments for social policies which determine citizenship (McNamara, 2005; McNamara and Roever, 2006, Kunan, 2009, a, b). Under the officially claimed purpose of promoting integration and social cohesion, what language tests can actually assess becomes not so much the language construct, but the social and cultural identity that determine an outsider from an insider (McNamara, 2005). Kunan (2009) questions the ability of the naturalisation test to assess what it claims to assess and any beneficial value of the test to society.

Language tests have also been used as powerful tools for curriculum innovation, particularly when high-stakes purposes are attached to the test scores. The idea of exam reform being a 'lever for change' (Pearson, 1988, p. 101) comes from the optimistic view of the possibility of asserting a positive influence through a test's powerful effect and has been widely adopted by educators. Yet not until the last decade were there empirical studies in both general education and applied linguistics that investigated the consequences of the reforms.

Empirical studies, however, reveal less optimistic findings. The findings show that tests as tools for curriculum innovation have had an influence on some aspects of teaching and learning, but not others (Andrews et al., 2002; Cheng, 1997, 2005; Qi, 2004, 2007; Wall, 1996). Wall's (1996) study of the revised Sri Lankan "O" level English exam has shown that there was washback on the content of English lessons and the ways classroom tests were designed, but not on the teaching methods or how teachers marked pupils' test performance. Cheng's (2005) study on the changes to the Hong Kong Certificate of Education Examination (HKCEE) English language paper with the

intention of engineering washback has presented similar findings. She discovered washback on 'what' teacher taught, but little washback on 'how' teachers taught and concluded that 'the changes tend to happen at the obvious and format level' (1997, p.52). Similarly, Qi (2004, 2005, 2007) noted that the National Matriculation English Test (NMET) in China did not bring as much change on ELT teaching and learning as intended by the test reform. She argued that a test may not be a good lever for change since 'the very function that empowers the test is likely to be in conflict with its intended washback effect, making it too blunt an instrument for promoting desirable changes in teaching and learning' (2005, p.164). Andrews et al. (2002) also noted that although students' performance showed improvement with the introduction of the Use of English (UE) oral examination in Hong Kong, washback on learning outcomes was of a superficial level.

Policy makers' and test designers' belief that tests are powerful enough to bring about changes in teaching and learning in the form they intend is probably overly optimistic, because the manifestation of washback is indirect and unpredictable (Andrews et al., 2002). In addition, the context in which the innovation takes place needs to be taken into full account, as there are factors other than the test itself that can mediate or prevent the intended washback from happening. Nevertheless, even with careful planning and implementation, it is still likely that test effects on teaching and learning will not turn out exactly as intended (Wall, 2000; Andrews et al., 2002).

This study aims to explore further the relationship between government policies and language tests. The educational context of this study is different from those of the abovementioned studies in two ways. First, the graduation requirement does not

introduce a new test or make changes to an old test to promote innovation in teaching. Second, unlike studies within multilingual and multicultural contexts, there is less problem of language tests creating *de facto* policies that suppress diversity in a nation with English as the main foreign language, which is not in conflict with the official language-Mandarin. However, the graduation requirement for English proficiency is indeed a recent implementation with the intention of promoting university students' English proficiency and their motivation for learning English (1.3). In addition, the alignment of test scores of different English proficiency tests stated in the requirement with a certain level in the Common European Framework (CEF) (1.3.1) remind me of what Shohamy (2007) has problematised concerning the framework:

There are therefore doubts as to whether such broad and generic testing descriptions are relevant and valid for different language learning contexts and uses...this shows the problems that arise when test criteria such as rating scales affect language policy, and definitions of 'what it means to know a language' when such rating scales presuppose a hierarchy of both development and performance, adhere to generic descriptions and claim to be universally applicable, detached from the contextualised nature of language and language performance in multilingual environments (2007, p.125).

The current study is thus built on similar concerns and attempts to explore how the alignment of the requirement with the CEF works, particularly in the tertiary context in Taiwan. This study also attempts to follow Shohamy (2001) in an attempt to challenge the imbalanced power between the government, the universities, the teachers and the learners by raising their attention to the learners' voice and their concerns towards the graduation requirement.

In the next section, I will turn to the one of the most debated issues in washback and

impact literature, namely whether the effect of tests is positive or negative.

2.2.3 Direction of Washback: Washback as positive, negative or neutral

Washback is perceived to vary in its direction, positive or negative (Alderson and Wall, 1993; Hughes, 2003; Cheng et al., 2004), and also in the strength of its manifestation, which Cheng (1997, 2005) refers to as ‘washback intensity’. In this section, I will review studies on washback direction as beneficial or detrimental.

Washback, produced by high-stakes examinations, used to be widely perceived as being negative (Alderson & Banerjee, 2001), in a great number of studies in both general education and applied linguistics (e.g. General education: Vernon, 1956; Wiseman 1961; Madaus, 1988; Bradfoot et al., 1990; Smith 1991a, b; Herman and Golan, 1993; Hargreaves, 1997; Morrison and Tang, 2002; Applied Linguistics: Spolsky, 1995;; Bailey, 1999; Shohamy, 2001;). The most common criticisms of tests are that they narrow the curriculum (Smith, 1991 b; Herman & Golan, 1993; Spolsky, 1995; Hargreaves, 1997; Morrison and Tang, 2002; Stecher et al., 2004) and encourage “mechanical, boring and debilitating forms of teaching and learning” (Oxenham, 1984, in Shohamy, 2001). Smith (1991b), in her longitudinal qualitative study in two US primary schools, is eminent in providing empirical evidence on the assertions of such criticisms. She discovered that testing programs had resulted in the reduction of the time available for instruction, the narrowing of the curriculum and the limitation of teaching methods, and also the reduction of teachers’ capacities to teach content and use methods and materials that were incompatible with standardised testing formats. These findings are echoed through other studies such as the focus on the contents, the formats and the subjects tested (Hargreaves, 1997; Morrison and Tang, 2002; Stecher et al.,

2004), the neglect of developing skills not tested in the exam (Wall, 1999, 2005) and instructional plans affected by students' test performance to reflect more tests objectives and contents (Herman and Golan, 1993). Other negative effects of high-stakes tests include test score pollution (Haladyna et al., 1991) with unethical test preparation practices, and the suppression of minority language and disadvantaging students of minority backgrounds (Shohamy, 2001, 2007; Shih, 2007) Test may also result in a high level of student anxiety and pressure, which can be detrimental to their learning motivation (Paris et al., 1991; Jones et al., 2003; see Harlen and Deakin Crick, 2003 for relevant studies).

On the other hand, some researchers believe that well-designed tests can be levers for change, changing formerly bad practice to good teaching and learning (Pearson, 1988; Davies, 1990; Hughes, 1988, 2003; Bailey, 1996). They advocate that efforts should be made on test design features to engender intentional positive washback (Hughes, 1989, 2003; Bailey, 1996; Chapman and Synder, 2000). Hughes (2003) argues that certain criteria need to be met in order to achieve positive washback, for example, using direct testing and making the test criterion-referenced. Bailey (1996), on the other hand, emphasises the importance of having communicative language tests and providing detailed score reports to test takers to promote positive washback.

Alderson and Wall (1993), however, claim that the term 'washback' is a neutral one and that there is possibility for badly designed tests to have positive washback and vice versa. Alderson and Wall further argue that the relationship between a test and its washback may be much more complex than the assumed linear relationship, and that "the quality of the washback might be independent of the quality of the test" (p.118).

Messick also points out the possibility of the ‘bidirectional nature of washback’ (cited in Cheng & Curtis, p.7), by defining washback as “the extent to which a test makes teachers and students do things they would not otherwise do that *promote* or *inhibit* language learning” (p.241). Green (2007b) reviews the literature based on Chapman and Synder’s (2000) suggested test design features for positive washback, and proposes a basic model of washback direction which captures the bidirectional nature of washback (as seen in Fig. 2.1). Green argues that what determines the direction of washback is the ‘overlap’ between both test and curriculum and the construct to which they are directed. In other words, there is greater potential for positive washback if the test characteristics reflect the focal construct, as understood by the stakeholders (the greater the overlap). On the other hand, there is greater potential for negative washback if there is a smaller overlap.

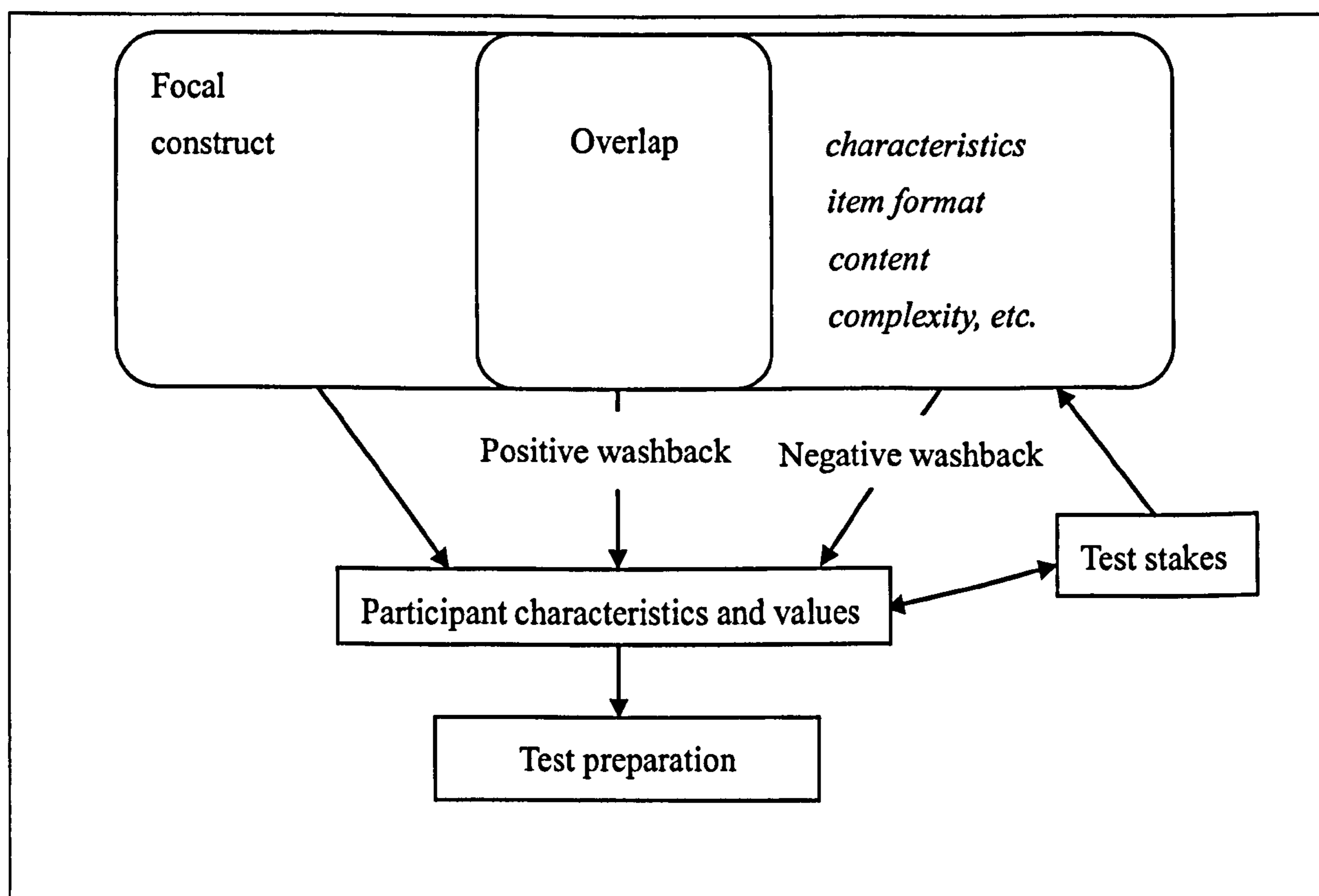


Figure 2.1 A basic model of washback direction

(Green, 2007b, p.17)

Green also discusses the more debated issues of test purpose and test stakes in relation to washback direction. He argues that there is still no consensus as to whether test purposes and associated test stakes can determine the direction of washback, because test effects beneficial to some may be considered detrimental to others. Thus, he emphasises the consideration of individual difference in characteristics and values along with test stakes in determining washback direction.

The above discussion thus leads to the focus of the next section, washback variability, that is, the 'differences between participants in how they are affected by a test' (Green, 2006b, p.339).

2.2.4 Washback Variability

The empirical studies conducted after Alderson and Wall's (1993) appeal for more empirical investigations of the washback phenomenon with classroom observation, and reveal that washback is more complex than the seemingly simple statement of 'a test will influence teaching and learning'. It may be understood that washback is likely to be elusive and unpredictable even in the context of tests intentionally used for curriculum innovation and with careful planning. An important finding is the differences between the participants in the ways they are influenced by the tests and hence, washback variability.

Firstly, washback variability is evident in how the teachers respond to test preparation and changes in tests. Both Alderson and Hamp-Lyons (1996) and Watanabe (1996) have discovered that teachers, instead of the tests themselves, are the reason behind the different extent and type of washback manifested in the classes. Alderson and

Hamp-Lyon's study of TOEFL preparation reveals that the distinctions in the instructions between test and non-test courses are less significant than the distinctions in the instructions between teachers. With similar findings, Watanabe (1996, 2004) has found out that a range of factors related to the teachers including their educational backgrounds, experiences, their beliefs about teaching, their concerns for students' proficiency levels and their psychological factors explain why there is washback in some teachers' classes but not in other teachers' classes. Thus, both argue that teacher factors may account for the variations in washback on teaching and that there is the need to extend the exploration of how these factors contribute to the washback process.

Burrows (2004) studied the Certificate in Spoken and Written English in Australia and found that teachers' responses to the new test differed. She deliberately selected teacher participants who revealed different responses in the interviews. From the classroom observations, she further found that there were different degrees of change among the teachers whose teaching practices had manifested washback. She categorised the teachers in her study into four models, according to the extent of the changes. Drawing on Markee's (1997) models of response to educational change and McCallum et al.'s 'models of teacher assessment' (1995), Burrows proposes that there are four models of teacher in response to test preparation and teachers can be *resisters*, *adopters (partial)*, *late adopters* and *adapters*, thus varying in the extent of washback they mediate. In her study, one particularly interesting participant combined his old teaching practice with the aspects of new test and curriculum he chose to adopt. Such an individual stakeholder who 'takes from the new system as she or he chooses' (*ibid.*, p.125) is what she referred to as an 'adapter'. She also argues that her findings showing one teacher as a resister and another as an adaptor demonstrate teachers' free

will in allowing the extent of washback to occur, which has been less explored in the previous literature.

Secondly, there is also washback variability among the learners. Gosa (2004) and Tzagari (2006), in the Romanian and Greek contexts respectively, have both found that similar to the teachers, students also responded to test preparation differently, as shown in their diaries. Gosa (*ibid.*) points out that learner variables such as their feelings, attitudes, beliefs, learning styles, expectations and anxiety can also be the reasons why there is more washback on some than on the others. Similarly, Tzagari (*ibid.*), argues that learners's views, feelings and attitudes play a major role in the presence or absence of washback.

Drawing on Burrow's models of responses among teachers (2004), and taking into consideration the varied responses different stakeholders may have, Scott (2005) argues that models similar to Burrows can be built to influence other stakeholders. In regards to the context of her study concerning English as Additional Language (EAL), she points out that both pupils and parents can be influenced by a number of factors. Young learners can be affected by their own individual characteristics, age, cognitive and linguistic stage of development, preferred learning styles, and also their teachers' and parents' response to the tests, while parents are likely to be influenced by their personal experience of education, their awareness of the tests and the purposes of the tests. The current study is different from Scott's, in that it targets adult learners instead of young learners. The role that parents play in mediating washback on adult learners may be very different from Scott's findings. It will be interesting to see how the parents' differing responses to the tests stated in the requirement may affect university

students' test preparation.

The findings of empirical studies using classroom observation demonstrate that tests alone do not cause washback, and factors from stakeholders especially teachers and students can mediate the differing extent of washback on teaching and learning (see also 2.3.1, 2.3.2). In addition, the evidence of washback variability in these studies points out the need to include more participants for classroom observation in exploring the complex role that participant factors play in washback on teaching and learning (e.g. more teachers teaching the same course) .

In 2.3.1.3, I will discuss the issue of washback variability presented in previous empirical studies (2.3.1, 2.3.2) and how their findings of washback variability inform this study.

2.2.5 Strength of washback: washback intensity

Besides varying in direction and manifestation, washback can also vary in strength, washback intensity' (Cheng, 1997, 2005). Cheng(1997, 2005) uses this to refer to the degree of washback on an area or a number of areas of teaching and learning (e.g. high washback intensity on teaching contents but low intensity on teaching methods. See2.3.1). Watanabe (2004) and Green (2007) extend Cheng's definition of the term to a more general reference to the degree of washback associated with a test. It includes not only the extent of washback on different aspects of teaching and learning but also the extent to which individual participants will adjust to test demands. In short, 'washback intensity' encompasses intensity in areas of teaching and learning' and on participants'. For the purposes of this study, an extended definition is adopted.

The level of washback intensity are often said to be indicated by the stakes of a test (Watanabe, 2004; Green, 2007b). The higher the stakes of a test, the stronger washback will be (Alderson & Hamp-Lyons, 1997; Shohamy et al., 1996). Tests are considered as high-stakes when test scores are directly used for admission, promotion, placement or graduation (Madaus, 1990). However, evidence of washback variability among participants suggests that it may not be the real stakes a test entails but the participants' perceptions of test stakes that determine the extent of washback on their behaviours (Madaus, 1988; Gipps, 1994; Tsagari, 2006). Chapman and Synder (2000) in their argument of using high-stakes testing for educational change, clearly state the significance of participants' perceptions of test stakes.

'...it is not the examination itself that influences teachers' behaviour, but teachers' beliefs about those changes. As Madaus and Kellaghan (1993) point out, the power of tests to influence instruction is a perceptual phenomenon – if you believe it does, then it does. The effect is produced by perception, regardless of the reality of the importance of the linkages.' (p.462)

Some researchers also address the issue of what determines the strength of washback, other than test stakes. Hughes (1993) brings up the notion of test importance, stating that for washback to fully work as intended, participants' success in the test should be of real importance to them. Gates (1995) lists a number of factors that will influence washback intensity, including *prestige*, *accuracy*, *transparency*, *utility*, *monopoly*, *anxiety* and *practicality* (p.102). Besides *accuracy*¹ and *transparency*², which are more related to test reliability and test construct, the other four factors can be more or less linked to participants' perceptions of test importance. In determining the strength

¹ *Accuracy*, in Gates' (1995) definition, refers to the perceptions of the stakeholders who use the test scores on the reliability and accuracy of the tests.

² The resemblance of the test construct to real-life language use (final language needs of learners) is what Gates (*ibid.*) considered as *transparency*.

of washback, Gates argues for the need to consider the reputation of the test developing organisation (*prestige*), the degree of dominance a test is in the marketplace (*monopoly*), the extent of stress a test can place on learners (*anxiety*), opportunities that the score of a test can provide (*utility*) and the degree of *practicality*. Referring to the context of this study, in Gates' sense, the washback of any tests accepted by the graduation requirement may be diluted like IELTS for the application of British universities. IELTS is not the only test that determines the entrance into British universities. Likewise, there is a wide range of tests that allows the students to graduate from universities with graduation requirements, reducing the degree of monopoly of any tests accepted by the requirement. Since students are given the freedom to choose a test that suits their situations and inclinations, hence washback of each test may be diluted. However, there are other factors that might influence washback intensity or even test monopoly, and thus, all of those factors should be taken into consideration. It will be interesting to see how the findings of this study may reflect Gates' ideas stated above.

Washback intensity is also considered to be affected by participants' perceptions of test difficulty. An important study that sheds light on this issue is Watanabe's (2001) study, which explores about washback on motivation for test preparation. He argues that tests of appropriate difficulty, as perceived by the learners will result in positive washback on learners' motivation for test preparation with (2.3.2.1). Green (2007b), drawing on Crooks (1988) and Mehrens (1998), argues that the relationship between washback and test difficulty is not linear. Only when standards are attainable but challenging will teachers and learners devote themselves to test preparation to meet the standards. If the standards are perceived as either too easy or too difficult to

achieve, teachers and learners will be less likely to adjust their teaching and learning behaviour to the test demands. He proposes a model of washback intensity in which he specifies how washback intensity can vary in relation to participant's perceptions of test importance (incorporating test stakes), and test difficulty (Fig 2.2).

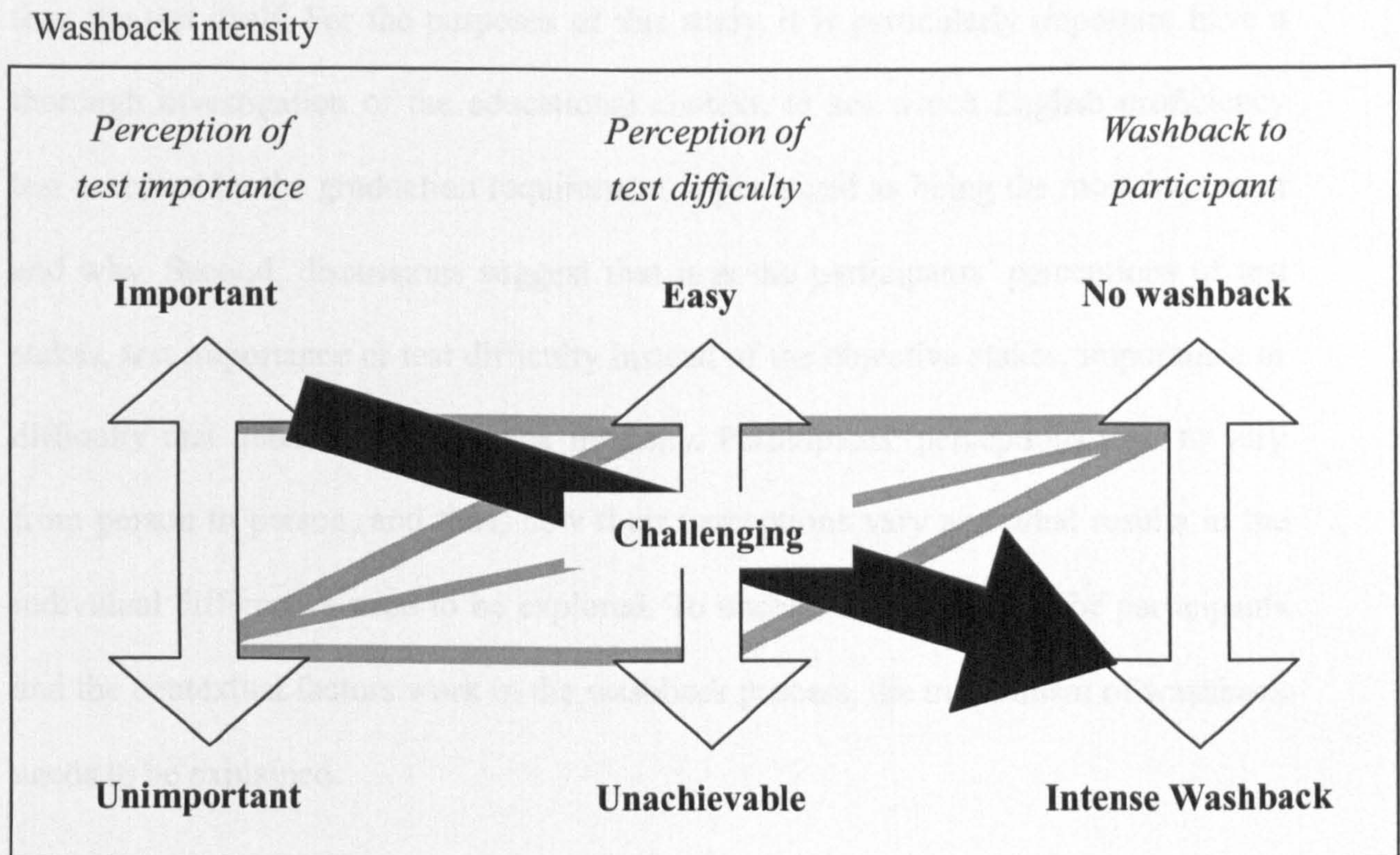


Figure 2.2 A model of washback intensity

(Green, *ibid.*, p.24)

Green (*ibid.*) suggests the level of washback intensity will be the highest where participants:

- 1) value success on the test above developing skills for the target language use domain;
- 2) consider success on the test to be challenging (but both attainable and amenable to preparation);
- 3) work in a context where these perceptions are shared (or dictated) by other participants (p.25).

The above discussions on washback intensity inform the present study in two ways.

First, the majority of the factors that Gates (1995) indicates as being linked with test importance are, in fact, the contextual factors a test entails. The reputation of the test developing organisation, the degree of monopoly a test holds and the opportunities a test score can provide are all more related to the context within which a test is used than the test itself. For the purposes of this study, it is particularly important have a thorough investigation of the educational context, to see which English proficiency test accepted by the graduation requirement is perceived as being the most important and why. Second, discussions suggest that it is the participants' perceptions of test stakes, test importance or test difficulty instead of the objective stakes, importance or difficulty that determines washback intensity. Participants' perceptions tend to vary from person to person, and thus, how their perceptions vary and what results in the individual differences need to be explored. To understand the role of the participants and the contextual factors work in the washback process, the mechanism of washback needs to be explained.

2.2.6 Mechanisms of washback and its conceptualisations

The fundamental step in investigating washback is to understand the mechanism of washback. The mechanism of washback unpacks how washback works and helps identify the contextual factors and stakeholders included in the washback process. The traditional view of washback, as represented by Burrows (2004) (Fig. 2.3.1), sees the relationship between testing and teaching as a linear, 'stimulus-response' relationship. It holds the assumption that washback is a definite by-product of a test and the quality of the test determines how the teachers will all have similar changes in their teaching under the test influence, be it positive or negative.



Figure 2.3.1 Traditional washback theory: A stimulus-response model.

(from Burrow's models of washback, 2004, p.126)

Alderson and Wall (1993) are the first to question the over simplistic assumptions. In their seminal paper "Does Washback exist?", they not only argue, as stated above, the non-linear relationship between the quality of a test and that of washback, but they also 'lay out the territory' for future washback studies with their 15 Washback Hypotheses (See below).

Washback Hypotheses

(1) A test will influence teaching. This is the Washback Hypothesis at its most general. However, a second partly different hypothesis follows by implication from this first one, on the assumption that teaching and learning are related, but not identical:

(2) A test will influence learning.

Since it is possible, at least in principle, to separate the content of teaching from its methodology, then we need to distinguish the influence of a test on the content of the teaching from its influence on the methodology. Thus:

(3) A test will influence what teachers teach; and

(4) A test will influence how teachers teach; and therefore by extension from (2) above:

(5) A test will influence what learners learn; and

(6) A test will influence how learners learn.

However, perhaps we need to be somewhat more precise about teaching and learning, in order to consider how quickly and in what order teachers teach and learners learn. Hence:

(7) A test will influence the rate and sequence of teaching; and

(8) A test will influence the rate and sequence of learning.

Similarly, we may wish to consider explicitly both the quality and the quantity of teaching and learning:

(9) A test will influence the degree and depth of teaching; and

(10) A test will influence the degree and depth of learning.

If washback relates to attitudes as well as behaviours, then:

(11) A test will influence attitudes to the content, method, etc. of teaching and learning.

In the above, however, no consideration has been given to the nature of the test, or to the uses to which scores will be put. Yet it seems not unreasonable to hypothesize:

(12) Tests that have important consequences will have washback; and conversely

(13) Tests that do not have important consequences will have no washback.

It may be the case that:

(14) Tests will have washback on all learners and teachers.

However, given what we know about differences among people, it is surely likely that:

(15) Tests will have washback effects for some learners and some teachers, but not for others.

(Alderson & Wall, 1993, p.121)

They tease out the complexity of the washback phenomenon by including test effects on different aspects of teaching and learning, and on the attitudes that teachers and learners will have concerning the changes. They also point out the positive correlation between test consequences and the stakes and the possibility that test will influence some but not others. Alderson and Hamp-Lyons (1996) elaborate on this possibility, suggesting that the amounts or types of washback on teachers and learners vary according to ‘the status of the test, the extent to which is counter to current practice, the extent to which teachers and textbook writers think about appropriate methods for test preparation and the extent to which teachers and textbook writers are willing and able to innovate’ (1996, p.296).

The original and the refined hypotheses can be partly illustrated by Burrows’ (2004) “black box” model (Fig. 2.3.2), which suggests that teachers will have individual, different responses to a test because of their beliefs, assumptions and knowledge. Burrow further proposes a “curriculum innovation model” (Figure, 2.3.3), suggesting that there may be patterns in the teachers’ responses to the introduction of a new test just like the models of responses teachers have under curriculum innovation in general education. However, Burrows’ models, unlike Alderson and Wall’s Hypotheses only take teachers into account.

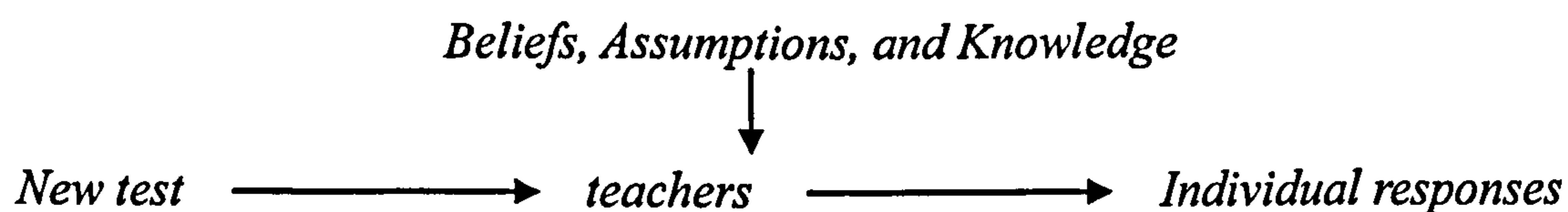


Figure 2.3.2 1990s view of washback: A “black box” model

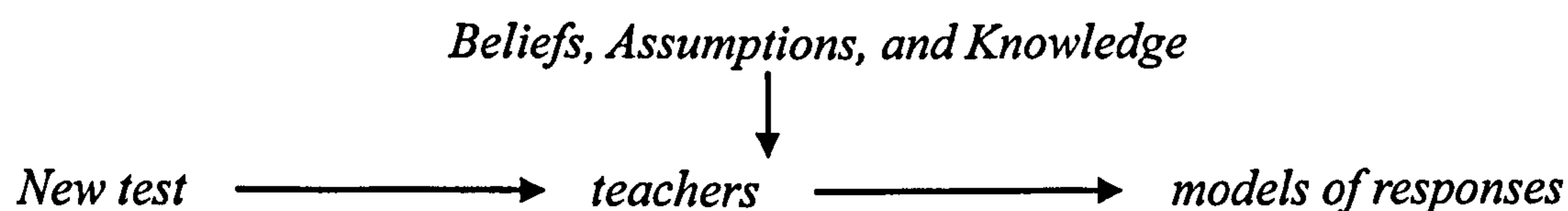


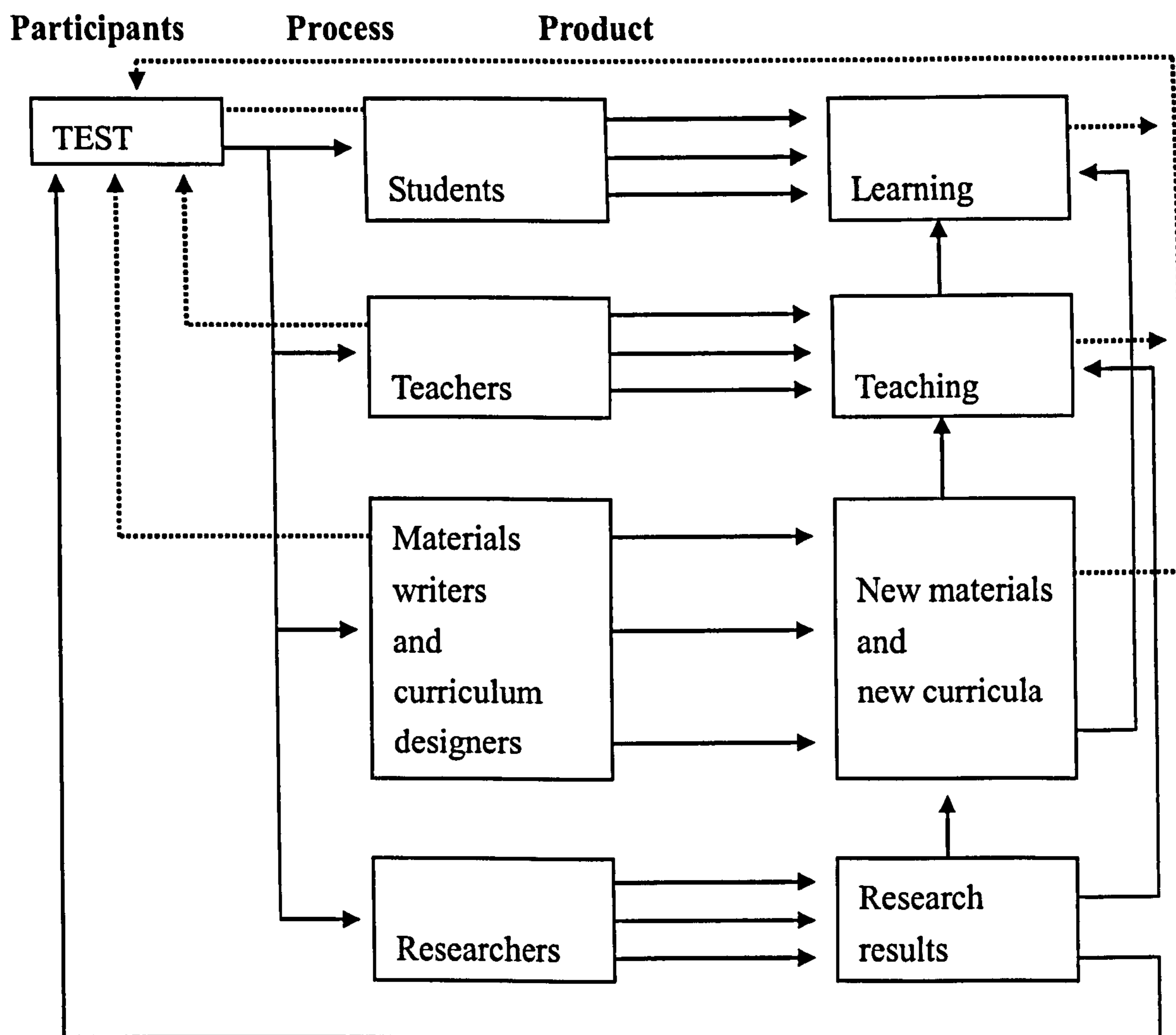
Figure 2.3.3 Proposed view of washback: A curriculum innovation model

(From Burrow’s models of washback, 2004, p.126)

Another influential conceptualisation of washback other than the Washback Hypotheses is Hughes’ (1993) trichotomy of participants, process and products in explaining the mechanism of washback in an educational context. Hughes (*ibid.*) defines the participants as “all of whose perceptions and attitudes towards their work may be affected by a test”, process as “any actions taken by the participants which may contribute to the process of learning” and product as “what is learned and the quality of the learning” (p.2). Hughes further explains how the trichotomy constitutes washback, as follows:

The nature of a test may first affect the perceptions and attitudes of the participants towards their teaching and learning tasks. These perceptions and attitudes in turn may affect what the participants do in carrying out their work (process), including practicing the kind of items that are to be found in the test, which will affect the learning outcomes, the product of that work. (*ibid.*).

Drawing on Hughes’ framework, Bailey (1996) proposes a model of washback (Figure 2.4).



**Figure 2.4 Bailey's model of washback
(1996, p.264)**

In her model, she specifically identifies the participants as students, teachers, materials, writers, curriculum designers and even researchers. She also identifies the type of products those participants will produce, and how other participants' products can feed into the ultimate product of washback, 'learning of the construct being measured' (Bailey, 1999, p.11). The most interesting aspect is the dotted lines in her model, signifying the possibility of "washforward" (Bailey, 1996, p.265, citing van Lier, 1989), of the influences participants and their products may in turn have on the test itself. Bailey also differentiates between "washback to the learners" and "washback to the programme". She specifies that 'washback to the learners' is limited to the influence of test-derived information on learners, but that 'washback to the programme' include

such influence on other stakeholders. She then uses the differentiation to examine Alderson and Wall's (1993) Washback Hypotheses and suggests that hypotheses 2, 5, 6, 8, 10 go under "washback to the learners", while hypotheses 1, 3, 4, 7, 9, 11 go under the other. Bailey's model places more emphasis on showing the interactions between the test, the stakeholders, teaching and learning. It does not specify the stakes of a test, or the similarity and difference among individuals and thus, the rest of Alderson and Wall's hypotheses (12 – 15) are not linked to the model.

The recent conceptualisation that takes into account the complex interaction between the stakeholders, test stakes, washback variability, washback intensity, washback direction and the test construct is Green's model of washback (2007) (See 2.2.3, 2.2.5 for details). Fig. 2.5 is the full model of washback Green proposes.

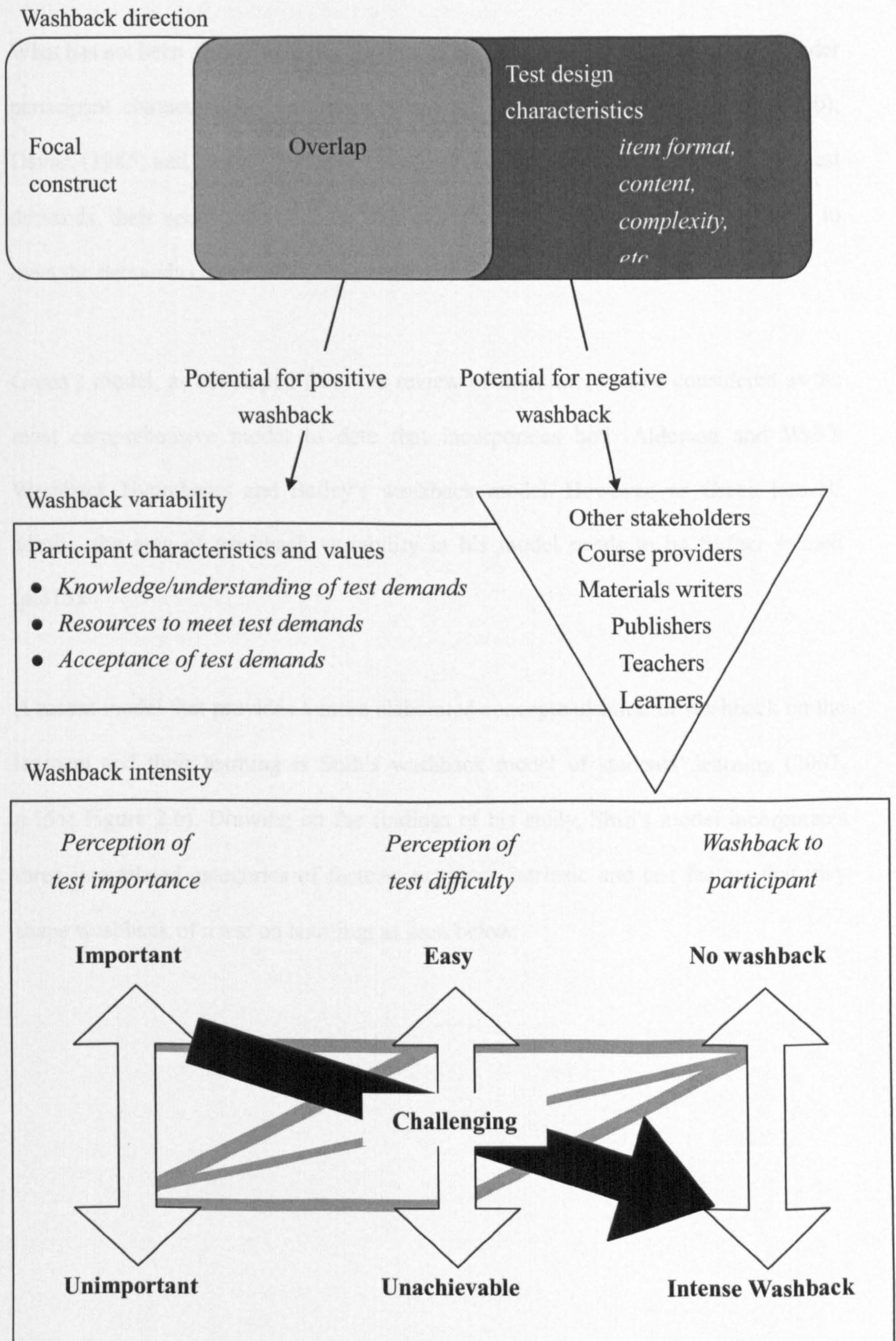


Figure 2.5 Green's model of washback
(2007, p.24)

What has not been discussed in the previous sections is the three points that come under participant characteristics and values. Drawing on Hughes (1993), Brown (2000), Davies (1985) and Smith (1991a,b), Green argues that stakeholders' awareness of test demands, their acceptance of those demands and the resources available to them to meet the demands can all affect the extent of washback realised.

Green's model, as developed from his review of literature, can be considered as the most comprehensive model to date that incorporates both Alderson and Wall's Washback Hypotheses and Bailey's washback model. However, as Green himself admits, the area of washback variability in his model needs to be further refined (p.315).

A recent model that provides a more elaborated conceptualisation of washback on the learners and their learning is Shih's washback model of students' learning (2007, p.151; Figure 2.6). Drawing on the findings of his study, Shih's model incorporates three interrelated categories of factors: extrinsic, intrinsic and test factors that may shape washback of a test on learning, as seen below.

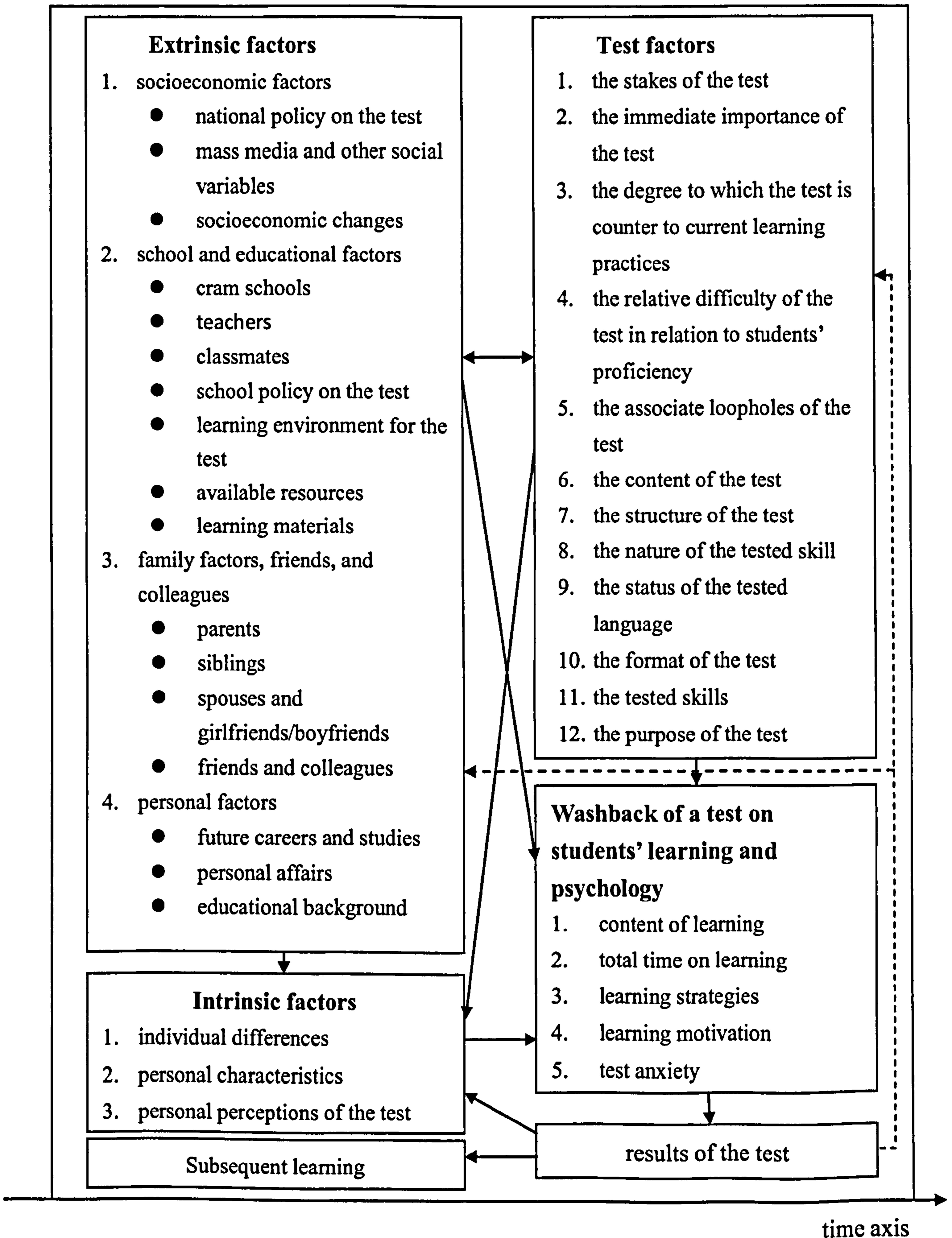


Figure 2.6 Shih's washback model of students' learning (2007)

Shih's model has indeed shed some light on the many possible factors that may result in the manifestation of washback. However, since the model has been developed, mainly based on the findings of his case study, compared to the categories of extrinsic factors and test factors, the category of intrinsic factors is much less developed. Thus, his model is also limited in the area of washback variability, particularly on the role of learners in mediating washback on their own learning. The present study takes account of the above mentioned conceptualisations of washback mechanism but attempts to explore more on washback variability among the learners.

This section, and the previous sections, discuss where the present study is situated in the washback and impact literature and also the theoretical issues concerning test influences. The following section then turns to reviewing related empirical studies and discussing how the findings of those studies inform this study.

2.3 Empirical studies of washback and impact

This review section centres on the washback and impact studies in language education. There are quite a number of such empirical studies conducted within different educational contexts. Those studies have shed light on the nature and the scope of washback but most findings are related to washback to the teachers and washback on teaching. Washback to the learners has received less attention, and thus there is insufficient evidence to understand in depth how learners' attitudes and behaviours may be affected by a high-stakes language test. Bailey (1999) argues that learners should be singled out from the other stakeholders because washback to the learners will affect their learning and the learning outcomes directly, while others will contribute to the processes involved in promoting language learning. Bailey's statement above indicates

that it is important to develop a more sophisticated conceptualisation of washback to the learners, taking into account washback to the other stakeholders, who may bring influences to the learners.

The central stakeholders in the investigation of washback to the programme are the teachers as they have more direct contact with the learners than other stakeholders. They also contribute most to promote their students' learning. Thus, the review below will first provide a synthesis of the findings on washback and the impact on teachers and teaching, and then to the limited findings of washback and impact on learners and learning.

2.3.1 Washback and impact on teachers and teaching

This review section of washback and impact on teaching focuses on the three most discussed aspects of test influences: teaching contents, the use of test-preparation materials and teaching methods.

2.3.1.1 Teaching Contents

In considering Wall's definition, teaching contents here refer to the type of knowledge teachers try to transmit to their students ('e.g. the form of a specific grammar structure, or facts relating to a particular topic') or to the general skill teachers focus on ('e.g. reading, listening') (Wall, 2005, p.16). Washback and impact on teaching contents specifically indicates the abovementioned knowledge or skill that seems to be related to a test.

Wall and Alderson (1993), in their landmark study on the introduction of the new

O-Level exam in Sri Lanka, conducted classroom observation to see whether there would be empirical evidence of washback of the exam on classrooms. They found out that much more time was spent on developing reading and writing skills, which were tested while untested skills like listening and speaking were paid less attention to, which they considered as negative washback. They also found evidence of the “narrowing of the curriculum”, whereby teachers spent most of the time on test preparation with the approach of the test.

The “narrowing of the curriculum” was also reported in study of an assessment-driven reform on teaching writing by Stecher *et al.* (2004). Through teachers’ surveys, there was evidence of the focus on tested content and format. Since the reform was a deliberate move, the washback effects were considered positive because the performance-based assessments with multiple-choice questions indeed increased student writing opportunities. However, teachers also reported the allocation of more instructional time on subjects tested, at the expense of untested subjects, and also on the focus only on the writing genres tested.

The relationship between stakes and washback on teaching contents is also reported in Shohamy *et al.* (1996). In the surveys and interviews conducted with the teachers, they reported the teachers’ claims of focusing their teaching exclusively on the oral skills and tasks tested in high-stakes English as foreign language exams and of allocating more class time to test preparation. Contrarily, little test preparation was done for the low-stakes Arabic as second language exam.

Difference in the degree of washback on lesson content can also result from school

difference. Read and Hayes (2004) studied IELTS preparation courses with different aims and structure. The course in School A was specifically aimed at IELTS test preparation, while the course in School B was more topic-based and had the aim of developing general and academic English skills alongside IELTS preparation. Through observation on the two courses, they found that test-related activities were held more in the intensive course in School A, while activities in School B were more balanced between the four skills and more effort was directed to developing learners' overall language proficiency.

Alderson & Hamp-Lyons (1997), from their study of TOEFL preparation, pointed out that differences in individuals may also cause test influences to differ. They observed two teachers teaching both TOEFL and non-TOEFL classes. They discovered that although there was evidence of washback on the content of the TOEFL classes, the differences between TOEFL and non-TOEFL classes were not as significant as the differences between the two teachers. Teacher A spent more time on test taking and used metalanguage more in his non-TOEFL classes than in Teacher B's TOEFL classes. They thus suggested that without the mediation of other contextual factors like administrators, materials writers and teachers, a test alone would not cause washback.

The extent of washback on test preparation can also vary because of the interaction between those who had more test awareness and those who had less, as demonstrated in Mickan and Motteram's (2008) ethnographic case study of an IELTS preparation program in the Australian adult education context. They documented classroom practice for 24 teaching hours (3 hours per week for 8 weeks) to investigate the teacher participant's classroom discourses, and found out that test preparation throughout the

whole program was dynamic, rather than static. What was incorporated in the test preparation program changed over time, as the teacher gained more awareness of the IELTS test, influenced by the researcher's presence and the discussions among them. This study provided evidence as to how participants' awareness of test demands can influence the degree of washback, and also highlighted the role of the researcher as an agent of impact that can provide the teachers with the necessary professional development to achieve intended washback of a test preparation program.

Even though the extent of washback on teaching content differs for different reasons, Cheng (1997, 2005) claims that the teaching content is "an area of high washback intensity" (p.50). In other words, teaching content has been influenced most by intended test change. In her study, she observed teachers teaching two cohorts of students, one taking the old exam and the other taking the new one (HCKEE in English language). The comparison between the two cohorts revealed that reading aloud activities, which were related to the old exam, were replaced by role play and group discussions, reflecting the new exam content.

2.3.1.2 Teaching materials

In this section, I will review studies that have discussed washback on the use of teaching materials, especially the contents of test-related teaching materials in the classroom.

Previous studies show that there was also evidence of washback on teaching materials. In the Sri Lankan case of Wall and Alderson (1993), the exam was intentionally designed to reinforce the textbook series launched earlier for curriculum innovation.

Thus, the study found that teachers relied on the textbook in the first two semesters and used past papers and commercial publications for test preparation when the test was approaching. Cheng (1997) reported similar findings. She noted that teachers followed the new syllabus 'simply by adherence to the new textbooks' (p.51) as the textbooks for the revised HKCEE were the most direct teaching support for them. She explained that the detailed teaching and learning activities with suggested time frames of a lesson might be a reason for the teachers' reliance on the textbooks. However, she argued that this reliance on the textbooks demonstrated more of a 'cosmetic change' rather than a substantial change in teaching as intended (2005, p. 122). From observations of teachers and school principals attending publisher seminars, she discovered that they preferred textbooks clearly labelled for the revised exam to those without, disregarding whether or not those textbooks reflected the changes in the HKCEE.

Alderson and Hamp-Lyons (1997) found that most teachers relied on the in-house TOEFL-oriented textbooks, and did little or no preparation for their teaching. It was perhaps because of their negative attitude towards the test itself and even teaching it that they did what the book asked them to, without considering whether it is a good way to teach TOEFL. The researchers thus argued that "it may be difficult to untangle test effects from textbook effects" (p.282).

The fine line between test effects and textbook effects was also evidenced in Chen's (2002) study of the Basic Competence Test (BCT) in English in the Taiwanese junior high school context. From the interviews with the teachers, she found out that the addition of oral and aural activities in teaching as the teachers claimed was not due to the BCT, in which listening and speaking skills were not assessed, but due to the change

of textbook content.

The study of Wall & Horak (2006) on TOEFL impact in Central and Eastern Europe also revealed the mediating role of preparation coursebooks in washback. The teachers interviewed claimed that their students expected them to use coursebooks and other than those books, there was insufficient training and resources for them. Since the coursebooks provided full coverage of what was needed in test preparation, they were often used as the syllabus, and the format and the content of the courses also derived directly from them. Despite the importance of the coursebooks, the researchers warned that they were not selected “because of their pedagogic value, but because of price or other pragmatic considerations” (p.112).

Some studies have gone further, to investigate how the commercial test preparation materials reflected test influences. Hamp-Lyons (1998), Hawkey (2006) and Tsagari (2006, 2009) developed a framework or an instrument for the systematic analysis of test preparation materials for TOEFL, IELTS and FCE, respectively. Both Hamp-Lyons and Hawkey found that the tests, indeed, exerted strong washback on these materials. However, Hamp-Lyons argued that reliance on the materials might lead to curricular alignment in a negative way, since teachers and learners might find themselves teaching and learning discrete chunks of language rules and vocabulary items without any specific context. Findings in Hawkey (2006)’s study, however, indicated that some teachers might not limit themselves to using only test preparation materials. Teacher surveys and classroom observations both revealed teachers’ use of materials from ‘within and beyond the textbook’ (p.112) The findings further suggested the importance of the teachers’ role in mediating the extent of washback on teaching materials in their

teaching.

Tsagari (2006, 2009) discovered that the teaching and learning materials varied in the way they reflected the FCE exam specifications. Some materials demonstrated positive washback by including a wide range of sources of input and language elements, while others showed negative test influences through the inaccurate reflection of some test features or the tight alignment to the exam, which could not fully represent the range of language skills or tasks needed at the level. From the diaries collected from learners (See 2.3.2 for details), Tsagari also found evidence stating that teacher washback was in fact a reshape of the FCE washback on teaching materials. Her teacher participants used the textbook in her lesson but incorporated extra techniques or structured the lesson in her own way. Thus, she argued that test washback might be mediated through teaching materials shaped by how publishers and writers perceived the needs of teachers and learners, and teachers then played an essential role in mediating between those materials and the learners.

The use of systematic analysis instruments for test preparation materials in the above studies yielded fruitful results that not only revealed the differences among test preparation materials, but also provided triangulation for both interview and observation data. Nevertheless, the present study did not adopt or develop an instrument as such to analyse test influences on teaching materials. The instruments were developed based on the test construct and design characteristics of one particular test, i.e. IELTS or FCE. However, as shown in 1.3.1, even though the GEPT might likely to be one of the most influential tests on the university students and their English curriculum, the GEPT was not the only test recognised by the graduation requirement

that could have washback on the teaching materials. It was thus not realistic to use an instrument restricted to the analysis of only GEPT washback.

The studies reviewed above have shown that washback on teaching materials can be found on the contents of the test-related materials and the use of those materials in teaching and learning. The studies have also suggested the central role of material publishers and teachers in mediating different degree of washback on teaching materials. However, they did not go on and explore whether washback on teaching materials have influenced students' learning.

Andrews et al. (2002) investigated whether washback mediated by published materials would be manifested in students' learning outcomes. They recorded three cohorts of students' performance of the Use of English (UE) oral examination and analysed the students' speech by a list of functions and forms derived from the textbooks. They further analysed and compared the frequencies of the language features in the list and the contexts in which they appeared among the three cohorts. It was found that the textbooks indeed mediated influences on students' learning outcome but at very superficial level, such as "familiarisation with the exam format, and the rote-learning of exam-specific strategies and formulaic phrases" (p.220). Andrews et al. disregarded the above as a meaningful internalisation, and argued that 'the students appear to have learnt which language features to use, but not when and how to use them appropriately' (p.221).

These studies reviewed in this section highlighted the mediating role of writers or publishers of teaching materials and the teachers in the washback process. However, in the context of the present study, the use of teaching materials seems to be more complex

than the use of materials in the abovementioned cases. To my understanding, the English department in each university may have different rules in selecting teaching materials, with some universities having uniform teaching materials for the same courses while other universities allow teachers to make their own choices (See rationale for case study in 3.5.1. See also 4.3). Thus, how uniformity and individuality in the choice of teaching materials inform washback on teaching materials, and also washback in teaching in the two case universities, should be considered. In addition, the deliberate choice of teaching materials related to a certain English proficiency test but not other tests stated in the graduation requirement may provide evidence of which test the decision makers (either English departments or teachers themselves) consider as the most relevant to teaching and learning in that particular context.

2.3.1.3 Teaching Methods

In the following section, studies of washback and impact on how teachers teach will be reviewed.

Shohamy *et al.* (1996) noted that under the influence of the high-stakes EFL exam, teachers incorporated simulations of test situations in their classes and used techniques that would help develop the exam skills. Stecher *et al.* (2004) found that teachers changed the way they taught writing after the Washington Assessment of Student Learning was introduced. The most significant changes were the “increases in the use of rubric-based approaches and in commenting on student writing in different content areas” (p.64). Saif (2006), in a rather different context, also discovered changes in the teacher’s teaching methods. Saif was involved in designing a high-stakes performance test based on the needs of a group of international teaching assistants. Her study was

about whether the aim of the test to bring positive washback on teaching and learning was reached. The teacher she observed adapted her teaching and her choice of activities to the test contents and its goal. The reason for this adaptation, as Saif speculated, might be because of her involvement in the testing development process:

the teacher's enhanced awareness of the test caused by her involvement in the test administration process, interaction with other raters, understanding of the rating process and the ability components of the rating instrument (*ibid.*, p,29).

However, the majority of studies on exam innovation reported it was futile to use exam innovation to bring intended positive washback on teaching methods. The textbook that the new O-level examination (Wall and Alderson, 1993) was intended to enforce had the aim to promote a more communicative approach in teaching and encourage more student participation in the normally teacher dominated classrooms. However, they discovered that no change had been found in how teachers taught after the introduction of the test throughout the two years of their research. They argued that factors such as resources, management practices, teacher training, teachers' beliefs and their commitments might prevent washback on teaching methods from appearing and a test itself could not "reinforce an approach to teaching the educational system has not adequately prepared its teachers for" (p.67).

In the Hong Kong context, the intended washback of changing its Certificate Examinations in English was to promote life-like, task-based teaching approaches. However, by observing three teachers teaching before and after the introduction of the HCKEE, Cheng (1997, 1999, 2005) reported that although there was significant test-related change in the teaching content, little change was found in terms of teaching

methods. In comparing the lessons in the two years of this research, the teachers carried out the lessons similarly. The interaction pattern of the lessons in both years and the amount of teacher talk did not reveal much change. Cheng (1999) further pointed out that differences were found “between teachers, not within the teachers themselves” (p.268). Thus, Cheng (1997) claimed that the manifestation of the intended washback was superficial, and that changing high-stakes public examination might only “change the form of teaching, and not the substance of teaching” (p.52).

Similar findings may be seen in the study of the National Matriculation English Test in China (Qi, 2004; 2005; 2007). The NMET was introduced with the purpose of pedagogical reform, so that a communicative approach and language use would replace the traditional focus of language forms in teaching English. Qi’s (2007) interviews with test constructors confirmed the intention of promoting writing in a communicative way. However, the teacher and student questionnaires and the observation of classroom behaviours all showed that teachers and students still focused on linguistic accuracy rather than the communicative context of writing expected by the test constructors. One important reason for the lack of washback on teaching methods was “the conflict between the selection function and the directing function” of the test (Qi, 2005, p.164). The function of the NMET in selecting students for tertiary education made teachers and students emphasise achieving high test scores and neglect the communicative approach the test was directed to. Qi also claimed that teacher factors such as their educational background might be another aspect that hindered the intended washback.

Watanabe (1996; 2004), in his research on Japanese university entrance examinations, stressed the role of teacher factors in mediating washback. He first observed whether

the translation tasks in major university entrance examinations resulted in the use of grammar translation method in classroom teaching. Like Alderson and Hamp-Lyons (1997) and Cheng (1999), the findings of his research revealed that differences between teachers were of greater importance than differences between courses. He concluded that teacher factors such as their beliefs, educational background and past learning experience might be the reason why washback didn't have the same effects on all teachers. Watanabe (2004) further explored the teacher factors in his later research on the same topic. He observed five teachers from three different high schools. The results added to his former argument of teacher factors as the mediating force of washback. He suggested that teachers' psychological factors, such as their concerns for students' proficiency levels, their biased perceptions (i.e. what they did in class based on their assumptions of test preparation might not be an accurate reflection of the actual test contents) and the degree of their familiarity with a range of teaching methods would mediate washback. He further suggested that teachers' psychological factors could be both 'debilitating' and 'facilitating' (p.141). He also suggested that school cultures might be another mediating factor.

Burrows (2004), had a similar view in her washback study on a new classroom-based assessment in New Zealand. She maintained that in order for curriculum innovation to work, teacher "variables" should be taken into account in implementation strategies. Drawing from the field of curriculum innovation in general education, Burrows claimed that patterns of teachers' responses to a new test can be placed on a continuum, from resister, the least affected by the change, to the adopter, the most deeply affected (See 2.2.3).

The above discussions suggest that washback seems to have the lowest intensity in teaching methods. Although tests used as the tool for curriculum innovation in teaching methods in most cases, how teachers teach has shown little change, or not as much change as intended. Findings of previous studies also point out that a number of factors such as resources, professional training, teacher factors and school factors may all be the reasons why intended washback has not occurred. In the present study, the exploration of washback on teaching methodology is peripheral to the understanding of the impact of the graduation requirement in the universities since curriculum innovation is not one of the purposes why it is implemented. Besides, none of the tests stated in graduation requirement is aligned to the English curriculum for the majors, there is much difficulty in relating teaching methodology to a number of tests with different test purposes and test constructs.

2.3.1.4 Implications for this study

The above sections (2.3.1.1, 2.3.1.2, 2.3.1.3) show that there are indeed a great number of empirical studies exploring washback on teachers and teaching. Only a small number of studies have attempted to cover washback to the learners, which will be described in further details below (2.3.2). This corresponds to the previous claim that the majority of washback studies put their focuses on the teachers, but not the learners.

Previous studies have provided evidence of ‘washback intensity’ (Cheng, 1997, p.43; 2.2.4) and ‘washback variability’ (Green, 2006b, p. 339; 2.2.3). Washback seems to be the most intensive on the format and the content of language classes, but few studies found changes in teaching methods under the influence of a test. Some claim that school factors or teacher factors may exert greater influences than the test *per se*

(Alderson & Hamp-Lyons, 1996; Watanabe, 2004) and it may be the reason why evidence of washback can be seen in some classrooms, not in others, even in similar contexts. This reflects ‘washback variability’.

The previous studies also provide methodological implications for the present study. What needs to be noted from the review is how research methods employed in those studies may have influenced the findings. Survey studies (e.g. Shohamy et al., 1996; Stecher et al., 2004) tend to find washback on teaching methods as commonly agreed by most teachers. On the other hand, those taking the more ethnographic approach (e.g. Watanabe, 1996, 2004; Burrows, 2004) and using classroom observation (e.g. Wall and Alderson, 1993; Watanabe, 1996; 2004; Cheng, 1999; Burrows, 2004; Qi, 2004, 2005, 2007) often reveal teachers’ different attitudes towards the teaching methods the test tries to promote and also reveal teachers’ different behaviours in the classrooms. Therefore, washback intensity and washback variability can be better captured using the ethnographic approach and conducting classroom observations.

2.3.2 Washback and impact on learners and learning

Since teachers, according to Bailey (1999), are the most visible participants for being the “front-line-conduits for the washback processes related to instruction” (p.17), many of the previous empirical studies have explored washback to the teachers while having learners as “peripheral” in the design of washback studies (Green, 2006a, p.114). Past studies which have attempted to reach out to the learners through questionnaires and interviews (Shohamy et al., 1996; Cheng, 1998, 2005; Ferman, 2004) are very limited in the scope of their exploration of learners’ perceptions, views of test preparation and their learning. Watanabe (2001) was one important study that probed into how learners

are affected by high-stakes testing, focusing on the relationship between test influence and learner motivation. The few latest studies (Gosa, 2004; Scott, 2005; Green, 2006a, 2007b; Tsagari, 2006) have further brought learners under the spotlight by making them the main participants and the important stakeholders in the exploration of the washback phenomenon. However, there are limitations in terms of the methods they have adopted, and also in the extent to which they investigate the factors that can explain the individual differences among learners.

Table 2.1 Overview of studies exploring washback and impact on the learners and their learning

Researchers	Context	Test	Research methods	Main findings
Cheng (1998, 2005)	Hong Kong	Revised Hong Kong Certificate of Education Examination (HKCEE)	Learner questionnaire	<p>a) The new integrated and task-based HKCEE does not have a fundamental impact on learning and the changes in learning remain superficial.</p> <p>b) Although learners report that exam scores have the strongest impact on their motivation to learn, their reasons to learn English reflect mainly instrumental purposes.</p>
Shohamy et al. (1996)	Israel	Arabic as second language (ASL) test and EFL Oral Matriculation Test	Learner questionnaire	<p>The learners were more anxious about the EFL Oral Matriculation test than the ASL test and more of them considered the results of the former could affect success in their future studies. The reason was because the EFL test was perceived as high-stakes while the ASL was low stakes and the perceptions of stakes reflected the status of the language.</p>
Ferman (2004)	Israel	EFL Oral Matriculation Test	Learner questionnaire	<p>a) The learners with the average ability level experienced the highest level of anxiety.</p> <p>b) Parental influence was also important in shaping washback to the learners, by urging them to learn for the test and employing tutors for them.</p>

<p>a) Washback on learners' motivation is a complex picture. Depending on test importance and test difficulty, learners' attitudes to test preparation and washback on their motivation on test preparation differed.</p> <p>b) The findings suggest that the test of the appropriate difficulty can have a positive effect on their motivation to prepare for the test. However, it is the learners' perceived difficulty of the test that counts, not the objective difficulty.</p>		<p>Learner interviews</p>	<p>University entrance examination</p>	<p>Japan Watanabe (2001)</p>
<p>a) The differences between the course contents the learners expected to receive and the course outcomes showed that the differences were not driven by learner expectations, but by teachers' adoptions of distinctive aims on different course types.</p> <p>b) Teachers and the courses may shape washback to the learners.</p> <p>c) Learners' perceptions of tests and test preparation may differ from those of teachers.</p>	<p>Learner questionnaire at course entry and exit</p> <p>Same questionnaire filled by the teachers</p>		<p>IELTS academic writing module</p>	<p>UK Green (2006b)</p>
<p>a) Learners' personal environment appeared to be influenced by BAC in a more substantial way than in the classroom.</p> <p>b) Learners play a major role in the presence or absence of washback and learner variables may explain individual differences among the learners.</p>	<p>Learner diaries analysed with both deductive (predicted washback effects) and inductive (grounded theory) procedures</p>		<p>Bac (Bacalaurate) Romanian school-leaving exam</p>	<p>Romania Gosa (2004)</p>

<p>a) Learners play an equally important role in the washback process as other stakeholders and their views, attitudes and feelings towards tests and test preparation may explain the presence or absence of washback.</p> <p>b) It is learners' perceptions of the degree of importance and test status instead of the test per se that motivate learners in test preparation.</p> <p>c) Learners are motivated to learn English for instrumental purposes than for the desire to communicate or integrate with L2 speakers.</p> <p>d) Washback is a dynamic and interactive process between the test, the textbook, teachers and learners.</p>	<p>Learner diaries</p>	<p>First Certificate in English (FCE)</p>	<p>Greece</p>	<p>Tsagari (2006)</p>
<p>a) Parents play an equally important role as the teachers in shaping washback to the young learners.</p> <p>b) There are aspects of washback specific to the EAL context, such as time constraints on addressing issues related to language and helping EAL learners to catch up with monolingual learners.</p>	<p>Case studies of learners: learner interviews, observations, interviews of teachers, parents</p>	<p>Statutory National Curriculum Assessment</p>	<p>UK</p>	<p>Scott (2005)</p>

2.3.2.1 Empirical studies on washback on learners and learning

Cheng (1998, 2005), Shohamy et al. (1996) and Ferman (2004) have extended beyond the use of learner questionnaire data in triangulating teacher's perceptions by exploring learners' perceptions on test influences on their learning.

Cheng (1998, 2005) investigated learner attitudes towards their learning in English and on the renewed HCKEE from two cohorts of learners in 1994 and 1995, through a questionnaire. The findings suggested that there had been limited washback of the new exam on students' learning as there was not much difference between the perceptions of the first and the second cohort of learners on whether or not they were affected by the public exam. Cheng thus argued that the changes brought by the renewed exam had been superficial. The study also discovered that learners' motivation to learn was mostly related to instrumental purposes. In addition, it was noted that although learners did not think exams could reflect accurately all aspects of their learning, exam scores still had the largest impact on their motivation to learn.

Shohamy et al. (1996) also used a questionnaire to explore learner's perceptions of test influences. They found that learners experienced a higher level of anxiety towards the EFL oral test than towards the Arabic as Second Language (ASL) test, as they perceived the former to be more important and hence higher-stake than the latter because the English language received higher status than Arabic in society.

In the same context as Shohamy et al. (*ibid.*), Ferman (2004) focused on the new version of the EFL matriculation oral test and investigated learners' perceptions of the test on their learning. The questionnaire data showed the following findings. First, the learners did self-learning for the test to compensate for the parts of the test not studied

in class. Second, the extent of teaching and learning towards the test was dependent on the ability of the learners. The weaker the learners' ability, the more there was test preparation. However, the extent of test preparation did not equal to the level of anxiety the learners experienced. Another interesting finding was that learners with the average ability had the highest anxiety level and were more likely to suffer from potential test failure. The last finding concerning the learners revealed how parents could shape learner washback in urging them to learn for the test or employing tutors to help them.

Although the above studies indeed attempt to draw in learners' views of test influence and test preparation, learners are still considered to be the peripheral participants in their studies. The majority of items in the learner questionnaires are still largely devoted to the triangulation of the learners' perceptions with the teachers' perceptions. In addition, the limited but varied findings of those studies suggest that learner washback, as with washback on teaching aspects, can be very complex. Therefore, I believe it is inadequate to use only a few items to elicit learner perceptions if one is exploring the complex ecology of washback on learners and their learning.

Watanabe's (2001) study of Japanese university students on their test preparation practices (which is part of his larger washback study of Japanese university entrance examinations) is one among the few studies that focused on learner washback. He was particularly interested in how learner motivation came under the influence of a test. Through interviews with the students, Watanabe found that the relationship between motivation and test preparation was complex. Students' attitudes to test preparation and the impact of the university exams on their motivation on test preparation varied, depending on the importance and the difficulty of the exam. Watanabe concluded that test of the appropriate difficulty to the learner can have a positive effect on their

motivation on test preparation and what is important is that learner's perception of difficulty of the test, not the objective difficulty, may be what cause washback.

Watanabe's study is significant in shedding more light on learner washback, especially on the complexity of test impact on learner motivation for test preparation. However, learner motivation to prepare for a test can only be considered as a very small dimension of the washback on learners and their learning. The following studies are those to date that have pulled the learners to the centre and explored to a fuller extent the influences of tests on learners and their learning.

Green's (2006b, 2007) study of IELTS preparation attempted to link washback to the learners to the learning outcome, namely their IELTS results. He gave the Chinese students on IELTS preparation and pre-session courses a questionnaire at entry, asking about the expectations they brought to their courses, and another questionnaire at the end, to ask for their retrospective perceptions of the course with an undertone of evaluation. Students from both courses had a shared concern with formal register and effective written communication, and did not expect IELTS-like tasks to be the major focus of their courses. All of the students placed greater emphasis on developing their writing skills than on preparing for the test. However, students from the two courses differ in their retrospective accounts of the courses. The students' experiences reflected the fact that the courses were shaped by what teachers viewed as priorities, rather than by the students' expectations. Green (2006b) concluded that these learners' accounts seemed "to reflect to a far greater extent shared beliefs about learning" (p.130), yet, what they had learned differed because of teachers' intentions to differentiate the two courses. Green suggested that washback to the learners may be mediated by the teachers and the courses. He also suggested that since perceptions of the learners on

tests and test preparations might be quite different from those of the teachers, both should therefore be taken into consideration in examining how the interaction between teachers and learners constitutes washback to learning. Green's study indeed drew a more extended picture of washback on learners and their learning, by including the teacher factors, the time factor (i.e. at the beginning and at the end of the courses), the learners themselves and how those factors together contribute to the learning outcome. However, the use of questionnaires, as Green himself acknowledged, which is more appropriate in his study in building a statistical model for comparison between two groups of students, is limited in its ability to provide an in-depth understanding of the complexities inherent in washback on learning. Thus, to probe deep into this matter, qualitative methods such as interviews and classroom observations should also be used to capture the subtlety and complexity of how washback on teaching interacts with washback on learning.

Instead of using questionnaires, Gosa (2004) and Tsagari (2006) went further towards the qualitative end by conducting diary studies in the Romanian and Greek contexts respectively. Both of them asked their learner participants to keep a learning diary over an academic year. Both studies revealed that learners varied in their reactions towards the tests, and the individual difference might result in the presence or absence of washback. Gosa's (2004) study of the Bacalaureate (Bac) concentrated more on the individual dimension, while Tsagari's (2006) three-phased study of the First Certificate in English (FCE) linked teaching, teaching materials and learning together. Gosa found that learners also played a major role in the presence and absence of washback, and that learner variables such as emotional or cognitive factors or their beliefs rooted in their cultural and educational backgrounds. Tsagari, on the other hand, not only confirmed the important role of the learners in shaping washback on their own learning, but also

explored further the relationship between test and learner motivation. She found that it was the status and the importance of the test as perceived by the learners, rather than the test itself, that motivated them to prepare for the test. In addition, learner motivation on English learning was stimulated more by instrumental purposes instead of purposes such as communication or integration with L2 speakers. Thus, she argued that it was unlikely for such type of motivation to be sustained once the stimulus, the test, was over. Tsagari also went further, to propose a model of washback that attempts to explain the circuitous nature of washback that involves the interaction between the test, the textbook, teachers and learners.

The use of diaries in the two studies is useful in collecting data relating to the washback phenomenon over time, as opposed to a 'snapshot' approach used by previous studies (Tsagari, 2006). In addition, the data generated by students' introspective thoughts, without the researcher's intrusions, provides depth and richness and this quality is helpful in accentuating individual differences among learners. Nevertheless, there are several pitfalls with the diaries: limited number of participants, uncontrollability of diary contents and difficulty in differentiating washback from influences of other factors. The abundance of data diaries may generate and the voluntary willingness to keep a diary makes it difficult to involve a large number of students (Gosa = 10, Tsagari = 29). As diary keeping in the studies is not a highly-controlled activity, diarists differ as to what they chose to write or to omit and thus, there is difficulty in determining if absence of predicted washback effects is the result of no washback or just plain omission. Furthermore, as both researchers admit, to identify features that might count as washback can be a tricky and complicated matter. This is especially difficult in Gosa's case, as the learners had to take the university entrance exam as well as the Bac and the specifications of the two tests did not differ vastly.

The other limitation of both studies is the lack of other forms of data (Tsagari, 2006, p.318). Student diary is the only method used by both Gosa (2004) and Tsagari (2006) in studying washback to the learners. The data could only be received at the end of their projects. Therefore, the researchers' later inquiries on the diary contents could not be answered and the weakness of diaries as uncontrollable magnifies. In addition, the influence of washback on teaching on learner washback found in Green's (2006a) study with both teacher and student data, will not be captured by the sole use of student diaries.

Scott (2005)'s case study of two primary school learners in the English as Additional Language (EAL) context has not only taken account of the perceptions of related stakeholders, along with the perceptions of the learners, but also probed further into young learners' actual learning in their classrooms. Her study suggested that in addition to teachers, parents also played an essential role in shaping washback to the learners and washback on their learning, especially young learners. This finding corresponded to that of Ferman's (2004) study in Israel, in which the parents might determine the extent of washback on their teenage children's learning towards the English oral test. From the analysis of data collected from the two EAL learners including their perceptions along with perceptions of teachers and parents, their involvement in class and their test papers, Scott also found out aspects of washback specific to the EAL context, such as time constraints, in addressing language issues, providing scaffolding through negotiation and focus on form, and having the EAL learners catching up with monolingual English peers. The investigation of learners' performance through classroom-based assessment was able to provide an in-depth exploration of washback on learning, especially for the young EAL learners, as test scores might not be a reliable measure of their learning. However, the study was limited in terms of a small number

of cases; unlike the above mentioned diary studies, the individual differences between students in the same context could not be explored.

2.3.2.2 Implications for this study

The above review informs the present study in several ways. First, in the studies reviewed above, 'learning' was conceptualised in different ways and different aspects of learning were investigated. Some studies focused on learners' perceptions of washback, while others considered learning as learners' performances, the products of learning. Aspects of learners' perceptions of washback and what counts as learner performance also varied from study to study. Therefore, I believe it is very important to state upfront what aspects of learning or learner perceptions are being investigated in the present study (2.5.3, 5.6.3). Second, similar to Alderson and Hamp-Lyon's (1997) finding in their TOEFL study, Green's (2006b, 2007) study suggests that there is discrepancy between the accounts of teachers and learners. Teachers' beliefs and claims as to what learners want and how they feel do not always correspond to students' needs and feelings in reality. Other studies also showed that learners, like teachers, may also manifest 'washback variability'. They may have different responses toward test change or the implementation of a new test because of their backgrounds and beliefs. Thus, the perspectives of learners should not be neglected in order to understand a fuller picture of washback. The review also showed that washback on learning may be influenced by washback on teaching, which include the perceptions and attitudes of other stakeholders. Accordingly, washback on learning should not be studied only through learners. The present study investigates washback at the classroom level, and thus, at least, investigation on the teachers and the school contexts should also be brought into the research framework.

Third, although these studies, especially those which use qualitative methods such as interviews and learner diaries, show washback variability among learners and argue for learner variables to be considered in the presence or absence of washback, there is a lack of studies to systematically explore the learner variables. Gosa (2004) has suggested possible variables, ranging from personal, emotional or cognitive differences to learners' beliefs in relation to their cultural and educational background. However, her suggestions were made out of speculation instead of empirical data since she did not interview her diarists after her analysis of the diary entries. Thus, in an attempt to fill this gap, the present study will investigate one particular learner variable, namely students' learning power (see 2.5), to examine washback variability

Lastly, the above review also informs the methodology of the present study. In acknowledging the limitations of using only one method in studying washback on learning, either by diaries or questionnaires, a mixed method approach incorporating classroom observation, questionnaires and interviews is adopted in this study. Due to time constraint, the study is not longitudinal and thus, collecting learner diaries within a short period of time will yield less fruitful results. Semi-structured interviews and classroom observations are capable of producing rich data as diaries but with more control at the researcher's end. The observation data and the teacher interview data serve as the primary source for the exploration of washback on teaching (Chapter 4) while the learner interview data and their learning power as measured by ELLI (Effective Lifelong Learning Inventory) will be used to relate to the former are used mainly for the exploration of washback on learning (Chapter 5).

2.4 Studies related to the educational and research context in this study

In this section, recent studies in Taiwan related to the educational and research context of the present study will be reviewed, including studies that have investigated GEPT washback and impact, and also studies related to the graduation requirement for English proficiency in Taiwanese universities.

Wu and Chin (2006), from Language Training and Testing Centre, the centre that developed the GEPT, reported part of an impact study of the intermediate-level GEPT on teaching and learning in Taiwanese senior high schools. The study reported findings of 11 teacher interviews and 71 teacher questionnaires as preliminary piloting, triangulated by information obtained from classroom observations (not reported in details in this study). The findings suggested that teachers had positive views about the implementation of the test. It was considered to be able to raise the importance of listening and speaking in the English curriculum of the first and second year. This was considered as evidence of washback because listening and speaking were not tested in the English section in Subject Competence Test (SCT-E) and Assigned Subject Test (AST-E), two high-stakes tests directly related to university entrance. In addition, students with higher proficiency level were perceived to be more motivated to prepare themselves for the test. Nevertheless, observations of teaching in the classrooms revealed that there was in fact limited GEPT washback since little time was given in class for direct GEPT preparation, due to largely the constraints on instructional time and the much stronger impact of the SCT-E and AST-E.

The following two studies were conducted in the same context of the present study, with the former focusing only on the GEPT washback on university English teaching

and learning and the latter on a more general reference to the adoption of standardised English language proficiency tests as requirement for graduation.

Shih (2006, 2007, 2009, 2010) conducted a case study of two universities of technology, one with the 'GEPT requirement' while the other without. He investigated GEPT washback on teaching and learning, with a particular focus on English majors. He adopted only qualitative methods including interviews, classroom observations and review of documents and records regarded this graduation requirement. He interviewed department chairs, two or three teachers, 14 to 15 English majors, and some of their family members. For observation, he not only observed one GEPT-related class for each teacher interviewed, but also observed activities in the self-study centres. The analysis of the department meeting minutes and the interviews with department chair and teachers revealed that there were complex factors that influenced the decisions on implementing the GEPT requirement (see Shih, 2010). In the university with the requirement, he discovered that there was GEPT washback on teaching but the extent of washback varied from teacher to teacher. Washback variability on learning was also evident in his findings. However, the degree of washback was limited in both universities as very few students had long term preparation for the GEPT. Shih argued that there were many factors that might determine the extent of the washback on the students' learning. He categorised the factors into extrinsic factors, intrinsic factors and test factors. The extrinsic factors included socioeconomic, school and educational factors, family, friends and colleagues and personal factors. Intrinsic factors referred to learners' individual differences (i.e. varied reactions to tests), personal characteristics (i.e. personalities or other inherent characteristics) and personal perceptions of the test under study. The most elaborated were those factors directly related to the test, encompassing test stakes, immediate importance, difficulty, content, structure, format

and test purpose. Shih argued that these three categories of factors and the interaction between them mediated washback on learners' learning and psychology, which further affected the learning outcome.

The latest study to date is Tsai and Tsou (2009). They probed into learners' viewpoints on the adoption of standardised English language proficiency tests as a tool to assess their English competence for graduation. They also chose to conduct the study in a 'technical university', which is equivalent to what Shih (*ibid.*) refers to as universities of technology. As distinct from Shih, they collected questionnaire data from 520 university students of different schools, including the school of humanities and management, nursing, environmental and life science, and medical and health science. One important finding was that the majority of the learners considered using only such tests in evaluating their English proficiency for graduation as inappropriate. They did not consider the tests to be the best tool to assess their proficiency and evaluate the effectiveness of the English instruction. Another finding was that most students perceived that English requirements for graduation would make classes become test-oriented, only enhancing their test-taking skills instead of communicative competence. They also found that learner's attitudes towards the requirement, the burdens of pressure they perceived the requirement brought onto them and self-perceived proficiency levels would determine whether they considered the tests as an appropriate assessment or not. Thus, Tsai and Tsou suggested that instead of tests as the only tool, there should be multiple measures in assessing university students' English proficiency for graduation. They also suggested that teachers should encourage learners who are more motivated, and who have a higher proficiency, to take standardised English proficiency tests for their own benefits.

Both Shih (2007) and Tsai and Tsou (2009)'s studies have shed light on university students' views and perceptions of the GEPT and the graduation requirement for English proficiency; both studies chose to focus on students in universities of technology. Nevertheless, as Tsai and Tsou argue, students in those universities are less confident in English and have insufficient training since they provide more professional training and focus more on English for Specific Purposes than general English language proficiency. It is also important to consider what students in general universities may think of the English graduation requirement. Shih's study is also limited, in that he only included English majors in his studies. He acknowledged that the limited GEPT washback might be because it was difficult to differentiate between English majors' regular English learning for their courses and preparation for the GEPT. Most English majors only intensified what they did regularly, without much additional preparation. However, this is not the case with non-English majors, since in most universities, non-English majors only have one to two years of compulsory English courses, unlike English majors' compulsory exposure to English over their four years of study. Tsai and Tsou do not limit their participants to English majors, but include university learners from a wide range of backgrounds. However, the questionnaire they designed contained only 9 items, which is thus limited in depth in its exploration of learners' perceptions of the graduation requirement and the impact it may bring, or have brought to them.

The present study attempts to fill the gaps in the above studies by investigating the impact of the graduation requirement for English proficiency in Taiwanese general universities, with a particular focus on non-English majors. Instead of a questionnaire with a limited number of items, semi-structured interviews are conducted for more in-depth investigation of non-English majors' perceptions towards the graduation

requirement (3.7.4).

Furthermore, this study has attempted to explore how learner variables can be used to explain the learners' varied perceptions on the impact of the requirement and test influences. Learner variables in this study are related specifically to the assessment of learning power as represented by the Effective Lifelong Learning Inventory (ELLI) (Deakin Crick et al., 2004).

The next section will centre on rationalising the use of the ELLI in the present study. The construct and the dimensions of learning power, measured by the ELLI, will then be provided in detail. A study that has attempted to link learner's learning power with test impact will also be reviewed.

2.5 Approaches to Learning and Assessment: The Use of Effective Lifelong Learning Inventory (ELLI)

This section (2.5.1) begins first with situating the use of the ELLI in the discussion of the relationship between assessment and approaches to learning. The section then continues with the construct the ELLI measures: the concept of learning power (2.5.2), the seven dimensions of learning power (2.5.3) and also the use of ELLI in the discussion of test impact (2.5.4). (See Appendix B for the ELLI).

2.5.1 Approaches to Learning and Assessment

When discussing learning, learner's learning styles (i.e. how they prefer to learn) have been considered as important variables that show differences at an individual level.

Coffield et al. (2004)'s review of inventories measuring learning styles, classified the different inventories on a spectrum based on 'the degree to which the underlying

assumptions about learning styles were ‘fixed’ (Deakin Crick and Yu, 2008, p. 388). At one end of the spectrum was the assumption of fixed learning styles. The ELLI, according to Deakin Crick and Yu (*ibid.*), is located at the other end of the spectrum, representing the assumption that learning styles are changeable instead of fixed. At the same end of the spectrum is the concept of approaches to learning and the inventory Entwistle (1988) has developed.

Approaches to learning consist of three types of approaches with embedded motivation:

- deep approach (focus on understanding, motivation: interest in the subject matter)
- surface approach (focus on recall and reproduction, motivation: fear of failure)
- strategic approach (focus on both academic contents and demands of assessment, motivation: competitive achievement)

(see further definitions and features of the approaches in Entwistle, 1988 and Entwistle et al., 2001).

The relationship of assessment and approaches to learning has been widely discussed in the general education literature. Several instruments measuring the concept of approaches to learning including Entwistle’s have been adopted and adapted to study how assessment affected students’ approaches to learning. Studies such as Entwistle and Entwistle (1992), Scouller and Prosser (1994), Scouller (1998), and Thomson and Falchikov (1998) have all discovered that assessment indeed shaped students’ approaches to learning. For example, Scouller’s (1998) study has discovered that students were more likely to adopt surface learning approaches for exams of multiple choice questions but for assignment essays, deep learning approaches were employed

instead. The different approaches they employed for different types of assessment were related to their perceptions towards the levels of cognitive processing needed for each type of assessment.

The inventories or instruments in such studies have shed light on how assessment affects approaches to learning. However, less attention has been given to how learning styles, approaches to learning or similar concepts might explain how individual learners experience washback. Previous studies of washback on learning have pointed out how learner variables may shape washback on the learners and their learning but few have focused on one particular learner variable. In this study, the ELLI, which measures one specific learner variable, learning power, will be employed to discuss how such variable can explain the varied perceptions of learner washback among learners. The construct of which the ELLI is developed upon and the dimensions of learning power it measures will be described below.

2.5.2. Learning how to learn and learning power

The ELLI is developed based on the concept of learning how to learn and learning power. The concept derives from the recent shift of education from merely teaching learners knowledge and skills to equipping them additionally with aptitudes and attitudes to become good real-life learners (Carr and Claxton, 2002). The reason for the paradigm shift in education towards a 'relational and transformative model of learning' (Deakin Crick, 2007, p. 137) is due to the contextual challenge of the information age to the traditional acquisition, mastering and application of knowledge. Learning in this era concerns not only 'know-what', the knowledge itself, but also 'know-how' and 'know-why', the context, the purpose and meaning of learning (*ibid.*). In order to meet the demands of the networked society, it is important for the learners to have a sense of

themselves as learners, and also to show flexibility, creativity and dynamics in learning by being able to adapt what they learn and extend their learning in the real world.

The concept of learning how to learn can be explicated by Deakin Crick's (2007) definition attempts to provide a broader framework that captures its complexity:

Learning how to learn involves the person who is learning, and requires motivation, a sense of direction and desire, and a sense of agency and self-regulation. This implies a sense of time and direction: a person chooses a particular goal, or desired outcome which is achieved over time. (p.138)

In short, an important feature of learning how to learn is not a learner's capacity to learn in a short period of time, but the ability to evolve as a better learner over time, and hence, the capacity for life-long learning. The ELLI is thus developed with the goal of fostering learners with such capacity in learning which can help them meet the demands in this ever-changing world.

What is measured by the ELLI is a learner's 'personal power to learn', or 'learning power'. Carr and Claxton (2002) argue that a learner's learning power consists not only of learning capabilities but also learning dispositions. Even though learning capabilities are necessary in providing learners with the skills and abilities required by learning, it is not sufficient for learners to be good ones if they are not disposed to learn, to be willing to take opportunities in learning. Expanding Carr and Claxton's argument, Deakin Crick states that the concept of learning power must be understood and contextualised as 'part of a complex system in which the formation of a learning identity, personal power to learn and competencies for managing life in the post-mechanical age are as important as the acquisition of knowledge' (p.136). Learning power described in the

current study is a form of consciousness or critical subjectivity (Heron and Reason, 1997, in Deakin Crick, 2007), characterised by a set of dispositions, values and attitudes, which is part of a complex learning journey, 'with a lateral and temporal connectivity' (Deakin Crick, *ibid.*, p. 138). In other words, the dimensions of learning power measured by the ELLI are the dispositions, values and attitudes that portray learners' self-awareness and intentionality in learning, with learning taking place in the context of learning relationships over time. The complexity of the concept lies in the fact that learning power, on the one hand, is deeply personal and autogenic as it reflects backwards to the learners' identity, desire and motivation. On the other hand, it also reflects forward as mobilised scaffolding towards the acquisition of skills, knowledge and understanding for learners' development of competency. See Figure 2.5 for the illustration of such complex and embedded description of learning power as characterised by dispositions, values and attitudes.

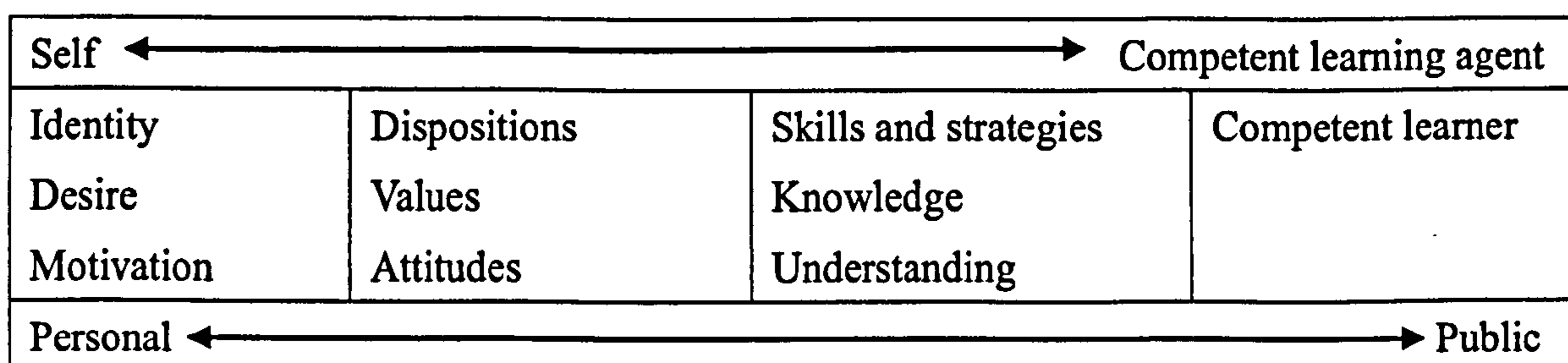


Figure 2.7 Complex and embedded description of learning power and learning dispositions

(From Deakin Crick and Yu, 2008, p.389)

2.5.3 The seven dimensions of learning power

The seven dimensions of learning power ELLI measures and the items in the inventory are derived from Deakin Crick et al.'s (2004) exploratory factor analysis study. In the study, the items for ELLI were designed based on Harlen and Deakin Crick's (2003)'critical review of studies on the impact of testing on students' motivation for

learning. The review has presented a conglomerate of variables that can affect individual's capacity and motivation to learn, hence, the potential dimensions of learning power (See Figure 2.6). The variables are related to:

- a) what a learner feels and thinks about oneself (e.g. "self-esteem, self-concept, sense of self as a learner, attitude to assessment, test anxiety, learning disposition")
- b) what drives a learner to undertake the task (e.g. "effort, interest in and attitude to subject, self-regulation")
- c) how a learner perceive one's capacity in undertaking the task (e.g. "locus of control, goal orientation, self-efficacy") (p.182)

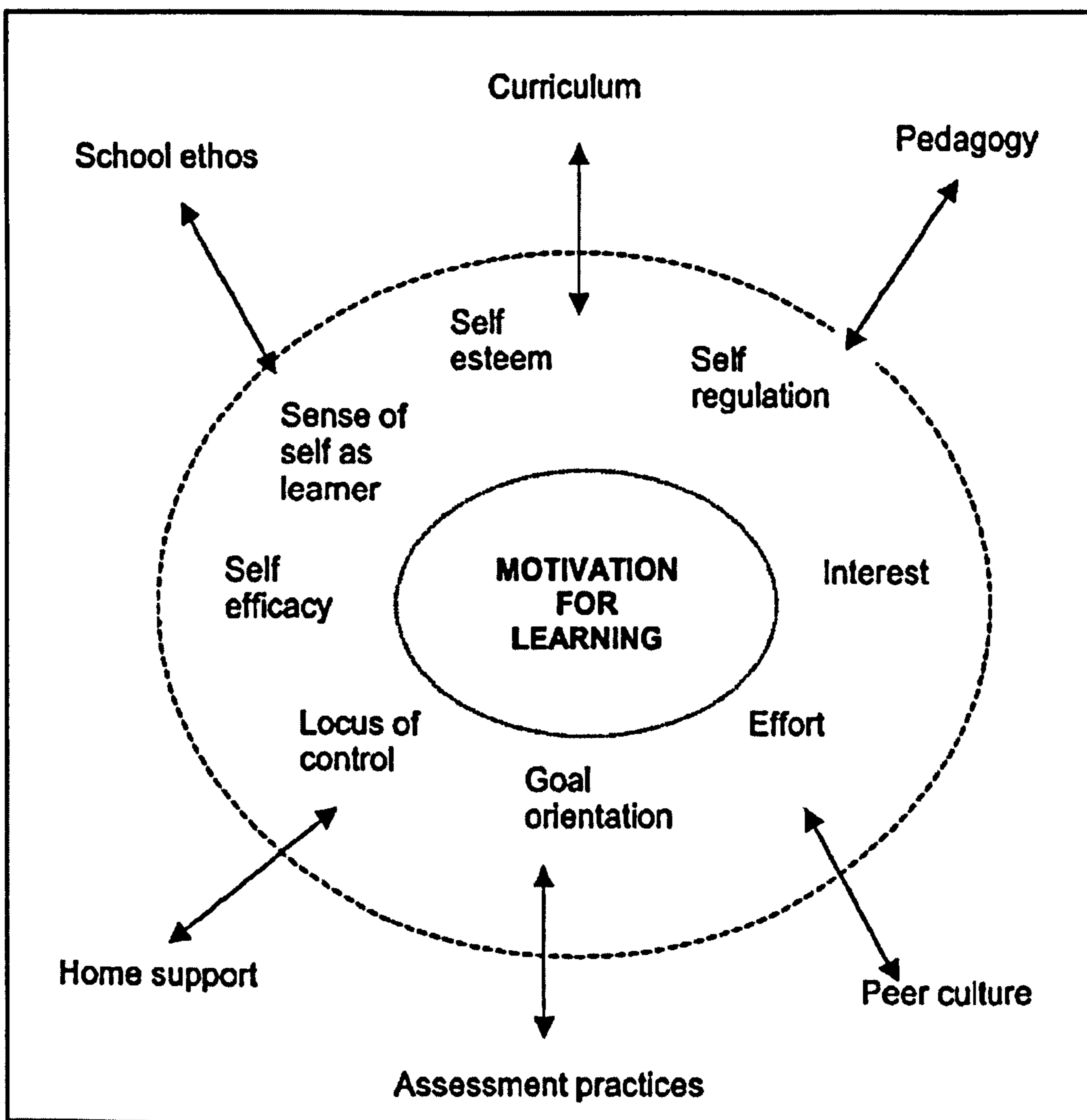


Figure 2.8 A map of some of the variables integral to motivation for learning and learning power (Harlen and Deakin Crick, 2003, p.183)

The exploratory factor analysis study that Deakin Crick et al. (2004) conducted on the items came up with seven dimensions. The seven dimensions of learning power the ELLI identifies will be described in detail below. For each dimension, a sample item in ELLI that measures that particular dimension will also be provided. Further information of the number of items relating to the seven scales and sample item for each scale is provided in 3.8.5, Table 3.4.

Changing and learning: Effective learners think of learning itself as learnable. They believe that similar to the growth of their bodies, their minds can get bigger and stronger. They have the energy to learn as they gain pleasure and self-esteem from expanding their ability to learn. With a sense of history and hope, they are able to grow, change and adapt as learners, which make them better at learning over time. The opposite of changing and learning is 'being stuck and static'. Less effective learners believe that their learning ability is fixed and cannot be expanded. 4 items are included in this dimension. A sample item is:

Q32. I'm continually improving as a learner.

Critical curiosity: Learners who have the desire to question 'received wisdom' and get at the truth underneath the surface of things are effective learners at this dimension. They are more likely to challenge what they are told and they prefer coming to conclusions about things on their own. They are more willing to show their uncertainties and doubts in public. On the contrary, some learners show more 'passivity' in learning. They can be less thoughtful of what they are told and are more likely to believe in the 'received wisdom' without active speculation. There are 8 items, including the following:

Q6. Getting to the bottom of things is more important than getting a good mark.

Meaning-making: Effective learners enjoy knowing how they can relate new things they learn to their previous knowledge because they gain pleasure from seeing how new learning fits within the big picture. They also enjoy learning about what matters to them. The opposite pole is ‘fragmentation’ (Deakin-Crick, 2007) or ‘data accumulation’ (Deakin-Crick and Yu, 2008), referring to how some learners tend to view each learning situation piecemeal, without looking for the associations between them. There are 7 items, including the following:

Q18. I like it when I can make connections between new things I am learning and things I already know.

Dependence and fragility: Dependent and fragile learners are more easily frustrated and discouraged when they get stuck or make mistakes. They prefer less challenging situations and are less able to persevere. They depend on others for their learning as ‘they are passive imbibers of knowledge, rather than active agents of their own learning’ (Deakin Crick, 2007, p. 141). The contrast is learners with ‘resilience’, who are willing to take on challenges even if they are not certain about how they should go through the challenges and also the outcomes. . They embrace risks in learning even though they encounter frustration and anxiety. There are 17 items, including the following:

Q47. Sometimes when I start a task I don’t know what I am going to do until I see my friends getting on with it.

Creativity: Effective learners are able to take different perspectives when they look at things. They enjoy making use of their imaginations, visual imagery, pictures and diagrams while they learn. They are less restricted to rules and believe that playfulness is as important as purposeful, systematic thinking in learning. The contrast pole is

'rule-bound'. These learners have less sense of security when more creativity is required because they prefer to be provided with clear-cut information and routines. There are 10 items for this dimension, including the following:

Q 67. I like to imagine how other people might feel and think about things.

Learning relationships: Learners who have higher scores on this dimension are good at 'managing the balance between being sociable and being private in their learning' (Deakin-Crick, 2007; Deakin-Crick and Yu, 2008). That is to say, they value the chance to learn from others as well as the chance to learn on their own. They are neither isolated nor dependent. On the contrary, some learners are either too isolated or too dependent. They are isolated as they do not engage with other people or they depend too much on others in leading them. There are 12 items that assess learning relationships, including the following:

Q5. I prefer to work on a problem on my own.

Strategic awareness: Effective learners demonstrate more sensitivity in their learning. They are more aware of themselves as learners. They like to take control of their own learning and they are good at self-evaluation. They also enjoy trying out different approaches to learning to see what happens. The contrast is 'being robotic', referring to learners who have less self-awareness of being learners. There are 13 items for this dimension, including the one below:

Q69. I like to find my own ways of doing things even if everybody else is doing it a different way.

As one aim of the current study is to have a better understanding of test impact on the learners and their learning, it is important to look at any possible learner variable that

may explain test impact. The concept of learning power is not limited to a particular context but is embedded in social, historical, cultural and ethical trajectory. It takes into account learners' sense of agency, intentionality, and capability in varied real-life contexts, communities of achievement and lifelong learning (Deakin Crick and Yu, 2008). This complex and embedded description of the dimensions of learning power, provided by ELLI, is 'important and indicative of their sense of agency and of their learner identity' (Deakin Crick and Yu, *ibid.*, p.390). Few previous washback studies on the learners and their learning have explored learner variables that can encompass the complexity of how learners see themselves as learners. In addition, previous studies of ELLI (Deakin Crick et al., 2004; Deakin Crick, 2007) have shown that ELLI a robust instrument which can indeed be used to differentiate between efficacious, engaged learners and those who were more passive and dependent. Deakin Crick and Yu (2007) have also shown that the scales in ELLI are able to demonstrate a significant degree of stability, reliability and internal consistency over time. Thus, the instrument can indeed provide information on what the learners say about themselves in a particular context, in a particular trajectory in time. Therefore, the use of ELLI in this study attempts to show how this complex concept of learning power can explain the learners' perceptions towards the impact of the graduation requirement.

In the next section, I will review a study that has attempted to incorporate the ELLI instrument in discussing the impact of a language test.

2.5.4 ELLI and impact of language test

Rea-Dickins et al. (2007) conducted a post-test impact study that explored the affective and academic impact of IELTS scores on successful IELTS students after they started their academic programmes. The 'impact' in the study was specifically

conceptualized as students' learning identity, the process in which they manage membership in the academic programme community and negotiate new learning challenges. They further conceptualised learner identity with three facets of individuals: intra-personal identity and knowledge, socio-historical constructions of self, and constructions of self in relation to significant others. The use of the ELLI in the study was to inform the facet of intra-personal identity and knowledge by indicating internal factors referenced to growable accounts of learning power (learning capacity, motivation and orientation). The study attempted to explore how this intra-personal capacity would further inform the learning processes of the students in the negotiation of membership in the academic programme community.

The analyses of the learners' narrative accounts in relation to their ELLI profiles and IELTS scores revealed that this research approach could offer some insights into understanding test impacts in several ways. From the first cohort of students who have taken the ELLI twice, at the beginning and the end of the track period, all except the student who had high IELTS scores had a decrease in the dimension of 'meaning making'. The student also turned out to be the highly 'successful' one who had progressed in the programme with excellent academic performances. The researchers thus argued that students with higher language proficiency were more likely to cope with the challenge of the information load provided by the intensive academic graduate programmes. On the other hand, students with weaker IELTS scores were more likely to be overwhelmed by information load without being able to see how the accumulated data fit into the big picture. The findings of the data from the second cohort (took ELLI only once) suggested that the students with lower IELTS scores and having weaker academic performances were those with weaker ELLI profiles. Their scores on the 'meaning making' dimension were generally lower than those

with high IELTS scores and their ELLI profiles were likely to reveal a particular weakness in some dimensions. Their narrative accounts revealed their frustrations as they struggled in learning and adapting to the new community. In fact, all students struggled as they achieved participation in the new community. However, the learners who were successfully engaged, were more likely to effectively deal with struggles and anxiety and see feedback as opportunities to learn.

Rea-Dickins et al.'s study has sought to shed light on the positive link between the learning power, language test performances and academic performances of subject learning. The use of the ELLI in their study has also demonstrated how the ELLI can contribute to a further understanding of a learner's learning process through intrinsic learner variable. The current study is different from theirs, particularly in relation to contexts³. Nevertheless, it will be interesting to see how the findings of this study on students' learning power and their learning can relate to or contradict the findings of their study (see 6.4.3 for discussions).

2.6 Summary

In the first part of this chapter, I have provided definitions of the key terms in this study, such as washback and impact. I have also discussed several key issues of washback, including the direction of washback, washback variability and washback intensity, mechanism of how washback operates and different conceptualisations of washback in the current literature.

The second part has reviewed work on the washback and impact on teaching and

³ a) Post-test impact v.s. washback. b) International students v.s. local students. c) intensive, short programme v.s. extended time in preparation for the graduation requirement

learning. The review has revealed two important issues for consideration in the present study. The first is that unlike the abundant studies on washback and impact on teaching, test influences on the learners and their learning remain to be more fully explored. In addition, washback and impact on learning should not depend on the accounts of the teachers, but should be looked at from the perspective of the learners themselves and also how their perspectives have been shaped by other stakeholders. The second issue is related to the methodological concerns for washback and impact studies to draw a fuller picture of a phenomenon. The review has shown that qualitative research methods such as ethnographic approach and classroom observations are more likely to capture washback variability among participants. Furthermore, a combination of a survey and qualitative methods can provide an opportunity to use a large sample to understand general patterns as well as individual differences in their perceptions towards test influences. Following this was a review on a limited number of studies conducted in the same educational context of the present study in relation to GEPT and the English requirement for graduation.

The final part presents the rationale for measuring the student participants' learning power using ELLI, and the seven dimensions of learning power assessed by the learner profile.

CHAPTER 3 Methodology

3.1 Introduction

This chapter presents the methodological considerations of the current research. To begin with, I describe my philosophical position underlying the development of the research design (3.2). Then, I present in Section 3.3 the research questions and describe the pilot study (3.4) and how its results inform the decision of the research approach. The overview of the research design is in Section 3.6, followed by detailed descriptions of data collection (3.7). The methods of analysis for each method are then described in Section 3.8. The chapter concludes by discussing the trustworthiness of the research design and the ethical issues arising from it (3.9~3.10).

3.2 Philosophical position

Before describing the research procedures and methods in this chapter, the philosophical underpinnings of the present study need to be presented first, as Guba and Lincoln argue that ‘questions of method are secondary to questions of paradigm, which we define as the basic belief system or worldview that guides the investigator, not only in choices of method but in ontologically and epistemologically fundamental ways’ (1994, p.105). I also agree with Miles and Huberman (1994)’s argument that researchers should make their orientations explicit, because ‘how a researcher construes the shape of the social world and aims to give us a credible account of it is to know our conversational partner’ (p.4). Although, as they have observed in actual research, researchers with different orientations seem more alike than different and the lines between different paradigms have become blurred, I still believe that by explaining how this study has been influenced by which philosophical school of thought will make the essence of this study clearer.

The language testing field, as McNamara and Roever explain, is ‘an interdisciplinary endeavour between applied linguistics and psychometrics’ (p.1) and thus with measurement and psychometrics as its basic foundation, language testing has been long regarded as ‘an inherently positivist enterprise’ (Lynch and Hamp-Lyons, 1999, p.83). However, the main philosophical position that this study leans towards is not that of a positivist, but of a postmodernist, because postmodernist ideas provide a better explanation for the following: the social aspects of the washback concept, the unpredictable and varied nature of washback, and the power relations between the stakeholders involved in this study and in the washback process. Within the context of this study, however, it is difficult to position this study philosophically within only one particular paradigm, because I believe not one single school of thought can fully explain a research with mixed methods. Therefore, even though this study shows major postmodernist inclination, there is a part of the study that remains positivist, e.g. the assessment of students’ learning power. It is only by delineating both the postmodernist influences that underline the study and also the positivist elements that the study retains that I will be comfortable to say that the essence of the study has been made clear.

In the next section, I will start with the postmodernist ideas upon which this study is developed.

3.2.1 Postmodernist influences in this study

Postmodernism is a complex school of thought. Reflecting one of its characteristics, heterogeneity, Derrida’s (1982) idea of society and culture as text with no correct reading but interpretations, Foucault’s (1982) notion of power relations in knowledge construction with power generating its own forms of resistance and Lyotard’s (1984) view of knowledge as language games each with its own rules incompatible with others

are all considered as being significant arguments in developing postmodernism (Derrida, 1982; Delanty, 2005; Foucault, 1982; Lytorard, 1984; Smith, 1998,). Drawing on Delanty's (2005, p. 113) synthesis, the main ideas of Postmodernism are as follow:

- 1) society can be interpreted as a text
- 2) the deconstruction of agency involves a shift in emphasis from structure to culture, with the literary text becoming a cultural discourse
- 3) an anti-foundationalism approach involving indeterminacy and contingency, stressing the absence of any one correct viewpoint
- 4) a concern with identifying diversity and resistance

Postmodernism rejects the idea of a single reality, and suggests plural realities or no realities, but interpretations, and there are no standards in judging any interpretations as superior to any others. Knowledge is never objective, but cultural, local, individualistic and momentary.

The postmodernist ideas that influence this study are its embrace of plurality, diversity and subjectivity, and also Foucault's alignment of power relations and resistance with the creation of knowledge. How the above ideas influence this study will be illustrated below through three aspects: the concept of washback, the nature of washback and the power relations in this study.

3.2.1.1 The concept of washback

The concept of washback is often linked to Messick's revolutionary framework of validity. Messick's means of considering value in test constructs and score interpretation and of considering the social meaning a test can entail caused ripples in the conservative field of language testing. Among the two concepts of Messick's challenges to the traditional validity framework, washback is directly related to

consequential validity, and the social consequences of a test. McNamara argues that the most important influence Messick has brought to the field is 'a revolution in epistemology that has led to challenge to the positivist research tradition in the social science' (2006, p.40). He also argues that the increasing awareness of language testing as social practice is not only the influence of Messick, but also 'the intellectual changes triggered by postmodernism, where models of individual consciousness have been reinterpreted in the light of socially motivated critiques' (McNamara, 2001, p.333).

3.2.1.2 The nature of washback

Although washback research started from a simple washback hypothesis, namely that a test will influence teaching and learning (Alderson and Wall, 1993), more and more researchers realise the complex nature of washback that can never be explained by simple causal relationships. Studies of washback show that negative or positive test consequences do not necessarily correspond to the quality of the test, and show that washback, if used for curriculum innovation, may not bring about the intended positive consequences. Furthermore, the reports of washback studies show that cultural factors, institutional factors and individual factors (e.g. the teacher factor) cause washback to manifest itself differently in different contexts. The complexity of washback can be summed up by Rea-Dickins and Scott (2007) in their editorial for the special issue on washback, stating that washback 'can be viewed as a context-specific shifting process, unstable, involving changing behaviours in ways which are difficult to predict' (p.9). Thus, washback researchers are limited to only unpacking part of the complexity by providing a picture of the phenomenon in a particular context. The above statements reflect the tenets of postmodernism. Washback is context-specific rather than context-free. Washback is a phenomenon that may be displayed differently in time instead of being one set picture that transcends time. Washback is unpredictable and

uncontrollable.

3.2.1.3 Power relations

In this decade, Foucault's influence in the field of language testing can be seen with the debated discussions of the power relations among different stakeholders. The gradual emphasis on test takers can also be seen as one of his influences. Lynch (2001) argues that to determine if the impact or consequences of test interpretation and use is good or bad should include the 'multiple perspectives that make up the assessment setting' (p.366). He points out that the value of test effects would be a 'negotiated consensus' among different stakeholders in the assessment setting (*ibid.*). He further argues for evolved power relations to empower the participants. His arguments on impact and power relations are similar to the concept of 'Critical Language Testing' by Shohamy (2001) who emphasises language testers' responsibility to investigate test consequences, misuses or unethical uses, monitor power and value the voice of test takers. Foucault's ideas are evident in both Lynch (2001) and Shohamy (2001). Shohamy takes on Foucault's (1979) view of tests being used for exercising power and control, and therefore, if misused, tests can bring detrimental consequences. Lynch goes forward and emphasises Foucault's (1982) notion of free and ethical power relations.

In the context of this study, the power relations between stakeholders are especially visible. The implementation of the requirement is a result of the interaction between the government, the university assessment units, the universities, the teachers, and students (See 1. 3). How washback of a test in the graduation requirement works involves the power relations between some other stakeholders like test designers and publishers and teachers and curriculum designers in the universities. To explore the impact of the

graduation requirement on students and their English curriculum, the power relations between the stakeholders should be taken into consideration. In addition, despite being recognised as important stakeholders, learners are still being given the least attention in previous studies on washback. As I have mentioned in chapter 1, decisions as to the graduation requirement are often made without considering what the consequences will be for the learners, and without taking account of learners' viewpoints on the implementation. Another postmodernist influence in this study is to pull learners from the periphery to the centre of attention and to look at the requirement from their eyes.

3.2.2 The positivist elements

Although postmodernist influences are evident in the present study, this paradigm alone cannot fully explain the epistemological and methodological considerations in this study. Not only is it difficult to place a mixed method study in only one paradigm, but the postmodern views of indeterminacy and uncertainty can be contradictory to being a school of thought itself. Therefore, despite the postmodernist influences on epistemology, this study retains several positivistic elements such as the ways in which the research has been conducted, the ways the research are presented in this dissertation and also the assessment of learners' learning power. Besides the basis of the traditional social science research, the assessment of learning power is also developed and analysed in a positivistic fashion. The reasons why, methodologically speaking, the assessment of learning power remains positivistic are twofold. First, there is a large sample of students in the case universities of this study. A relatively larger scale of questionnaire study on the students and their learning provides a balance to the large amount of classroom observation data that focused more on teaching. The second is that the semi-structured interview data and the interviewees' questionnaire data also complement each other by valuing each learner's voice and getting an idea of what kind

of a learner the learner is from a validated assessment tool.

3.2.3 Philosophical underpinnings as manifested in the study

Post-modernistic influences in this study are revealed through the concept and the nature of washback, and also the considerations of the power relations between different stakeholders involved in the implementation of the English graduation requirement and the washback process. In addition, students in this study have been given the opportunity to provide their perceptions of the graduation requirement, and the empowerment of the test takers also hints at post-modernist ideas.

Nevertheless, although the study is substantially influenced by postmodernist ideas, in terms of research methodologies, one paradigm cannot fully explain the philosophical underpinnings of this study. Therefore, the study remains positivistic in the execution of the research, the organisation and presentation of data and findings and most importantly, the incorporation of survey data and statistical analysis to provide a pattern of learners' learning power in each case university, and to complement the qualitative data that allow findings to emerge.

3.3 Research questions

The research questions of the present study are as follows:

- 1) What are the effects of the English requirement for graduation on the English for Academic Purposes curriculum for non-English majors in Taiwanese universities?

Sub-questions:

- a) What are the effects of specified tests stated in the requirement on the English curriculum?
- b) In particular, what are the effects of the GEPT on the EAP curriculum?

- c) Is there a difference between the effects of the different tests on the EAP curriculum? What are the differences and why??
 - d) What are the teachers' perceptions towards the requirement?
 - e) What are the teachers' perceptions towards the effects of the tests on the EAP curriculum?
- 2) What are the effects of the English requirement for graduation on the non-English majors in Taiwanese universities?

Sub-questions:

- a) What are students' perceptions towards the requirement?
 - b) What are the effects of the tests (GEPT and other tests) on the students?
 - c) Is there difference between the effects of the different tests on the students?
What are the differences and why?
- 3) To what extent students' learning power explain their perceptions of washback on the learners and their learning?

3.4 Pilot study

An exploratory pilot study was conducted in summer 2007 with three tertiary EFL teachers from three different universities in Taiwan. Although the purposes of this pilot study were to establish a baseline for the main study, as well as trying out the interview schedule, the emphasis was put on the former. Watanabe (2004) stresses the importance of the 'baseline' in any type of research, including washback studies, so that the assumption of "the presence of a phenomenon which may not, in fact, exist" can be avoided (p.24, citing LeCompte and Pressle 1993, p.120). Since little research has been conducted concerning how the English graduation requirement has influenced teaching and learning in universities, the pilot study was intended to explore whether there was washback of any test stated in the requirement on teaching and learning as perceived by

the teachers before exploring the nature and the scope of the phenomenon in the main study.

The three English teachers were selected to represent, to some extent, the teacher participants in the main study, as explained below. All of the teachers have taught non-English majors in the past few years. The universities they work with are different, in the sense of how they have responded differently to the language educational policies as stated in 1.2. Two universities have implemented the English graduation requirement, while the third university does not require its students to provide any evidence of English proficiency for graduation. Semi-structured interviews were carried out, asking the teachers to reflect on their perceptions toward the graduation requirement and towards the impact that the graduation requirement brought to the university English curriculum for non-English majors.

The results of the pilot study reveal that the GEPT was perceived by all the teachers as the most influential test of their students, because they considered it to be the most recognised test in terms of graduation requirements, the workplace and society. However, these three universities had different regulations in their requirement, and thus, within each context, different impacts. The implementation of the requirement resulted in different curriculum changes such as the increase of English EAP courses or the addition of a test preparation course or a remedial course. Besides the contextual differences, there were also individual differences in the teachers' attitudes towards the requirement, and thus, their perceptions of GEPT washback on their teaching differed, with some incorporating aspects of the GEPT according to the courses, and others insisting on prioritizing the development of students' English proficiency over preparing them for the test. In addition, they also pointed out that there was a different

degree of GEPT washback on their students, and it was possible that the students' needs in test preparation would in turn, shape washback on their teaching.

Besides setting the baseline for this research, this exploratory pilot study also informed this study in several ways. First, the findings confirmed that there was much complexity in the educational contexts embedded in the English graduation requirement, and because of this complexity, the implementation of the requirement in different contexts might have different impacts on teaching and learning. For example, teaching remedial or test-preparation courses in one university was fundamentally different in purpose and nature from teaching EAP courses in another university, despite both being the results of curriculum change from the implementation of the requirement. Therefore, in order to understand how the impact of the requirement manifested in each context, this study adopted a case study approach. Second, the individual differences of the teachers in their perceptions of the requirement also informed the study so as to include more teacher participants, thereby enabling the way in which teacher factors shape washback on teaching to be discussed. Similarly, albeit indirectly, the pilot study also pointed out that there was evidence of washback to the learners. In order to explore this aspect in depth, a number of learner participants should be included in order to understand potential individual differences of washback to learners. In summary, the pilot study informed the main study in the choice of case study approach to encompass contextual differences, in the involvement of multiple participants to explore participant factors and in providing an incentive to explore washback to the learners.

3.5 Research context and research approach

3.5.1 Ethnographic case study approach

The research adopts an ethnographic case study approach to research design. Yin (2003) suggests the use of a case study, 'especially when the boundaries between the phenomenon and context are not clearly evident' (p.13). Cohen et al. (2001) argue that 'contexts are unique and dynamic' and that case studies can capture 'the complex dynamic and unfolding interactions of events, human relationship and other factors in a unique instance' (p.181). The reasons for choosing this approach is the important role each university context plays in investigating the impact of the graduation requirement and the washback of different English proficiency tests mediated by different stakeholders. Institutional context is important because how each university establishes and implements the English graduation requirement varies. The differences then result in different influences on the English curriculum, teaching and learning in the university. Thus, the case study approach can best encapsulate the complexity and the depth of washback produced by different tests in the requirement, and the inter-relationships of the two important stakeholders, teachers and students.

Two universities were selected as cases for this study. According to Stake's (2005) typology of case studies, this study is 'instrumental' and 'collective' (or 'multiple'). The two universities (3.4.1, 3.4.2) were regarded as separate cases not for the sake of the cases alone ('intrinsic'), but for the purpose of understanding how in two different contexts, will the graduation requirement impact on English teaching and learning. The two cases were chosen as a result of 'opportunistic sampling' (Patton, 2002). They represent two different contexts, one with the implementation of the graduation requirement, the other without. Multiple case studies with more than one case may provide an opportunity for comparison (Johnson, 1992; Yin, 2003). Yin (*ibid.*)

analyses the replication logic of multiple experiments to multiple case studies. Cases with similar results are ‘literal replications’ while those with contrasting results are ‘theoretical replications’ and both replications can lead to a more robust study with theoretical inferences. From a different position, Stake (2005) warns that the uniqueness and complexities of each case may be glossed over by the focus on comparison. However, he still acknowledges that similar or dissimilar cases, ‘with redundancy and variety each important’ (*ibid.*, p.446) are illustrations of ‘how a phenomenon occurs in the circumstances of several exemplars’ and can provide ‘valued and trustworthy knowledge’ (*ibid.*, p.459). Stake’s position seems to fit more into the present study. It is difficult for Yin’s ‘replication’ or any measured comparisons between the two cases of this research, because they are neither similar nor contrasting. Although there are still components of comparison in the quantitative part included in the study, what is more important is to explore how washback effects manifest in each context, and the reasons behind those manifestations.

In addition to case study approach, this study also takes on Watanabe’s (2004) recommendations in designing washback studies: the ethnographic approach. Watanabe accommodates LeCompte and Preissle (1993)’s characteristics of ethnographic or qualitative research, and explains why they are relevant to research into washback. The first is that the approach ‘elicits phenomenological data that represent the worldview of the participants being investigated and participants’ constructs are used to structure the research’ (p.22). That is to say, it is important for washback studies to take into consideration what the test users are concerned of within the context the test is used. Secondly, the approach asks the researcher to ‘employ participant and nonparticipant observation to acquire firsthand, sensory account of phenomena as they occur in real-world settings’ (p.23). In washback research, how the test actually impacts

on teaching and learning in the 'real', non-experimental classroom is essential. The third characteristic of the ethnographic approach is that it asks the researcher to 'construct descriptions of total phenomena within their various contexts and to generate from these descriptions the complex interrelationship of causes and consequences that affect human behaviour toward and beliefs about particular phenomena' (p.23). In other words, using the ethnographic approach in washback studies can assist in describing the total phenomena by taking into account factors other than the test itself in the context that affect what happens in the classroom. Lastly, researchers using the ethnographic approach make use of a variety of research techniques to collect their data, and using methods that can complement each other helps to understand from different perspectives how washback work. In addition, it is very likely that data may emerge during the process of conducting the research, and the ethnographic approach embraces all research-related data, even if the collection of the data has not been planned beforehand.

This study incorporated Watanabe's recommendations with the case study approach. It attempted to explore the impact of the university graduation requirement for English proficiency in each case university, by taking the whole context into consideration, by conducting observation in the real classroom setting (See 3.7.3 for details), and by trying to consider factors other than the test in the context. Different research methods were conducted to elicit data on the impact of the graduation requirement in the classrooms and from different stakeholders. Furthermore, unplanned data, e.g. private talk with Becca, were also collected as data contributing to a deeper understanding of washback on teaching in the requirement-related, remedial class (see 4.3.2 for further details).

The characteristics of the two universities including their backgrounds, students English proficiency level will be described respectively in the next sections (3.4.1, 3.4.2). The regulations concerning English language education for non-English majors in both universities are presented below.

3.5.2 The two case universities

Case A

Case A is the university where I have previously worked for four years. This private university, located in Southern Taiwan, was established nine years ago. The university is a teaching-oriented university, with only undergraduate students in both day and night divisions. Students in this study were limited to those in the full-time day division. The majority of the academic departments are related to business, finance and management, with only a few related to information technology and language. Excluding the English Department, there are fifteen non-English departments altogether. Non-English majors are required to fulfil 8 credits of 'English Integrated Skills Training (I)~(IV)' sequentially in their first and second year. Considering students' overall English proficiency, only English majors of the day-division are required to provide evidence of English proficiency in order to graduate. Non-English majors are not asked to meet the demand for English due to the fear that most students will not be able to meet the graduation requirement. Among the regulations concerning English education for non-English majors in University A, credit exemption is the only regulation that is related to language tests. It says that students have to provide evidence of reaching a B1 level in the Common European Framework (CEF; Manual for relating language examinations to the CEFR, 2009) for the abovementioned 8 credits to be exempted.

Case B

Case B is a national university located at the Northern Taiwan with over hundred years of history, providing both undergraduate and post-graduate degrees. It is a teacher-education institution. Most academic departments in the university are related to primary education. The basic requirement of English education for non-English majors is 4 credits of English, to be completed in their first year. The degree requirement is applied to all students, English and non-English majors, with the criteria of English majors higher than the non-English majors. The 'Regulations for the Implementation of Promoting Students' English Proficiency' states the graduation requirement in details.

Regulation no.3:

All students who are receiving the Bachelor's degree, besides those who can have exemption according to Regulation no.5, are required to take 2 semesters of English remedial courses in their third year...English remedial courses for non-English departments are 'English Conversation and Listening' and 'English Reading and Writing'. (Note: Contents of the regulation not related to non-English majors are omitted.)

Regulation no.4:

The English remedial courses are required courses with no credits. The fail and pass grade is 60. Those who failed the courses have to re-take the courses and only those pass all the courses can be graduated.

Regulation no.5:

Students in this university can receive exemption from English remedial courses by reaching one of the standards listed below and by providing documents of proof before the end of the add and drop period (i.e. when students decide which course to take) in their third year.

1.... (Contents of the regulation unrelated to non-English majors are omitted.)

2. Non-English majors:

- (1) GEPT Intermediate Level 1st Part and above
- (2) TOEFL (paper-version) 457 and above
- (3) TOEFL CBT 137 and above
- (4) TOEFL iBT 47 and above

- (5) TOEIC 550 and above
- (6) IELTS 4 and above
- (7) Other proficiency tests or standards approved by the Office of the Academic Affairs and related departments.

In short, non-English majors are required to pass the course without getting any credits (no credit is granted for the English remedial course) if they do not display proof of reaching the standards listed in Regulation no.5.

The purpose of the above section is to provide information that forms the basis of the institutional context each case entails. In the next section, the procedures of the data collection and the difficulties encountered in the process will be discussed in detail.

3.6 Data collection

A review of empirical washback studies reveals that a variety of research methods, quantitative and qualitative, have been used to study the washback phenomenon. Earlier washback studies used questionnaires and interviews (e.g. Herman and Golan, 1991; Shohamy et al., 1996) to collect stakeholders' perceptions about the influences of a test. However, as washback is a complex phenomenon, 'the whole context wherein the test is used' should be taken account of (Watanabe, 2004, p.22). From the account of the teachers and students, only an introspective view of washback can be provided. In order to capture a fuller picture of washback, researchers advocate ethnographic approaches and triangulations of the reported perceptions of the participants with classroom observations (Alderson & Wall, 1993; Alderson & Banerjee, 2001; Bailey, 1996; Cheng, 1997; Watanabe, 2004). The classroom is where teaching and learning occurs, and it 'embodies a specific set of functions and values from the teacher and the learner, and also from the point of social setting and institutions at large' (Cheng, 2005, p. 68). Classroom practices and events can reveal whether teachers and students'

perceptions of washback on teaching and learning for example, collected from questionnaires and interviews, observations correspond to what they actually do in the classroom, and if not, classrooms may help to uncover unintended washback.

The present study employed a mixed method approach, including ethnographic classroom observation (Section 3.5.2), interviews (3.5.3) and a questionnaire (3.5.4) under the overarching approach of case study. Before discussing in detail the rationale, aims, and the procedures of each method, I will start with the unforeseen difficulties I have encountered and what compromises I have made in searching for participants for classroom observations.

3.6.1 Identification of teacher participants

Watanabe (2004) argues for the need, when designing observation research, to exclude all possibilities other than tests that may influence teaching and learning. He points out that ideally, washback can be evidenced by the absence of test influences in non-exam courses when compared to exam preparation courses. In addition, he suggests observing at least two teachers teaching both courses, with the idealistic assumption that if washback exists, similarities should be found between the teachers and only the differences between the two courses will surface. Green (2006b) further suggests that washback studies should have a higher number of teachers so that there can be a deeper and further understanding of how teacher factors mediate washback (See 2.2.4, washback variability). Thus, to have room for the unpredictable drop-out of participants during the observation process, this study has the goal of observing 6 teachers in each university. Among the six, there should be at least two teachers in each university who have taught a course related to the degree requirement and a course that aimed at developing general English proficiency. However, due to the restrictions of the

circumstances described below, only 4 teachers and their classes in Case A and 3 teachers and their classes in Case B were observed, and none of the teachers taught both types of course.

Case A: As the teachers in Case A were acquaintances of mine, none of the five teachers I contacted beforehand seemed to reject the idea when I asked their permission to observe their classes several weeks before I started the observation. After I received approval from the head of the English Department, emails were sent to my former colleagues, stating the details of classroom observation such as the classes I would like to observe and the duration of the whole observation, leaving them room to change classes and time to inform their students. Since I received only one reply of confirmation and no other replies, I assumed that the participants were confirmed. However, when I arrived at the university, two of them revealed a reluctance for observations being conducted in their classes, and did not participate. As the time I could spend in each university was limited, I had no choice but to start observation with three teachers first. In the meantime, I contacted other possible teacher participants and fortunately, one professor agreed. Thus, in Case A, 4 teachers and their classes participated in this study.

Case B: The situation was more complicated and the problems more serious in Case B. Although one of the assistant professors there is my long-term friend, I had foreseen difficulties in conducting research in a university with little familiarity. The preparation I made to enter this field site began two months ahead. My friend assisted me in circulating my email among her colleagues and provided information of the classes they taught. However, as I was relieved to receive six teachers' approval, I realised that the timetable of English courses for non-English majors in Case B made it impossible

for one researcher to accommodate all six teachers for observation. All the general English classes were scheduled on Tuesday morning from eight to ten and ten to twelve while the timeslot for the third-year 'English Reading and Writing' was Tuesday afternoon, one to three. With the consideration of observing more than one lesson per teacher to avoid the 'observer effect' and capture patterns that teachers may display in successive lessons, the decision to observe two teachers in the morning and one in the afternoon for a few weeks was made.

With the information my friend in Case B had provided, one participant taught both general English course and 'English Reading and Writing', unfortunately, went earlier for her maternity leave.

In summary, what I had achieved deviated greatly from what I had previously proposed. In the end, I observed four teachers in Case A teaching English Integrated Skills, and two teachers teaching general English and one teaching 'English Reading and Writing' in Case B.

Teacher profiles

The following is a brief introduction of the teacher participants in this study. All of the teachers were given pseudonyms. For the purpose of convenience, those in University A had names that begin with A, while those in University B were given names beginning with B.

Adam:

Adam was a full-time professor of linguistics in University A. He had received his last degree in linguistics from the United States. He had more than twenty years of teaching

experience in tertiary education. At the time of the study, he had only been in University A for almost two years and had just started to teach English Integrated Skills Training course to non-English majors.

Alice:

Alice was a full-time instructor who had been at University A for six years. Her latest degree, from the U.S., was a Masters degree in educational administration. She had an English degree for her undergraduate studies in Taiwan, as well as several years of experience teaching English in junior high schools. She had been teaching non-English majors since she came to the university.

Amy:

Amy had about ten years of experience teaching non-English majors as a full-time instructor in two tertiary institutions. She was the only teacher who received both her undergraduate and masters degree in TESOL from the U.S.

Anna:

Comparatively young, Anna had six years of formal experience both as a full-time instructor and as a teacher of non-English majors. She also obtained her masters degree in TESOL from the U.S,

Becca:

Becca was a full-time associate professor in another university, teaching only one class in University B at the time the study took place. The class she taught was 'English reading and writing', which was related to GEPT preparation. Her doctorate in TESOL was obtained in Taiwan, but she had devoted herself in the EFL field for more than

twenty years. She owned a chain of local English cram schools for children and a local publishing house which developed EFL materials, targeting mostly young children. In addition to EFL materials for young learners, in recent years when the GEPT started to gain popularity in the society, she has been involved in the development of GEPT preparation materials and has since published several sets of GEPT related mock tests and teaching materials with her publishing house.

Ben:

As a full-time associate professor of educational administration and management in the same university, Ben taught only part-time for the English department. With his doctoral degree in educational administration from the U.S. and his genuine interest in teaching English, he had been teaching non-English majors for more than five years.

Betty:

Betty was part-time lecturer in University B, teaching only two classes as she was a full-time instructor in another university. She had over twenty years of teaching English to non-English majors. Her latest degree was a masters degree in drama from the U.S..

3.6.2 Classroom observation

The primary aim of the classroom observation in this study was to seek observable evidence of washback to the programme and washback on teaching. As mentioned in 2.3, teachers are central in the exploration of washback to the programme. Thus, in this study, classroom observation focused mainly on the teachers. The observation provided information on: what teaching English really looked like in the university classrooms, how teaching differed between test-related courses and general English courses, whether there was difference in teaching between different teachers. It would

also reveal what the influences of the tests accepted by the graduation requirement on classroom practices were.

The approach

Cheng (2005)'s review of literature in her HKCEE project has pointed out that the two approaches that have been used for classroom observation are the systematic observation approach and the ethnographic approach. The systematic approach uses observation schemes to 'reduce classroom behaviour to small-scale units under pre-determined categories suitable for tabulation and statistical analysis' (*ibid.*, p.93). Observers use the schemes to record classroom events quickly by marking, normally either by 'event sampling', which describes the frequency of the observed events, or by 'time sampling', which provides 'the distribution of the particular phenomenon throughout the class (Dornyei, 2007). On the other hand, the ethnographic approach uses participant observation, which requires researchers to immerse themselves in the context of their research. The researcher collects data through recording detailed field notes and interacting with the participants. This approach 'acknowledges the complexity of the classroom situation and uses a holistic framework, basing the observation not on pre-determined categories but according to the context in which the teaching is occurring' (Cheng, 2005, p.93).

Many washback studies that incorporate classroom observation adopt the systematic approach. The instrument that has been used or adapted in these studies (Burrows, 2004; Cheng, 2005; Green, 2006; Read & Hayes, 2004; Watanabe, 2004) is the Communicative Orientation to Language Teaching (COLT) observation schedule (Spada & Frohlich, 1995; in Green, 2006b, 2007b). The use of the COLT has proved successful in capturing elements of washback among classroom interactions. However,

as with other observation schemes, those aspects of teaching and learning which are important to a particular research but not captured by the scheme will be neglected (Nunan, 1992). For example, Hayes and Read (2004) in their study of IELTS washback in test preparation courses, incorporated another observation instrument that looked at specifically IELTS text and task types and test-related activities to complement the macroscopic description COLT provided. However, they realised that neither of the structured instruments had yielded completely satisfactory results for the purposes of their study. Their acknowledgement also somehow reflected what Dornyei (2007) pointed out as a weakness of the systematic approach, namely that it reduces the complexity of the observed lessons and is less sensitive to emergent information which is context specific.

The present study took on an ethnographic approach to classroom observation due to the nature of the study and also the complexity the graduation requirement encompassed. The holistic framework of the ethnographic approach fit into the overarching case study approach of this study. The detailed descriptions of the educational context in Chapter 1 and the results of the pilot study (3.4) both showed the significance of context in understanding washback. This was the most important reason why I conducted two case studies instead of a large scale survey study. Similarly, the ethnographic approach for observation should be better than the systematic approach for allowing context specific evidence of washback and impact to emerge in data collection. In addition, unlike most washback studies which often studied washback of one particular test, the graduation requirement for English proficiency in this study accepted scores from several tests. Hayes and Read's (2004) example has already revealed the difficulty for the systematic approach to be comprehensive enough for capturing washback of only one test. I believe it may be extremely difficult for any

structured observation schemes to encompass test-related features from a number of English proficiency tests. Besides, direct teaching to the test should be quite rare except in preparation courses for a specific test such as the GEPT. It would be difficult for observation schemes to capture the more subtle and covert forms of washback effects. Therefore, the classroom observation in this study was conducted with the ethnographic approach.

Although the classroom observation was ethnographic in nature, I had specific focuses on several aspects of classroom practices. Firstly, I would focus on any explicit evidence of washback from teacher talk such as the teacher mentioning a specific test or pointing out the importance of a test. The second focus was on the similarity or difference between the content of teaching, the teaching materials used, and how the lessons were taught. In other words, I would focus on the similarities or differences between test-related and general English EAP courses and whether they manifest influence of a test. Explicit evidence of washback might be easy to detect but it was difficult to pin down less explicit washback by classroom observations alone. Thus, in order to determine what could be considered as washback, the ‘total phenomena’ including the observation data, the supplementary field notes, the private talks and interviews with teachers should be examined.

As mentioned in 3.5.2, the non-English majors in University B were required to pass the first stage of the GEPT Intermediate level in order to receive their undergraduate degree. Expected GEPT washback effects, if any, might be the following:

- 1) Teachers explicitly used preparation materials, test papers, or practice test items of the GEPT intermediate level for teaching in class. Likewise, teachers’ explicit instructions on test strategies for GEPT test items could also be regarded as

washback.

- 2) The first stage consists only of the listening and reading components. Expected washback would be the relative longer time spent on developing listening and reading skills than on speaking and writing skills in the lessons. In general EAP courses for non-English majors, the development of writing skills is usually not a focus. Thus, the deliberate prioritising of reading and listening over speaking could be considered as GEPT washback. Expected washback effects also included the development of listening and reading skills based on the GEPT task types. For example, there might be GEPT washback if a teacher taught listening particularly by asking students to listen and select which best described a given picture (See Appendix A). However, what needs to be noted is that the GEPT is not the only test that might exert influences on the general EAP curriculum. GEPT task types for listening and reading components may not be too different from those components in other English proficiency tests. Therefore, it is very important to use teacher interviews as triangulation to what has been observed in the classes. In this study, only when teachers revealed that their focus on listening and reading or on certain task types were preparations towards the GEPT, would such teaching behaviours be regarded as GEPT washback.

Collected data

The classroom observation data collected for this study comprised 17 lessons, 13 for Case A and 4 for Case B. Originally, according to the proposed design, two classes each teacher teaches were to be observed. This criterion was met roughly in Case A with only one exception. The professor's class happened to be at the same time as the teacher I had started to observe earlier. Thus, only one class that the professor taught was observed. In Case B, I was only able to observe one class for all the teachers, due

to the availability of the teachers and the timetable constraints (3.5.1). Besides the timetable constraints, two of the participants were part-time teachers in the university. In addition, one of them taught only ‘English Reading and Writing’. Another deviation from the proposed design was the duration of the observation. Originally, I planned to observe at least three lessons per class in both universities. However, for a variety of reasons such as official leave, personal leave and holding mock tests, I observed only two lessons for most classes and only one lesson for one particular class. The table below presents an overview of the observation data collected in comparison with the original plan

Case	Planned			Achieved		
A	6 teachers	12 EAP classes	24 lessons	4 teachers (Adam, Amy, Alice, Anna)	7 EAP classes	13 lessons (10 video, 3 audio)
B	4 teachers	8 EAP classes	16 lessons	2 teachers (Betty and Ben)	2 EAP classes	3 lessons (all audio)
	2 teachers	2 EAP classes	4 lessons	1 teacher (Becca)	1 test-related class	1 lesson (audio)
		2 test-related classes	4 lessons			

Table 3.1 Planned and achieved observation data collected

The EAP course for non-English majors in University A was called ‘English Integrated Skills Training’ (EIST). All the teachers in this study taught the same course. The only difference was that Alice and Anna taught both EIST IV and EIST II (different levels of proficiency) while Amy and Adam only taught EIST II.

The table above also shows that only audio data were collected for some classes and their lessons. These teachers declined to be video-recorded, because they thought it

would be intrusive. The lesson of English Reading and Writing (test-related remedial course) was a special case. Becca had been a researcher herself and was familiar with the observation method. She suggested not revealing my identity until the lesson was finished for me to capture the moments in the classroom with less observer effect.

During the classroom observations, I also took field notes. They were written down in the form of a note-taking sheet, which included two main parts: the basic information of the class and classroom and the notes on the lesson observed (see Appendix C for an example). The first part included the name of the lesson, date and time, the teacher, the number of students in class, the kind of classrooms (e.g. language lab with monitor on every students' table, or classrooms without monitors), the seating arrangements of the students, blackboard input (what teachers wrote on the blackboard) and other details of the setting. The second part, which was more related to the research focus, followed some broad categories of washback such as explicit evidence of washback, teaching materials, teaching content (what teachers teach) and teaching method (how teachers teach). Interesting issues that surfaced during the observation concerning the teacher and the students, and the interactions between them that did not fit in the categories above were also written down in detail. As the digital camcorder could only be set up either in front or at the back of the classroom due to the constraint of space in language labs in Case A and the audio recorder on the teachers' desks or on my desk (in classes where video-recording was not allowed), it was difficult to cover all the participants in the classroom. Thus, the field notes also included issues that I considered important, but which were not captured by the video and audio recorders. In addition, I also wrote down questions concerning what I had observed, to be asked in the teacher interviews. In terms of data use, as the observation was conducted using the ethnographic approach, the main purpose of the field notes was to complement the video and audio data instead

of being a separate set of data for analysis.

3.6.3. Interviews

Semi-structured interviews were conducted with the teachers who had been observed and with some of their students. All of them were interviewed after the completion of the observation sessions. Altogether, 7 teachers (4 in Case A and 3 in Case B) and 18 students (9 in Case A and 9 in Case B) were interviewed. Patton's 'interview guide approach' (2002) was adopted when conducting the interviews. Before carrying out the interviews, I prepared an interview guide for listing several issues I wanted to explore further. The guide served as a 'framework within which the interviewer would develop questions, sequence those questions, and make decisions about which information to pursue in greater depth' (p.344). This approach provided more flexibility than structured interviews, yet maintained a level of control on the issues to be covered. I roughly described the issues for discussion to the participants before the interviews were conducted so that the participants, especially the students, could feel less uneasy by knowing what would be discussed beforehand. All of the interviews were recorded using a digital voice recorder by prior permission.

Teacher interviews

The purposes of the interview were three-fold. Firstly, it was to elicit the teachers' perceptions of the ways the graduation requirement has influenced (University B) or would influence (University A) their teaching and their students. As Tsagari (2007) argues, it was problematic to assume that what teachers think about their students would corroborate with students' feelings and what they think about their own learning. Thus, in this study, students' perceptions were of priority concerning the impact of the requirement on them. However, it was also interesting to take teachers'

perspectives into account, to see whether there was any discrepancy between the students. Their attitudes concerning the degree requirement were also probed into. The initial interview protocols prepared before classroom observations were as follows:

- 1) What do you know about the graduation requirement for English proficiency in this university? (For Case B)
- 2) Do you think the graduation requirement for English proficiency should be implemented in your university? (For Case A) Why?
- 3) Which test accepted by the requirement do you consider is the most influential to your students? Why? (A &B)
- 4) Do you think the requirement has influenced your teaching: what you teach, how you teach, the activities you use and the choice of your teaching materials? If yes, in what ways? If no, why? (B)
- 5) Do you think the requirement will influence your teaching: what you teach, how you teach, the activities you use and the choice of your teaching materials? If yes, in what ways? If no, why? (A)
- 6) Do you think any test accepted by the requirement has influenced your teaching: what you teach, how you teach, the activities you use and the choice of your teaching materials? If yes, in what ways? If no, why? (B)
- 7) Do you think any test accepted by the requirement will influence your teaching: what you teach, how you teach, the activities you use and the choice of your teaching materials? If yes, in what ways? If no, why? (A)
- 8) Do you think the requirement has influenced your students' English learning? If yes, in what ways? If no, why? (B)
- 9) What is your attitude towards the implementation of the requirement? Why? (A&B)
- 10) Do you think the requirement has brought about the effects it has intended to? Why? (B)

The other purpose of the teacher interview was to ask the teachers to comment on specific episodes of their lessons. Watanabe (2004) advocates post-observation interviews as an important source of evidence to understand teachers as mediating factors of washback. During the interviews, the rationale for their choice of teaching materials, teaching activities and methods of assessment were discussed, to understand what factors influence their teaching and what role the teachers play in the presence or

absence of washback.

In addition to more formal interviews, I also had some informal private talks with some teachers after the observation sessions. The private talks contributed to a fuller picture of the impact of the graduation requirement on the EAP curriculum and the learners.

Student interviews

Nine students in University A were interviewed. The students were from three classes, each from different department. I asked them about how they evaluated their EAP courses and the teaching. I also asked them to talk about what their expectations of those courses were and the factors that motivated them to learn English. Lastly, they revealed their perceptions of the implementation of English graduation requirements, and which English proficiency test in the requirement was most influential for them.

The interview protocols are as follows:

- 1) What do you know about the graduation requirement?
- 2) Do you agree with the idea of the graduation requirement implemented in your university? Why?
- 3) Which English proficiency test in the requirement will you take? Why?
- 4) Do you think the implementation of the requirement will motivate you to learn more English? Why?
- 5) What factors in the English curriculum will motivate you to learn more English? (The lesson contents? How teachers teach? Test preparation? Etc.)
- 6) What are your expectations of an EAP course? Do you think the current course and lessons have met your expectations? Why?

In University B, 9 students from two classes were interviewed. Different from University A, an EAP class in University B comprised students from several departments. The student interviewees were from 5 different departments. They were

asked to comment on whether they considered the courses as helpful in promoting their English proficiency, whether their expectations of the course were met, what their attitudes were toward the implementation of the degree requirement and how their motivation to learn English had or had not been influenced by the degree requirement. None of the 9 students were selected from the test-related course, 'English Reading and Writing'. Only one lesson of the course was observed, and I was requested by the teacher, Becca, to conceal my identity and my research purpose (3.7.3). In addition, the time after the lesson ended and before their next class began was used to fill in the ELLI questionnaire (See section 3.5.4). This was considered to be a limitation of the study, since all of the interviewees in University B were in their first year, and therefore did not face the immediate threat of not receiving their degrees. Only when they were in their third year might they have to take the remedial course, or to provide formal evidence of test results to meet graduation requirements. As a result, no questions directly linked to manifestations of washback in the remedial course were discussed in the interviews. What follow are the initial interview protocols:

- 1) What do you know about the graduation requirement for English proficiency? Can you tell me what you need to do to meet the requirement?
- 2) What is your attitude towards the implementation of the requirement? Why?
- 3) Which English proficiency test in the requirement have you taken or will you take to fulfil the requirement? Why?
- 4) Which other English proficiency tests will you be considering taking? Why?
- 5) In what ways do you think the implementation of the requirement has influenced you?
- 6) Do you think the implementation of the requirement has motivated you to learn more English? Why?
- 7) What factors in the English curriculum will motivate you to learn more English? (The lesson contents? How teachers teach? Test preparation? Etc.)
- 8) What are your expectations of an English course? Do you think the current course and lessons have met your expectations? Why?

Unlike most washback studies (2.3.1), the student interviews were not triangulated with the classroom observations and the teacher interviews. The student interviews were conducted mainly to understand their perceptions of the graduation requirement and its impact. The student interview were analysed with reference to the students' learning power, as measured by ELLI (see below).

3.6.4 ELLI questionnaire for students

The rationale for the use of the ELLI has been presented in 2.5. This section, thus, focuses on the administration of the instrument in this study.

Although ELLI has been routinely administered online, I used a paper version of the instrument. The original ELLI questionnaire uses simple English. However, according to my experience teaching in University A, I speculated that the language might still be too challenging for some students. Therefore, I used a Chinese version of the ELLI (see Liu, 2007 and Appendix, D). In order to enhance readability and understanding of the questionnaire, the simplified Chinese version of the ELLI was converted into traditional Chinese. There were some minor adjustments of certain phrases or words that would seem confusing to Taiwanese students who read traditional Chinese. One of my former colleagues, a researcher in another Taiwanese university, and a fellow PhD student in the Graduate School of Education who were all familiar with both simplified and traditional Chinese helped to refine the traditional Chinese version.

Five hundred and five completed questionnaires were collected from students in both universities (246 in University A, 259 in University B). Any missing data or wrong data (e.g. two answers for a single item) could distort the mean scores of each dimension in ELLI and thus, these cases were removed from the final dataset. Four hundred and

fifty four valid questionnaires were used in the analysis.

What needs to be noted is that in this study, the ELLI data was mainly analysed and discussed in relation to the student interviews. The major purpose of employing ELLI in this study was to explore whether learning power could be used to explain how individual students experienced washback. Although the 454 ELLI data from the two case universities were analysed for comparison, the analysis of the 18 ELLI profiles from the student interviewees was the focus of the study on learner washback.

3.7 Methods of analysis

There were three main sets of data. The classroom observation data, teacher and student interviews were analysed to explore how the graduation requirement influenced the students and their EAP curriculum. ELLI data, first of all, were analysed for a comparison in the learning power of the students between the two universities. The comparison was used to find out whether the two cohorts showed similarities or differences in terms of the seven dimensions of learning power. The learning profiles of the student interviewees were then analysed with their interview data, to understand to what extent their learning power may affect their perceptions of graduation requirements.

Below is an overview of the research design and the methods of analysis.

Research Questions	Participants	Data Collection Procedure	Data Collected	Methods of Analysis
What are the effects of the English requirement for graduation on the EAP curriculum for non-English majors in Taiwanese universities?	Teachers and their classes	Classroom observation	17 video- or audio-taped lessons & field notes	Reiterative process of analysis Transana
		Teacher interview	7 interviews and some private talks	Inductive analysis
What are the effects of the English requirement for graduation on the non-English majors in Taiwanese universities?	Learners and teachers	Learner interview	18 interviews	Inductive analysis
		Teacher interview	7 interviews	Analysis relating to learner interview data
To what extent students' learning power explain their perceptions of washback on the learners and their learning?	Learners	ELLI paper questionnaire	454 questionnaires	Independent T-test Pie charts of seven dimensions
		Learner interview	18 interviews	Analysis relating to ELLI data.

Table 3.2 Overview of Research Design and Methods of Analysis

3.7.1 Transcription

Transana 2.3 (Fassnact and Woods, 2005) was used to store, transcribe and code the classroom observation and interview data. The data were then transcribed in full verbatim so that at the stage of data analysis 'information-rich' episodes which 'would not be missed at the transcription stage' (Scott, 2004, p.104) could be selected from data. One good feature of Transana is its compatibility with a number of languages other than English, including Chinese. Thus, the transcriptions could be kept as close to the original data as possible, even with code switching made by the participants. In Transana, I was able to make time codes for the transcriptions to be synchronised with the video and audio data. Thus, interesting episodes could be segmented from the full video data as clips, by highlighting the two time codes. This feature is especially useful in coding and presentation. For coding, Transana allows the creation of keyword groups, and the segmented clips can be coded according to the keywords. Graph 3.1 is an example of a segmented clip under the keyword of 'vocabulary' in the keyword group of 'language development', which is displayed within the whole video data with full verbatim transcripts. This clip, starting at 0:33:47 and ending at 0:35:05, shows that the teacher explains four vocabulary words listed in the teaching material. The visualization column above the transcripts shows a color bar where the clip that is coded with a specific keyword is situated in the whole video data. In addition to keyword coding, another feature of Transana was used in this study. The software can provide reports and graphs that give a general description of a series of video and audio data. In order to get a picture of the coding of each lesson, I chose to present a 'series keyword percentage graph' in Chapter 4 (See 4.3.2). The series keyword percentage graph and the timelines for the observed lessons can complement the observation conducted in the ethnographic approach by providing structure and systematicity.

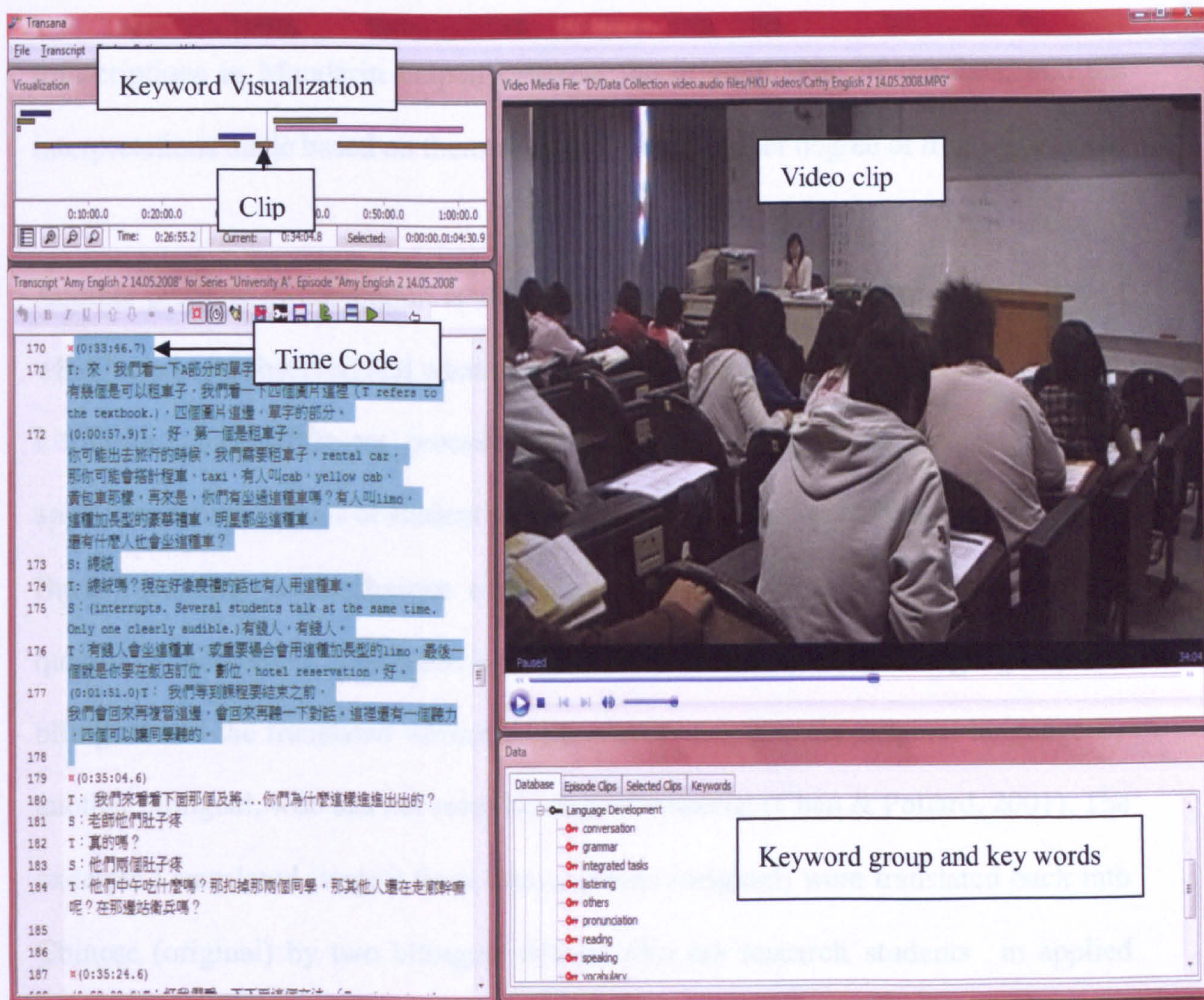


Figure 3.1 Transana clip screenshot

3.7.2 Translation

Phillips (1960, as cited in Birbili, 2000, p. 2 and Temple & Young, 2004, p.165) points out that the ‘unsolvable problem’ of translation has in research is to gain ‘conceptual equivalence’ because ‘almost any utterance in any language carries with it a set of assumptions, feelings, and values that the speaker may or may not be aware of but that the field worker, as an outsider, usually is not’. Although I am an insider for Mandarin, I acknowledge the fact that even a term with direct lexical equivalence ‘might carry ‘emotional connotations’ in one language that will not necessarily occur in another’ (Birbili, 2000, p.2). To keep to the principles of the ethnographic case study, the data were transcribed verbatim, and in the original language. Only the ‘information-rich’ episodes selected for analysis were then translated into English. The full verbatim

transcriptions in Mandarin helped preserve the original taste of the data, and the interpretations made based on them would provide a higher degree of trustworthiness.

Another problem of translation is whether my translation is a good re-presentation of what have been observed and what the participants have said. To tackle this problem, I followed three different procedures. First, the back-translation technique was applied to three excerpts of student interviews and two excerpts of teacher interviews. Back translation is a technique often used in cross-cultural studies to ensure the quality of translation. The material was translated into the target language by one bilingual and the translated version is translated back into the original language by another bilingual, who has not seen the original material (Chan & Pollard, 2001). The excerpts I translated (target) from transcriptions (original) were translated back into Chinese (original) by two bilingual friends who are research students in applied linguistics and drama education. Several divergences were noted. Therefore, I decided to give the two translators all the extracts that were cited in this dissertation in both the original Chinese version and my translation. They were asked to examine the equivalence between the two versions and to provide suggestions on how to improve the translations. Furthermore, the teachers in this study, who all had a good level of English proficiency, were asked for respondent validation of the revised translations of the extracts. The extracts in Chapter 4, 5, and 6 are the final versions resulting from the above-mentioned procedures.

.

3.7.3 Data analysis for observation and teacher interview data

The analysis of classroom observations and teacher interviews went through a reiterative process. How each type of data was analysed and triangulated at each stage was specified below. The analysis started with a more grounded approach, involving

an inductive coding of the recorded observation data and the field notes, the establishment of links from teacher interview data to the observation data and document analysis of the teaching materials collected from the teachers.

The main data for exploring the impact on the EAP curriculum for non-English majors was the observation data, including the video data and the field notes. Pertaining to the ethnographic approach to classroom observation, a more grounded approach was taken for the first stage of data analysis. Although this study was inquiry-driven, 'the pull of the data' was allowed and unavoidable, as in O'Brien et al. (2000) and Scott (2005)'s studies. One reason was that as with both studies, this study adopted a case study approach and thus, how firmly the analysis was based on the huge amount of raw data was utterly important (Blaxter et al. 2001). The other reason was the nature of the English graduation requirement. Since the graduation requirement did not target one specific English proficiency test, a more grounded approach towards analysis would allow not only the macro impact of the requirement, but also the micro impact from different tests on the curriculum to emerge. Thus, at this stage, the full verbatim transcripts of the observation data and the field notes were subjected to inductive coding, to see whether there were influences from the GEPT or other English proficiency tests in the lessons, and also what patterns each teacher's lessons showed. The field notes were not analysed as a set of independent data, but as supplementary data to the observation data, which provided details not being able to be captured by video and audio-taping. The teaching materials used in the observation sessions were also examined, to see whether those materials could be linked to any test influence included in the graduation requirement. Inductive coding was also applied to the teacher interview data to pick up those which were relevant to the above data and impact the curriculum, and to check whether what the teachers said in their

interviews matched what was observed in their classrooms.

The process of the analysis was similar to Holliday's (2002, p.100) description of the gradual process from 'messy reality', the corpus of raw data, to forming thematic organisations of data which are used to develop and support arguments, as shown in Figure 3.2. The corpus of raw data used to answer the research question in this chapter consisted of the observation of data, field notes, and teaching materials and teacher interview data, all compiled together.

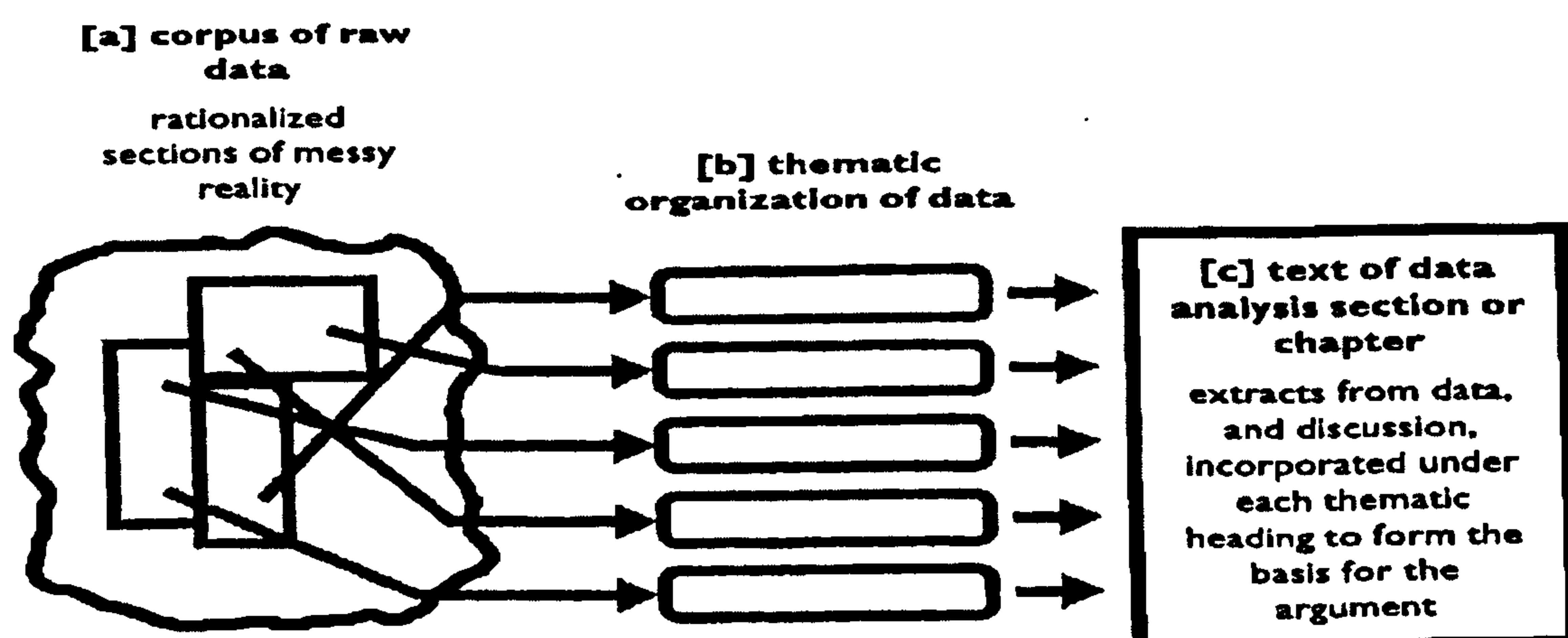


Figure 3.2 Holliday's (2002) process of analysing qualitative data

During the first stage of coding and data analysis, I realised that there was implicit evidence of GEPT washback that was revealed by the teachers in University A (4.3.4) but not sufficiently captured by the initial data collection, because of the limited time spent on the research site. In order to determine whether the materials were indeed products of GEPT test influence, I further collected electronic files of mid-term and final test papers used in University A. Although the test papers were not used during the time I was in the university, they were indeed used in the semester during which I conducted this study. Altogether, I received two sets of test papers, including midterm test papers used for both English Integrated Skills Training 1 and 2 for that semester (Appendix H).

The second procedure after the initial data analysis was to code the observation data with Transana, the software for analysing qualitative video and audio data (3.7.1). As mentioned in Chapter 3, transcripts of video and audio data in Transana were coded with keywords. Then, the way in which each observed lesson in one case university was coded was displayed in a 'series keyword percentage graph'. The third procedure was to make a timeline that depicted the chronological flow of each observed lesson. The above two procedures were used in a supplementary manner to the initial data analysis. One purpose in doing this was the attempt to add 'rigorous flexibility', the goal for systematic analysis of qualitative data, pointed out by Dornyei (2007, p. 245). The other purpose was to improve the presentation of the large amount of observation data, in order to provide a platform for comparison between the two case universities.

Before providing details of the coding keywords and the timelines, the way in which the observed lessons were segmented and time coded in Transana and presented in the timelines should be explained, as this was crucial for transcribing and presenting the observation data in a more systematic way. Instead of using the one minute interval real time coding by the original COLT (Frolich, Spada and Allen, 1985) or Cheng's (2005) idea of a 'segment', 'a stretch of classroom discourse having a particular topic and involving participants (both the teacher and students) in carrying out an activity or task through interaction' (Mitchell, Parkinson, & Johnstone, 1981, p.12-14), the lessons in this study were segmented according to related activities. In other words, one segment consisted of several activities which were related by either a section in the teaching material or might be fragmented, being viewed as several isolated activities. For example, an episode of a lesson in University A might include students listening to the audio clip of a short conversation, the teacher asking the students to read aloud the conversation lines along with the re-playing of the conversation, and

the teacher explaining vocabulary words and sentence structures in between. Since the keyword coding and the timeline were considered as supplementary to the main analysis, the lessons segmented in this way did not reduce the complexity of the data, yet provided a structure to the video and audio data. In sum, the observed lessons were time coded and cut into various segments according to the above principle in Transana. Then, the specific time codes were noted down in a timeline, so that the structure of a lesson could be clearly shown.

For the second procedure, the segments were coded with keywords. The keywords first derived from the initial inductive analysis and literature review, and were revised after several times of recoding and refining. The majority of keywords were related to ‘contents’, which refers to ‘the type of knowledge that teachers were trying to transmit to their students (e.g. the form of a specific grammar structure, or facts relating to a particular topic), or to the general skill they were focusing on (e.g. reading, listening)’ (Wall, 2005, p.16). The reason was that the ‘contents’ of an observed lessons showed what the lessons were about, and also what aspects the teachers chose to teach, which could reflect the influence of a test.

Lastly, the observation data were presented in the form of a timeline, in order to show the chronological flow of the lessons (see 4.2 for further details).

3.7.4 Analysis of interview data

Although both teacher and student interviews were analysed inductively, coding processes were different to some extent. The teacher interviews were used in three ways: to triangulate classroom observation and learners’ perceptions of the impact of the requirement on them, and to reveal their perception of the graduation requirement.

I first picked out extracts that were related to the classroom observations and their perceptions of the impact of the requirement on learners.

For the other part of the data, transcripts were read repeatedly for themes to emerge. Then, the transcripts were coded according to themes. This bottom up feature of inductive coding worked better with the ethnographic approach. Inductive coding prioritised the participants' voice in semi-structured interviews instead of the researcher's presumptions of what one should find out from the data. This emphasis on the emic perspective is one important feature of the ethnographic approach (Hammersley & Atkinson, 1995, 2007).

3.7.5 Statistical analysis of the ELLI questionnaire data

The ELLI data were analysed using SPSS 16.0. The results of the ELLI questionnaire displayed a student's seven distinct dimensions of learning power: changing and learning, critical curiosity, meaning-making, fragility and dependence, creativity, learning relationships and strategic awareness. There are 72 items in the ELLI questionnaire. The following table (Table 3.4) reports which items are used to calculate the seven dimensions.. 'Not at all like me' is 1 point and 'A little like me' is 2, while 'Quite a lot like me' scores 3 points and 'Very much like me' scores 4.

Items Corresponding to Seven Dimensions	
Changing and learning	Q32, Q34, Q42, Q48
Critical curiosity	Q1, Q2, Q4, Q6, Q14, Q16, Q25, Q35, Q36, Q37
Meaning-making	Q7, Q18, Q21, Q27, Q29, Q62
Fragility and dependence → Resilience	Q10, Q13, Q15, Q17, Q24, Q28, Q31, Q43, Q45, Q47, Q50, Q54, Q57, Q61, Q63, Q68, Q70
Creativity	Q8, Q11, Q19, Q22, Q23, Q33, Q40, Q56, Q66, Q67
Learning relationships	Q3, Q5, Q26, Q30, Q38, Q39, Q41, Q44, Q51, Q55, Q58, Q72
Strategic awareness	Q9, Q12, Q20, Q46, Q51, Q49, Q52, Q53, Q59, Q60, Q64, Q65, Q69, Q71

Table 3.3 ELLI items corresponding to the seven dimensions

Among the seven dimensions, ‘fragility and dependence’ is the only one that represents a negative orientation of learning power. In order to present all the dimensions on the same platform, the opposite pole of ‘fragility and dependence’, ‘resilience’ is used instead in this study (5.6.3), as in other ELLI studies (Dickin Crick et al, 2004; Dickin Crick, 2007; Dickin Crick & Yu, 2008). The scores of the 17 items that relate to ‘fragility and dependence’ were then converted into scores for ‘resilience’, with ‘Not at all like me’ becoming 4 points, ‘A little like me’ 3, ‘Quite a lot like me’ 2 and ‘Very much like me’ 1.

Lastly, the students’ mean scores were compared between University A and University B. The statistical procedure used for the comparison was an independent sample T-test. The comparison between courses which were originally proposed was abandoned, for two reasons. Firstly, an initial T-test of students’ ELLI scores in the GEPT-related ‘English reading and writing’ course and the non-GEPT-related ‘English II’ course revealed no significant difference on any dimension. Secondly,

other than the students in the observed classes, about half of the ELLI data collected in University B were non-English majors in their second year. As the second-year non-English majors did not need to attend English courses, there was no course variable for them. As a result, the T-test was only conducted to examine whether there was a difference in the seven dimensions between students in University A and those in University B. A university profile was then created, with 7 pie charts. Each pie chart was made up of three areas in traffic light colours, red, yellow and green, indicating the percentile of students who reported themselves as having low levels (red), moderate levels (yellow) and high levels (green) of one dimension (See Figure 5.1).

3.7.6 Integrated analysis for ELLI and student interview data

The ELLI scores of the 18 student interviewees and their interview data were analysed in an integrated way, in order to understand how their learning power as assessed by ELLI related to the ways in which graduation requirements might have influenced them. Each student had a learning profile with 7 scores put together in a spider diagram, which described the kind of learners they were by showing their strengths and weaknesses on the 7 dimensions (See Figure 5.3). The analysis focused on the relationship between the students' learning profiles and their perceptions of the graduation requirement and the washback of a language test.

3.8 Trustworthiness of this research

Since the present study was a mixed method study with mainly qualitative data, trustworthiness, the quality of qualitative studies proposed by Lincoln and Guba (1985) was more appropriate as the main criteria for this study than the traditional validity and reliability. They specify four criteria under trustworthiness: credibility,

transferability, dependability and confirmability, which can be considered as equivalent to internal validity, external validity, reliability and objectivity in quantitative studies.

The original version of the ELLI questionnaire is an instrument that has gone through several stages of adjustment and validation (Deakin Crick, *et al.*, 2004, Deakin Crick and Yu, 2008) and has been used in several large scale studies within different educational contexts in the United Kingdom (Deakin Crick, *et al.*, 2004). The results show that the ELLI has been able to provide learning profiles. The simplified Chinese version of ELLI was tried out and validated in a comparatively small-scale study by Liu (2007). The idea of piloting my traditional Chinese version, which contained only about 5% difference from the simplified Chinese one, was found unfeasible (See Appendix D. Changes are highlighted.). This was because most Taiwanese could understand writings in simplified Chinese to a certain extent, and taking one version before the other would make taking the other version too similar to notice the differences. Thus, only the examinations from fellow researchers and teachers were made to provide validation. As for the qualitative interviews and classroom observations, the measures proposed by Lincoln and Guba (1985), namely respondent validation (member checking), triangulation and thick description were taken.

3.8.1 Respondent validation

Respondent validation refers to having ‘participants confirm or disconfirm data, analysis, and conclusions’ (Qi, 2004, p.177). Ideally, respondent validation in this type of study should be conducted at two stages. The first stage, as previously mentioned (Section 3.6.2), was for the translation of the transcripts. The second stage was for the teacher and student participants to read the analysis, and the

interpretations and conclusions. As most of the teachers in this study expressed willingness and expectations to read the analysis, respondent validation from the teachers could be achieved. However, due to the constraints caused by the level of English proficiency of students and limited time, it was not possible to receive respondent validation from students, which was a limitation of the present study.

3.8.2 Triangulation and Thick description

Watanabe (2004) considers triangulation as being important in establishing credibility and dependability in research. This research employed methodological triangulation and data triangulation in several ways. Methodological triangulation involved teacher and student interviews with classroom observation to shed light on the washback of language tests on teaching and learning. The quantitative data from ELLI were complementary to the more in-depth student interview data in investigating students' perceptions of the impact of the graduation requirement on their English learning. Data were triangulated with perspectives from both teachers and students because 'it may be the case that some aspects of washback exist for learners but not for teachers, whereas other aspects exist for teachers but not for learners' (Watanabe, 2004, p.29).

Watanabe (2004) also suggests that 'thick description' of the context where the test is used will be helpful for readers to assess whether the results of the study can be transferred to other contexts. As this study took on the case study approach and placed emphasis on the complex contexts that the graduation requirement entails, thick descriptions were essential. The ethnographic approach of collecting observation data and the iterative process of coding and analysing also allowed for a more narrative account of the washback phenomenon.

3.9 Ethical issues

In gaining access to University B before the research took place, I was asked to provide the head of the English department a form or any evidence of how human subjects protection would be achieved in my study along with my research proposal and approvals from some of the teachers. I went through the ethical procedures of the Graduate School of Education (GSoE), as specified in the ethics form. I discussed the ethics guidelines with my supervisors and the ethics group in the school. Negotiated access was granted when the ethics form was presented to the head of the department with a letter from the ethics group of the GSoE, stating the completion of the procedures (See Appendix E).

Ethical guidelines for the purpose of this research were drawn up based on the British Educational Research Association ethical guidelines (2004), the recommendations on good practice in applied linguistics by British Association of Applied Linguistics (2006) and International Language Testing Association Code of Ethics (2000). The guidelines can be divided into the following categories: Informed Consent, Confidentiality and Anonymity, Participant's Rights and Researchers' Responsibilities.

3.9.1 Informed consent

Two written consent forms (see Appendix F&G) were distributed to the teacher and student participants who were involved in classroom observation. The forms contained the purpose of the research, the research methods, the requirements of the participants and their rights and the intended use of the collected data. In addition to the consent forms, the reasons why I was in their classrooms were either announced by the teachers at the first observation, or presented by myself as requested by some

teachers.

Most consent forms were returned, with only a few students being intimidated by video-recording. They expressed concerns about their faces being captured by the camcorder, which related to the issue of confidentiality and anonymity. The negotiations made between the students and me were described in detail in the next section. Students who were only involved in completing ELLI were given a short-form consent at the beginning of the questionnaire, which also stated the research purpose, the intended use of ELLI data and reassured them about issues of confidentiality and anonymity.

3.9.2 Confidentiality and anonymity

Participants' rights to confidentiality and anonymity should be respected. The consent forms clearly stated the confidential and anonymous treatment of the collected data. Anonymity could be assured by using pseudonyms for all teacher and student participants who had been interviewed. However, in the present study, maintaining the participants' full confidentiality was difficult to achieve in the form of video data. Thus, several attempts were made to help participants feel more comfortable in revealing their identities in the videos. First, their willingness to reveal their identities was requested in the consent. Second, the positions where the camcorder was set were negotiated in some classrooms. In one classroom, the camcorder was originally positioned at the front of the classroom and was moved to the back of the classroom because students sitting in the first row felt uneasy with the camcorder capturing their full faces. Furthermore, as the camcorder was set at a fixed spot, those who did not wish to be videotaped were provided the chance to move to the blind spots the camcorder could not capture.

3.9.3 Participants' rights

In addition to the rights mentioned above, participants had the right to withdraw at any point during the time the research was carried out. For most classes, the observation lasted two to three weeks, and student participants were not required to attend all sessions. The student interviewees were asked beforehand for their permission to reveal their names, the departments they belong to, and to record the interviews. In the consent forms, all participants' rights to discontinue their participation in the research at any stage were ensured.

3.9.4 Researcher's responsibilities

It is a researcher's responsibility to ensure the following: that no harm will be brought to any of the participants throughout the research process, and that protection of the collected data is ensured. I was fully aware of the potential harms this study might have on the case universities and the participants. Therefore, attempts were made to avoid these possible harmful effects as much as I could. Although the case universities were kept anonymous, certain clues provided by the institutional contexts essential for this study could be linked to the identification of the universities. As a result, information such as the exact ranking of the universities and their results of university assessment, which might influence their reputations but not essential to the study, were not included. For the participants, the intrusion of an observer in the classroom and the intimidation caused by videotaping might result in uneasiness and discomfort. Besides the abovementioned measures taken on the camcorder, I held conversations with students during breaks, and answered their questions concerning advanced study abroad, so that gradual familiarity would alleviate their uneasiness and discomfort.

3.10 Summary

In this chapter, I have attempted to make my philosophical position of this study clear (3.2). I have also presented the research questions this study tried to answer (3.3) and also discussed how the pilot study had informed the main study (3.4). In addition, I have given an account of the research context and the research approach (3.5) and described in detail the research methods used (3.7). I have then explained the method of analysis for each set of data (3.8). This chapter ends with a discussion about the quality of the study and issues related to ethical considerations. The next chapter presents the findings of Research Question One: the effects of the graduation requirement for English proficiency on the EAP curriculum for non-English majors.

CHAPTER 4 Impact on the EAP Curriculum for Non-English Majors

4.1 Introduction

This chapter explores the impact of the graduation requirement for English proficiency on the English for Academic Purposes curriculum of non-English majors. Such a curriculum is different from the English for Specific Purposes curriculum for students within specific departments (e.g. business English courses for the Department of Finance in University A), or the extended and specialised English curriculum for English majors. Although the graduation requirement has brought about discussions concerning its implementation in Taiwanese academia, there have been few formal discussions of its influence on the EAP curriculum for non-English majors, and subsequently, on the learners, the non-English majors themselves. To understand in what ways learners have been influenced by the graduation requirement, first, the effect of the requirement on the EAP curriculum is discussed. The General English Proficiency Test (GEPT) has been the most popular test taken by university students among English tests accepted by the requirement (See 1.3.1). Thus, one aim of this chapter is to see if the test has indeed brought about the strongest washback to the EAP curriculum and whether there were influences from other language tests as well. It is hoped that the findings derived largely from close observations of the English classrooms can allow readers to understand what has happened after the implementation of the policy (see 1.3.1) and how and why teaching may have changed or not changed. By focusing on test washback, it is also expected that the findings can contribute to the on-going discussions of washback in different educational contexts. The complexity of how washback works under a policy that sets

a benchmark for several tests instead of one designated test sets this study apart from previous washback studies in the literature, which often focus on one particular language test only.

The title and the theme of this chapter is derived from the notion of washback to the programme, which Bailey (1996) refers to as the effect of test-derived information on stakeholders other than test-takers, including teachers, administrators, counsellors and curriculum developers (See 2.2.6.). It may be differentiated from washback to learners, which is the focus of the next chapter. However, for the purposes of consistency, 'curriculum' was used instead of 'programme'. Teachers are considered as central in studying washback on the EAP curriculum in this study, with stakeholders other than teachers and students as peripheral because teachers have direct interaction with the students in the classrooms. The classroom observations with supplementary field notes, teachers' explanations in the interviews on what they did and why they did so in the lessons, private talks with some teachers and also teaching and testing materials collected from the teachers were the main sources of data for understanding the impact of the graduation requirement on the EAP curriculum. In particular, this chapter addresses the first research question:

What are the effects of the English requirement for graduation on the EAP curriculum for non-English majors in Taiwanese universities?

The sub-questions are as follows:

- a) What are the effects of specified tests stated in the requirement on the English curriculum?
- b) In particular, what are the effects of the GEPT on the English curriculum?

- c) Is there difference between the effects of the different tests on the English curriculum? What are the differences and why??
- d) What are the teachers' perceptions towards the requirement?
- e) What are the teachers' perceptions towards the effects of the tests on the English curriculum?

In the next section, after a brief summary of the method for data analysis, I will describe the timelines used for presenting classroom observations in 4.2. I will also present the timelines of one lesson of each teacher. Then, an analysis of GEPT washback will be presented under four aspects: teaching materials in 4.3.1, teaching contents: language focus and skill development in 4.3.2, explicit reference to the test in 4.3.3 and washback on 'testing and assessment' in 4.3.4. Teachers' view on GEPT washback, which derived from the interview data, is presented in 4.4. In 4.5, the reasons for the presence or the lack of GEPT washback are discussed. Finally, a summary of this chapter is presented in 4.6.

4.2 Presentation of classroom observations in timeline

As mentioned in 3.7.3, the classroom observations in this study were analysed inductively, and went through a reiterative process of analysis for the themes to emerge. For a better presentation of the large amount of observation data, an additional procedure was conducted. I made timelines with specific time codes that depicted the chronological flow of each lesson. In addition to a 'series keyword graph' (see 3.7.3 and Figure 4.4, 4.5), the timelines also provided a platform for comparison between lessons from different teachers, and between the two case universities.

The timeline presented information including the teaching contents, the activity type, the teaching materials and whether there was evidence of test influence. Each activity was separately described, with a phrase indicating what the activity was about (e.g. reading aloud, dictation, role play). How students were organised as participants in classroom interaction was only mentioned along with the descriptions of the activity if it was not intended for whole class participation. The initial inductive analysis of classroom observations in this study revealed that the majority of the lessons were highly teacher-centred. In addition, unlike Cheng's (2005) study with a focus on participant organisation as an intended washback of the new HKCEE, participant organisation was not an intended effect of the graduation requirement. The English proficiency tests accepted in the graduation requirement were not tailored to the EAP curriculum and there was no intended washback of those tests to promote more practice opportunities for students in their English classrooms. The short descriptions of the participant organisation in the timelines were thus sufficient to show whether the lessons were teacher or student-centred, and the overall pattern of the participant organisation in the lessons. The information of teaching materials used in the lessons was described according to two dimensions adopted from COLT (Frolich, Spada and Allen, 1985): type of materials and source/purpose of materials. The types of materials were sub-divided into written, audio, and visual. The materials were also noted according to their source and purposes, depending on whether they were pedagogical (i.e. materials specifically designed for L2 learning), semi-pedagogical (i.e. modified real life materials for pedagogical use) or non-pedagogical (i.e. materials originally intended for nonschool purposes) (*ibid.*). Lastly, the timeline also pointed out explicit evidence of washback in the lessons, including teachers' direct reference to any English test, or direct links between the lessons and a test.

The timelines for one lesson of each teacher in this study is presented below:

University A

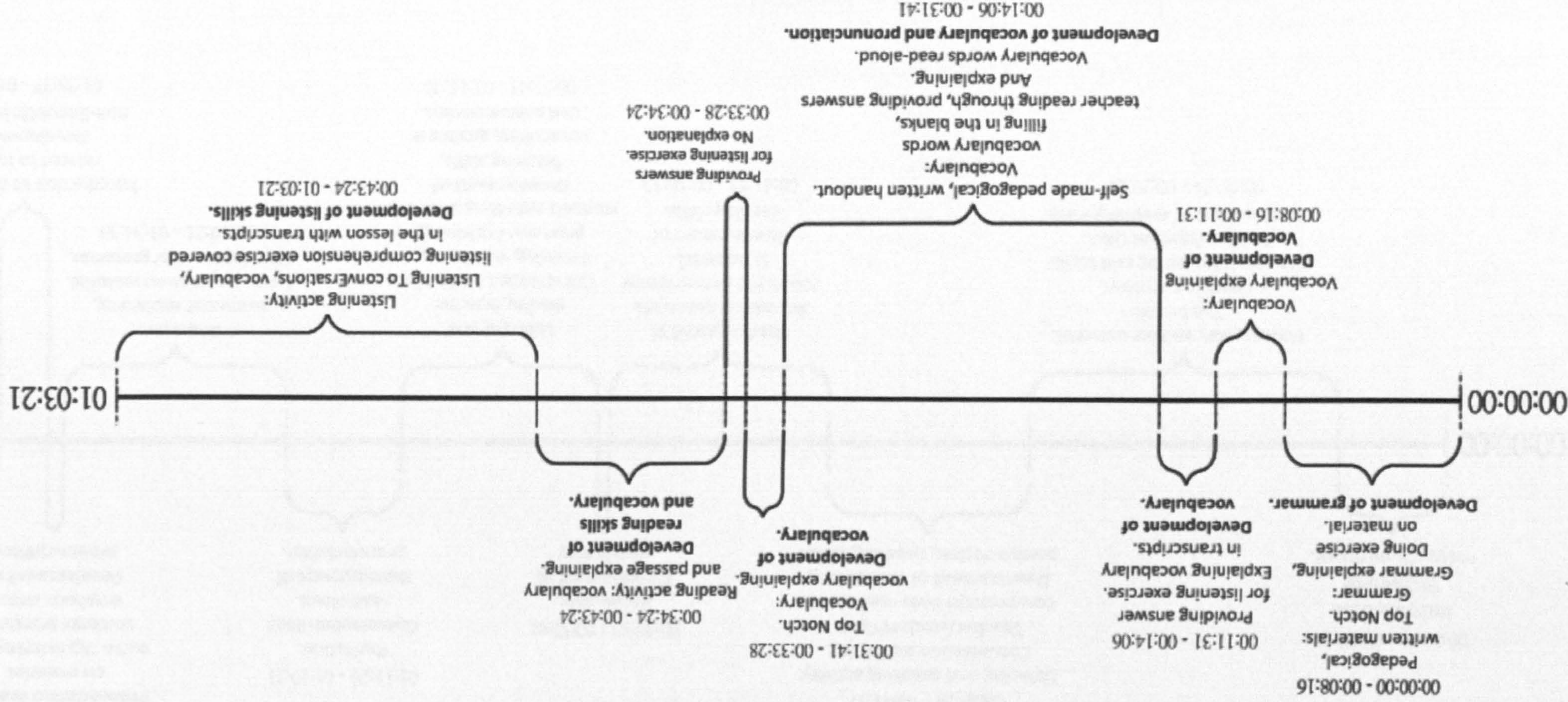


Figure 4.1 Timeline of Adam's lesson for English Skills Integrated Training 2008.05.21 (Note: Teachers with pseudonyms beginning with 'A' refer to teachers in University A and vice versa.)

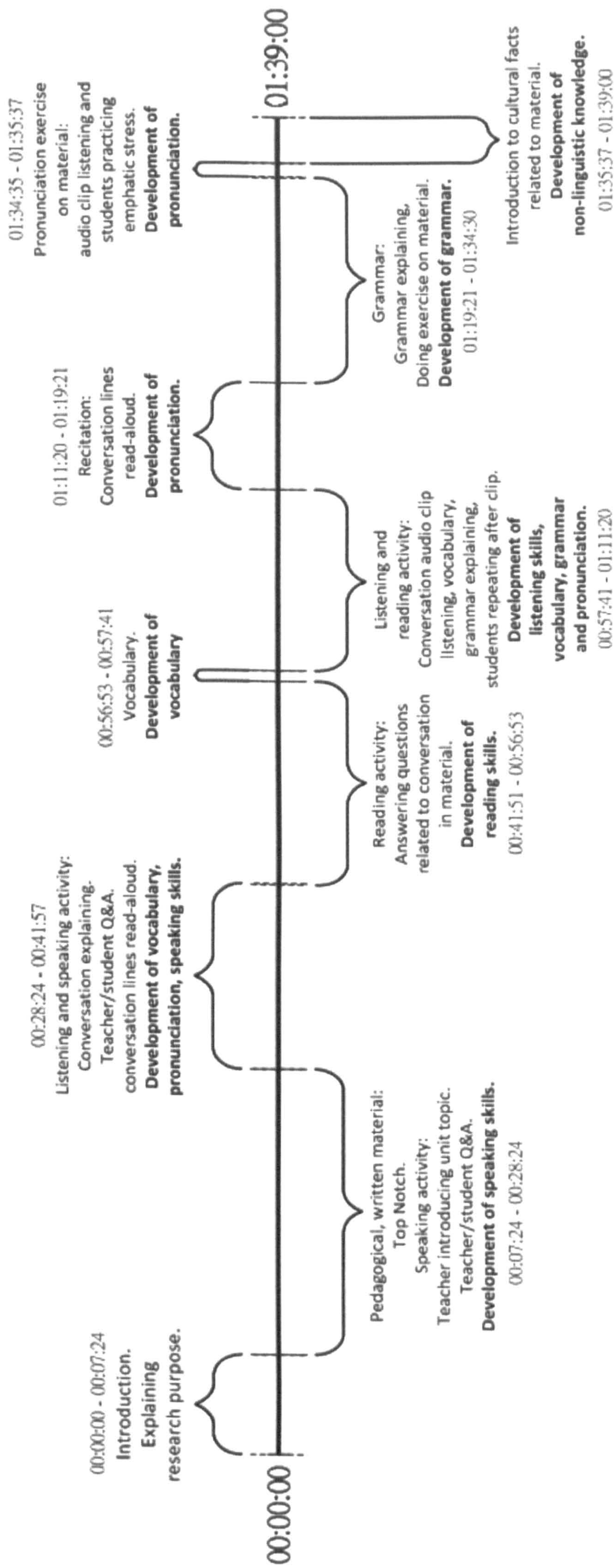


Figure 4.2 Timeline of Alice’s lesson for English Skills Integrated Training 2008.05.09

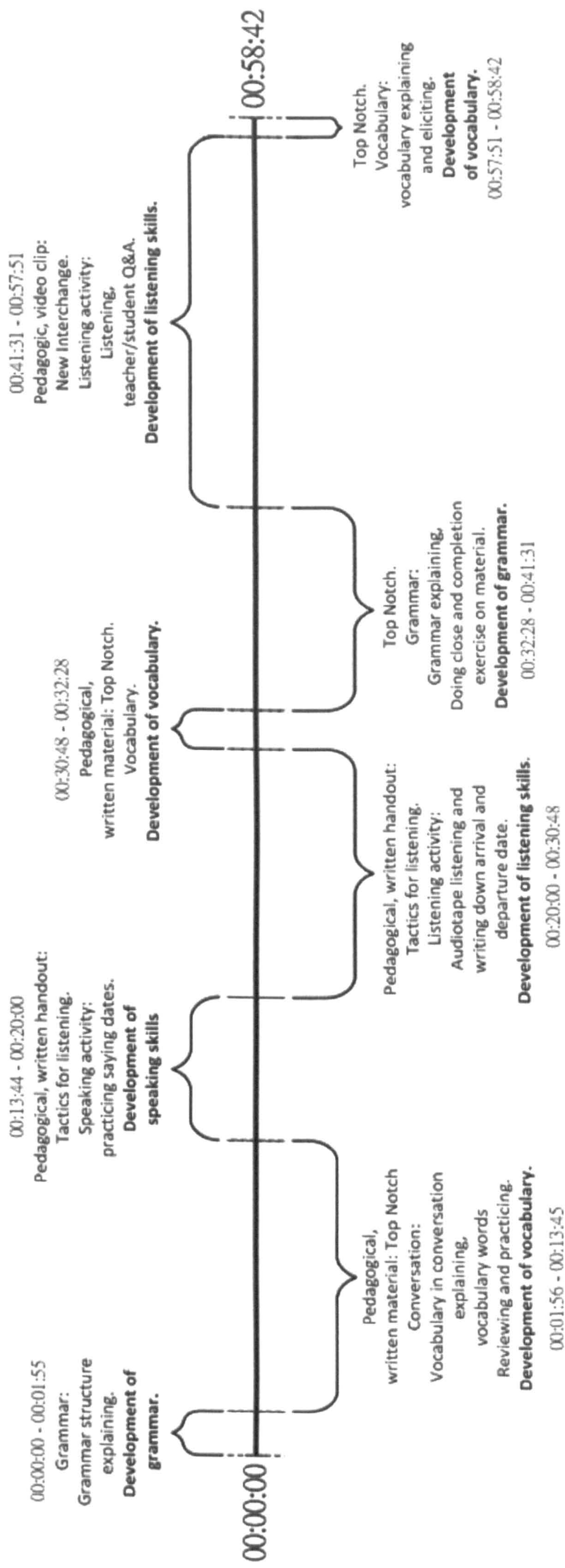


Figure 4.3 Timeline of Amy's lesson for English Skills Integrated Training 2008.05.14

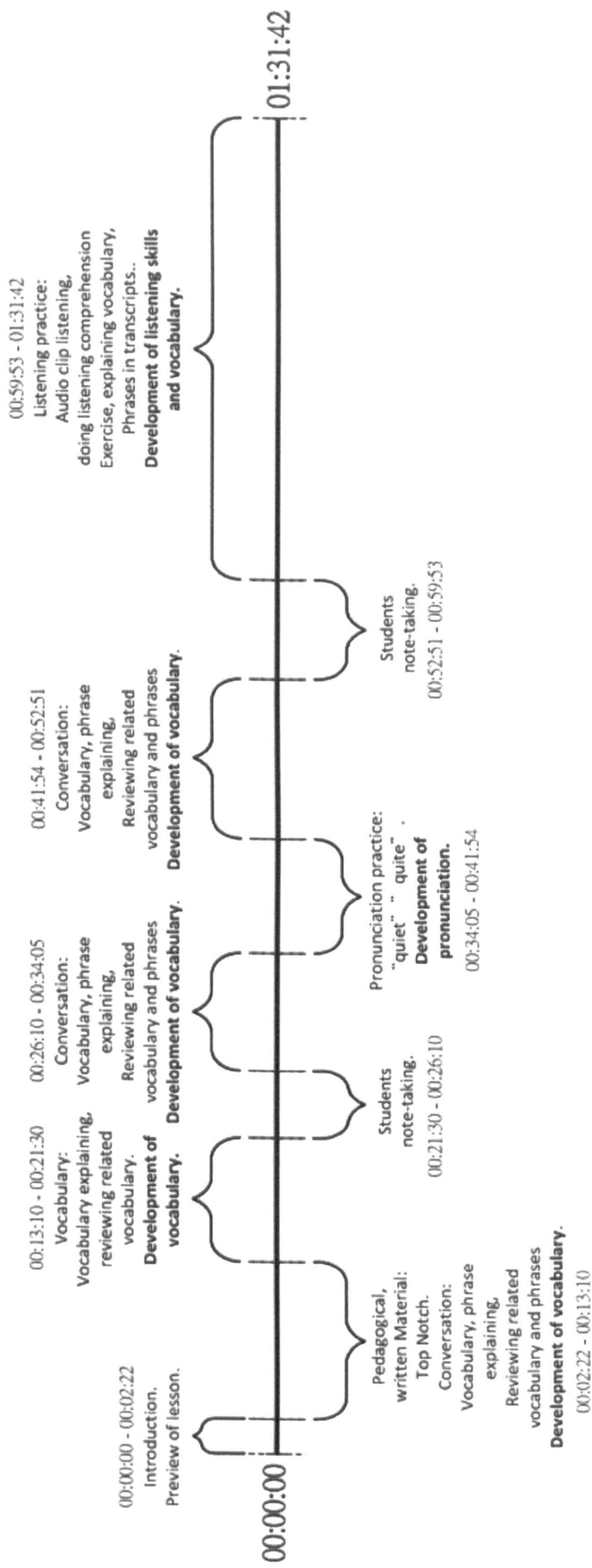


Figure 4.4 Timeline of Anna's lesson for English Skills Integrated Training 2008.04.28

University B

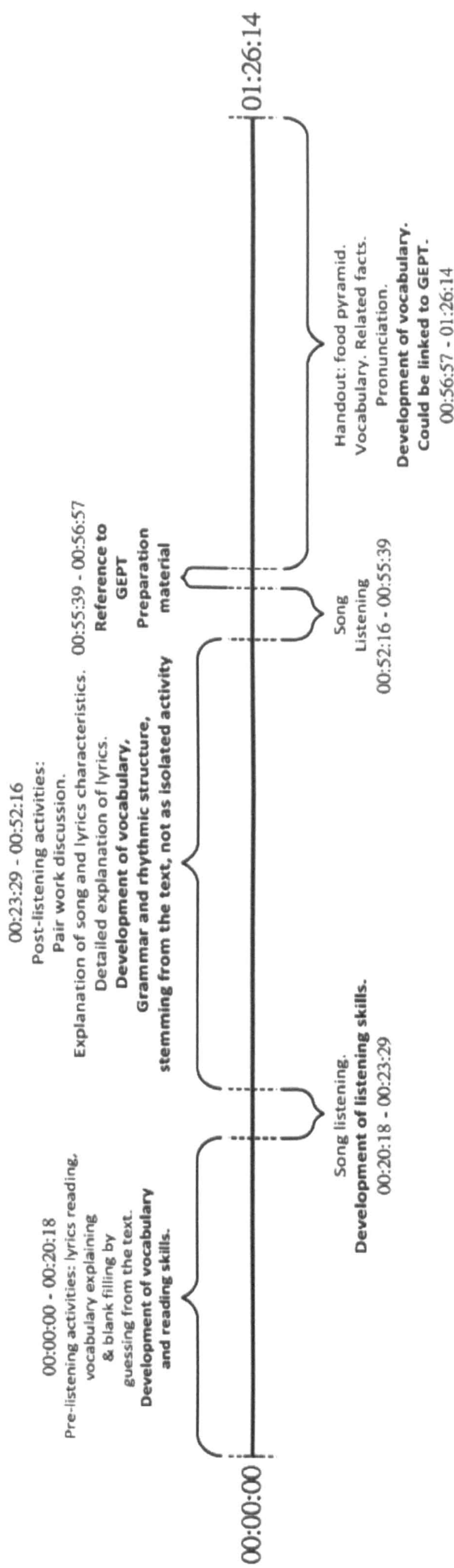


Figure 4.5 Timeline of Becca's lesson for English Reading and Writing 2008.05.27

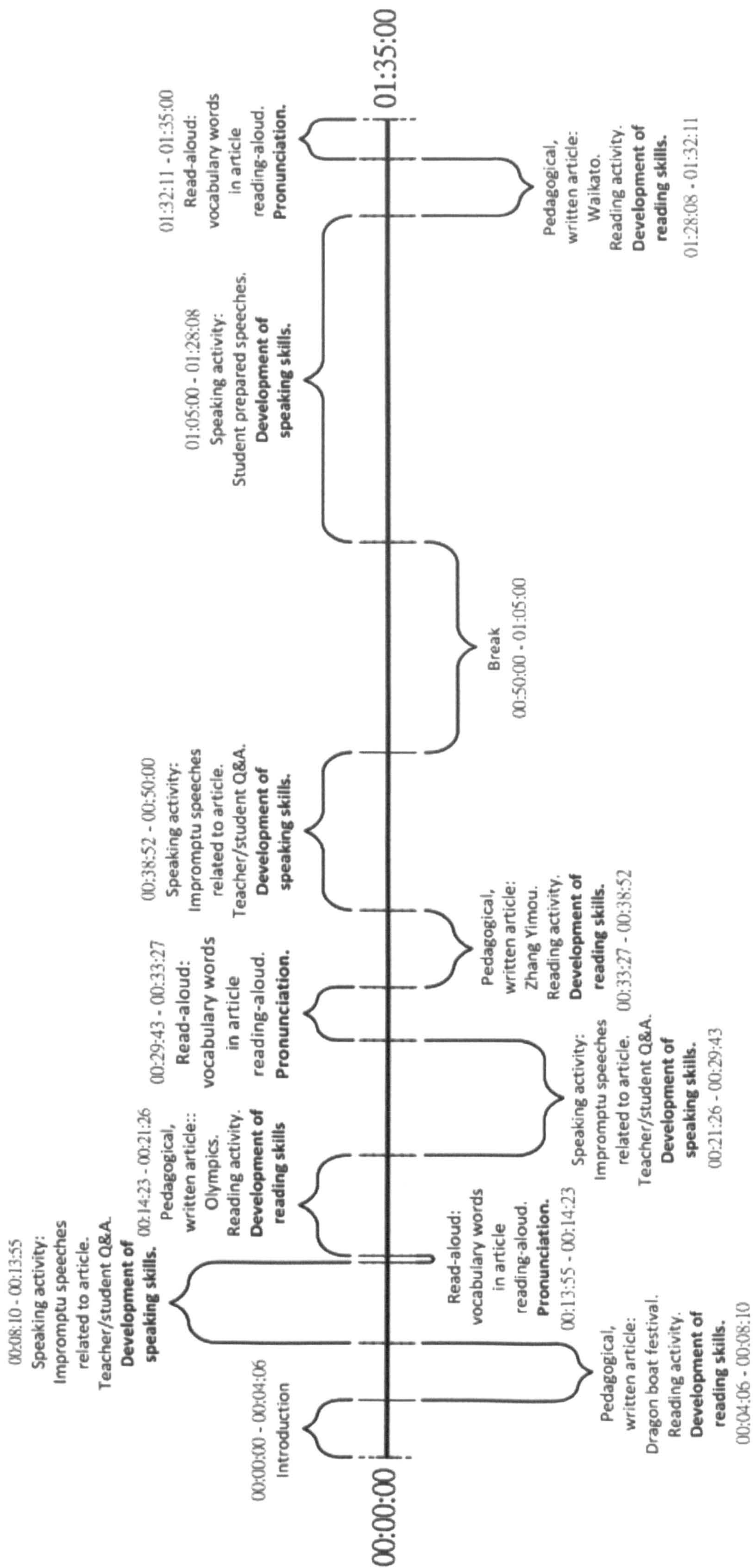


Figure 4.6 Timeline of Ben’s lesson for English II 2008.06.03

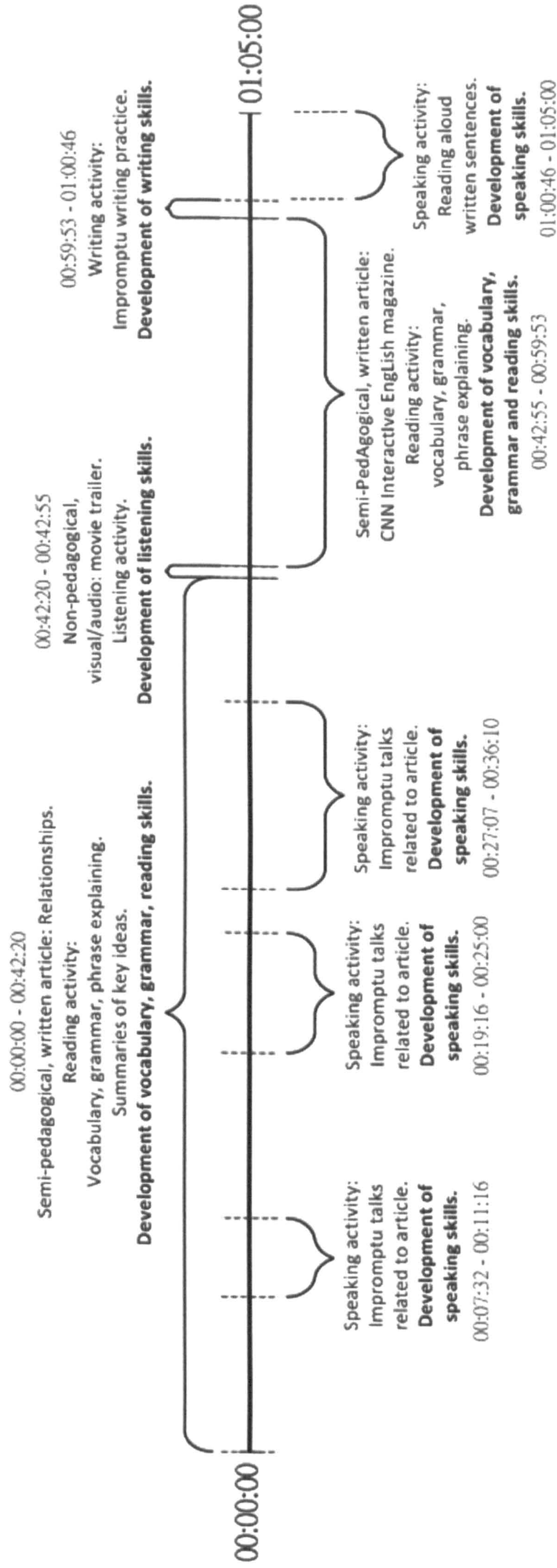


Figure 4.7 Timeline of Betty's lesson for English II 2008.06.03

In the next section, the results of the analysis will be presented in the following parts:

1) impact on teaching: teaching materials, teaching contents, explicit reference to the test, washback on testing and assessment 2) teacher's perceptions of the impact of the requirement and lastly, 3) factors that shape the impact and washback on the EAP curriculum for non-English majors.

4.3 Impact on teaching

In the following sections, the impact of the graduation requirement on the EAP curriculum for non-English majors will be discussed in relation to: 1) teaching materials 2) teaching contents 3) direct reference to tests 4) testing and assessment.

4.3.1 Teaching materials

This section reports on what materials were used and how they were used in the observed lessons, as evidence or lack of evidence of the influence of any English proficiency test. Table 4.1 and Table 4.2 below provide an overview of the use of teaching materials and how many of the materials were covered in the observed lessons.

University A

Name	Class	Material	Material Type	Coverage
Adam	English Integrated Skills Training (EIST) II	Top Notch 1 Unit 8	Pedagogic. Audio, written.	p.100~108 (page number of Top Notch)
		Self-made handout with fill-in-the-blank items of vocabulary words and Chinese translations of the words covered in the units for the final exam.	Pedagogic. Written.	All
Alice	EIST II, EIST IV	Top Notch 2 Unit 8	Pedagogic. Audio, written.	p.86~90.
Amy	EIST II	Top Notch 1 Unit 8	Pedagogic. Audio, written.	p.104, 105, 109
		Handouts from 'Tactics for Listening'.	Pedagogic. Audio, written.	All
		Video clips from 'New Interchange'.	Pedagogic. Visual, Audio.	All
		Video clips from 'True Colors'.	Pedagogic. Visual, Audio	All
Anna	EIST II, EIST IV	Top Notch 2 Unit 7	Pedagogic. Audio, written.	p.76~89, p.87, 90.

Table 4.1 University A: teaching materials

University B

Name	Class	Material	Material Type	Coverage
Becca	English Reading and Writing	Photocopied handout of lyrics 'Sixteen Going on Seventeen' with fill-in-the-blank items.	Semi-pedagogic. Audio, written.	All
		Photocopied handout of food pyramid.	Non-pedagogic. Written.	All
		'Get the Point': Intermediate level GEPT reading and writing.	Pedagogic. Written.	Referred to.
Ben	English II	'Studio Classroom', May issue.	Pedagogic. Written.	Articles: Dragon Boat Festival, The Olympic Spirit, Yi-Mou Chang.
		'English Digest', June issue.	Pedagogic. Written.	Articles: The Orsay Museum, Monthly News Report.
Betty	English II	Photocopied handout of article 'Journeys of the Heart'.	Semi-pedagogic. Written.	All
		Photocopied handout of article 'Indiana Jones and the Kingdom of the Crystal Skull' from CNN magazine, June issue.	Pedagogic. Written.	All
		Video clip of Trailer 'Indiana Jones and the Kingdom of the Crystal Skull', provided by CNN magazine, June issue.	Non-pedagogic. Visual, Audio.	All

Table 4.2 University B: teaching materials

Table 4.1 shows that the teaching materials used in the lessons in University A were limited in both number and variety. The document analysis of the materials revealed that there was little evidence to link these materials to the GEPT or other English proficiency tests. According to the teachers, teaching materials were assigned by the English department for their classes in the General English curriculum. The teaching material assigned for both Year 1 and Year 2 non-English majors were Top Notch, Level 1 and 2, depending on the proficiency level of the students in the classes. Top Notch was international English language material for adults and young adults, developed by Saslow and Ascher(2006), and published by Pearson Longman. In the 'To the Teacher' section, which introduces the material, it says that Top Notch is material that promotes communicative competence for both native and non-native speakers of English, and develops both linguistic and cultural proficiency of English with various topics and treats English as an international language.

According to the data from Table 4.1 and the timelines (See 4.2), all of the teachers in University A made extensive use of the book, except for Amy. The material for every segment was Top Notch or directly linked to Top Notch. The extensive use of the book was particularly evident in Adam and Alice's lessons. Alice went through almost every section in Top Notch unit 8 in the order arranged by the book, including a small section for pronunciation on emphatic stress. Adam's routine was very similar to Alice. What was special was that he even covered a supplementary section on the last page of the book, playing the song deliberately written for each unit in the material and working through the lyrics, explaining vocabulary and the meanings of the lyric lines. The only supplementary material he gave to the students was a self-made handout of fill-in-the-blanks that included all the words and phrases he had taught in the units that would be included in the final examination, and the Chinese translations

for each word and phrase. Thus, although the material was made by the teacher himself, the content was directly linked to Top Notch. Slightly different from the above two teachers, Anna was more selective in her choice of sections to go through, covering only conversation models and listening comprehension sections, but still no materials outside Top Notch were used. The reason why their lessons were highly textbook-oriented was perhaps because all the students in their classes had the same test for their mid-terms and finals, provided by the Taiwanese publisher Tung-Hua, which represented and promoted the Top Notch series. Alice told me in the post-observation interviews that this was the main reason why she tried to cover every part of the book:

Extract 4.1

“They have the same test and that is why all of the textbook contents are covered. But I also think that the textbook has a systematic arrangement of contents and if I can cover all of them, eventually students’ proficiency can be quite good.” (Alice, interview, 26.05.2008)

Among the teachers in University A, Amy was the only teacher who used other materials in addition to Top Notch. Besides going through sections in the book, she provided the students with listening comprehension exercises and video clips related to the topic discussed in Top Notch units from other pedagogical materials, such as ‘Tactics for Listening’ or ‘New Interchange’. As with Top Notch, these materials were pedagogic materials developed by international publishing houses for ESL/EFL teaching.

During the time I conducted observation in her lessons, she had just finished Unit 8 and started Unit 9 in Top Notch Level 1. The two units had related topics, with one about travelling and the other about taking transportation. The following episode shows her use of a video clip about travelling after covering some sections in Top Notch.

Extract 4.2	
Amy (Amy, EIST II classroom observation, 14 May 2008)	
Amy:	Now, let's show you, a clip. The husband thinks that he's very tired and feels the need to take a vacation. Let's look at the video clip. Where do they travel to? How do they plan? Let's try. It will have words we use here ('here' means the unit in Top Notch) ...(omitted) You'll see afterwards this couple is going travelling. Which country do they plan to go? What dates do they plan to depart and return? Have they booked a place? Let's try, if we can hear answers to the three questions. Ok?

In this episode, Amy played a video clip on the topic of travelling, provided by 'New Interchange', another international pedagogic ESL/EFL material. Before she played the video clip, she had already covered a conversation model in Top Notch which did mention dates of travelling, and she had also carried out several exercises for her students to review the English use of 'dates'. This was why she said that the students would hear words in the video clip that they were using in the previous activities, and it explains how she integrated other materials with the assigned material.

The above analysis shows that the majority of the teachers in University A used the assigned teaching material extensively. However, since there was no link between the assigned material and any English proficiency tests, no evidence of washback on teaching materials was found in University A, which does not have English language proficiency as a prerequisite for graduation. Top Notch was an internationally recognised textbook series developed by two professional ESL/EFL teachers and textbook writers from the United States. It did not consider the local Taiwanese contexts in its content selection, which was claimed by the GEPT developer as an important trait of the test as a locally developed English test. Although the purpose of the material was to promote communicative competence, and GEPT claimed to be

developed from the idea of promoting communicative competence, it was difficult to establish the link between the materials and the test. The reason was that among the various English proficiency tests which were designed to evaluate test takers' communicative competence, there was no specific evidence to say that Top Notch was a material that could be used for GEPT preparation and not for other tests. Since most teachers in University A made extensive use of the book, there was no washback of the GEPT on their teaching materials. Even for Amy, who incorporated materials from other pedagogic materials, it was very difficult to view her choice of materials as influences of GEPT, as these materials, like Top Notch, were also ESL/EFL materials that were developed, published and recognised internationally. There was no evidence of washback from other English proficiency tests either, for the same reasons.

It was a very different case in University B, in which the graduation requirement had been implemented. The use of teaching materials showed variability among teachers, and there was indeed evidence of GEPT washback on the materials used in some teachers' classes. Although Ben and Betty were teaching the same course, English II, the materials they used were different. They told me in the interviews that there was assigned material for the course, but there was no sign of the use of the material in the lessons I observed. Betty used an article she downloaded from a resource website of a project she did with fellow teachers in the other university she worked at. The resource website, developed by her colleague, contained a pool of reading articles written in authentic English, which was selected for guided reading. In her lesson, the article chosen, 'Journeys of the Heart', was originally from a counselling centre in the United States. Although this article was originally a non-pedagogic, authentic material, it was adapted for pedagogic use by being segmented into 8 paragraphs and

by putting numbers in places that needed to be noted. The other article she adopted for the lesson, 'Indiana Jones and Kingdom of the Crystal Skull', and the supplementary trailer of the movie described in the article was provided by CNN Interactive English. This is a monthly-issued English learning magazine developed locally for self-study, with articles adapted from CNN news reports. The articles provided the chance to learn English related to up to date news, covering different areas in life, but they were not written to follow a specific English syllabus or a plan for improving English language proficiency. Thus, unlike Top Notch, it was not tailored for a curriculum or institutional use. In the interview, she told me that she had already gone through the units in the assigned material before I arrived and she preferred to use articles that could stimulate students' interest.

There was no evidence of washback on teaching material in Betty's class. The articles Betty selected for her lesson were either authentic (semi-pedagogical reading article from counselling centre) or adapted from authentically written CNN News reports. Although many tests that aimed to assess communicative competence, including the GEPT, encouraged authenticity, it was not evidenced that the articles could be linked to any test. Furthermore, her decisions were made from her assumptions of what might interest her students, rather than how the articles were related to the GEPT or the effort to prepare her students for the test.

On the other hand, the analysis of the teaching material used in Ben's class revealed indirect influences of the GEPT. Monthly-issued English learning magazines from local publishers were the main source of Ben's materials, as he no longer used the assigned material either. He requested that his students purchase the magazines he recommended, Studio Classroom, intermediate and higher-intermediate and English

Digest. Both magazines came with an audio-visual CD, which they called MP3 interactive CD and Studio Classroom even provided free daily radio programs which discussed daily articles and sections in the magazines. Aimed to promote all four skills of English, the contents of the magazines were long articles, long conversations, cloze exercise, song lyrics, guided writing, etc. As with CNN Interactive English, the magazines were intended more for personal use than for institutional use. Among the different contents in the magazines, Ben only focused on long reading articles or shorter passages of news reports for his lessons. In the lessons I observed, there was no use of any other materials except for the magazines.

What was interesting was the explicit reference to the GEPT in both magazines Ben used despite his denial of having the intention to help his students meet the graduation requirement. Both magazines were first published in the 1980s, before the development of the GEPT. I recalled that in my high school years, the English teacher either used the magazines as supplementary materials or suggested we should read them at home. At that time, the test we were required to take was the Joint College Entrance Examination, rather than the GEPT. Thus, it was less than appropriate to say that the magazines were specifically used for GEPT preparation. However, one particular feature of a magazine was its nature to evolve through time. In recent years, the GEPT has become the most popular English proficiency test among Taiwanese citizens (see Roever and Pan, 2008) and one goal of the GEPT is to promote life-long learning in English, which coincides with one of the aims of the magazines. The majority of those commercial magazines are not targeted at a certain age group, but at all citizens in Taiwan. Therefore, in order to promote sales to the larger population, the magazines aim to cater for their needs in the context of English language learning in Taiwan. The phenomenon by which almost all the English self-study magazines

developed locally now have included certain aspects of the GEPT in their contents reflected the commercial value, and also the status of the test as viewed in society. This could be considered as the social impact of the GEPT.

The magazines he selected for his class was at an intermediate and higher-intermediate level, and there were mock GEPT Reading items for the two levels in the Studio Classroom, as shown in Figure 4.3.

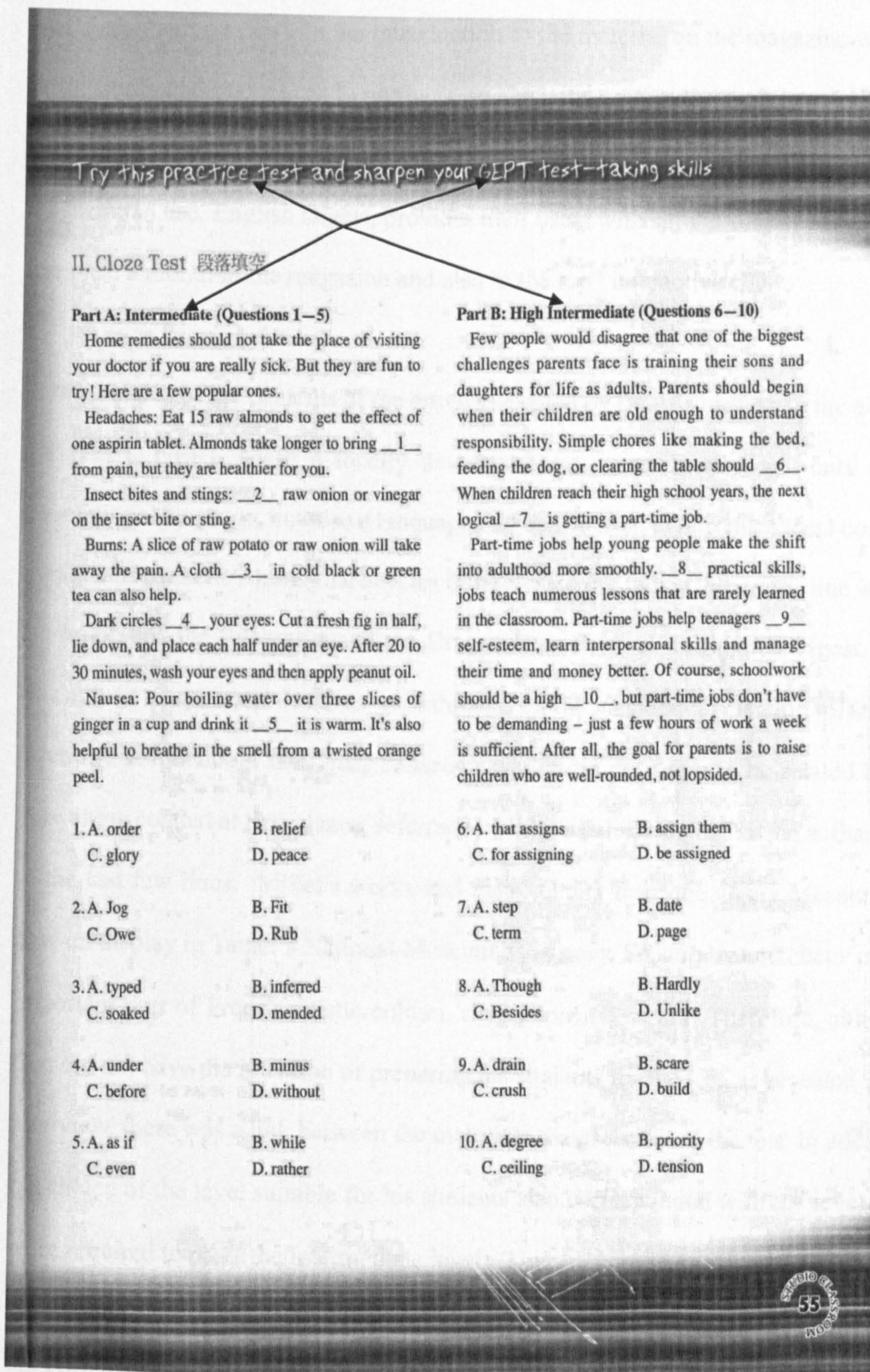


Figure 4.8 Monthly-issued English learning magazine: reference to GEPT preparation

As observed from Figure 4.3, there was an explicit reference to the GEPT with a statement at the top of this page, saying that ‘try this practice test and sharpen your

GEPT test-taking skills'. In the introduction to the material on the magazine website (http://studioclassroom.com/h_products_sc.php), it says that the mp3 interactive CD that comes with the magazine also includes practice tests for the GEPT. The other magazine in use, English Digest, provides their users with similar offers, with a mock test once a month in the magazine and also in the mp3 interactive CD.

Some features of the contents of the articles selected by Ben also reflected the aims of the GEPT. GEPT, being a locally developed test, treats English not only as an international language, but also a language that can be used to introduce local contexts to English speakers. Among the articles Ben selected in Studio Classroom, one was on the origin and the celebrations of the Dragon Boat Festival. In English Digest, there was a short news report that Ben went through on the presidential election in Taiwan. Even the article about the Orsay Museum, which did not seem to be related to the Taiwanese context at first glance, referred to Millet's exhibition in Taiwan at that time in the last few lines: 'Millet's works and many others from the Orsay Museum are now on display in Taipei's National Museum of History. So stop by and check out an important part of French artistic culture, right here in Taiwan!' Therefore, although Ben did not have the intention of preparing his students for the test, as revealed in his interview, there was a link between the materials themselves and the test. In addition, his choice of the level suitable for his students also corresponded with the level they were required to pass (the intermediate level). Thus, to some extent, there was indeed GEPT washback on the teaching materials in his lessons. The interesting part was that the GEPT washback in Ben's lessons was in fact a result of the social impact of the test on those English learning magazines (see 6.2.1 for further discussions).

It was anticipated that there would be some extent of test influence in Becca's lesson,

as the course she taught (namely English Reading and Writing) was part of the curriculum innovation, and was a direct response to the English language proficiency requirement for graduation. However, what was unexpected was the role the English department played in moulding the course which had a 'generic' name into a preparation course for a specific test. Among all the EAP courses for non-English majors in University B, it was the most direct link to the graduation requirement as the course was designated for students who had not fulfilled the requirement before they reached their third year. When Becca was asked about the use of teaching materials for her course, she reported that she was informed by the Department of Children English Education, which was in charge of the General English curriculum, to choose a material specified for GEPT preparation instead of materials for other English proficiency tests or materials that develop general English proficiency:

Extract 4.3

"They (the department of Children English Education) told me to use GEPT textbooks so I chose 'Get the Point' from Dun Huang (a Taiwanese publisher), 'Get the Point', reading and writing, intermediate level GEPT." (Becca, interview, 03.07.2008)

'Get the Point', from the publication information, is a series of materials developed especially to prepare for the GEPT intermediate level. Becca's English Reading and Writing class had a specific focus on writing. The introduction of the material to the teachers claims that it uses topic-related reading passages to guide the writing process and supplement the main content with grammar structures that are tested in the GEPT. In the beginning pages of the book, the level descriptions of the intermediate GEPT and the test construct are also included. Moreover, some of the pages provide information on several test taking techniques for items in the writing section, translation and English composition.

Despite having a material directly linked to the GEPT for her class, in the only lesson

I observed, other materials took up most of her lesson while the GEPT material was limited to a reference in one sentence:

Extract 4.4	
Becca (Becca, English Reading and Writing classroom observation, 27 May 2008)	
Becca:	<p>Now, let's work on Unit 6 on page 46, The book 'Get the Point'. We're going to discuss about eating well, eating right. So as you can see, the reading is the guessing game Is that right? So the very first hour I want you to guess, the words from the text. 用前後文來了解字的意義 文章的意義 今天是歌詞 (Trans.: Using the surrounding texts to understand the meaning of the word, the meaning of the article. Today, they are lyrics.) Now, when we talk about healthy life, you know, the first thing you will.. What is the first you will think of? When we talk about healthy life, About diet, food, exercise. What else? (T distributes a handout.) What is this? A triangle, isn't it? A triangle? Yes? Have you been to Egypt? Have you been to Egypt? The Middle East? Egypt? 埃及. (Trans.: Egypt) Have you been there?</p>
Ss:	(inaudible) .
Becca:	<p>No? I've been there. There you can see many, many what? Shaped like a triangle?</p>

	It's called the?
	What's this?
	Pyramid

In the extract, she referred to a teaching material for the class, which she revealed to be the GEPT preparation material in the interview (see below). She pointed out that the unit they were going to cover afterwards was about healthy eating. She then distributed a handout of the food pyramid and started teaching words of the food included in the pyramid till the end of the lesson. When she was asked why the appointed material for GEPT preparation was not used, she said that her limited use of that material was due to several reasons:

Extract 4.5

“That material, ‘Get the Point’,...the cover says...it is for the intermediate GEPT. So sometimes, if a material has not been used for several times, it is not easy to...judge whether it is able to cater to their (students) needs by its cover and the GEPT written on the cover. It is in fact not easy, not easy. So I think, I would rather find materials students will be interested in than focusing on reading and writing. I realise that the book has a lot of stuffs on writing, like topic sentence, supporting sentences, concluding sentence, these writing techniques. However, I found that its articles are not written according to those techniques. So it made me feel unconfident to teach (the content of the book). Its topic sentences are badly written. So I think, this made me feel that it is not easy to choose a material that is appropriate for the students.”

(Becca, interview, 03.07.2008)

That was why instead of the GEPT material, she first used a semi-pedagogical material (see 4.2 or Frolich, Spada and Allen, 1985), by adapting authentic song lyrics to fill-in-the-blanks listening and reading exercise and later, she turned to the food pyramid, an authentic, non-pedagogical material. Yet, she still related her own materials to the topic of the book on healthy life, which she justified:

Extract 4.6

Becca: Um, why did I use the ‘pyramid’? Because that lesson was about ‘how do you...uh...keep, how do you keep healthy?’, a ‘topic’ something like that. So I think I should provide (students with) the concept of a ‘pyramid’

Researcher: Is it chosen according to the topic in the material?

Becca: Yes, the material has that topic. It was just that I haven't covered that yet.
.....(omitted)
That pyramid was mainly saying that, for the next lesson, I would cover the text content. (text referring to 'Get the Point').

(Becca, interview, 03.07.2008)

GEPT washback on teaching material was imposed on Becca's class by the purpose of the course and by the request of a GEPT preparation material from the English Department. Although in her only lesson I observed, there was not much reference to 'Get the Point' because of Becca's personal dissatisfaction with the book, it was still her central teaching material and her non-textbook materials were all somewhat related to its unit topic. This was evidenced in the interview, which she said the food pyramid was to introduce the students to the content in the assigned material. In a private talk with her after the lesson ended, when she first revealed to me her dissatisfaction with the material, she told me that if she had known the quality of the material beforehand, she would have used GEPT materials she had edited (3.7.2). Thus, it was quite apparent that the only lesson I observed in which GEPT washback on teaching materials were evidenced to the minimal extent, could not be counted as a typical lesson in that course and perhaps more observation sections would reveal more GEPT washback on teaching material in Becca's class (See 7.3.2 for limitations of this study). Nevertheless, it was the decision by the English department to make this course aligned with GEPT preparation that shaped Becca's choice of the main teaching material and the direction of the course at the first place.

It might be seen from the above analysis on materials that under the influence of the graduation requirement, there was indeed GEPT washback on the teaching materials. Materials directly linked to the test were found in Becca's assigned material, 'Get the

Point', a GEPT preparation material for writing. Despite the little reference to the material in the lesson observed, it was the main material for the course of 'English Reading and Writing'. The two magazines Ben required for his lessons could also be linked to GEPT in that there was GEPT practice test in either the magazines, or in the CD that came with the magazine. In addition, the contents concerning local Taiwanese contexts also reflected the aim of the GEPT. The exception was Betty, whose materials could not be linked to the GEPT. Another important finding emerged from the analysis was that the varied choice of teaching materials among the teachers in University B echoed Watanabe's argument (2004) that teacher factors could mediate washback or could be the reason why washback was not manifested. For the same course, Ben and Betty decided for themselves what materials to use in their classes and while Ben's choice reflected GEPT washback, Betty's choice did not. Becca's personal dissatisfaction with the GEPT preparation material and her favour over the use of materials from multiple sources also decreased the extent of washback in the lesson I observed despite strong washback was anticipated from GEPT preparation class.

On the contrary, there was no evidence of GEPT washback on the materials all the teachers in University A used. Most of them used exclusively Top Notch, the internationally developed ESL/EFL material and the exception, Amy, despite her integration of audio and video materials from other ESL/EFL materials, there was no sign of GEPT washback. Teaching materials, as demonstrated in previous washback studies (Cheng 2005; Qi 2007; Wall 2005), are considered to be one aspect that is likely to reflect higher degree of washback intensity. The sharp contrast between the evidence of washback on teaching materials in University A and B thus supports previous studies which show that the higher the stakes, the stronger the extent of the

washback.

4.3.2 Teaching content

The teaching ‘contents’ of the observed lessons, what teachers taught and which linguistic skills were targeted are presented in Series keyword percentage graph from Transana and described in details in the following section. Whether the teaching contents reflected influences of the GEPT or other English proficiency tests are then discussed.

The percentage of which language focus and language skill(s) were found in the observed lessons in University A is shown in Figure 4.9. The figure shows the percentage of different areas of language development in each observed lesson, regardless of the total length of the lesson.

.

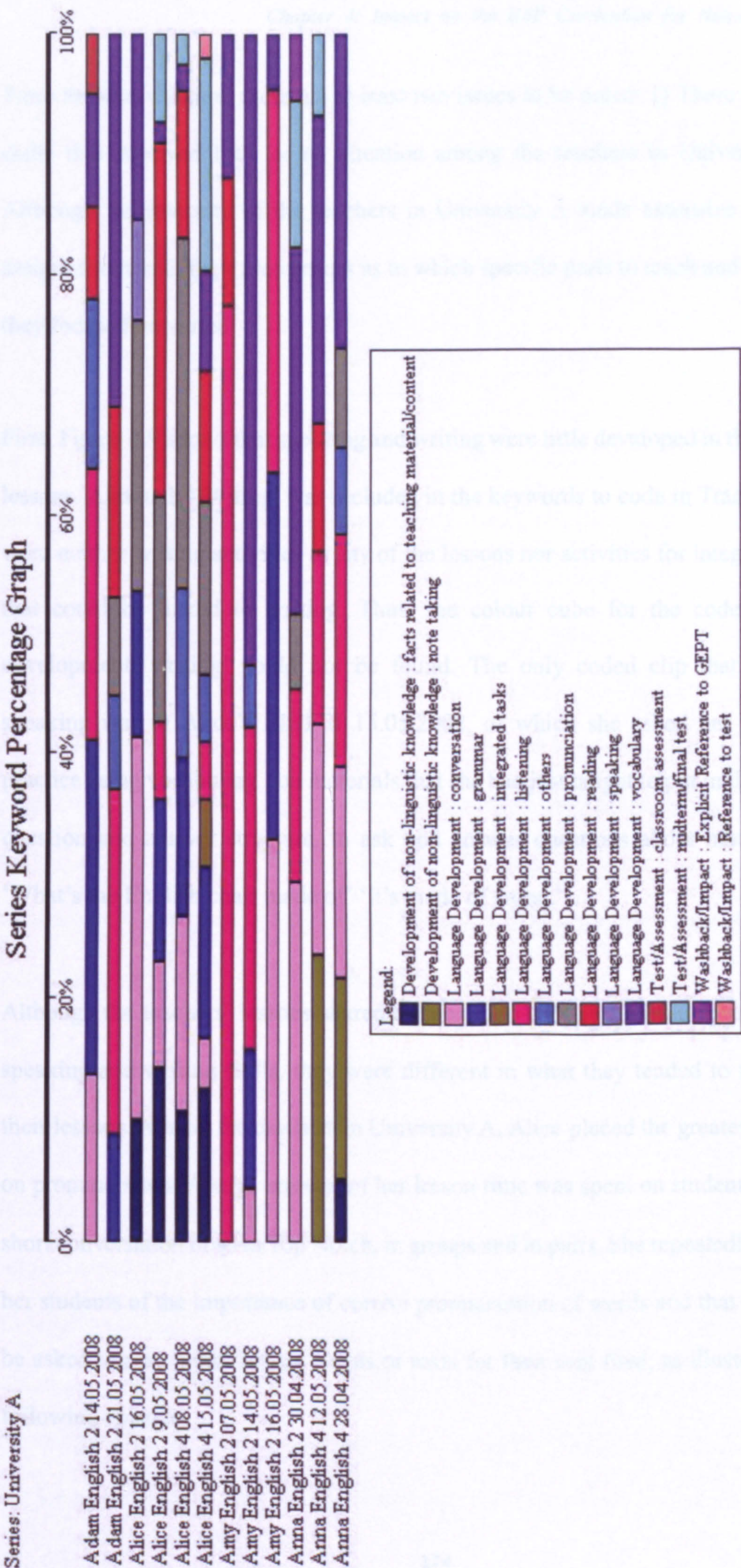


Figure 4.9 University A: Transana series keyword percentage graph

From the above figure, there are at least two issues to be noted: 1) There were some skills that received little or no attention among the teachers in University A. 2) Although the majority of the teachers in University A made extensive use of the assigned material, their preferences as to which specific parts to teach and what skills they focused on varied.

First, Figure 4.9 shows that speaking and writing were little developed in the observed lessons. Although 'Writing' was included in the keywords to code in Transana, there were neither writing activities in any of the lessons nor activities for integrated skills that could be linked to writing. Thus, the colour cube for the code 'language development: writing' could not be found. The only coded clip that concerned speaking was in Alice's EIST II 15.05.2008, of which she asked her students to practice using vocabulary on materials that she had just taught to put in the provided question and answer structure, to ask and answer questions about materials (e.g: 'What's the English chair made of' 'It's made of wood.').

Although the teachers' lessons shared the common trait of the little development in speaking and writing skills, they were different in what they tended to focus on in their lessons. Among the teachers in University A, Alice placed the greatest emphasis on pronunciation. A large amount of her lesson time was spent on students repeating short conversation lines in Top Notch, in groups and in pairs. She repeatedly reminded her students of the importance of correct pronunciation of words and that they would be asked to read aloud certain words or texts for their oral final, as illustrated in the following episode.

Extract 4.7	
Alice (Alice, English Integrated Skill Training 4 observation, 15 May 2008)	
Alice:	<p>Next, students, please look at p. 92.</p> <p>Attention, students.</p> <p>The oral test.</p> <p>Here, you need to read aloud.</p> <p>Teacher will stress about this again.</p> <p>One word wrongly pronounced will result in 2 points less.</p> <p>The full mark is 50, oral test 50, alright?</p> <p>So students need to take care of this part.</p>

Adam followed the sequence of the assigned material with limited variation in his two lessons I observed. However, he showed his emphasis on vocabulary by a handout he made by himself by putting together all the words and phrases in the units he taught to help students review for the final exam. An interesting difference was also found between his first and second observed lesson. Instead of playing the audio clips for short conversations and listening comprehension section beforehand, as was the case in his first lesson and in most teachers' lessons, he worked through all the conversation lines and listening comprehension transcripts as texts before all the related audio clips were played, towards the end of the second lesson.

In Anna's lessons, only sections of short conversation, listening comprehension and vocabulary were covered despite Top Notch was the only material she used. The sections of grammar structure and exercise and the long reading articles were omitted. She revealed in the interview that it was from a practical point of view that she focused on listening, short conversations and vocabulary and she believed that those materials were sufficient for the non-English majors with limited English proficiency she taught. One special feature of her lessons was the recurrent review of words with synonyms, antonyms and related usage of words, as illustrated in the following

episode.

Extract 4.8	
Anna (Anna, English Integrated Skill Training 4 observation, 28 April 2008)	
Anna	<p>Next, it says 'blue', ('blue' is used in short conversation in Top Notch) . 'Blue', we mean sad and depressed. What are some synonyms? Other than down, blue, What other words are related to sad and depressed?</p> <p>S?(m): 'sad'</p> <p>Anna: No need to take your dictionary out. You know how to say (the words). 'sad'. What else? Not happy, English for not happy</p> <p>S?(m): 'unhappy'</p> <p>Anna: 'unhappy', ok, next? What else?</p>
S?(m)	'sad'
Anna	No need to take your dictionary out. You know how to say (the words). 'sad'. What else? Not happy, English for not happy
S?(m)	'unhappy'
Anna	<p>'unhappy', ok, next? What else? Not in a good mood means your mood? Unstable. So, will use this word, word for mood, 'mood'. So, 'moody', is the same. You can use it here.</p>

Her students were required to write down the synonyms, antonyms, related phrases and usage of words as notes, which would be checked and the contents be tested in the final oral examination.

Among the four teachers in University A, as illustrated in the previous section, Amy was the only teacher who adopted materials other than the assigned material and who designed her lessons based on a specific topic rather than being textbook-driven. When she was asked the rationale of using the additional materials, she reported that she did not consider those materials as supplementary to the assigned material but just a way to reinforce the ideas and to provide more opportunities for practice:

Extract 4.9

“I just want to give more chance for practice. Just to reinforce, not as supplementary...(omitted)...I'd rather pick this part out of this unit than browsing through all the sections in the unit. You will have deep impression on this part because everything, listening or speaking, are related to this. This can be considered as topic-wise.”

(Amy, interview, 22.05.2008)

Her emphasis on listening could be shown from her choice of additional audio and video materials and the way she used them. Without giving her students transcripts for those materials, she made them practise listening to keywords by asking them questions, providing them with key words beforehand, and playing the audio/video clips repeatedly.

From the analysis above, we may see there was little evidence of GEPT washback on the language focus and skills development in the observed lessons. First, there was no overt test preparation such as taking mock tests or practicing mock items. Nor were there any activities that could be related directly to the development of test taking skills or techniques. Second, it was difficult to claim that the exercises or practices would be helpful only for GEPT preparation but not for the preparation of other English tests. For example, some of the grammar exercises in Top Notch, such as cloze tasks and sentence reconstruction tasks, might be similar to a multiple-choice cloze section in GEPT reading comprehension and sentence reconstruction tasks in the GEPT writing test. However, these were commonly seen tasks in grammar exercises. Taking into consideration that Top Notch was not aligned purposely with any English proficiency test (4.3.1), I think its link to the GEPT was very weak.

The individual difference on language focus and skill development was also found among the teachers in University B, as illustrated in Figure 4.10.

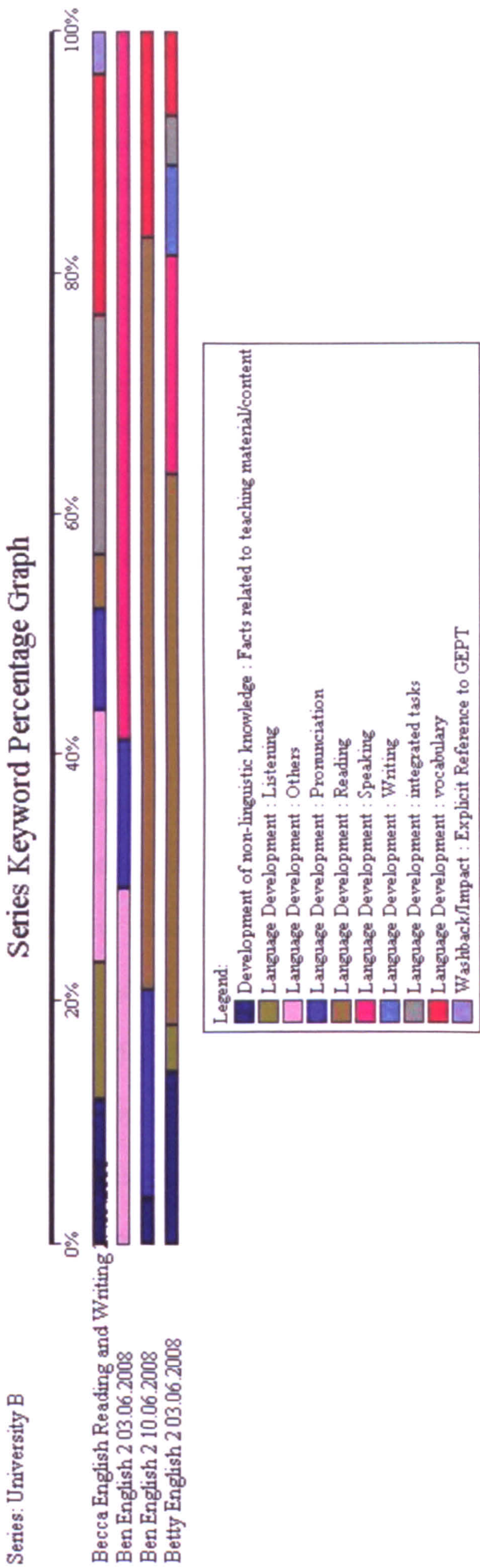


Figure 4.10 University B: Transana series keyword percentage graph

The above figure showed that the two skills that Ben and Betty focused on were reading and speaking, as confirmed in their interviews:

Extract 4.10

'The most basic thing, the most important is reading. Reading specifies speed and comprehension. They (students) are too passive. Besides reading, after you understand the ideas (of the passage), you need to be able to discuss about the ideas. The next (stage) is oral expression). You bring up your views on what you read and understand and have interactions with others.' (Ben, interview, 10.06.2008)

Extract 4.11

'Basically, we (teachers) try hard to teach students the skills to be able to speak and read.'
(Betty, interview, 17.06.2008)

However, even though their choice of materials was all reading articles, their lessons differed to some extent. Ben's first lesson I observed was spent mostly on read-aloud and student impromptu speeches. It was difficult to see how reading and vocabulary were taught, as Ben did not actually go through the articles. Instead, he read the articles aloud himself, or he asked students to read and then, he picked students to come up to the front to say something about the topic. As long as the content was related to the topic, the students' impromptu speeches did not necessarily have to be about the contents of the articles, which meant comprehension of the reading passages was not essential for the speaking afterwards. Betty, however, went through her articles thoroughly, by displaying on the screen her powerpoint file with definitions for the words in English and by explaining the meaning of each phrase and sentence in both English and Chinese. Ben's second lesson was similar to Betty's lesson, in which he provided explanations of words, phrases and sentence structures as he read through the articles. When a student from his class was asked about her opinion of the lessons, she revealed that the second lesson instead of the first lesson was more typical of what he usually did in his lessons. In this sense, the observer effect might have taken place in

his first lesson I observed. In addition to the impromptu speeches, speaking was realised through students' prepared presentation in the second half of Ben's lessons; while in Betty's lesson, it was more of teacher-student discussions on certain ideas in the articles and experience sharing.

Despite teaching the course which is specifically aimed to develop reading and writing skills, Becca spent less time on reading than Ben and Betty. In addition, Becca did not strictly stick to the purpose of the class, which was preparation for the GEPT. In that particular lesson I observed, there was no explicit sign of test preparation. Instead of activities that developed reading and writing skills, she incorporated listening practice and devoted much time on vocabulary. She used a song and lyrics instead of a reading passage to train students to listen to words in the blanks and to get to the answers with the surrounding text. For the second half of the lesson, to keep the contents related to the topic of the textbook unit, she talked about the food pyramid and provided the vocabulary for every food shown in the pyramid. The reasons why her lesson was not restricted to reading and writing development in relation to the GEPT was because she wanted to increase students' motivation to learn:

Extract 4.12

"I think if we focus on reading and writing every time and we do some exercises after reading, it is indeed quite boring. So I provide them with these, something that can motivate them to learn."

(Becca, interview, 03.07.2008)

The analysis of teaching contents in the observed lessons in University B did not reveal much evidence of the GEPT washback either, as with the case in University A. The regulations in the degree requirement for University B stated that non-English majors only need to pass the first stage of the intermediate level GEPT, which consisted of only the listening and reading tests. However, all of the lessons by the three teachers did not

show a strong focus on developing listening and reading skills students need to meet the requirement. One might ask if the focus on reading skills in some teachers' lessons were the product of GEPT washback. However, without the evidence of an increase of the focus on reading as the GEPT test date came near, it was difficult to establish the relation between the emphasis on reading skills in the lessons and the GEPT. In addition, both Betty and Ben did not agree that they were influenced by the graduation requirement on their teaching as they believed reading and listening skills were the basic skills for university non-English majors. This could be explicated by the following extract. Betty pointed out the other skill she focused on was not tested in the first stage of intermediate GEPT (her alignment of the test to the graduation requirement):

Extract 4.13

'But the graduation requirement tests the students on their reading and listening and it has nothing to do with speaking.' (Betty, interview, 17.06.2008)

On the other hand, even though reading and writing skills were assigned to be the major skills to be developed in Becca's class, she made her own decision not to restrict her course to mainly test preparation. She tried to incorporate activities that promoted other skills as well. In the lesson I observed, there was indeed very little evidence of GEPT washback on the teaching contents. Although it was not possible in this study to compare Becca with another teacher teaching the same course, it was with little doubt that the data suggested that teacher factor resulted in the limited GEPT washback on teaching contents. Pan (2009) had some concerns that GEPT's two-stage tests on receptive skills (reading and listening), which is selected by most universities as graduation requirement, then productive skills (speaking and writing) could bring some negative effects. Such concerns were not evidenced in this study. Becca (2.2.4) may be a typical "adapter" as what Burrows (2004) referred. She incorporated

activities not directed to test preparation (i.e. song listening) and activities that developed skills other than the targeted skills of her course. Yet, her course did not deviate from GEPT test preparation. The findings above have demonstrated the importance of teacher factors in the presence or absence of washback, as Watanabe (2004) has strongly argued.

The above analyses have shown that there was little evidence of test influences on the teaching contents. Teacher's individual differences contributed to the different focus on different language skills in different classes for the same course. Curriculum narrowing, that might be caused by test design or test preparation, were not found in the lessons. The analyses of the classroom activities also suggested that it was not evident that teachers were teaching to any tests (2.3.1.3). This finding confirmed my speculation and justified my decision to consider 'washback on teaching methods' peripheral for this study.

4.3.3 Explicit reference to the test

Despite the fact that the graduation requirement for English language proficiency was implemented in University B and all the teachers mentioned GEPT in the interviews when asked about the requirement, there was no explicit reference to the test, the requirement or any other tests stated in the requirement in any of the observed lessons. However, the analysis of the full verbatim transcripts of the lessons in University A unexpectedly revealed two instances in which two teachers explicitly referred to English proficiency tests. In the following extracts, I present how these teachers related the linguistic aspects they were covering in the lessons to GEPT or to the pursuit of a certificate from any English proficiency test.

While explaining the phrase ‘associate with’ on the introduction page of Unit 7 in Top Notch, Anna specifically indicated that the knowledge of this phrase could be helpful for someone taking the GEPT.

Extract 4.14	
Anna (Anna, English Integrated Skill Training 4 observation, 28 April 2008)	
Anna	Next, the third one, ‘Which color do you associate with happiness?’ Please underline the phrase ‘associate with’. This is a phrase that is often used. If you can use this phrase when you write your cover letter or when someone takes the GEPT.

Her direct reference to the test echoed with what she revealed in the interview, stating that the GEPT was the test her students most likely would take, with or without the need to fulfil the graduation requirement.

Extract 4.15

‘I believe it should be the GEPT because most of the students know the test. Without the graduation requirement, it is still the test most students will take.’ (Anna, interview, 07.05.2008)

Unlike Anna’s overt reference to the GEPT, Alice stressed that it was important for her students to be able to use passive voice. She said that it was a must for her students to know this grammar structure in order to get a certificate from any English proficiency tests.

Extract 4.16	
Alice (Alice, English Integrated Skill Training 4 observation, 15 May 2008)	
Alice	The passive voice is a grammar structure that will be included in tests of which you can get a certificate for in the future. This grammar structure will definitely appear in English tests.

The reason why she did not specifically indicate the GEPT or other tests was because she believed non-English majors from different departments should take the English

proficiency test that they need most (See 4.6). Nevertheless, she emphasised the need for students to obtain a certified proof of English proficiency.

Since there was no requirement of English language proficiency for graduation in University A, the link the teachers made from what they were teaching to the GEPT or to any other English certificate could not be claimed as the evidence of the impact of the requirement, even though both teachers admitted in the interviews that their teaching would have been influenced if there were the requirement. However, the importance of a certificate for English proficiency in the society, institution-wise or career-wise, could be a reason behind Alice's emphasis on the importance of understanding passive voice. Anna's direct reference to the GEPT might also be a manifestation of the social impact of the GEPT which is arguably the most popular English language test taken by Taiwanese citizens.

4.3.4 Testing and assessment

The classroom observation data indicated that the washback effect of GEPT was mainly found on teaching materials (4.3). The teachers' interviews revealed another aspect of GEPT washback that could not be captured during my short-term stay in each university: its influence on testing and assessment. GEPT washback on testing and assessment was evidenced in both universities. In University A, teachers used test papers developed by the Taiwanese publisher as mid-term and final tests. The tests, which were modelled on GEPT elementary level, revealed strong influence of the test on the Taiwanese publisher which represented the international teaching materials. In University B, Becca made use of a GEPT mock test and a real GEPT test held by the test developer in her remedial course of English reading and writing for the GEPT.

As previously mentioned, the discovery of washback on this aspect in University A occurred during the initial inductive analysis (4.2). In the post-observation interview when Amy was asked whether she preferred oral test or written test for her students, she reported that the tests in fact used the format of the GEPT:

Extract 4.17

Researcher: So you will decide to include both tests, or just one according to the condition of the class?

Amy: But now, we are using GEPT.

Researcher: Using GEPT now?

Amy: Yes, the first years and the second years are now using the GEPT mode for their tests.

Researcher: Are the test contents in Top Notch?

Amy: Yes.

Researcher: Then, how is it related to the GEPT?

Amy: It's using the GEPT format, using its item types.

Researcher: Oh, really?

Amy: There is writing, listening. That is why the test takes one and an half hour (to complete). (It's the) Same for mid-term and final.

Researcher: Oh, I get it.

Amy: Item contents are those in the textbook. Item types are the GEPT item types. A mini version.

Researcher: To familiarise them with the GEPT item types?

Amy: Yes.

(Amy, interview, 23.05.2008)

She further explained that in University A, non-English majors attending English Integrated Skills Training (EIST) courses were required to take the same test for their mid-terms and finals. However, it was not the teachers who designed such tests. They were in fact developed by the Taiwanese publisher which was in charge of the local sales and promotion of the international textbook series. This piece of information was validated by Alice, who pointed out that all students taking the same test was why she needed to cover every part of the teaching material (Extract 4.1).

Once this was known, I asked Amy if I could get a copy of the test papers for the final exam. As the final was going to be held in three weeks at that time, my request was turned down. I received two copies of the tests which were used in the semester of which the study was conducted a few months after I left the university (See Appendix H for the test papers).

The analysis of the test papers, indeed, showed that the test was modelled on the test format of the GEPT elementary level (see Table 4.3 below). It was difficult to determine whether the GEPT was modelled on other available English proficiency tests which have been developed for a longer time and there was no formal discussion on such issue. Therefore, the analysis of the test papers directly reflected influences of the GEPT but not other tests.

	University A test papers		GEPT elementary level	
	Item type	Number of items	Item type	Number of items
Listening	Picture description	4	Picture description	10
	Question and statement response	10	Question and statement response	10
	Short conversation	4	Short conversation	10
Reading	Vocabulary and structure	10	Vocabulary and structure	15
	Cloze	10	Cloze	10
	Reading comprehension	5	reading comprehension	10
Writing	Sentence Writing:		Sentence Writing:	
	a) Sentence rewriting	5	a) Sentence Rewriting	5
	b) Sentence combining	5	b) Sentence Combining	5
	c) Sentence reconstruction	1	c) Sentence Reconstruction	5
	Paragraph writing	1	Paragraph writing	1
Speaking	N/A	N/A	Repeating	5
			Reading aloud	6
			Answering questions	7

Table 4.3 Comparison between University A test papers and GEPT elementary level

Table 4.3 clearly shows that the midterm and the final tests in University A were, indeed, fully modelled on the GEPT test, except for only the Speaking part. Detailed analysis of the test paper also revealed that even the instructions of the test followed exactly GEPT instructions. (See 6.2.1, Table 6.1).

The design of the test papers modelled on GEPT item types might suggest that GEPT was probably perceived by the Taiwanese publisher as more important than any other English proficiency tests available in the society (See 6.2.1 for discussions). Under the strong influence of the test, the local publisher which represented the international teaching materials thus incorporated what was important in the local context for the use of the international materials in Taiwan. The GEPT impact on the publisher in turn, had

an influence on the assessment system in the EAP curriculum for non-English majors in a university without the graduation requirement. Without interviewing the Taiwanese publisher, it was difficult to say if the publisher's intentions lay behind such practice. In addition to the perceived importance of the test, it could only be speculated that the test publisher's action probably reflected the commercial value and interests of the GEPT. With the popularity of the test and the implementation of the English graduation requirement in universities in Taiwan, the test could be used as a means of promoting the sales of their textbooks. Perhaps, if the head of the English department and the teachers in charge of material selection chose to use Top Notch not only because of the quality of the material but also for the sake of GEPT, then I would argue that this might be strong evidence of GEPT washback on testing in a university setting.

In University B, where the final exam was designed by the teachers themselves, washback of GEPT, if any, on mid- and final tests may not be so much influenced directly by other stakeholders as by the teachers themselves. Although Becca's teaching in the lesson I observed did not manifest overt washback, nor was the teaching content much related to the GEPT textbook, she reported explicitly her incorporation of the GEPT test into her course syllabus. A GEPT mock test was used as a pre-test in her first lesson. In her last lesson before the final exam, the Language Training and Testing Centre (LTTC) came and delivered a real GEPT exam, for LTTC's research purposes and also for the students to practice a GEPT test conducted in a semi-authentic context (i.e., authentic in how they would take the test in reality but they took the test in their own classroom and their performances were not officially scored and recorded):

Extract 4.18

"Because last semester was one teacher (another teacher teaching the course) and this semester I was new to the course. So I didn't have a basic understanding on the students and I was not clear about how to link the courses (in the previous and the latter semesters) together. But, in my class, I

had them do a pre-test, a GEPT mock test. I wanted to know what lessons would be most appropriate for students.”

.....(omitted)

“At the end, I did a post-test. The post-test was delivered by the LTTC. But I haven’t received the results yet. I did a real post-test for that purpose.”

(Becca, interview, 03.07.2008)

The purpose mentioned in the last line of the extract above referred to passing the intermediate level GEPT for the graduation requirement in University B. Despite the fact that her final exam was designed according to what she taught in class and her course English reading and writing was not officially named as a GEPT preparation course, she still blended test practice in her lessons to achieve the purpose of a remedial course for students who did not reach the requirement before their third year.

Becca’s case echoed the findings of Wall and Horak (2006)’s TOEFL impact study in that test papers similar to the target test were used for the purposes of screening, diagnosis and practice in the preparation course. The pre-test in Becca’s lesson functioned as a diagnostic test to give her an idea of her students’ overall English proficiency. The test at the end of the term might act as a practice test for both Becca and her students to see how the students had progressed. Wall and Horak (*ibid.*, p.112) questioned whether this type of tests from preparation course books mirror the real TOEFL tests. They also questioned the reliability of test results since the tests were often administered under conditions unlike the real TOEFL test. Similar questions could also be asked about the first mock test Becca used at the beginning of the term. However, the test at the end of the term was indeed administered by the official GEPT developer, with GEPT test researchers as invigilators and a real GEPT test developed by the organisation. However, the test results were simply for the students’ own assessment and for research purpose, the scores were not used as the students’ course grade, nor were they used to determine whether they could graduate or not. Thus, I still

agree with Wall and Torak's (*ibid.*) argument that the GEPT test Becca used in her course "can only give the students some indication of their level of preparedness" for the GEPT test 'rather than a true picture of their language proficiency' (p.112).

4.3.5 Summary

The impact of the graduation requirement for English proficiency on the EAP curriculum for non-English majors can be summarised from three points of view, as follows:

First, the analyses of the observation data along with the field notes and interview data have shown that among the number of English proficiency tests accepted by the graduation requirement, the GEPT has exerted the most influence, and there was little evidence of influence from other tests. In addition, the extent of GEPT washback manifested was related to the implementation of the graduation requirement. In University B, where the requirement was implemented, GEPT washback was more evident, with the remedial course served as the GEPT preparation course, leading to the GEPT washback on teaching and testing materials. On the other hand, such explicit means of test preparation were not found in University A. Although there were GEPT influences on the item types of the testing materials in teacher assessment, they were not mediated by the teachers themselves, but by local publishers representing and promoting international teaching materials.

Second, as in previous studies, this research found that washback may be more intensive in some aspects than on others. Washback intensity tended to be the highest on teaching materials and teacher assessment, and the least on language focus or activities in the lessons. The data also showed no sign of curriculum narrowing

resulting from test design.

Lastly, it was evident that individual differences among the teacher participants might be more important than the tests themselves, or the graduation requirement, in determining the degrees of test influences as manifested in their lessons. The next section will report on the teachers' perceptions of the impact of the graduation requirement, and explore whether the differences in their perceptions may have influenced their teaching.

4.4 Teachers' perceptions of the impact of the graduation requirement

Before discussing teachers' perceptions of the impact of the graduation requirement, it was important to understand which English language test they considered to be the most influential among the tests accepted by the graduation requirement. It was noted that GEPT was mentioned most frequently as the most influential test for most non-English majors.

Although all of the teachers acknowledged the importance of GEPT for their students and also viewed the test as equivalent to the graduation requirement, only some of them admitted that their teaching would be influenced by the test or the implementation of the graduation requirement. The two teachers (Ben, and Betty) who taught English II to the first-year non-English majors in University B did not think they were influenced. Betty clearly claimed that she had not been influenced by the GEPT. Her goal was to develop non-English majors' reading and speaking skills and the implementation and the GEPT did not change her goal. Similarly, when discussing the graduation requirement, Ben commented:

Extract 4.19

Ben: I am not sure to what extent the graduation requirement is implemented. But I know that non-English majors need to pass the elementary level, no, intermediate level, first stage. The intermediate level first stage is reading and listening.

Researcher: Yes.

Ben: When I teach, I don't care much about that (the requirement), because I think many students do not take it seriously...(omitted).
For me, as an instructor, my task is to keep them, no matter which proficiency levels they are in, to be interested in English and to have ongoing exposure to English. And to take it seriously. What I mean by seriously is not only to reach the requirement, but to understand good English benefits them... (omitted)
What I just say is to look at this apart from the graduation requirement.

(Ben, interview, 10.06.2008)

The extract above shows Ben's nonchalant attitude towards the implementation of the requirement, and also explains why he showed no sign of explicit GEPT preparation in his lessons, despite using GEPT-related teaching materials. What was interesting was the reason behind his attitude, as inferred from this extract. Ben claimed that his students' perceptions of the requirement were the reason why there was no washback of teaching in his class. This was similar to the findings of some past studies (Alderson and Hamp-Lyons, 1997; Green, 2006 a; Wall and Horak, 2006) which have shown unjustified claims from the teachers. Teachers in these studies claimed test preparation in their classes was what students wanted, but those claims were not justified by the students themselves.

Becca, as the only teacher who taught the course developed especially for students because of the graduation requirement, accepted the orientation of the course and incorporated some elements of the GEPT into her lessons, including her main teaching materials and tests. However, she argued that the effect of such a course was limited, especially for developing writing skills. She taught reading and writing, a continuous

remedial course after the course of listening and speaking in the previous semester. She believed that students' proficiency should reach a certain level to be able to understand what strategies should be used in writing. She considered the students were barely ready for writing instructions after their General English language course in Year 1. Furthermore, she thought it was almost impossible to teach them writing in the first year. Although her lessons manifested most GEPT washback, she insisted on incorporating different materials and other skills like listening in her lessons, which she thought would be more motivating to her students.

Extract 4.20

Becca: The effects are limited, very limited. Why? It is because I think that for reading and writing, students need to have a good level of English proficiency. You need have that level of proficiency to think about the strategies you need for writing.

(omitted)

Becca: You know, after the first year English lessons, do they have a good level of proficiency? I do not think so.

(omitted)

Becca: You have problems on vocabulary, and also sentence patterns. Then, the way you structure and compose when making a sentence may be different from what you actually mean. Thus, I don't think this (course) works.

(Becca, interview, 01.07.2008)

Teachers in University A varied in their perceptions of the impact of the graduation requirement and the GEPT washback. Adam and Amy claimed that their teaching would not change as a result of the implementation of the requirement. However, Anna and Alice admitted that there would definitely be changes in their teaching if the university began to request non-English majors to provide evidence of English proficiency for graduation. Alice had some successful experience of preparing English majors for GEPT, and she claimed that she would tailor her lessons to the requirement of non-English majors

Extract 4.21

“I will ask them to buy vocabulary books, and there will always be quizzes and tests in the class. I used to give them a small test on vocabulary every week when I taught intermediate level GEPT preparation to the English majors. I think that for those courses, teachers should be eager to push the students and should be strict so that the function of the test can be fully carried out. Last time, almost half of the class passed the GEPT.”

(Alice, interview, 26.05.2008)

On the other hand, some teachers referred to other tests when asked whether their teaching would be influenced by the English language proficiency requirement for graduation. Although Amy claimed that the GEPT washback on her teaching was limited, she revealed that she would use TOEIC in her lessons, whether or not the graduation requirement was implemented.

Extract 4.22

Researcher: Is your teaching influenced by the GEPT?

Amy: I don't think tests can (influence my teaching) so I won't get in that mode. Except for the business field, I will.

Researcher: TOEIC?

Amy: Um, TOEIC, sometimes when you teach business English, teach how you can make conversations with foreigners, it is very difficult to get materials for them.

Researcher: You will take TOEIC test items?

Amy: Items, yes, can be used for practice. If you can't grasp long conversations, if you can understand short ones, you will gradually, you can gradually listen to longer ones. I will also encourage them to take the TOEIC. It has more accountability. Some companies like airlines will need it.

(Amy, interview, 23.05.2008)

From the above extract, the reason why Amy preferred to incorporate TOEIC into her lessons was not related to the implementation of the requirement, but to the background and the career needs of her students. Similarly, the consideration for students with different backgrounds was why Alice referred to the TOEIC, in addition to the GEPT:

Extract 4.23

Alice: I think that Business Administration, they should take the TOEIC test.(One of Alice classes I observed was a class from the department of Business Administration.)

Researcher: So you think...

Alice: I think it's different.

Researcher: Dependent on the orientation of the department?

Alice: Yes, department, department.

(Alice, interview, 26.05.2008)

In sum, teachers' perceptions on the graduation requirement, the GEPT and their influence on the EAP curriculum and their teaching differed to some extent. Some did not think they would be influenced by the graduation requirement or the test that was considered as very important and influential. They focused on their own preferences on teaching while others might adapt themselves to the implementation and even felt motivated to prepare students for the test. These findings, similar to previous studies (e.g. Watanabe, 1996; 2004; Alderson and Hamp-Lyons, 1007; Burrows, 2004; Green, 2006), reflected how teacher factor could explain the presence or the absence of washback in their teaching.

4.5 Factors that shape the impact and washback on EAP curriculum for non-English majors

The reasons why GEPT washback manifested itself in certain respects but not in others might be due to the following factors.

The English department might be one of the reasons why there was washback in some dimensions but not in others. The department was in charge of developing the EAP curriculum for non-English majors and of determining the regulations of the graduation requirement. Washback on teaching materials was somewhat determined by this factor

because a GEPT related textbook was requested for Becca's English Reading and Writing Course by the English Department. However, it was perhaps because not every teacher was informed of the regulations in the graduation requirement that GEPT had only limited washback.

Extract 4.24

Researcher: Do you know how this school carry out the requirement? No idea?

Betty: They did not notify us. I think that they, our department, they do not have frequent contact with the part-time teachers. They rely on the assistants. I think the department should distribute those important regulations to all teachers.

(Betty, interview, 17.06.2008)

Betty revealed to me that she was actually unclear as to what the regulations of the degree requirement were. As a part-time instructor in University B, she only knew that the students had to take the GEPT to graduate. However, she had limited knowledge of other regulations such as the level of GEPT as threshold, other English tests and what students had to do if they did not reach the requirement. For her, her only task was to teach the first years English without taking into consideration the whole curriculum and the degree requirement.

The limited washback might be also due to the extent to which the department or the university was determined to execute the requirement. As one teacher pointed out:

Extract 4.25

“So that thing, I think it doesn't bring much constraint. I'm not sure whether the school takes this seriously. What I mean by seriously is that if it has not been reached, students have to do something, something that you must do. For example, something they must do means that if they cannot pass the test, they have a chance to retake it. If they fail again, they are not allowed to graduate. Then, it's a matter of great importance, right? But I, I'm not sure the school means it or not. I think the school has made general declaration of the requirement but concerning execution, I think the school does not mean to fully execute it.”

(Ben, interview, 10.06.2008)

What Ben argued was that the requirement did not enforce students to keep retaking the

test until they passed in order to receive their degrees. Instead, the university provided remedial courses in the third year for them and all they had to do to graduate was then to pass the course. That was why, as I previously mentioned, he did not take the requirement into consideration in his teaching.

The institution and department factor can also be applied to University A, which did not require its students to reach a certain level of English language proficiency to be able to earn their degrees. The textbook, the syllabus or even the tests were determined by the department. On one hand, teachers in the English department, the curriculum designers selected a textbook which had test items that followed GEPT format. On the other hand, with limited choice over what to cover in their lessons and what to include in the mid-term and final exams, the washback on teaching was also limited.

Teachers, as Watanabe (1996, 2004) argues, may also be an important factor that mediates washback. In this study, the teacher factor played a more significant role than the institution and department factor. Although most teachers did not have the choice to decide for a textbook they used, the supplementary teaching materials they used reflected much of their own choices. Becca followed the instruction of the English department, and chose a GEPT related textbook, but she actually used authentic materials which she thought would motivate students and which she preferred. Although both Ben and Betty were assigned a textbook for their courses, at the time I conducted the classroom observation, they had both turned to other materials in their lessons. Betty chose materials not related to GEPT, which she thought her students would be interested in. Ben insisted on his students using the monthly English learning magazines for self-study. Even though he claimed that he did not have the intention to prepare his students for the GEPT, his choice reflected GEPT washback to some extent,

as the magazines provided local a context for using the English language, which was what the test promoted. The magazines also provided materials for GEPT test practice. In University B, the design of the end of term examination was left at the hands of the teachers and thus, the teachers themselves determined what they would choose to test, rather than what GEPT or graduation requirements would push them to do.

Teacher's differing attitudes towards the test and test preparation could also result in whether or not there would be washback (4.7). Amy's personal preference for TOEIC over GEPT might result in the manifestation of TOEIC washback instead of GEPT washback in her teaching. On the other hand, Alice would have helped her students with GEPT test preparation if the graduation requirement had been implemented because of her successful experience of a GEPT course she ran for English majors, which helped most of them to pass the intermediate level GEPT.

In sum, either factor, whether the institution/department or the teacher, reflects the role of different stakeholders in the process of washback, as delineated in Bailey's model (1996). In this study, the institution or the English Departments played a determinant role in the extent of washback, because they took charge of the choice of teaching and testing materials, or the dissemination of information concerning the graduation requirement. Nevertheless, as evidenced in this study, the teacher factor contributed the most to the presence or absence of washback on teaching. It was the teachers who personally determined what to teach and whether or not to include test preparation in their lessons.

4.6 Summary

Roever and Pan (2008) have pointed out the need to explore "whether the introduction

of the GEPT as an graduation requirement leads to pedagogical changes in curricula that facilitate overall gains in communicative competence, or whether it results in teaching to the test, which is a common type of negative washback from test use' (p. 406-407). This study showed how test impact and washback works in a context where not everyone was obliged to take one appointed test was rather complex. It was evident that tests could influence EAP curriculum in two main areas: teaching materials and mid- and final examinations of the curriculum. GEPT washback was evident in the English Reading and Writing remedial course which used GEPT specific test preparation materials, a GEPT mock test and a real test. What were particularly interesting were the less evident ones: the use of monthly-issued English learning magazines in Ben's lessons and the midterm and final tests modelled after GEPT in University A. Both showed the complex nature of washback and impact. The magazines were a result of the social impact of GEPT, which in turn exerted an impact on the teaching materials Ben used in his lessons. However, other than the two areas, the influence of the GEPT or any other language tests on the observed lessons was rather limited.

The impact of the English degree requirement on the EAP curriculum was a complex picture; it might be predicted that the impact on the students would be at least as complex, if not more so. In the next chapter, I will report the impact of the requirement on non-English majors.

CHAPTER 5 Impact on the Non-English Majors

5.1 Introduction

In the last chapter, I reported on the impact of the English graduation requirement for EAP curriculum taken by non-English majors. This chapter will focus on the learners of the curriculum. Learners and their learning have always been peripheral in the previous washback studies (2.3.2). Until recently, the few washback studies on learners (Gosa, 2004; Green, 2006 a, b, 2007 a, b; Tsagari, 2006) have demonstrated that learners can play an equally important role as teachers and other stakeholders in the washback process. In the context of the present research, since the main purpose of the graduation requirement was to promote the learning of English among university students, it was essential to put the students at the centre in order to understand how the graduation requirement has influenced them. Due to time limits, this study was not longitudinal, and was not able to investigate how students' learning outcomes might be affected by the English language proficiency requirement. The focus of this study was thus to explore the impact of the requirement on the learners and their learning process.

It was important to find out what kind of learners the non-English majors are, so that the manifestation of different aspects of the impact of the graduation requirement on different learners can be understood. For this purpose, the Effective Lifelong Learning Inventory (ELLI) was used. ELLI is a self-report questionnaire with the purpose of assessing a learner's learning power (See 2.5 for a detailed description of ELLI) that can be used for self-diagnosis and also for showing patterns among a large group of learners. The general patterns across students in an institution and the individual learning profile of each student will be described in the later sections.

This chapter attempts to answer the following research questions:

Research Question 2:

What are the effects of the English requirement for graduation on the non-English majors in Taiwanese universities?

Sub-questions:

- a) What are students' perceptions towards the requirement?
- b) What are the effects of the tests (GEPT and other tests) on the students?
- c) Is there difference between the effects of the different tests on the students?
What are the differences and why?

Research Question 3: To what extent students' learning power explain their perceptions of washback on the learners and their learning?

For RQ2 concerning the impact of the graduation requirement on the students, the main source of data was the 18 student interviews. The interviews were coded inductively, and triangulated with the part of the teacher interviews which referred to the students.

The data for analysis for RQ3 consisted of the following three sets:

- 1) 454 ELLI data from non-English majors in both University A and B,
- 2) Among the 454 valid data (among 505 questionnaires collected, only 454 were valid), 18 ELLI profiles of student interviewees in the observed classes, and
- 3) 18 student interviews.

The 454 ELLI data were analysed with SPSS 16.0, and the students' seven dimensions of learning power were calculated, producing an ELLI learning profile for each student. The scores on the seven scales were then analysed using an independent T-test, in order to find out whether there was difference between the macro patterns of

learning power among students in University A and B. The mean scores of all the students in each university were presented in pie charts for the three sections, which indicate whether the students' overall performance on the seven dimensions were of a high, moderate, or low level (Figure 5.1, 5.2). The scores of the 18 student interviewees were presented in spider diagrams, so that the seven scales of each student would be clearly displayed. Lastly, the 18 learner profiles were also analysed in relation to the students' interviews to explore how their learning power could explain their perception of the impact of the requirement.

The report on the analysis in this chapter starts by providing information about student interviewees (5.2). From the analysis of student interviews, I will then discuss what the students knew about the graduation requirement for English proficiency (5.3), and provide evidence of washback from GEPT and other English proficiency tests (5.4). The main focus of this chapter is on the perceived impact of the graduation requirement on the students (5.5) from two perspectives – teachers and learners. Section 5.5.1 reports on how the graduation requirement influenced the learners from the teachers' perspectives. Section 5.5.2 reports on the learners' attitudes towards the implementation of the language proficiency requirement. The next few sections provide a detailed report of the analysis of the ELLI data in this study. 1) the ELLI university profile, presented in pie charts of the seven dimensions (see above for description of pie charts, see also 5.5.3, Figure 5.1, 5.2) 2) The T-test results between the two universities in relation to the seven dimensions (see Table 5.1) 3) the individual ELLI profile, as shown in a spider diagram (5.5.4, Figure 5.3). Lastly, the learners' perceptions of the impact of the requirement on themselves, along with their ELLI profiles, are presented in 5.5.6, followed by a summary of all the findings in 5.6.

5.2 The student interviewees

Altogether, 18 students were interviewed, 9 from each university. The interviewees had participated in every section of the present study, including classroom observation, the ELLI questionnaire and individual interviews. All the 18 students attended the EAP curriculum for non-English majors, entitled 'English Integrated Skills Training' in University A, and 'English' in University B. Due to the limited time for classroom observation, there was no student interviewee from the specified remedial course, 'English reading and writing'. The names representing each student in the data presentation were all pseudonyms. For the convenience of differentiating participants from the two cases, following the rule of names used for the teachers, students from University A were appointed with names beginning with 'A', and B for students from University B.

5.3 Learner's knowledge of the graduation requirement

As there was no graduation requirement for non-English majors in University A, the majority of student interviewees in University A only had a very rough idea of what the requirement was. They only understood the requirement as a policy that required them to provide official evidence of English proficiency in order to graduate and receive their bachelor's degrees. In University B, even though the regulations of the graduation requirement were publicly announced on its official website, very few of the students could recall the details of the requirement, for example, what English proficiency tests are accepted and the levels. Much to my surprise, one of them was not even aware of the graduation requirement.

Extract 5.1

Researcher: I would like to ask you if the degree requirement brought any influences to you?

Bryan: Graduation requirement?

Researcher: It has been there since you first came into this university.

- Bryan: Do you mean credits?
Researcher: English, English graduation requirement.
Bryan: What is that?
Researcher: The English graduation requirement is that your university will require you to take English proficiency tests and ask you to pass a certain level.
Bryan: Really?
Researcher: Don't you know?
Bryan: I am doomed. I won't pass.
(Bryan, interview, 10.06.2008)

The rest of the students were less ignorant of the requirements. However, they were still not able to elaborate beyond the first stage of the intermediate GEPT, what other English proficiency tests were accepted by the requirement and the levels required. This can be clearly shown from Bonnie's description of the requirement.

Extract 5.2

- Researcher: Are you familiar with the regulations of the English graduation requirement?
Bonnie: Not familiar.
Researcher: Do you know what test you have to take?
Bonnie: I only know that we have to pass the first stage of the intermediate GEPT and then, we do not need to attend another course in our third year.
Researcher: Do you only know about the GEPT? Do you know about other tests?
Bonnie: Yes, I do but I don't know whether they count as the requirement.
Researcher: You don't know what grades you need to get (for the other tests)?
Bonnie: No.
Researcher: TOEFL and TOEIC can also be considered in the requirement. But you don't know what grades are required?
Bonnie: No. I have absolutely no idea.
(Bonnie, interview, 10.06.2008)

The extract reveals that although Bonnie had knowledge of other English proficiency tests available in the market; she only recognised the graduation requirement with the GEPT and the level she needed to pass. This direct alignment of the GEPT with the requirement was also evident in the interviews of almost all other students in University B. None of them referred to English proficiency tests other than the GEPT, when the

graduation requirement was mentioned. For them, to meet the requirement was equivalent to passing the first stage of the intermediate GEPT. Other tests like TOEFL or TOEIC only came up when their future academic or career plans were discussed (5.5). The above findings thus suggested that what the students in University B knew about the graduation requirement implemented in their university reflected GEPT washback, to some extent.

In the next section, I will report on the evidence of washback from English proficiency tests, as emerging from student interviews.

5.4 Washback of GEPT and other tests

The graduation requirement accepts a wide range of English proficiency tests external to the university settings, including the GEPT, the TOEFL, the TOEIC, and IELTS (see 3.5.1). As with what was discovered on teaching in Chapter 4, GEPT had the strongest impact on learners, as evidenced from two points of view: (a) GEPT is the only test that is directly linked by the students to graduation requirement (5.3.1) and (b) the GEPT is the test taken or going to be taken by the majority of the students. In addition to fulfilling the requirement, most students chose GEPT over other language tests because (a) it was considered relatively less demanding, (b) it is more popular in Taiwan than other language tests, and (c) they were influenced by their parents. Despite GEPT being more influential than other English proficiency tests in Taiwan, some students would still choose to take other tests such as the TOEFL or the TOEIC, because of their internationally-recognised status and for purposes such as further studies in the foreign countries and student exchange.

As discussed in 5.3.1, the students' limited knowledge of the graduation requirement

and the alignment of the intermediate GEPT with the requirement already suggested that the GEPT exerted a level of washback. In addition, among all the tests stated in the requirement, GEPT was the test that was the most popular. In fact, 6 of the student interviewees (3 in University B and 3 in A) had already taken the GEPT before this research was conducted. Other students in University B also revealed that taking the GEPT was a planned goal before their third year, when they needed to take the remedial course. Similarly, most of the other students in University A, when asked which test to take if they were required to, also referred to the GEPT. The reasons are as follow:

First, the GEPT was considered by some students as a starting point for taking external English proficiency tests, because it was perceived as less demanding.

Extract 5.3

Researcher: (Have you thought of taking) TOEFL, TOEIC, or other tests?

Brad: TOEFL is more difficult. I'd start with the easier one.

Researcher: Because you think you can reach the level?

Brad: Yes.

(Brad, interview, 10.06.2008)

Another reason was the students' assumption about the important role the GEPT certificate played in job searching. The certificate of the test was considered to be a necessity in job applications because it was well considered by wider society, as explicated by Alex's comment:

Extract 5.4

Researcher: You think that it (taking the GEPT) is necessary. At which aspect do you think is it necessary?

Alex: For a job.

Researcher: Which test will you choose to take?

Alex: GEPT.

Researcher: Why GEPT?

Alex: It's more popular.

(Alex, interview, 23.05.2008)

As with Alex's comment, Alvin also revealed his assumption about the test as being essential in the work field. As a person who was reluctant to take language tests, he said that the only reason he would take the GEPT was if 'it is requested by the boss or the company' (Alvin, interview, 23.05.2008).

Because of the popularity of GEPT among other levels of education, April considered that she would also start with the GEPT.

Extract 5.5

April: Should be the GEPT.

Researcher: Why?

April: Because (we) start to take the test since junior high school. Also take it in senior high school. So...

Researcher: So did you take it in your senior high years?

April: No, I didn't.

Researcher: Did your school ask you to?

April: We registered on our own. If you want to go, you go. If you don't want to, then no. It wasn't compulsory.

Researcher: So you want to take this test because you always hear about the test?

April: Um.

Researcher: Do you consider taking other tests?

April: I will definitely take this (test) as the fundamental (step).

(April, interview, 07.05.2008)

It was implied in her comments that the GEPT had a strong impact on institutions of different educational levels, and that students would be encouraged by the institutions to take the GEPT, even though the test was not designed for their curriculum.

Parental suggestions were a further reason why some of them took the test, as illustrated by Bridget's case:

Extract 5.6

Researcher: Why do you take the GEPT, not other tests?

Bridget: Because...actually, it's my father who asked me to take the test. GEPT can be

considered as the most basic one and so I began with this.

(Bridget, interview, 10.06.2008)

The above extracts have two implications. First, the GEPT was perceived as less difficult and the more 'basic' test to start with, compared with other English proficiency tests available. This implication further suggested that students' perceptions of test difficulty as not too demanding might result in a higher extent of washback (see 6.4.2 for discussion). Secondly, the students' choice of the GEPT was shaped by the following factors: 1) the impact the test has had on the society 2) its well-known status, 3) the assumption of its necessity in job applications 4) its influence on the institutions of different educational levels, and 5) its influence on parents.

Nevertheless, the relative dominance of the test did not mean that there was no impact of other tests on the students. The TOEFL and the TOEIC were two popular alternatives to GEPT. However, the reasons why the students wished to take those tests were not specifically related to the graduation requirement. They planned to take TOEFL and TOEIC tests for their own academic plans. Some were concerned about the disadvantage of the GEPT being a test that receives only formal recognition locally, while others were discouraged by their previous failures with GEPT.

Some students in University B revealed their plans to take TOEFL in order to fulfil their personal desire for studies abroad, such as student exchange or advanced studies. Bess's case was especially interesting, in that she would take GEPT for the sake of meeting the requirement, but later, she would take the TOEFL for further studies abroad in her field.

Extract 5.7

Researcher: So most of you will take the test. Have you taken the test?

- Bess: I haven't, but I am going to.
- Researcher: Are you going to take the GEPT?
- Bess: Yes.
(omitted.)
- Researcher: Is English very important to your department?
- Bess: Very important. Because local design curriculums and the design field in Taiwan were not as developed. If you really want to be good in the design field, studying abroad is a must.
- Researcher: So which test are you going to take then?
- Bess: TOEFL. You will be selected according to the assessment of your English proficiency and then the evaluation of your results in the university
- (Bess, interview, 10.06.2008)

Bess' approach on which test to take for different purposes suggested that the GEPT washback was related to the graduation requirement but the washback of TOEFL was not.

One important reason why TOEFL or TOEIC was favoured over the GEPT was the drawback of the GEPT as a locally recognised test. Abel, being the only student who refused to take the GEPT, argued that in addition to the purpose of advanced studies abroad, he preferred taking the TOEFL or the TOEIC because the certificate of the two tests were 'internationally recognised' (Abel, interview, 10.06.2008) whereas the GEPT certificate was restricted to local use.

The certification of English proficiency that could be used internationally was also one reason why Archer went for TOEIC preparation. His interview revealed the most explicit evidence of test influence on students. He was the only student among the 18 interviewees who was taking external test preparation courses at the time of the interview, in addition to the compulsory EAP course in the university. The course he attended was a TOEIC preparation course, provided by a private language school. What

was particularly interesting was the other reason why he prepared himself for TOEIC: his past experience of failing the GEPT. As an education major with the intention of obtaining a masters degree in education, it was difficult to understand why he would want to prepare himself in business English. Besides favouring the international certification of TOEIC, Archer confessed that it was mainly because of his past failure to pass the second stage of the intermediate GEPT.

Extract 5.8

Archer: It's very easy to pass the elementary level. But I tried the intermediate level and I could only pass the first stage. I could not get through the second stage.
(omitted)

Archer: I am very keen on promoting my English proficiency so at the moment, I'm taking a TOEIC preparation course in a private language school.

Researcher: Why do you want to take TOEIC preparation course? Since the masters degree you are going for in the future was not related to business English at all, why TOEIC but not other tests on four abilities like GEPT?

Archer: There is only one stage in the TOEIC, only reading and listening. It is easier for preparation and it is also easier to get the certificate. There are two stages in the GEPT and it is difficult to pass the second stage. Furthermore, TOEIC is an internationally recognised test while GEPT is only accepted in Taiwan. That's why I choose to prepare TOEIC.

(Archer, interview, 7.05.2008)

Archer's anxiety towards the GEPT somehow challenged Brad and some other students' perceptions of the GEPT as a basic, less demanding test. Therefore, I would argue that the way in which students perceive a test, their test preparation, and their previous experiences in taking the test may affect the impact of the test on each individual student.

In summary, the GEPT was perceived as the most important test among the English proficiency tests accepted by the English language proficiency requirement for graduation. Its importance could be seen from the students' assumptions that the

English language proficiency requirement was equivalent to the GEPT requirement. In addition, they had little knowledge about other English proficiency tests they could choose from. Another piece of evidence was that the majority of them had either taken the GEPT or were planning to take the test. Nevertheless, the GEPT washback was less the result of the graduation requirement than the result of the test's social impact. This social impact could be seen through its high popularity, its influence on institutions at different levels of education, and its influence on parents. Nor did the impact of other tests such as TOEFL or TOEIC result from the graduation requirement. Instead, it was more closely related to the students' academic plans, to their past failures in GEPT, and to the weakness of the GEPT as not being accepted and recognised internationally.

5.5 Perceived impact of the graduation requirement on non-English majors

This section discusses in what ways the implementation of the graduation requirement for English proficiency influenced the non-English majors, from both the perspective of the teachers and the students. In the previous chapter (3.6.4), I have mentioned that the teachers' perceptions were used to triangulate students' perceptions of how the requirement influenced the students, and to see whether there was a discrepancy between the two. In addition to probing into the students' views, an important part of this section is to relate the full ELLI data and the learning power profiles of the 18 interviewees to their perceptions of the impact of the graduation requirement on them. The purpose is to see whether their learning power can be used to explain the individual difference that shapes the impact.

5.5.1 Teachers' Perspective

The teachers' perceptions of the ways the requirement influenced the students varied

with their own attitudes towards the requirement itself. Those who held a positive view of the requirement believed that the implementation, if carefully executed, would bring about good influences on students, especially their motivation to learn English. On the other hand, others were less positive, remaining sceptical as to the certainty that the requirement would promote students' English proficiency and their English learning. They had doubts as to how the requirement would be implemented, and also challenged the overly optimistic view that testing would promote the motivation to learn.

Betty was the only teacher who fully agreed with the idea of the English language proficiency requirement and its implementation. She regarded the graduation requirement as a positive tool to increase students' motivation to learn.

Extract 5.9

Betty: I think that the graduation requirement can urge the students to re-evaluate their language proficiency, to see if they need to enhance their ability. For some students, what they show is only 40% of their potential. They think they are rated Class B (not top students) and lack interest in learning. This provides a sense of pressure. You need to fulfil the requirement or else you can't graduate. This will fire up their motivation to learn. Being compelled to do something is also a type of motivation. Even though it is negative, it is still motivation. For them to graduate, they definitely need to study more. They will be willing to listen in class, to ask more questions. So I think this (requirement) helps to increase their learning motivation. It is also beneficial for students with higher proficiency as well. They once dominated in English but now since everyone else is working hard on English, they will realise that they also need to work harder.

(Betty, interview, 17.06.2008)

The extract above suggests that she believed that the benchmark level of the graduation requirement for the non-English majors in University B was not too high. Therefore, she argued that the implementation would motivate the students, whether or not they have lower or higher proficiency. She also revealed that she did not object to GEPT washback on the students because she considered the test as a publicly

recognised, fair test, which was promoted by the Ministry of Education.

Similarly, Alice was supportive of the implementation of such a requirement, arguing that there was indeed a difference for the students between having to pass an external test and having to pass an internal course. She also pointed out that teachers should play an active role in compelling students to reach the graduation requirement. However, as mentioned in the previous chapter (4.4, Extract 4.23), she disagreed with the alignment of the requirement with the GEPT. She suggested that the university should tailor the regulations to the needs of students from different backgrounds, so that the desired effect of the requirement can be fully realised.

Extract 5.10

Alice: I think (external) testing still make a difference. The university should require students from different departments to take different English test. For example, students majoring in business administration should take the TOEIC while those with backgrounds that are more likely to be in the civil service should take the GEPT. This will definitely increase students' learning motivation. As for the teachers, we should tailor our lessons to the requirement, such as preparing the students for the GEPT. We should actively engage the purpose in our teaching and strictly push our students so that the requirement will truly come into effect.

(Alice, interview, 23.05.2008)

Other teachers, however, were less confident that the graduation requirement could help in promoting students' learning and their motivation to learn, because they did not think the requirement would work as most students were passive learners. In addition, there was always a possibility for students to take advantage of the loopholes in the requirement or just gave up because they knew it would be too difficult for them to pass external English tests.

Anna was worried that the implementation of such a requirement would result in unintended malevolent effects:

Extract 5.11

Anna: I do think that the graduation requirement is important because my students are all passive learners. This may be able to push some of them to learn more English. But most students in our university have poor English proficiency and I wonder how many of them can pass. I'm afraid that this may cause 'adverse effect' for the university such as higher drop-outs.

(Anna, interview, 01.05.2008)

Some teachers with the experiences of teaching in a university with the graduation requirement did not think that the requirement was as effective as intended.

Extract 5.12

Amy: The English graduation requirement did not work well in the university I was previously in. Students did not feel alert nor did they sense much pressure. It's like, I'm ('I' referring to students) just not good in English. Even if you set a requirement, I'm not able to fulfil it... (omitted). I don't think that it works well. We had the experience before. What happen to those students who could not fulfil the requirement? The university has to find some ways to help them since the Ministry of Education has not yet announced that all university students are not allowed to graduate without an English certificate. Perhaps half of the students fail the tests and do not succeed after several attempts. Then, the university will provide a remedial test or course for them as substitutes. In the end, the substitutes may be far different from what the requirement originally asks for.

(Amy, interview, 22.05.2008)

The above extract suggests that Amy believed that students with low English proficiency who were not already motivated in English learning would hardly change through the implementation of the graduation requirement. In addition, she pointed out that it was possible that universities might be lured into reducing the number of students who failed to reach the requirement.

There were loopholes created by the universities, and also by the students themselves. Both Adam and Becca raised the issue of students using remedial courses to avoid taking an external language test.

Extract 5.13

Becca: Those who have better English proficiency think that it's alright if they do not take a test as long as they attend this course. One reason may be that perhaps they are not wealthy or what, so

they don't find it necessary to pay for such tests. I don't know them. Some people think differently. They just don't want to take external tests. I also doubt whether those with poorer English proficiency have ever tried taking the test. It is just much easier to attend a course and pass it.
(Becca, 03.07.2008)

What Becca suggested was that it was always easier for students to opt to take an extra course, in preference to taking an external English proficiency test, no matter what level of proficiency they had. This was why Adam argued that the universities should not accept remedial courses as a substitute for certificates from an external test.

Extract 5.14

Adam: Some students are under the impression that as long as they take and pass the remedial courses, they have fulfilled the graduation requirement. It is an easy way out. Thus, execution is very important. It doesn't really matter what level of the tests you ask the students to reach, intermediate or higher-intermediate GEPT. What matters is whether the university insists that the students can only graduate with an external proof of English proficiency.

(Adam, interview, 02.06.2008)

Although the teachers knew the problem of the loopholes hidden in the graduation requirement and the possible consequences, these were not easy to solve. Ben was particularly pessimistic about the extent to which the university could enforce the requirement.

Extract 5.15

Researcher: So do you think the graduation requirement should be strictly executed?

Ben: I don't think the university is able to do so. Because there is always another way out, such as the remedial course. At the latest, the university will definitely let them graduate in the summer of their fourth year. If they can't pass the intermediate level (GEPT) reading and listening, they still have that chance. I think the university is unable to execute the requirement strictly.

Researcher: I heard there are universities that do so but most don't.

Ben: I think they can't. Some students can fulfil the requirement. They take the requirement seriously and feel that more learning is useful for themselves. I think it depends on the students' learning attitude. At present, the university is

not going to warn by example. In other words, I wonder what the university would do if the students fail. The crucial thing is that there is no way the university can be very strict about this. They can only make the students to take another course. Where will they get the money for the additional course? They will probably protest against this if you make them spend more money.

(Ben, interview, 10.06.2008)

What Ben was trying to point out was that if the students themselves did not want to learn, the impact of the requirement on the students' learning and motivation to learn would be very limited, since the requirement could not be readily executed.

The above analysis revealed that the teachers' perceptions of how the graduation requirement influenced non-English majors were more negative than positive. Although it might be a good intention to pushing students to learn more English, most teachers did not believe that the intended impact on students' learning would be realised. It was likely that students' motivation to learn English would not be changed by the need to take an external English proficiency test. There was the possibility that the requirement might result in adverse effects, such as a higher drop-out rate for students with low English proficiency. Furthermore, students might choose the easy way out by skipping the test taking procedure and taking the remedial courses instead, regardless of their language proficiency. Therefore, what the majority of teachers suggested was that it was very likely that the implementation of the graduation requirement would not promote students' English learning or their motivation to learn English as intended.

5.5.2 The learner's perspective: attitudes towards the graduation requirement

Students in both universities showed individual differences concerning their attitudes towards the graduation requirement. Some of them held a positive view of the graduation requirement because they believed the implementation of the requirement

could push them to work hard in terms of improving their English proficiency, while others were intimidated by the pressure of not being able to graduate. Some others disliked the compulsory elements the requirement entailed, preferring to choose for themselves whether or not to take an English proficiency test. The remainder, on the other hand, did not see the requirement as either positive or negative, as they took English learning into their own hands and cared more about their own learning than merely fulfilling the requirement.

Those who approved of the idea of setting a benchmark for English proficiency in University A considered the requirement as essential and beneficial. For example, Alex regarded a requirement for English proficiency as being essential because of the status of the language.

Extract 5.16

Alex: I think it is necessary (to have the requirement).

Researcher: Why?

Alex: Because nowadays, everything needs English.

(Alex, interview, 23.05.2008)

Aiden and Abel both thought that the benchmark of English proficiency set for the students would be beneficial for them.

Extract 5.17

Researcher: Do you agree with the idea of graduation requirement?

Aiden: I agree, highly agree.

Researcher: Why?

Aiden: Because, (students) will have the standard when we leave (the university).

(Alex, interview, 23.05.2008)

As with Aiden, Abel strongly expressed his support for the implementation. He

commented that if he were unable to reach the level of English proficiency specified in the requirement, he would be willing to stay in the university and try again until he passed.

Extract 5.18

Researcher: Why?

Abel: Because I like it.

Researcher: You like it?

Abel: It is better to have that 'level'.

Researcher: Do you think you can reach that level?

Abel: If I can't, I just don't leave (the university).

(Abel, interview, 15.05.2008)

The positive view of the graduation requirement as something that would help students on learning English was shared by Bess.

Extract 5.19

Bess: Because it (the requirement) will compel me to learn English and English is very important.

(Bess, interview, 10.06.2008)

Brenda also applauded the implementation because she believed the graduation requirement would be helpful in promoting non-English majors' English proficiency for general communication.

Extract 5.20

Brenda: I think it (the requirement) is quite good because not everyone has good English proficiency. If the university didn't require and we only have English in our first year, even though we read academic books in the original language (English), we can't communicate with others after we graduate.

(Brenda, interview, 10.06.2008)

Even Bryan, who felt 'doomed' at having to take an English proficiency test, despite his poor level of English proficiency, accepted the idea of the requirement as fulfilling a practical purpose. Indeed, he considered the certificates of language proficiency as 'the materialisation of strength and ability' (Bryan, interview, 10.06.2008).

Although the majority of the students were not against the idea of the graduation requirement for English proficiency, this did not necessarily mean that they welcomed the implementation. From a practical point of view, Archer and Anson pointed out the need for the university to take students' proficiency level into consideration before implementing such a requirement. At the time of the interview, Archer had already passed the intermediate level of GEPT. However, he argued that it would be unrealistic for University A to implement the requirement with standards as high as the intermediate GEPT, since most non-English majors would not be able to reach. Anson shared similar views.

Extract 5.21

Anson: It depends on the threshold level of the graduation requirement and also the quality of the students.

Researcher: But the requirement is designed according to the university and the students' proficiency level.

Anson: Yes. If the university could arrange the requirement depending on students' level, I can accept that. But if the requirement is too high, I will think it's unrealistic. No one can reach it.

(Anson, interview, 14.05.2008)

Some students in University A disliked the idea of implementing the graduation requirement, believing that personal choice should be over and above the need to take an English proficiency test. Alvin reasoned that English was not an essential ability for everyone and thus, whether or not to have a proof of English proficiency was dependent on one's needs. Likewise, April was also against the compulsory element that came along with the graduation requirement.

Extract 5.22

Researcher: Speaking for your own, you do not need to consider others. If there were only you in the university, do you want the university to do so?

April: I think it's unnecessary. If they want to take the test, they will. You can provide

the opportunity but it is the students' choice to register (for the test) or not.

Researcher: So you think it does not need to be compulsory?

April: Yes.

(April, interview, 07.05.2008)

Billy, who needed to fulfil the requirement, also argued against the need to make test-taking obligatory for non-English majors. .

Extract 5.23

Billy : I feel ok if I study English for my own. But I don't think it is necessary to take a test for graduation.

Researcher: So do you think the university should not have this (requirement)?

Billy: (No.) Because those who want to study will study English on their own. Those who want to take English test will take the test.

(Billy, interview, 10.06.2008)

Aside from the compulsory factor encased in the requirement, some students were against the implementation because of their anxiety and fear towards the need to fulfil the requirement.

Extract 5.24

Researcher: Do you like the idea of English graduation requirement?

Andrew: English graduation requirement? Truly speaking, other schools can (have the requirement) but it is not good to have it imposed on me.

Researcher: Why?

Andrew: My English is not good. This will give me a hard time.

Researcher: So you want to graduate as soon as you can and you don't want the requirement?

Andrew: Yes. I don't want to go for further studies either.

You don't like to learn English?

Andrew: It's not that I don't want to learn English. I just think that if I don't need it (the English language), there's not much place to use it.

Researcher: So you think it's not necessary for you?

Andrew: I don't think I will need to use it in my daily life.

(Andrew, interview, 15.05.2008)

The extract above shows that Andrew did not want the graduation requirement to be

implemented for two reasons. He did not see English as essential and he was anxious towards the hard time he could foresee himself going through. This was also the case for Aaron and Alex as they knew they would not be able to reach the required standard of English proficiency.

As with the abovementioned students in University A, Bonnie revealed her anxiety by complaining about the link between external testing and degree completion. Being compelled to fulfil the requirement in University B, Bonnie considered test preparation in order to graduate as very troublesome, stating that ‘it is a great stress’ (Bonnie, interview, 10.06.2008).

In stark contrast, there were students in University B who were not anxious at all about fulfilling the graduation requirement. Their attitude towards the requirement was neither positive nor negative, as they cared less about the requirement than about their English learning. Brad and Bridget had already passed the level of the GEPT indicated in the graduation requirement. Blair and Bianca planned to obtain a certificate of English proficiency, even if the requirement was not implemented. One reason might be that they did not perceive the standard specified in the requirement as difficult to reach. This could be explicated by Blair’s doubts as to what the university intended to do with the requirement.

Extract 5.25

Researcher: Do you think that the graduation requirement bring stress to you?

Blair: Not really, because I think the standard (the requirement set) is so low.

Researcher: You don’t think it’s difficult to reach?

Blair: Um, because the intermediate level is supposed to be the level of high school graduates. And it (the requirement) only requires us to take the first stage. So, you know, many people have already passed that level in their high school.
(omitted)

Blair: Theoretically, I think the standard should be set higher.

Researcher: So it's better for you if the standard is higher?

Blair: Um, but, actually I think it depends on the individual. For me, I think that university doesn't really mean it.

(Blair, interview, 10.06.2008)

Blair's statement in Mandarin, 'the university doesn't really mean it' indicates her doubts on the university's intention to promote higher English proficiency. For her, the required English proficiency level to graduate was too easy to reach and thus, she was uncertain how serious the university took the requirement into consideration. Blair's scepticism was somehow similar to her instructor's (Ben) comments on the implementation of the graduation requirement. He questioned whether the university took the requirement seriously (Extract 4.25). Although Blair and Ben had different standpoints on this issue, both questioned why the university would implement the graduation requirement if it does not take it seriously.

The above analysis revealed that students varied in their attitudes towards the implementation of the graduation requirement in their universities. Those who viewed English language as very important welcomed the implementation. They also believed that the graduation requirement could compel them to learn more English and improve their English proficiency. However, because the graduation requirement is compulsory, some students were against the implementation of this policy. They argued that it should be their personal choice as to whether to take external language tests. Another important reason that determined students' attitudes towards the graduation requirement was the perceived stakes of the standard specified in the requirement. Those who did not consider reaching the threshold level as high stakes were nonchalant towards the existence of the requirement. On the other hand, others might view the benchmark scores as high stakes because of their lower English proficiency. Thus,

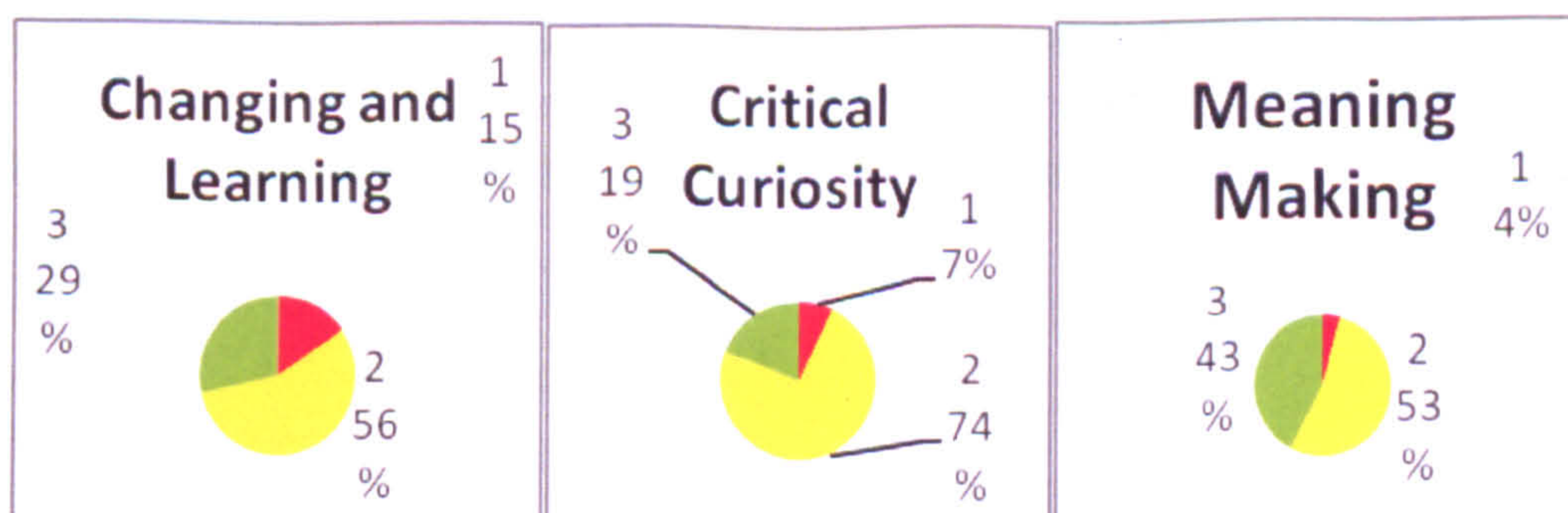
their anxiety resulted in their reluctance to learn English and take an English proficiency test.

This section has already pointed out the individual differences between learners, through a consideration of their attitudes towards the graduation requirement. In the next sections, I will discuss how their learning power could explain these differences. First, the learning power between students in the two universities will be compared to understand a general pattern of students' learning power in each university.

5.5.3 Learning power: the comparison between the two cases

University A

Two hundred and sixteen students ($n=216$) from University A completed the ELLI questionnaire. The students were from departments of Japanese language, business administration, information management and other business-related departments. The ELLI profile for University A students is presented in Figure 5.1. The pie charts are produced from the mean scores of each student on each dimension, with colours resembling traffic lights of red, yellow and green. The colours and numbers indicate the number of students reporting themselves as having a low level (red, 1, scores ≤ 33.33) of a particular dimension, moderate levels (yellow, 2, $33.33 < \text{scores} \leq 66.67$) or high levels (green, 3, $66.67 < \text{scores} \leq 100$). (Deakin-Crick, 2007).



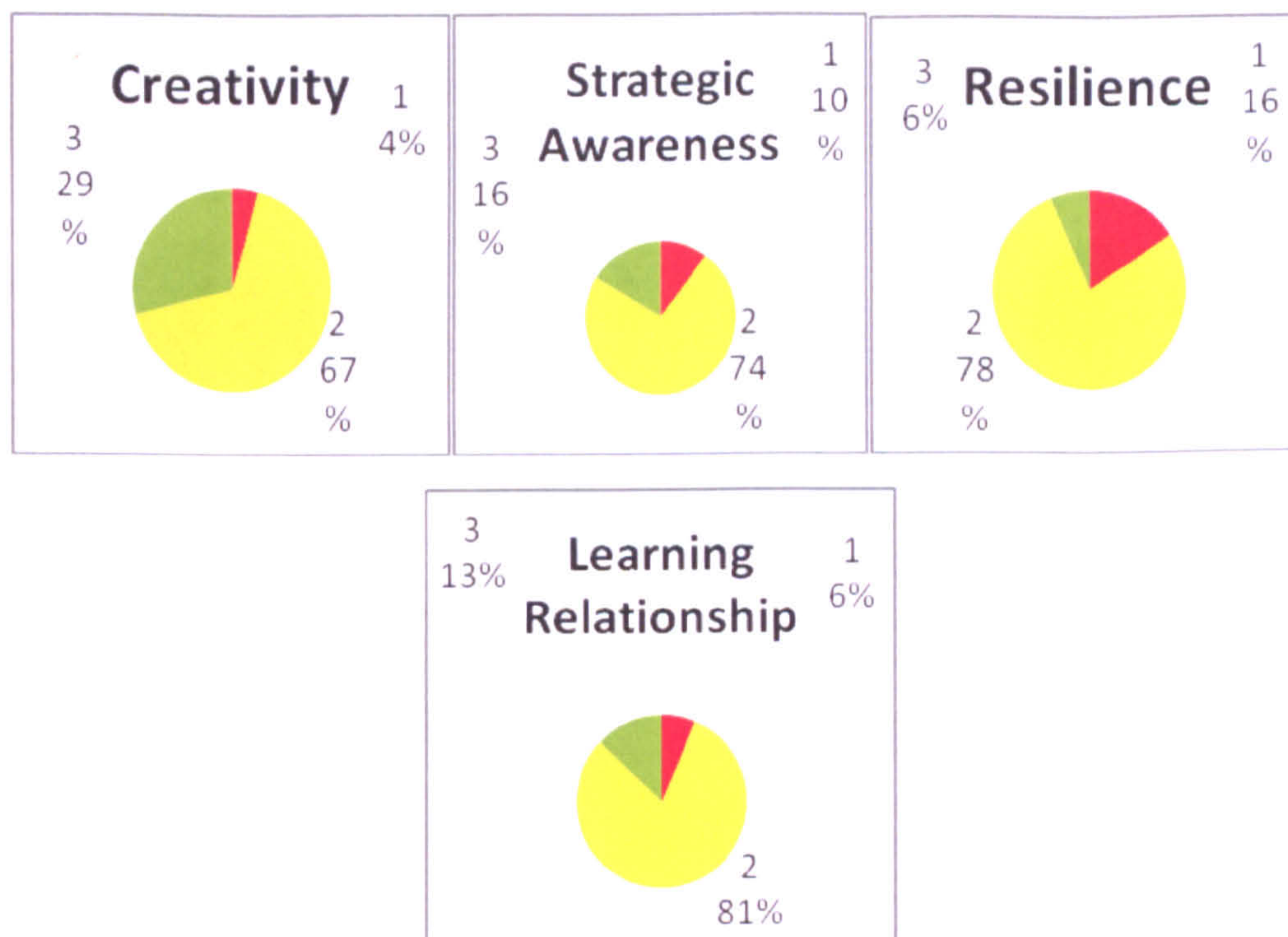


Figure 5.1 Mean Learning Profile for University A students (N=216)

The above profile suggests that the students in University A were reported to have relatively high levels in all seven dimensions. They were better at the active learning dimensions of curiosity, creativity and meaning-making and had positive learning relationships. However, their responses suggested that they had slightly less sense of changing and learning, strategic awareness and resilience.

University B

Two hundred and thirty eight (n=238) non-English majors in University B completed the ELLI questionnaire. Since University B was oriented towards education or early childhood education, the 238 non-English majors had backgrounds in different aspects of education, with a few less-related fields such as arts and design or social and regional development. The mean scores of the seven dimensions for University B are presented below.

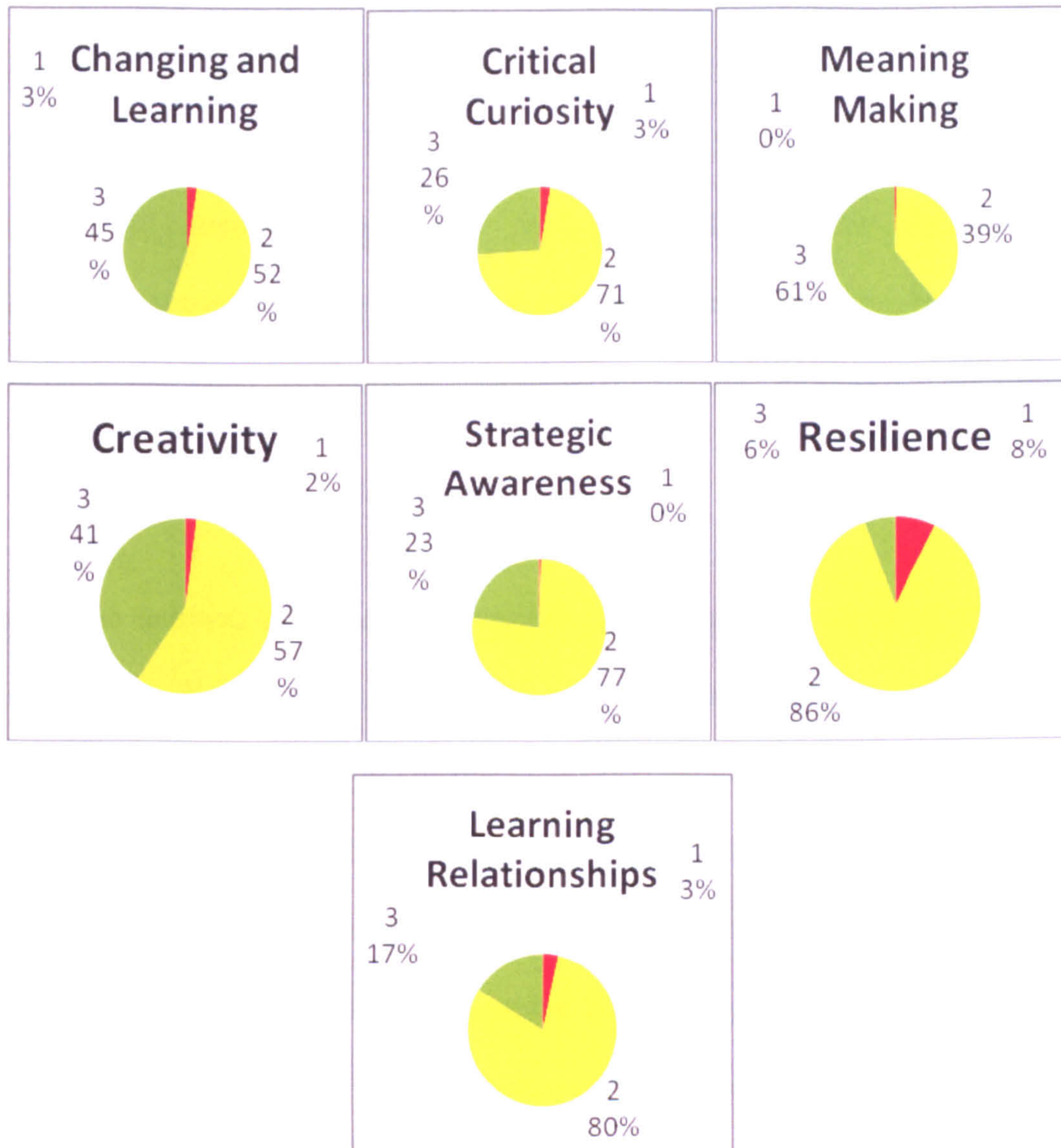


Figure 5.2 Mean Learning Profile for University B students (N=238)

The pie charts above suggest that most of the students in University B regarded themselves as moderately or highly effective learners in all seven dimensions. They were particularly good at the active dimensions of meaning-making and creativity, and they had a good sense of themselves as being able to grow over time. They also exhibited a good level of curiosity, strategic awareness and positive relationships. The only dimension that the majority of them considered themselves weaker in was ‘resilience’.

Comparison between University A and B

The comparison between the pie charts that represented University A and B revealed

the following similarities and differences:

Both university profiles suggest that 'resilience' is the weakest dimension of the students' learning power. Only 6 % of the students in both universities rated themselves as having a high level of resilience. Compared to the other dimensions, the percentage was very low.

Overall, the learning power of the students in University B on all dimensions was stronger than that of the students in University A. Excluding the 'resilience' dimension, less than 5% of the students in University B rated themselves as low on the other six dimensions. In other words, the majority of the students considered themselves as moderate to strong learners on those dimensions. On the other hand, for the dimension of 'meaning making' and 'creativity', under 5% of the students in University A rated themselves as low. Even for the dimension of 'resilience', twice the number of the students in University A considered themselves as weak on this dimension than those in University B. There were considerably more students from University B than from University A who reported themselves as being strong on three dimensions: 'changing and learning', 'meaning making' and 'creativity'.

The above comparison is made out of impression on the pie charts and it is important to see if the statistical analysis of the data echoes the similarities and differences. Therefore, an independent T-test was conducted to check whether there was significant difference between the 7 mean scores between University A and B. The results of the T-test (see Table 5.1) confirmed what the pie charts revealed. There was a significant difference in mean scores between University A and B on all the seven dimensions ($p < .05$). That is to say, overall, the students in University B had a higher level of

learning power, and were more effective learners than their counterparts in University A.

Dimensions	University	Cases	Mean	SD	T-Value	df	2-tail Prob.	Partial Eta Squared
Changing and Learning	A	216	60.0	20.3	-5.3	452	<.001	.057
	B	238	69.1	16.3				
Critical Curiosity	A	216	56.1	15.4	-3.4	452	.001	.026
	B	238	60.9	14.0				
Meaning Making	A	216	66.9	15.7	-5.6	452	<.001	.065
	B	238	74.6	13.6				
Creativity	A	216	59.8	15.6	-4.1	452	<.001	.036
	B	238	65.6	14.1				
Strategic Awareness	A	216	54.2	14.7	-4.6	452	<.001	.044
	B	238	60.0	12.0				
Resilience	A	216	47.9	14.4	-2.9	452	.003	.019
	B	238	51.5	11.2				
Learning Relationships	A	216	54.7	12.4	-2.0	452	.041	.009
	B	238	57.1	12.2				

Table 5.1 Differences in student mean scores on ELLI dimensions

The eta squared of each dimension corresponded with what the p value revealed as the effect size of ‘learning relationships’ was the smallest (.009) while the effect size of ‘meaning making’ was the largest (.065) (Meaning making: .065 > changing and learning: .057 > strategic awareness: .044 > creativity: .036 > critical curiosity: .026 > resilience : .019 > learning relationships: .009). The largest discrepancy between the students in University A and B was the dimension of ‘meaning making’ while the smallest was ‘learning relationship’. Referring back to comparisons of the pie charts, the T-test results indeed revealed that the students in University B had stronger learning profiles than those in University A and they particularly stronger in the ‘meaning making’ dimension.

In relating the above findings to the findings of the few ELLI studies that have been carried out, I can only say that one particular finding is consistent with Liu (2007)'s study with Chinese teenage and adult participants. Most students in my study and her study showed a particularly low level of resilience. She speculated that such low level resilience was a reflection of the negative effects resulted from high-stakes testing and the test-oriented culture. Students are likely to experience pressure and feel like escaping, in a culture where parents have high expectations of their children to glory the family, and where people give honour to the successful but look down at the defeated or failed. However, these were her speculations or observations that could not be completely verified. As what Deakin Crick et al. (2004) have pointed out, the ELLI results show only 'how students feel about themselves rather than representing any objective measure' (p.260). The only certain finding as demonstrated above was that students from University B had stronger learning power than those from University A. Therefore, it was more appropriate to relate individual ELLI results to their interviews than to the findings of other studies.

5.5.4 The learning profiles of the interviewees

Before going into detail regarding the ELLI profiles of the student interviewees and how their profiles can be related to their perceptions, I will first explain what the learner ELLI profile looks like and what information it provides.

Unlike the pie charts for the university profile, each student profile would be presented in the form of a 'spider diagram', as explicated in Figure 5.3. The seven axes that radiate from the centre, numbered 1 to 7, represents the 7 dimensions in the order of: changing and learning, critical curiosity, meaning making, creativity, strategic awareness, resilience and lastly, learning relationships. Each axis is cut by three points,

indicating three sections of mean scores, with 100 corresponding to ‘very much like me’ in the ELLI questionnaire, 67 to ‘quite like me’, 33 to ‘a little like me’ and 0 to ‘not at all like me’. The line in red indicates each student’s mean scores on each dimension, and also reveals his or her strength and weakness as the student sees himself or herself. In the sample shown in Figure 5.3, the mean scores of the person in the first dimension, ‘changing and learning’ reveal that she sees herself as a learner who is rather strong at evolving as a learner and expanding her ability to learn. Since there are seven scores in the spider diagram, the web it forms is of the heptagonal shape. In the example below, the dent on the left shows that the learner is weaker on dimension 6, which is ‘resilience’. If a student does not see herself as weaker on any dimension, the web profile forms will look like a straight heptagonal shape. In addition, how strong the person sees herself as, as a learner, can be shown by how big the shape is.

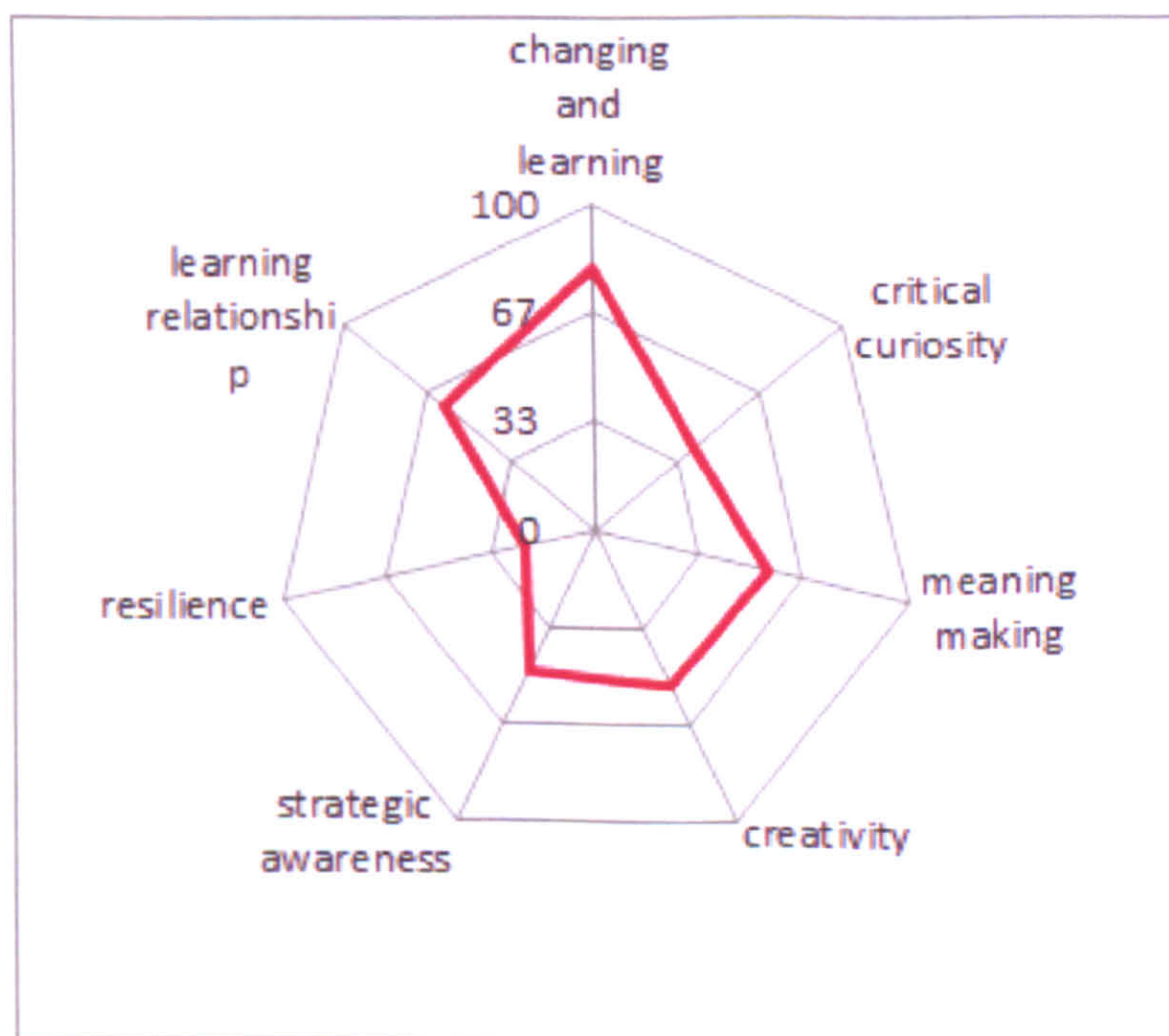


Figure 5.3 Sample spider diagram

5.5.5 Learner’s perspective

Student’s perceptions of the impact of the graduation requirement on them and their learning could be grouped into four groups, revealing different degrees of impact. The first two groups consisted of students who did not perceive much impact (5.6.5.1) while

the last two groups consisted of those who experienced anxiety and stress towards the graduation requirement (5.6.5.2).

5.5.5.1 Students who perceived little impact

The first and second group of students did not think that the graduation requirement impacted strongly on them. The first group were more active learners who were more concerned with their own learning, while the latter were less active learners, who were not interested in learning the language.

The first group consisted of students who took learning English on their own hands and were not much influenced by the requirement. Five students in University B and two students in University A were categorised in this group. What follows is a description of their ELLI profiles, along with their views on the impact of the requirement.

The seven students in this group, Brad, Bridget, Bianca, Blair, Brenda, Archer and April had comparatively stronger ELLI learning profiles than the other student interviewees.

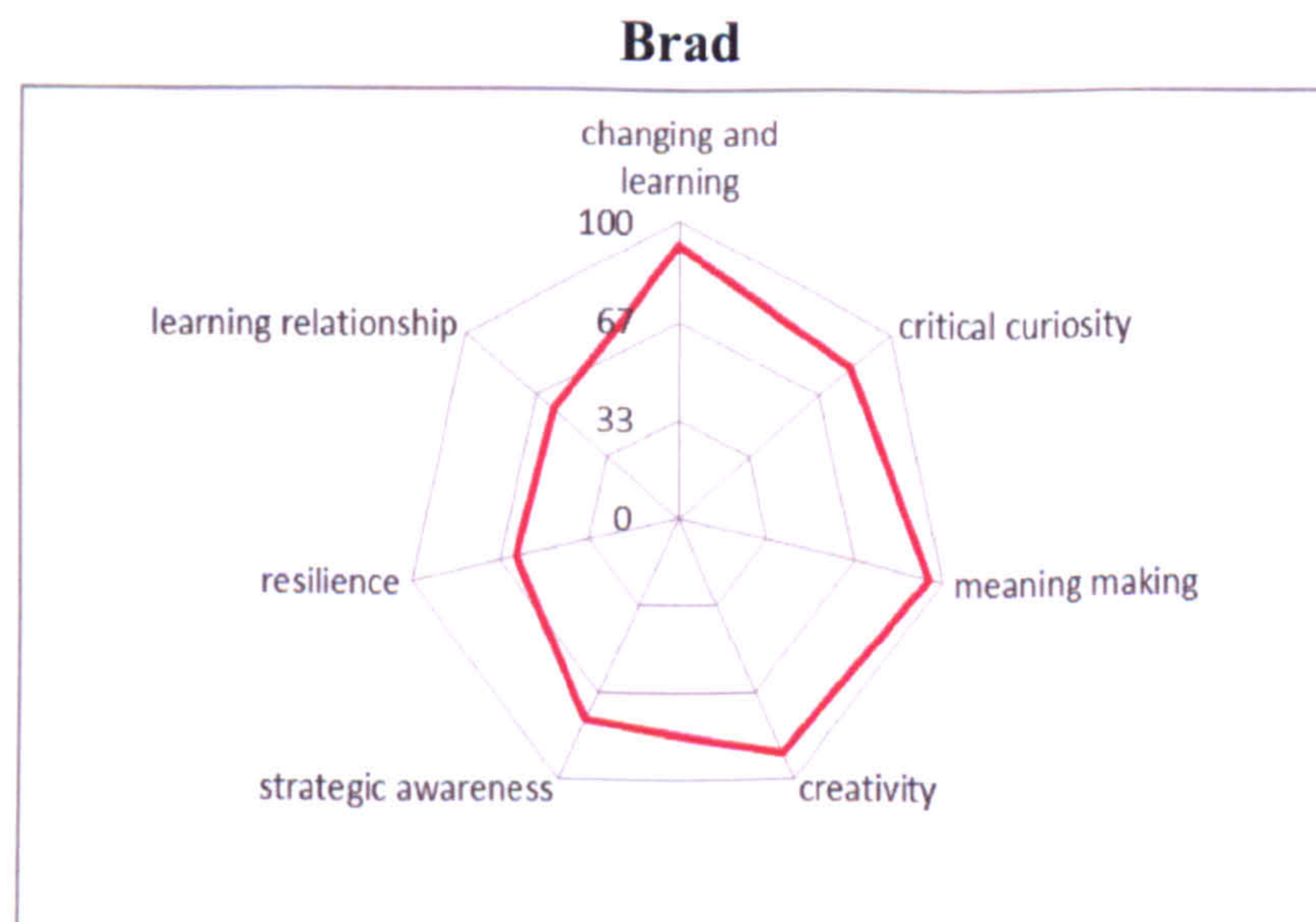


Figure 5.4 Brad's Learning Profile

Bianca

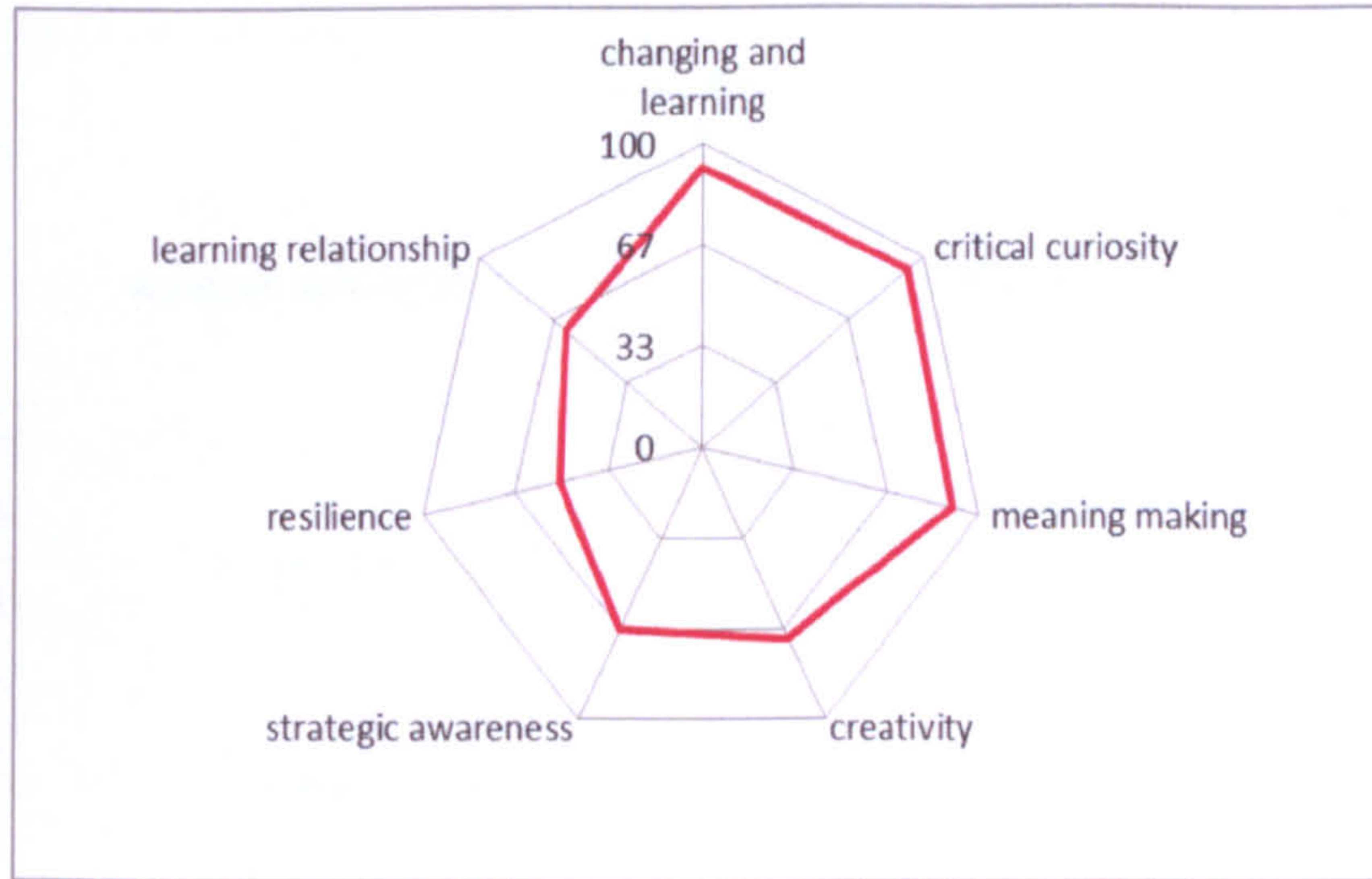


Figure 5.5 Bianca's Learning Profile

Blair

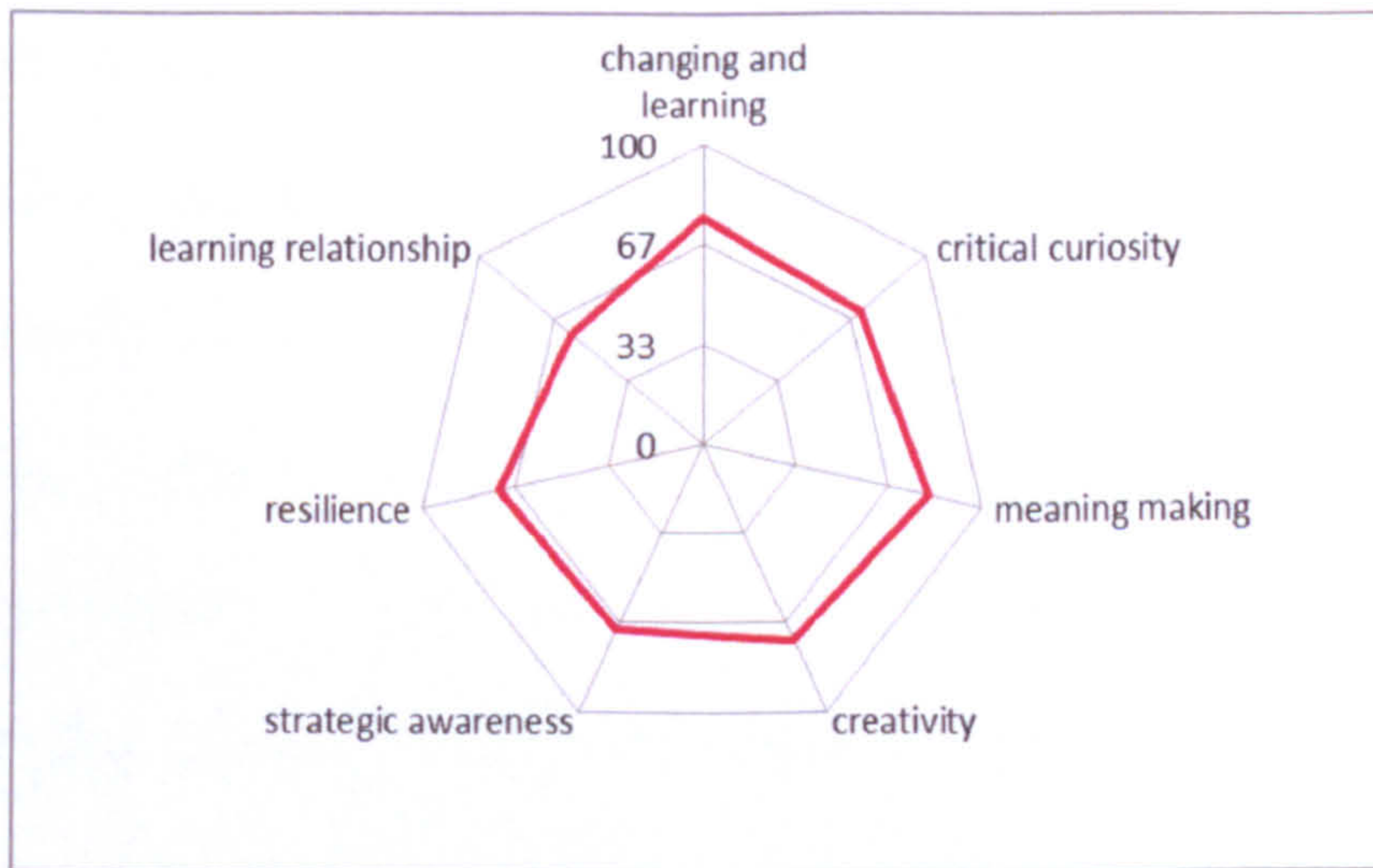


Figure 5.6 Blair's Learning Profile

Bridget

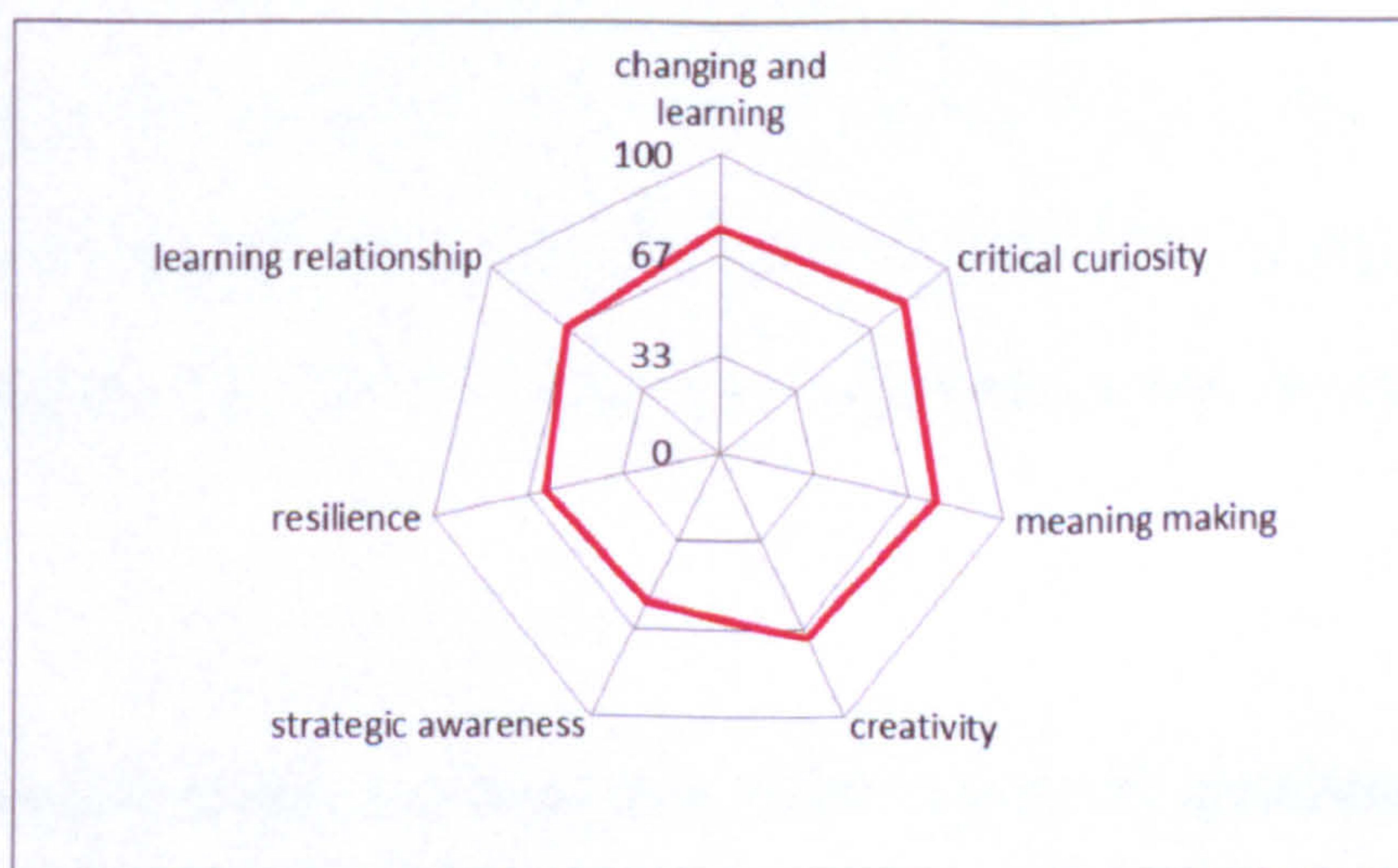


Figure 5.7 Bridget's Learning Profile

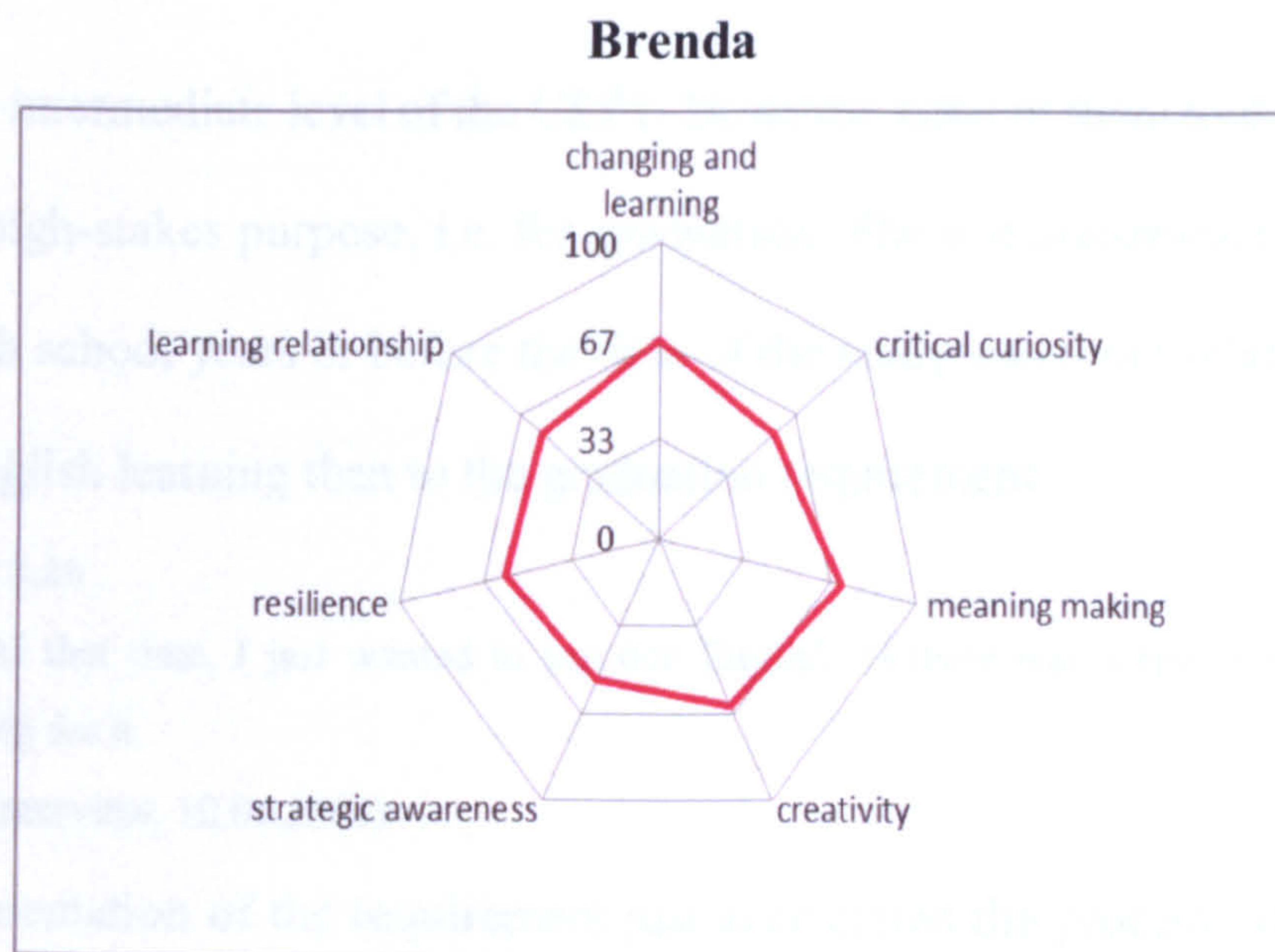


Figure 5.8 Brenda's Learning Profile

The above five diagrams suggest that all five students thought of themselves as relatively effective learners in all seven dimensions. Brad's learning profile was the strongest among the 18 ELLI profiles from both universities. He reported himself as greatly enjoying expanding his ability to learn, challenging the received knowledge, to be disposed to being aware of his own learning as well as being very creative. Bianca was similar to Brad in the first three dimensions but was weaker than Brad in seeing herself as creative. The others were slightly weaker than Brad on all the above aspects. Brad's relative weaknesses, with scores as high as or higher than most profiles, were on the dimensions of strategic awareness, resilience and positive relationships. Blair was slightly stronger than the others on 'resilience' while Bianca was the weakest among the five. As seen from the size of the red-lined web, Brenda's learning profile was the weakest among the five, but still displayed considerably strong learning power on all dimensions.

Most of them claimed that they were little influenced by the graduation requirement. Brad, Bianca and Bridget were the ones who had already fulfilled the requirement by

passing the intermediate level of the GEPT. However, none of them took the test for this particular high-stakes purpose, i.e. for graduation. The test preparation and test taking in their high school years or before the time of the study was more related to their own goals in English learning than to the graduation requirement.

Extract 5.26

Brad: At that time, I just wanted to practice English. If there was a test, I would think about preparing for it.

(Brad, interview, 10.06.2008)

The implementation of the requirement just accelerated the process, as explicated by Bianca.

Extract 5.27

Bianca: It just made me take the test earlier. That's all.

(Bianca, interview, 10.06.2008)

Some of them did not think that they were influenced by the requirement because their English proficiency level exceeded the threshold stated in the requirement. Blair looked down on the first stage of the intermediate GEPT as the level for high school graduates (Extract 5.12) while Brad directly pointed out that a higher level of the GEPT as the threshold level would have made him motivated to learn more English.

Extract 5.28

Brad: If it (the requirement) sets the level at high-intermediate, then I think there will be influence.

Researcher: The level it sets now is the intermediate, so you don't (think there has been influence).

Brad: Yes.

(Brad, interview, 10.06.2008)

This was probably why most of the abovementioned students were nonchalant towards the requirement, as has been mentioned previously. They did not perceive the standard that University B has stated in the graduation requirement as high stakes, which might

be an important factor that made them think that they were not influenced by the requirement.

In fact, they were more concerned with their own opportunities to learn English, such as the amount of courses available and the contents of the courses tailored to their needs. Brenda, Blair and Brad were the only students who thought that there were insufficient English courses for non-English majors.

Extract 5.29

Researcher: So, do you think that there are too few English courses?

Brenda: Yes.

Researcher: You think that there are too few. Then, how many (courses) do you think will be enough for you?

Brenda I think...

Researcher: Do you think you should have English courses for all four years?

Brenda: There's no need for all four years. But I think having English course for only the first year is just too few and there are still too few English courses, after adding the test-related one (referring to remedial course).

(Brenda, interview, 10.06.2008)

Blair considered that having English courses continuously throughout the four years would be best for them, but further argued that what the courses provided should be tailored to the students' needs.

Extract 5.30

Blair: I have to say that the course is very important. If we have these kinds of courses, then I think I'd rather not have any courses.

Researcher: So, you'd like to have different types of courses for you to choose from?

Blair: Yes. Some students prefer conversation. Others will like reading. I think it's good this way. There should be quite a lot of people who want to increase their English proficiency.

(Blair, interview, 10.06.2008)

Blair's comments were evidenced by the others, with Brad stating his preference on

courses of academic English, and Bianca and Bridget on group discussions of interesting issues in their English courses. These students cared less about the graduation requirement than what they could get out of the courses they were provided or wish to be provided. Hence, what concerned them was not so much the implementation of the graduation requirement but the development on their own English proficiency.

There were two students in University A who were also categorised in the first type: Archer and April. Although there was no graduation requirement for non-English majors in their university, both of them did not think the requirement would exert much influence on them either even if it were to be implemented. Their ELLI profiles were as follow:

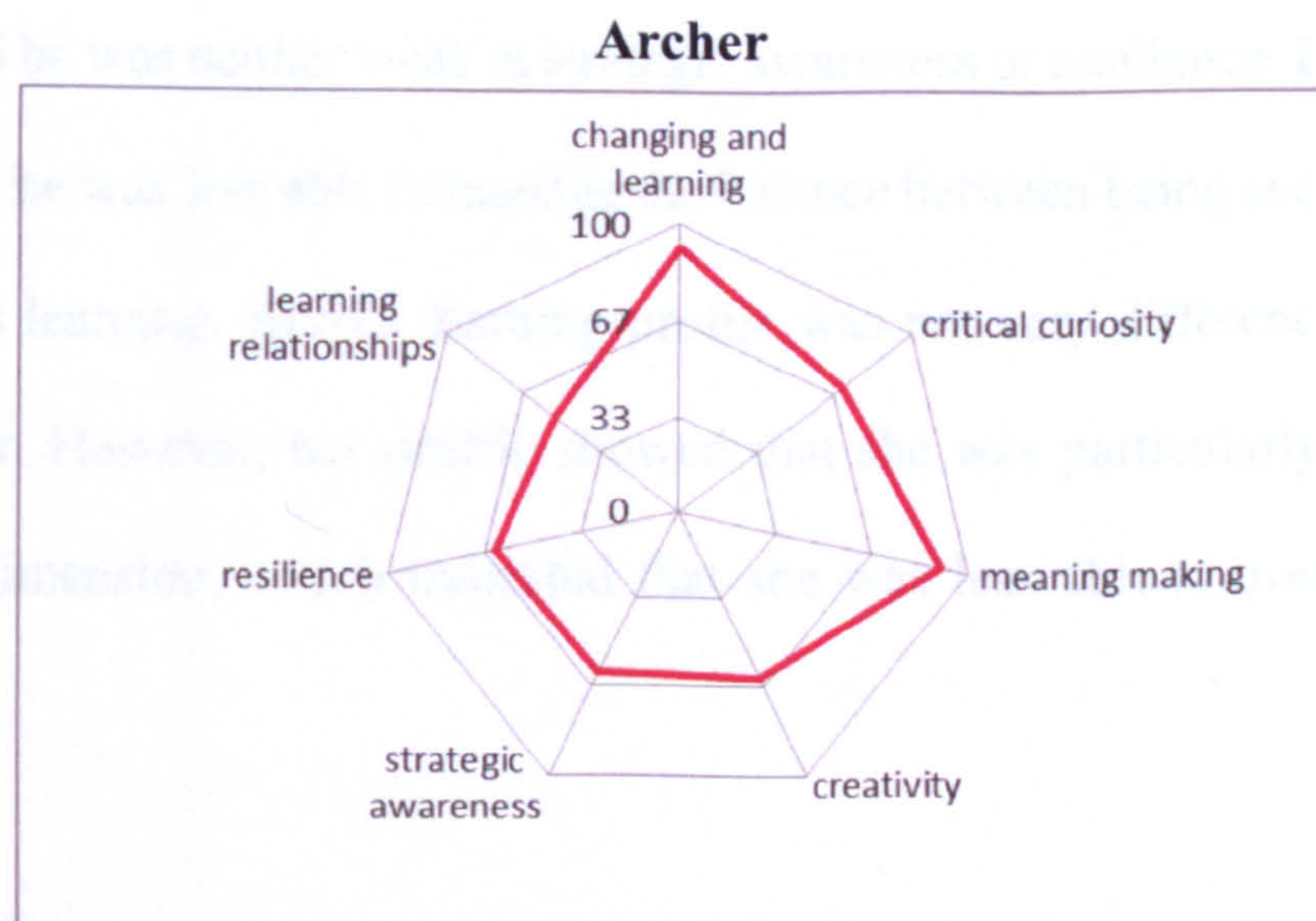


Figure 5.9 Archer's Learning Profile

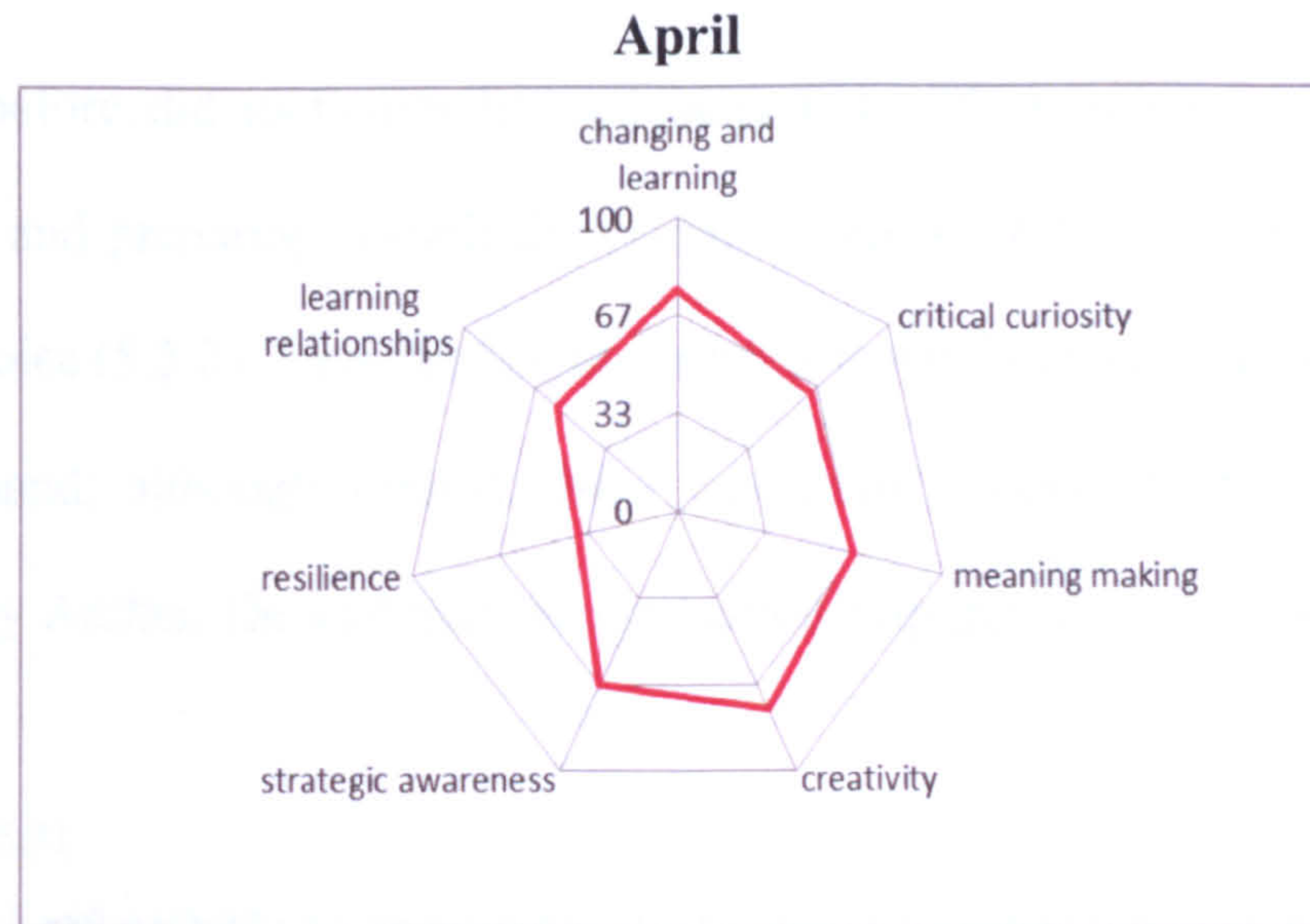


Figure 5.10 April's Learning Profile

Archer's profile shows that he considered himself as comparatively effective in most of the dimensions, which were very similar to the abovementioned five profiles. He had an especially strong sense of being able to change and grow over time and he was highly interested in seeing how new learning fit within the big picture he had established from the past learning experiences. He was not passive in active dimensions of curiosity and creativity and he was neither weak at strategic awareness or resilience. However, he did point out that he was less able to manage the balance between being sociable and being private in his learning. April's learning profile was not very different from the other profiles either. However, her profile showed that she was particularly weaker in the 'resilience' dimension, which indicated that she was less able to overcome learning frustrations.

As with those students from University B in this category, Archer and April considered that their personal learning goals instead of the need to fulfil such graduation requirement would motivate them to prepare for English proficiency tests. Both of them mentioned the necessity of taking English proficiency tests for their future academic plans, and said that they would do so without the graduation requirement. April, as I

mentioned before, did not favour the implementation of the requirement as she believed in studying and preparing oneself for external English proficiency tests out of one's personal choice (5.3.2). Thus, for her, the requirement did not have much influence. On the other hand, although limited, there was some impact of the requirement, as perceived by Archer. He said that the graduation requirement would make him work harder.

Extract 5.31

Archer: I will probably go on for further studies in the education field and the certificate of an English proficiency test is a necessity. The graduation requirement, if implemented, is not what pushed me to take the test but if there were a requirement, I will probably work even harder.

(Archer, interview, 7.05.2008)

The other similarity between Archer, April and the abovementioned 5 students was their personal concerns with opportunities to learn English. This could be explicated by Archer's effort in taking TOEIC preparation course in private language schools and also by his comments on the English course he was taking as not challenging enough.

Extract 5.32

Archer: The current course contents are more related to daily use, which is good, but for me, it is slightly too easy and I prefer the teacher using all English.

(Archer, interview, 7.05.2008)

The above analysis of the 7 students listed in this group suggests that their strong learning profiles corresponded to what they revealed themselves as an English language learner in the universities. Their ELLI learning profiles show that they were relatively effective learners in most dimensions. From the interview data, it may also be seen that they were active learners of English, since they learned English not just for the purpose of fulfilling the graduation requirement but for their own goals and plans. In addition, they were more concerned about additional opportunities to learn English such as increased number of courses for non-English majors and the variety of courses

or teaching contents tailored to their needs. They indeed seemed to take more control of their learning instead of being coerced by an external force. The graduation requirement would only make them work harder to reach the goals they had set up for themselves. Thus, for them, the implementation of the graduation requirement had little impact on them and even if there was impact, it was positive impact, e.g. working harder or doing more on their own learning in addition to the compulsory courses provided by the universities.

The second group consisted of only one student among the 18, Alvin. As with the first group, he perceived little impact of the graduation requirement on him and his learning. However, the reason was his lack of interest in learning the English language.

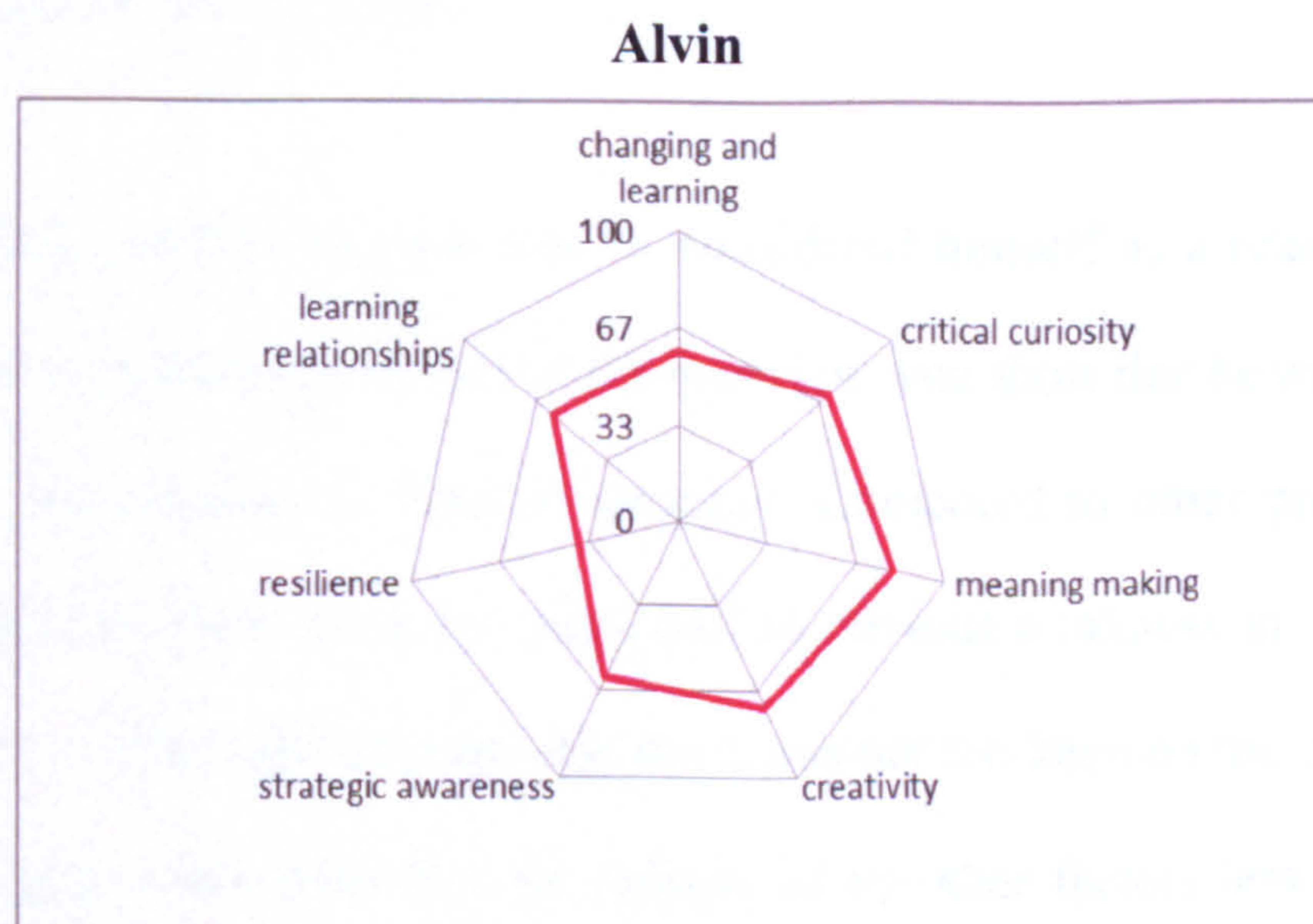


Figure 5.11 Alvin's Learning Profile

The spider diagram above suggests that Alvin rated himself as moderately good in six dimensions, with higher scores on the active dimensions of curiosity, meaning making and creativity and slightly lower scores on strategic awareness, changing and learning and learning relationships. One dimension that stood out among the others was the relative weakness he considered himself to have, namely, resilience.

With a profile not much different from April, Alvin was also similar to April in his belief that one's English learning was up to oneself and the requirement would have little influence over personal choice. However, unlike April, he was not keen on learning English. It was not what he needed, and thus, he did not think the implementation of such a requirement would be able to make him enthusiastic about English learning. Alvin confessed that only if his future work or career needed him to provide a proof of English proficiency would he then prepare himself for an English test.

Extract 5.33

Researcher: Is there anything that will make you want to take an English proficiency test?

Alvin: Work. If my job sees that as a must, then I will.

(omitted)

Researcher: Or you think that as long as you don't want to learn, the curriculum, the teachers, the requirement, nothing can make you want to learn more English?

Alvin: Um, yes.

(Alvin, interview, 23.05.2008)

Alvin's learning profile suggests that he considered himself as a relatively effective learner in most dimensions. However, his interview data show that he was not an active learner towards learning the English language. Compared to other profiles in above mentioned groups, both Alvin and April had an obvious weakness in the dimension of 'resilience'. This might explain why April was not too keen on the compulsory test taking, and why Alvin refused to be influenced by other factors less important than work on English learning and test taking.

5.5.5.2 Students who perceived stress and anxiety towards the requirement

As distinct from the first two groups, the groups in this section included students who suffered from the stress accompanied by the graduation requirement. However, the two

groups differed in whether they were supportive of implementing the graduation requirement or not.

The third group consisted of students (Bess, Abel, Anson and Aiden) who were willing to face the challenge, even though they felt pressured by the requirement.

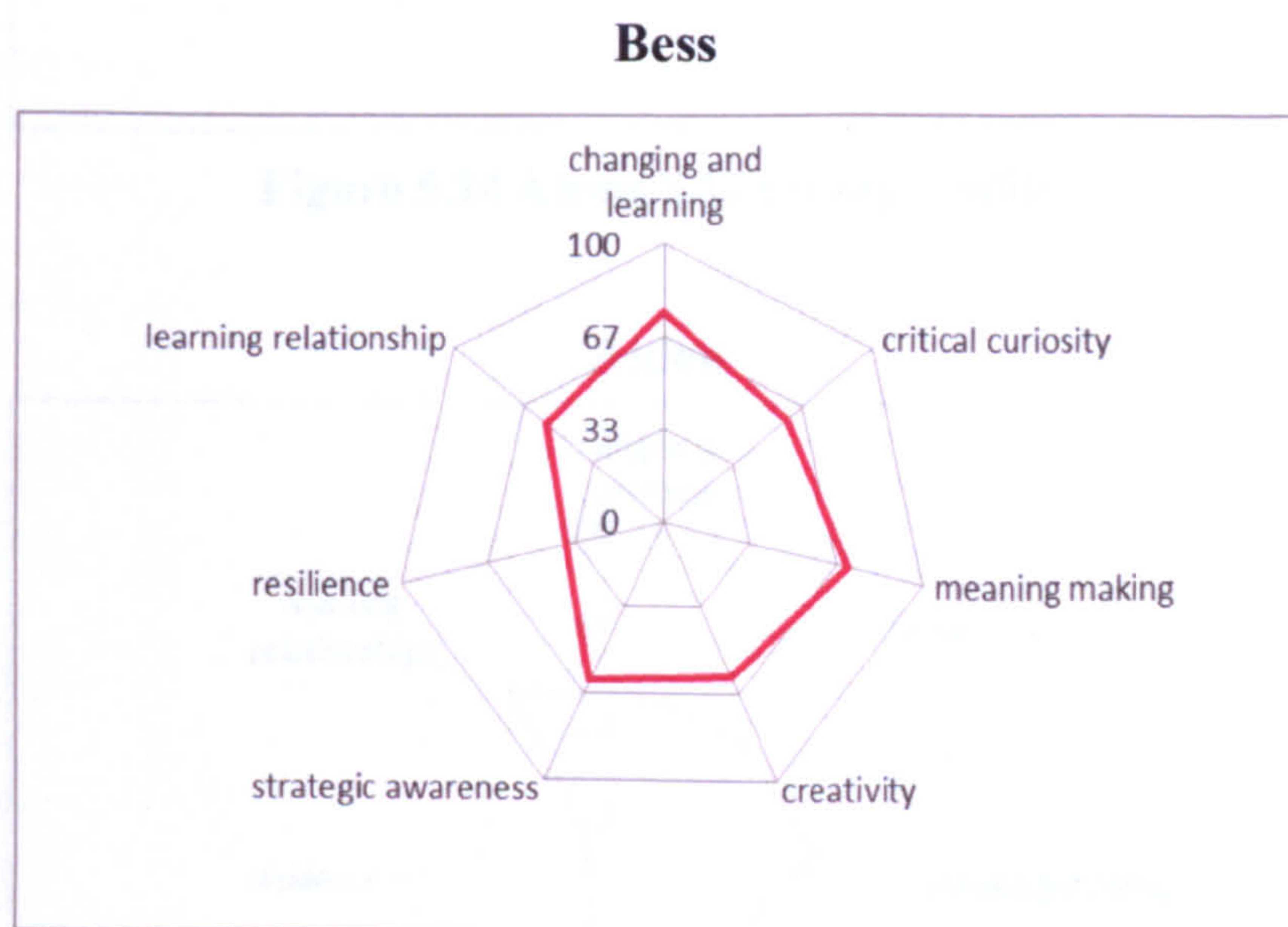


Figure 5.12 Bess' Learning Profile

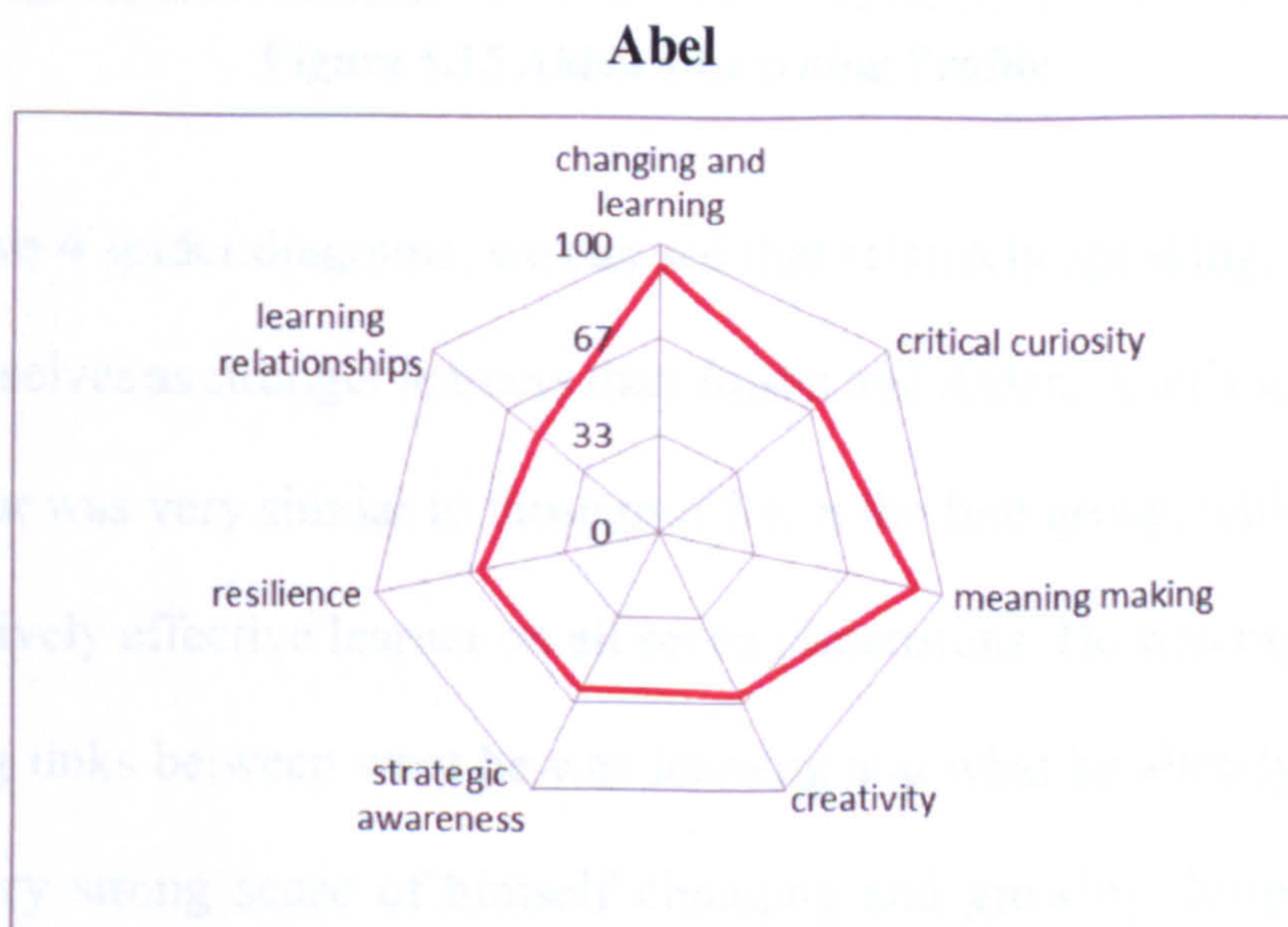


Figure 5.13 Abel's Learning Profile

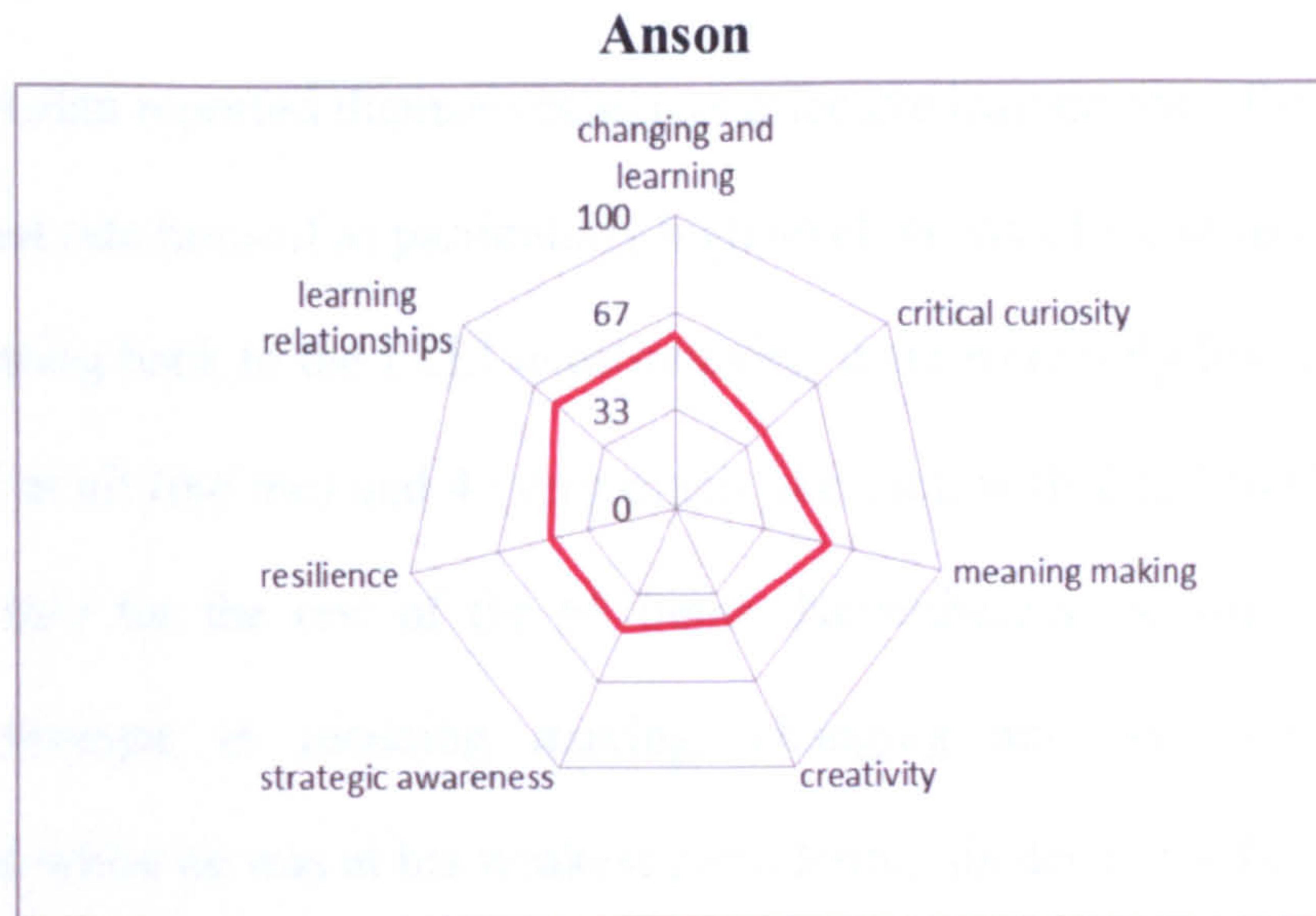


Figure 5.14 Anson's Learning Profile

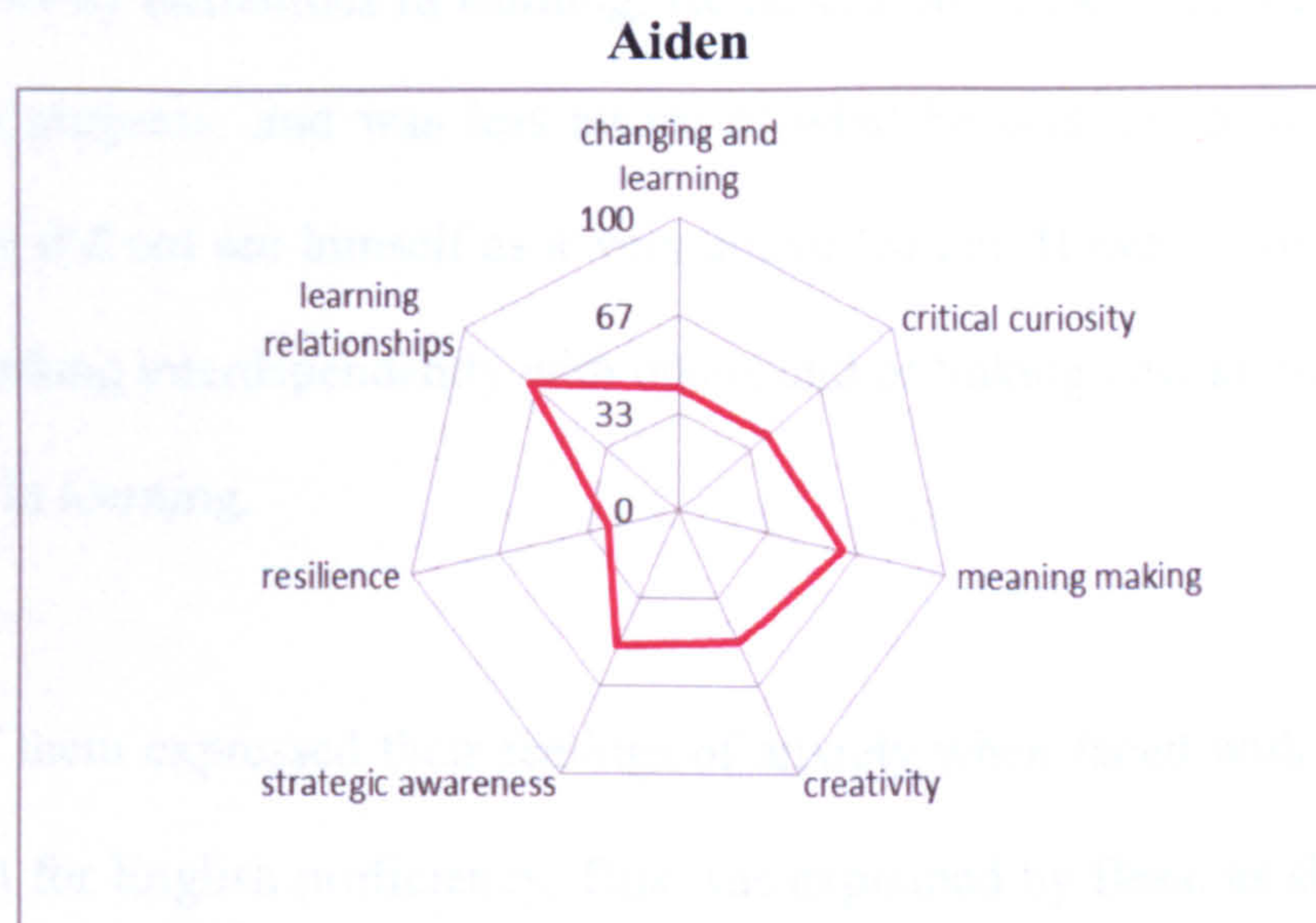


Figure 5.15 Aiden's Learning Profile

From the above 4 spider diagrams, we can see that relatively speaking, Bess and Abel reported themselves as stronger learners than Anson and Aiden. Abel's strongest profile among the four was very similar to those profiles in the first group, which implies that he was a relatively effective learner on all seven dimensions. He was especially strong at establishing links between what he was learning and what he already knew, and he also had a very strong sense of himself changing and growing during the learning process. Bess had a slightly weaker profile than Abel, with a stronger weakness in resilience.

Anson and Aiden reported themselves as less effective learners than Bess and Abel.

Anson did not rate himself as particularly high level on any of the seven dimensions. In fact, in referring back to the ELLI questionnaire, there were only four items which he rated 1 (not at all like me) and 4 (very much like me), with 2 (a little like me) and 3 (quite like me) for the rest of the 68 items. Nevertheless, he still saw himself as relatively stronger at meaning making, changing and learning and learning relationships while he was at his weakest considering his desire for finding things out. Lastly, Aiden thought of himself as significantly weak at handling negative feelings brought about by difficulties in learning. He lacked the sense of himself to be able to change and progress, and was less aware of what he was doing in the process of learning. He did not see himself as a very active learner. However, he had more of a sense of working interdependently with others and of linking new knowledge with old knowledge in learning.

All four of them expressed their feelings of anxiety when faced with the graduation requirement for English proficiency. This was explained by Bess, as she was worried that she could not fulfil the requirement in her university.

Extract 5.34

Bess: Stress? Yes, a bit, because...

Researcher: You're afraid you won't pass?

Bess: Yeah, I worry about not being able to pass. It's been a while since I studied English and I know my proficiency level was not as good as before.

(Bess, interview, 10.06.2008)

However, all of them held a positive view on the stress that came from the test-taking and the graduation requirement. Abel did not fear the possibility of not being able to graduate, even if he had to take an English test repeatedly (Extract 5.6). He considered

that the stress from the requirement was constructive to his English learning:

Extract 5.35

Abel: I think it's good to give myself pressure (with the graduation requirement).

(Abel, interview, 15.05.2008)

Anson was also willing to face the pressure from the implementation of the English requirement as long as the required proficiency level was not too high, considering the overall proficiency of students (Extract 5.4).

Both Aiden and Bess felt the need to be compelled by the requirement to learn English. Bess embraced the idea of being forced to learn English because she considered English as very important. Her remarks were shared by Aiden, as illustrated in the following:

Extract 5.36

Researcher: There is no requirement in this university. But if you are in any other universities, do you wish the university to have the requirement?

Aiden: Yes, I do.

Researcher: Would you be afraid?

Aiden: Super afraid (laughs).

Researcher: But you think the stress is necessary.

Aiden: Yes.

Researcher: If you think it's good to have the requirement, will you be forced to study English?

Aiden: Yes.

Researcher: Do you think it will make you want to learn more English?

Aiden: I don't know. Perhaps so. I will have to learn English in order to graduate.

Researcher: So you think students should be forced to learn?

Aiden: Yes.

(Aiden, interview, 23.05.2008)

Aiden's comments suggested that for some students, although they felt anxious towards the graduation requirement, they believed the anxiety was necessary to push them to work harder in learning English.

In relating their ELLI profiles to what they revealed in the interviews, 'resilience' might be a possible indication of the anxiety the students felt towards the requirement. Both Bess and Aiden, who expressed more anxiety and fear for not being able to 'pass', were weaker in the 'resilience' dimension than Abel and Anson. Aiden had the lowest score on 'resilience' among the 18 students and his overt statement of being 'super afraid' might be explained by the fact that he considered himself as significantly poor in resilience.

The above analysis of the four students' learning profiles and their interview data suggests that they sensed high level of anxiety towards the requirement. Those who were significantly weaker on the 'resilience' dimension in their ELLI profiles were more likely to feel a larger degree of anxiety. However, they viewed the anxiety and stress that came along with the requirement as positive and necessary and thus, they welcomed the implementation of the requirement. The way these students handled their anxiety concerning the graduation requirement was similar to what Betty considered as the positive impact the requirement could have on the non-English majors (5.6.1). However, only four students were positive about what they perceived as impact of the requirement. Others were less positive, as illustrated below.

The last group of students were those who felt highly anxious about what the graduation requirement demanded them to do, and who were less supportive of the implementation in their universities. The group consisted of two sub-groups. The former included student who did not support the implementation because they disliked the feeling of pressure and also the 'troubles' they had to overcome to graduate: Billy, Bonnie and Andrew. The latter, including Bryan, Aaron and Alex, were those who approved of the idea of English benchmark but did not support the implementation

because they lacked the confidence to face such challenge..

The learning profiles of the first sub-group were listed below.

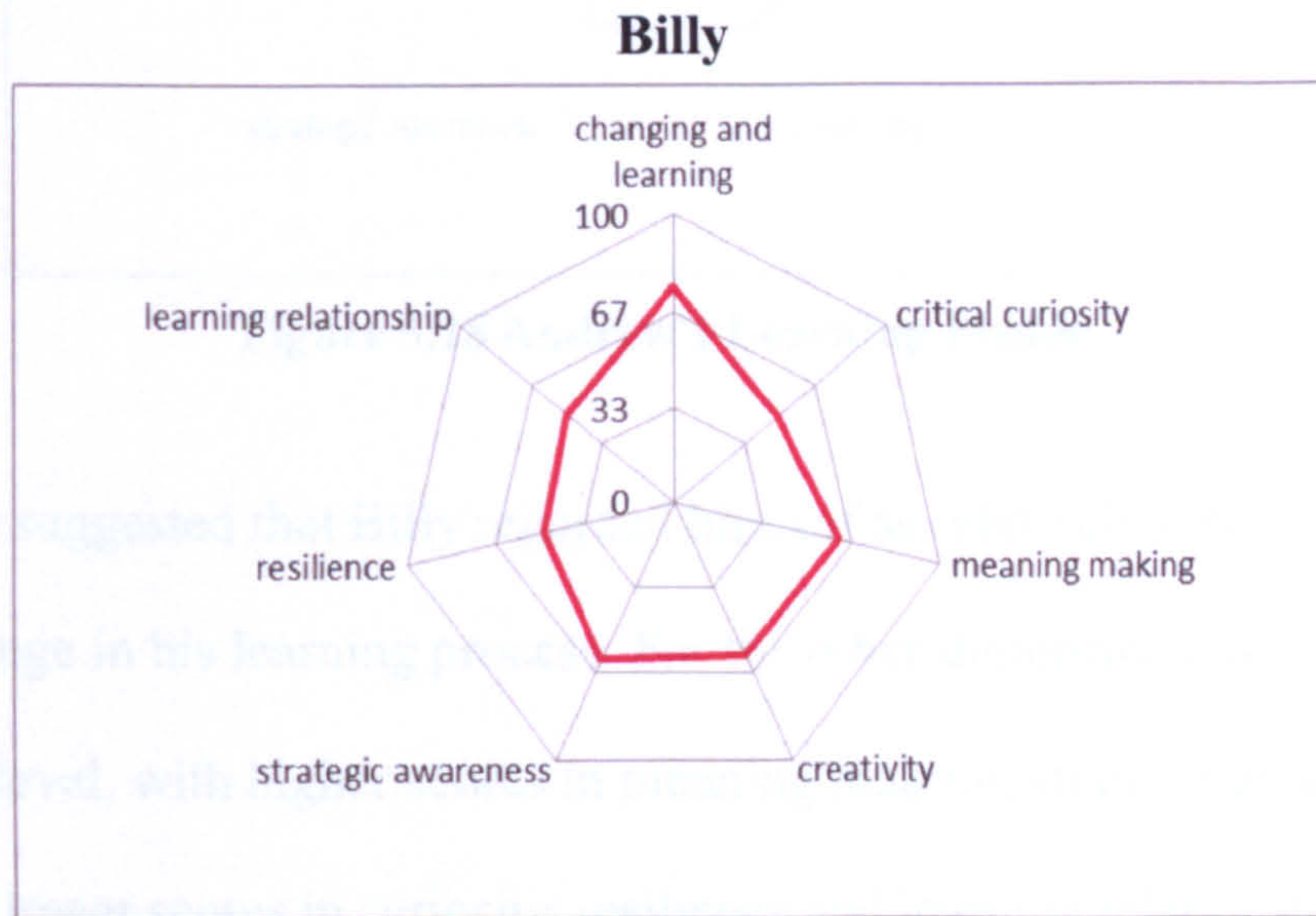


Figure 5.16 Billy's Learning Profile

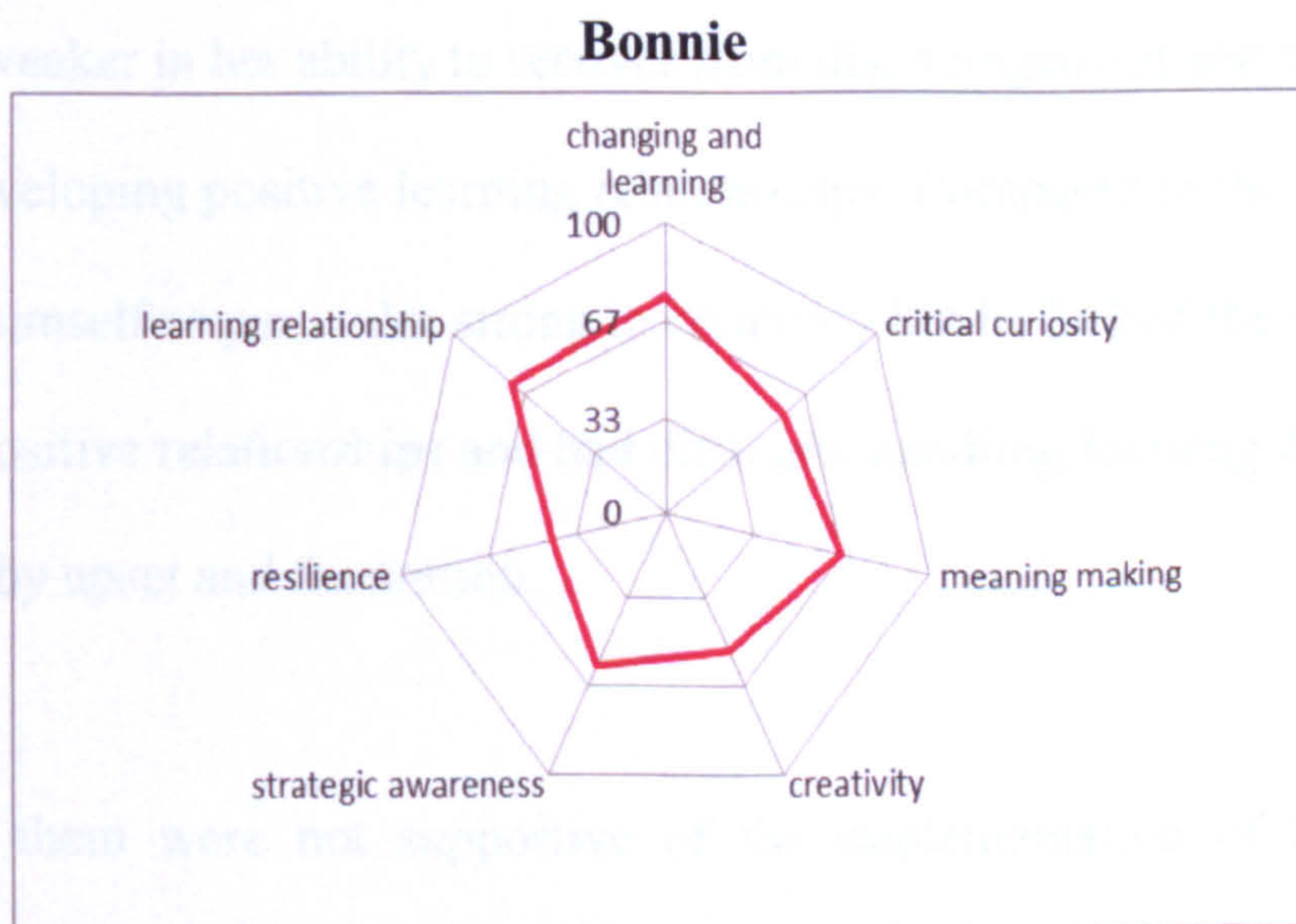


Figure 5.17 Bonnie's Learning Profile

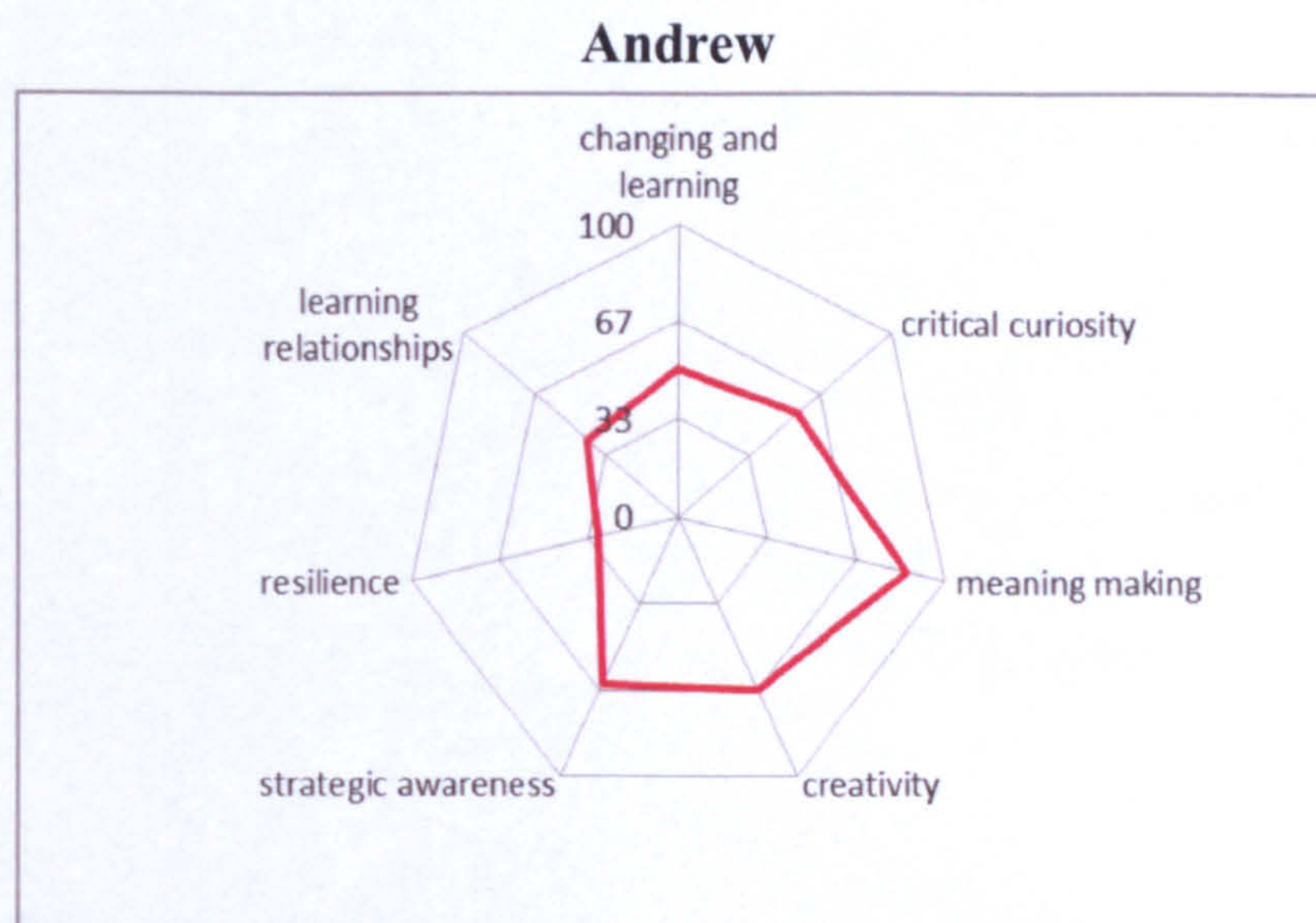


Figure 5.18 Andrew's Learning Profile

Billy's profile suggested that Billy regarded himself as relatively better at being able to adapt and change in his learning process. For the other dimensions, he rated himself at the moderate level, with higher scores in meaning making, strategic awareness and creativity and lower scores in curiosity, resilience and learning relationships. There was not much difference between Bonnie's profile and Billy's, except that she reported herself to be weaker in her ability to recover from discouragement and failure but stronger in developing positive learning relationships. Compared to the above two, Andrew saw himself as particular strong in creativity but he lacked the sense of maintaining positive relationships and had difficulty handling learning difficulties that accompanied by upset and frustration.

The three of them were not supportive of the implementation of the graduation requirement, because the requirement would make them do things they did not feel like to. For Billy, English was not something he needed and a certificate of English proficiency did not fall within his picture of the future. Thus, he disliked the idea of taking an English test solely for the purpose of graduation (Extract 5.10) and felt stressed by having to do so. Similarly, Andrew did not think English was a necessity

and he did not welcome excessive pressure the requirement would bring onto him.

Extract 5.37

Andrew: I believe that people are lazy and they always need to be pressured to learn to grow. If we really need to improve the quality (meaning English proficiency) of university students, the implementation of the requirement is useful. They (students) will learn in order to graduate. It will force them to learn.

Researcher: You just don't think it (the requirement) is good for you?

Andrew: No, it (the level) is too high to reach.

Researcher: You don't like being compelled to learn?

Andrew: I really think that I have bad English proficiency.

Researcher: So you'd feel greatly stressed?

Andrew: Yes.

Researcher: Will it make you less willing to learn English?

Andrew: Probably.

(Andrew, interview, 15.05.2008)

It is interesting to note that Andrew contradicted himself in the above extract. He believed that university students, who were not very active learners, should be compelled by the graduation requirement to learn English. Nevertheless, he personally rejected the requirement to be imposed on him as he thought his English proficiency was too low for him to reach the 'high' benchmark required by the university.

Bonnie showed her unwillingness to comply with the regulations of the graduation requirement of University B. She complained about the troubles she had to go through for English learning and for receiving her degree.

Extract 5.38

Researcher: Do you think the graduation requirement influences you in learning English?

Bonnie: I think it is troublesome.

Researcher: Why is it troublesome?

Bonnie: Because, I don't know why. I used to like English in my junior and senior high years. But I feel annoyed to have to read English in the university, especially when reading the books written in the original language (English.). It is a great stress.

Researcher: But it is related to your own field.

Bonnie: Yes. But I dislike that a lot. I just feel annoyed when I see English. So it is very troublesome because I even have to prepare for a test to graduate.

Researcher: So you feel stressed?

Bonnie: Yes, yes, yes.

(Bonnie, interview, 10.06.2008)

Her complaints were, in fact, based on how stressful she felt concerning the efforts she needed to make to learn English. The need to read books written in English already made her feel stressed and the fact that she was required to provide an evidence of English proficiency was even more stressful. The impact of the graduation requirement on her was the stress that came along with the need to prepare herself for an external English proficiency test. The stress was the trigger for her negative attitude towards the requirement.

The above three students all showed a rejection of the implementation of such requirement because one of its 'side effects' was the excessive stress to have to take an English proficiency test for the sake of graduation.

On the other hand, the other group of students were not supportive of implementing the requirement because they lacked confidence in their English because of their previous failure in English language tests. The last three students are categorised in this group: Bryan, Aaron and Alex.

Bryan

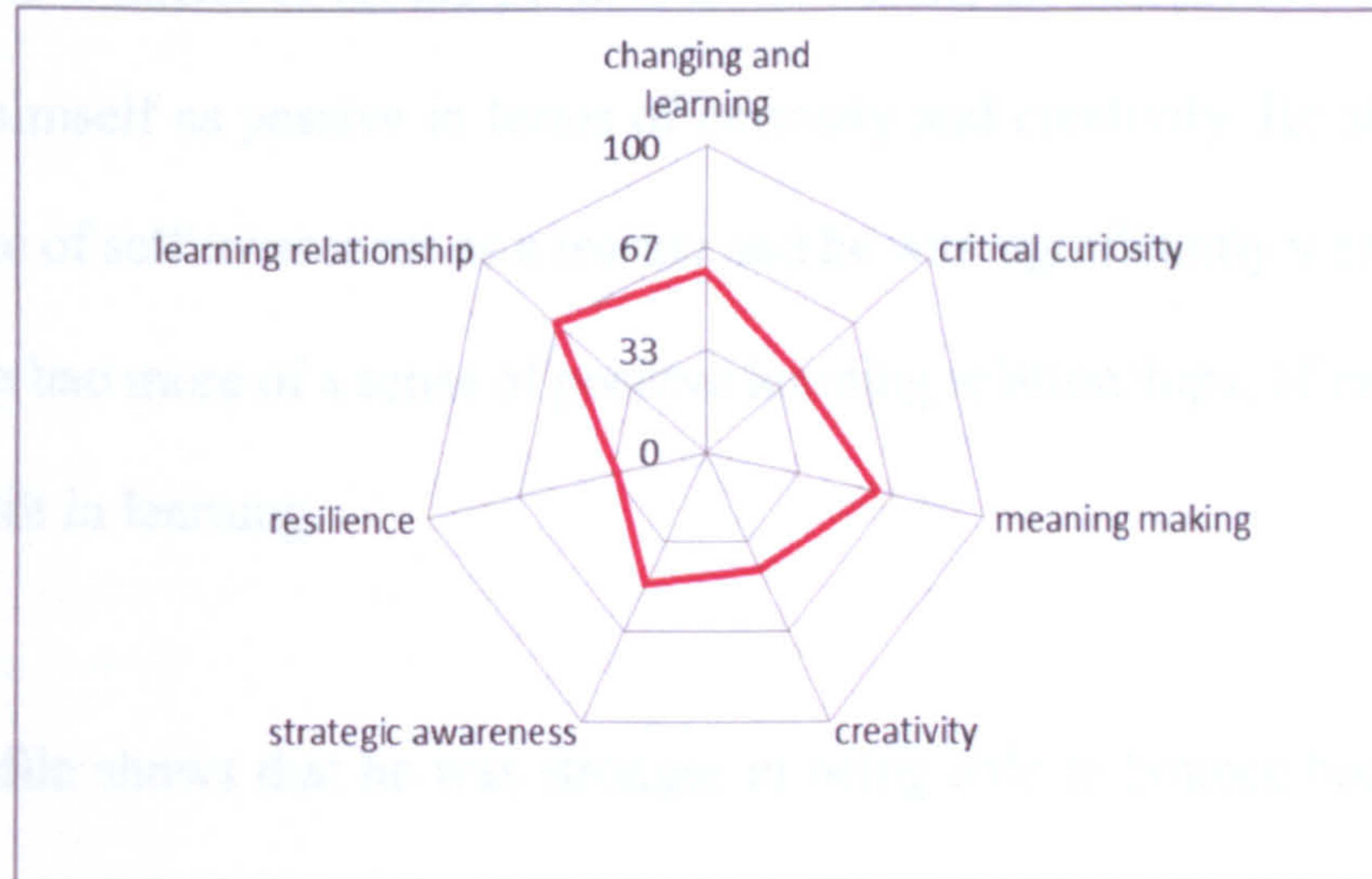


Figure 5.19 Bryan's Learning Profile

Aaron

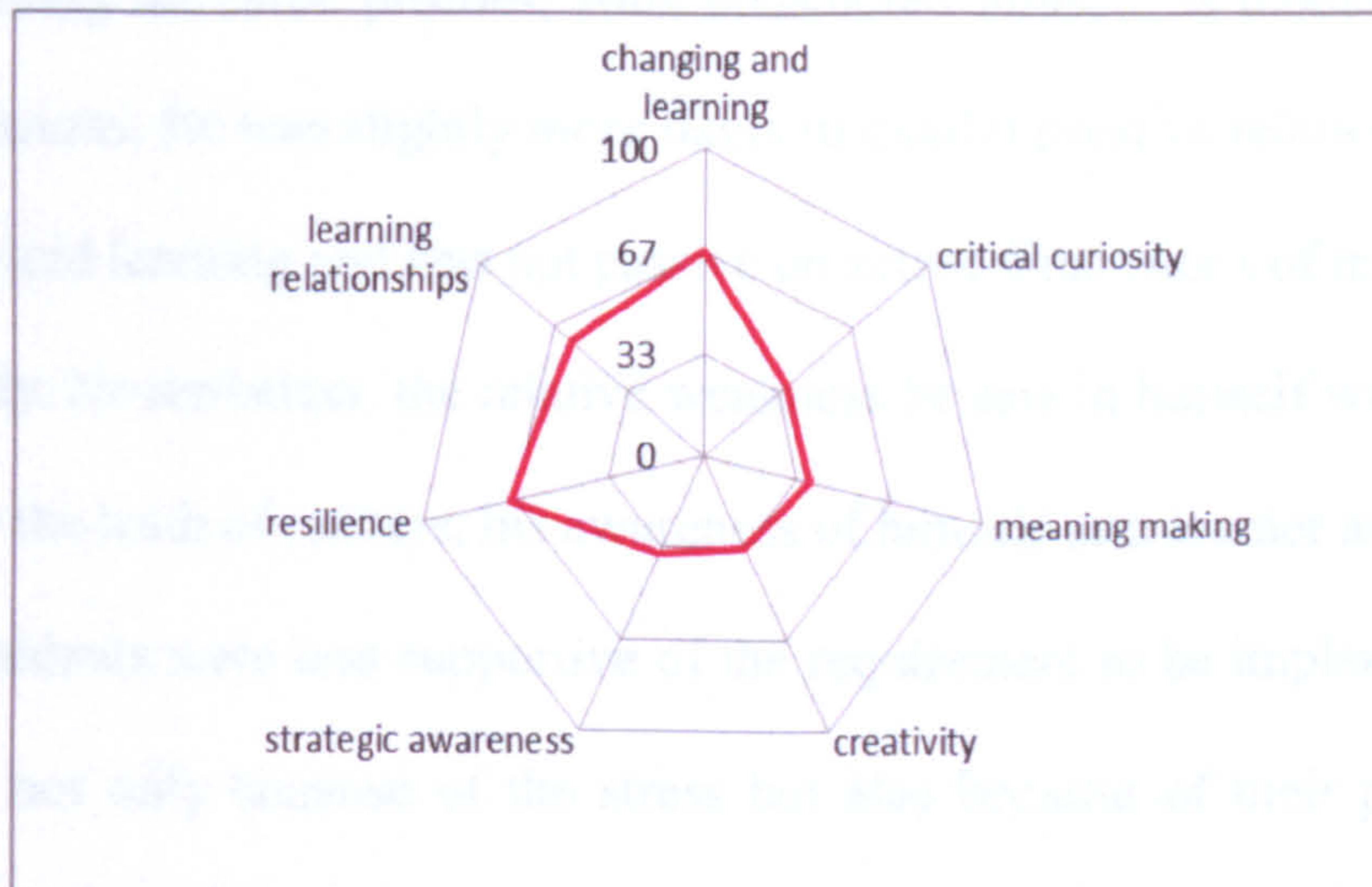


Figure 5.20 Aaron's Learning Profile

Alex

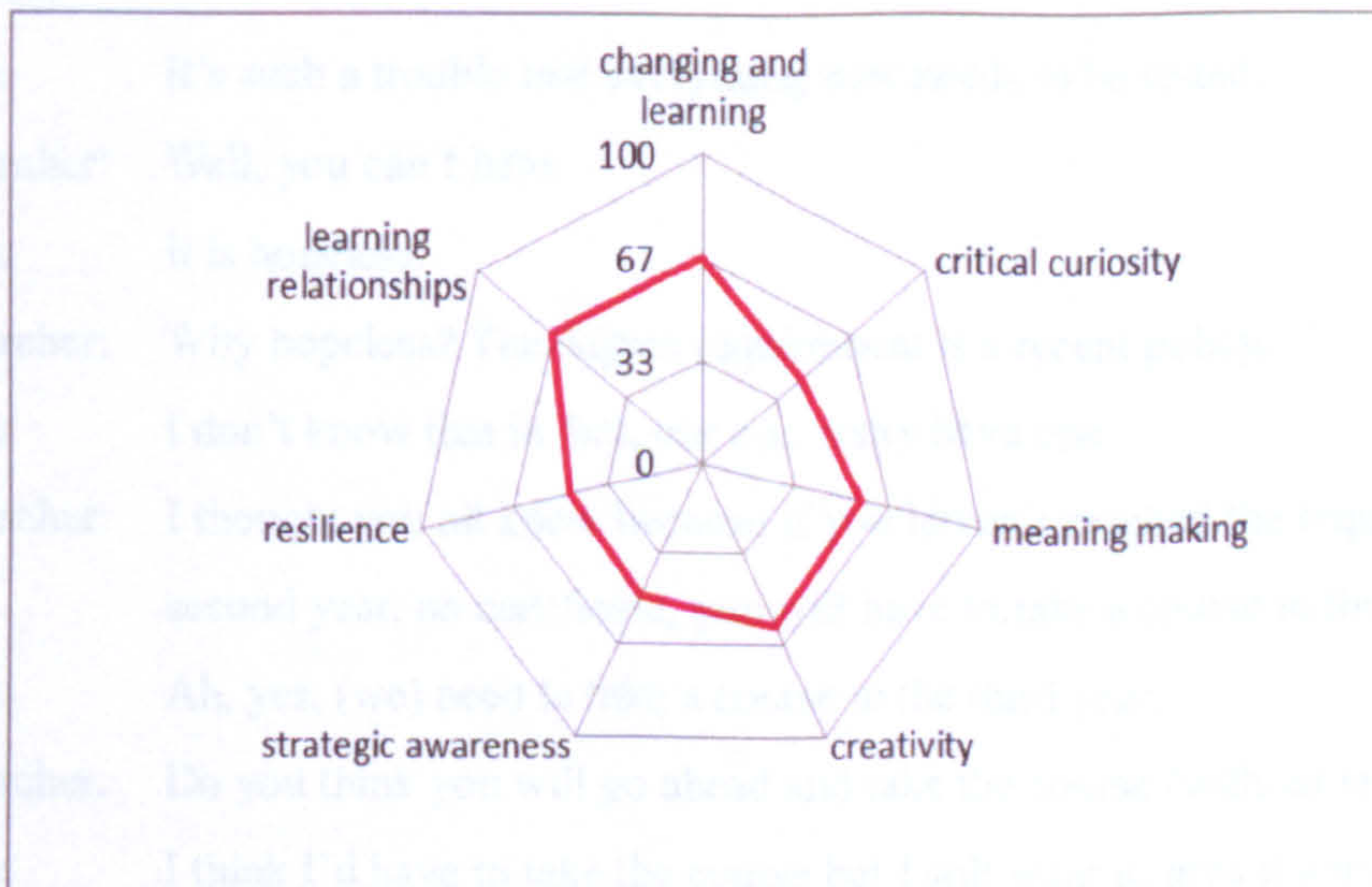


Figure 5.21 Alex's Learning Profile

Bryan has the weakest of all the profiles in University B. His profile suggests that he thought of himself as passive in terms of curiosity and creativity. He also exhibited a weaker sense of self-awareness as a learner and he was significantly weak at resilience. However, he had more of a sense of positive learning relationships, of meaning making and of growth in learning.

Aaron's profile shows that he was stronger in being able to bounce back from failure and anxiety than Brian but weaker in the active dimensions of curiosity, meaning making and creativity and also he had less awareness of his own learning. Being the strongest among the three profiles, Alex considered himself as moderate for all the seven dimensions. He was slightly more likely to exhibit positive relationships, a sense of changing and learning and was not passive on active dimensions of meaning making and creativity. Nevertheless, the relative weakness he saw in himself was his curiosity at getting to the truth of matters, his awareness of himself as a learner and resilience.

The three students were less supportive of the requirement to be implemented in their universities not only because of the stress but also because of their past failures in English tests. Their disappointing experiences made them fear that they would fail to reach the graduation requirement, as illustrated in Bryan's interview.

Extract 5.39

Bryan: It's such a trouble that everything now needs to be tested.

Researcher: Well, you can't help.

Bryan: It is hopeless.

Researcher: Why hopeless? The degree requirement is a recent policy.

Bryan: I don't know that in fact, our university have one.

Researcher: I thought you all knew because if you haven't reached the requirement in your second year, no certificate, you will have to take a course in the third year.

Bryan: Ah, yes, (we) need to take a course in the third year.

Researcher: Do you think you will go ahead and take the course (without test taking)?

Bryan: I think I'd have to take the course but I still want to give it a try.

- Researcher: Why do you lack confidence in yourself?
- Bryan: Because I only got 17 for English in the Department Required Test (university entrance examination).
- Researcher: Is it difficult?
- Bryan: But 17 is too low.
- Researcher: Maybe you were...
- Bryan: No. I got similar scores for my mock tests. At that time, I despaired of being good at English.
- (Bryan, interview, 10.06.2008)

Because of the extremely low scores for the university entrance examination, Bryan lost his confidence in English, and thus, although he would try to take a test for the requirement, he assumed he had to take the remedial course nonetheless. His high level of anxiety was also evidenced by using the words of 'despair' and 'doomed' (Extract 5.1) when he referred to getting good grades in English and having to fulfil the requirement. He also criticised the phenomenon of the wide use of certified tests in various areas as 'hopeless'. Nevertheless, different from what concerned the teachers such as using remedial course as easy substitutes (5.6.1), Bryan still showed his willingness to prepare and take an English proficiency test despite expected failures. In fact, none of the students in University B other than Bryan revealed their intention to take the remedial course.

Similar to Bryan, Aaron and Alex were uncertain if they could graduate if the requirement were to be implemented, because of their past failures to pass the elementary level of the GEPT.

Extract 5.40

- Researcher: Do you like to have the graduation requirement?
- Aaron: No.
- Researcher: No? Why not?
- Aaron: Because...I won't be able to pass.
(omitted)
- Researcher: Do you feel that this brings too much stress on you?

- Aaron: Eh..quite a lot of stress.
Researcher: Have you taken any tests before?
Aaron: Um, the GEPT. I failed.
(Alex, interview, 10.06.2008)

Alex was particularly discouraged by the fact that the elementary level of the GEPT was designed for the proficiency level of younger students who receive less English education:

Extract 5.41

- Researcher: Have you taken the test in your senior high years?
Alex: Failed to pass the elementary level. A very big frustration. Isn't it the level for elementary school students? Elementary level is for elementary school students. (Actually, Alex made a mistake concerning the level equivalence. The elementary level of the GEPT, as stated in its official website, is equivalent to the English proficiency level expected for junior high school students.)
(Alex, interview, 10.06.2008)

Relating their ELLI profiles to the above analysis, the relatively lower scores in resilience could explain Bryan and Alex's case but not Aaron's. Both Bryan and Alex rated themselves as the lowest on the 'resilience' dimension and both did reveal their strong anxiety towards the requirement caused by their past experiences of failing English tests. However, Aaron's highest score on resilience could not relate to his doubt of himself being able to fulfil the requirement. As ELLI was a tool to assess learning power not specifically relating to English learning, it was likely that Aaron simply lacked confidence in learning English but still saw himself as a learner who could overcome learning frustrations. Nevertheless, Aaron's profile was indeed the weakest among the students in University A and likewise, Bryan was also the weakest among his schoolmates. In other words, those who considered themselves not as highly effective learners could probably suffer from larger degrees of anxiety resulted from

the graduation requirement. They were less active in taking control of their own learning and they were likely to feel uncomfortable when faced with challenges.

Their attitudes towards the requirement were not negative as they considered the idea of such requirement a good one because of the status of the English language and the formal evidence of English proficiency those tests could give (5.4). However, quite contrary, they were not supportive of implementing the requirement in their universities as they had much fear that they would not graduate because of their relatively poor English proficiency. The high level of anxiety they experienced or anticipated from the implementation of the graduation requirement challenged what Betty considered as good impact of the requirement (5.6.1). In a university with the graduation requirement, students like Bryan, Alex and Aaron have no choice but to keep trying to fulfil the requirement for graduation, no matter how much anxiety and fear they have. Therefore, it is likely that there will be the adverse effect Anna warned about what the requirement might bring to University A (5.6.1).

5.6.6.3 Summary

In the previous sections, the impact of the graduation requirement for English proficiency on the non-English majors was discussed from the perspective of the learners. The learners perceived different degrees of the impact on themselves and their learning, with some who considered themselves to be little influenced and others who had more anxiety when faced with stress the graduation requirement brought along. The students who perceived little impact of the requirement were either more concerned about their own English learning than fulfilling the requirement or were not particularly interested in English learning. On the other hand, among those who experienced or anticipated fear and anxiety because of the requirement, there were

students who welcomed the implementation of the requirement and also students who were not supportive. The strength and weakness of their ELLI profiles corresponded to their perceptions of the impact of the requirement to a certain extent. The students who considered themselves as effective learners were less concerned about the requirement. The students who reported themselves as less effective learners and who were especially weaker on the 'resilience' dimension had higher level of anxiety. They were more likely to be discouraged by past unsatisfactory experiences of test taking and felt more intimidated by the implementation of the graduation requirement.

5.6 Summary of the chapter

This chapter has discussed the impact of the graduation requirement for English proficiency on the learners, in several aspects: 1) knowledge of the graduation requirement 2) attitudes towards it 3) washback from the GEPT and other English proficiency tests 4) the impact of the requirement on the learners and their learning from the perspectives of both teachers and learners.

First, the analysis of what the non-English majors knew about the graduation requirement revealed evidence of the GEPT washback. The majority of the students aligned passing a certain level of the GEPT to fulfilling the graduation requirement and they had little knowledge of either what other English proficiency tests or what scores and levels they needed to reach.

In terms of test influence on learning resulted from the implementation of the graduation requirement, the GEPT washback was the strongest. It was the test that was perceived as the most important among the English proficiency tests stated in the graduation requirement. Explicit washback of the GEPT was found in students'

assumptions of passing the GEPT as fulfilling the requirement without knowing there were other choices. The majority of them had either taken the test or were coerced to prepare themselves for taking the test in order to graduate. However, it was the strong social impact of the GEPT that brought about the manifestation of the GEPT washback in the universities. The students chose to take the test because of its popularity, its influences on different levels of education and on parents. On the other hand, there was evidence for the impact of the TOEFL and TOEIC but very limited.

Lastly, the analysis of the impact of the graduation requirement on the students from their own perspective (including the students' learning power profiles) and the teachers' perspective revealed the following. The teachers' focused more on whether the requirement would influence the students' English learning as intended. Although they knew the good intentions behind the idea of the graduation requirement, the majority of them were sceptical about the outcomes as there were loopholes for students to evade the burden and the stress of test taking. Secondly, students demonstrated individual differences in their perceptions of the requirement. There were students who held a positive view on the idea of the graduation requirement. They acknowledged the importance of the English language and what advantages they could have with certificates of English tests by fulfilling the requirement. On the other hand, there were others who viewed the requirement negatively as they disliked the compulsory element that the requirement entailed or they were intimidated by the idea of not being able to graduate unless they could meet the requirement. Some were nonchalant to the implementation of the requirement because they did not find themselves challenged by the threshold level required by their university as they had either fulfilled it or did not think it would be difficult.

As for the learners, they had different perceptions of the ways the graduation requirement influenced themselves. Their learning power could be used to indicate the differences. The students with stronger learning power profiles were more likely to be active learners who took control of their own English learning and perceived little impact from the requirement. Those with relatively weaker profiles or those with significantly low scores on the 'resilience' dimension were more likely to experience higher levels of anxiety and fear caused by the implementation of the requirement.

In the next chapter, I will present the discussions of the findings in relation to the impact of the graduation requirement on the non-English majors and their EAP curriculum.

CHAPTER 6 Discussion of the Findings

6.1 Introduction

The previous two chapters reported the impact of the graduation requirement for English proficiency on the EAP curriculum for non-English majors (Chapter 4) and the impact on the students (Chapter 5). In this chapter, I will discuss significant findings of this study from a macro level perspective, to understand the big picture that the data as a whole has presented and also how the findings of this study relate to and further contribute to the washback and impact literature. In 6.2, I will first summarise the major findings of this study in relation to the research questions (3.4). Then, in 6.3, I will discuss the eminent power of the GEPT in the Taiwanese society and how the power was reinforced by the graduation requirement. The next section will focus on the further conceptualisation of test influence on learners. Section 6.4 relates the findings of this study to the literature of washback on learning and learners. For the development of a more elaborated operationalisation of washback, the following findings were discussed: 1) The ways that learning power can help explain washback on the learners and their learning process, 2) findings from classroom observations and interviews. Finally, in 6.5, I will conclude with a brief summary of the issues addressed in this chapter.

6.2 Summary of major findings

In this section, the major findings of this study will be summarised, in accordance with the research questions for this study.

- 1) What are the effects of the English requirement for graduation on the EAP curriculum for non-English majors in Taiwanese universities?

- 2) What are the effects of the English requirement for graduation on non-English majors in Taiwanese universities?
- 3) To what extent students' learning power explain their perceptions of washback to the learners?

6.2.1 What are the effects of the English requirement for graduation on the on the EAP curriculum for non-English majors in Taiwanese universities? (RQ1)

In Chapter 4, the reiterative, inductive analysis of the observation data, triangulated with the teacher interviews, private talks, field notes in both universities and the subsequent collection and analysis of the test papers in University A (3.7.3), revealed the following findings.

First of all, GEPT seemed to be the most influential English test among all the tests accepted by the graduation requirement. This finding confirmed my speculation (1.3.1) and Shih's (2008) assumption that the English language proficiency requirement for graduation was almost equivalent to the requirement of the GEPT (2.4). The data indicated that the influences of the requirement on teaching were largely related to the GEPT (4.3.1, 4.3.3, 4.3.4). There were only two instances related to TOEIC. Two teachers stated their preference of the TOEIC over the GEPT for students from the business fields (4.4, Extract 4.22, 4.23). GEPT had more profound impact than other English proficiency tests on university English education for non-English majors.

Secondly, there was significantly more evidence of the GEPT washback in University B than in University A. University B, which had implemented the graduation requirement for a few years, included a remedial course that was directly linked to the GEPT. The test affected the choice of teaching materials and course planning (4.3.1,

4.3.4). On the other hand, in University A which did not impose the graduation requirement on non-English majors, there was little explicit evidence of GEPT washback on its EAP curriculum for non-English majors. The only exception was the washback on testing materials for mid-term and end of term examinations, mediated by the local Taiwanese publisher that introduced the teaching material (4.3.4).

Thirdly, the GEPT washback was relatively intensive on some areas of washback but not on others. There was explicit evidence of GEPT washback on the teaching materials the teachers (Becca and Ben) adopted for their classes in University B. They included GEPT preparation material (Becca) or commercial monthly-issued English learning magazines that incorporated GEPT elements, i.e., GEPT practice items, topics and contents related local culture (Ben, see 4.3.1). Washback intensity was also found on testing and assessment in both universities (4.3.4). There was the incorporation of the GEPT tests as pre-test (a mock test) and post-test (a test delivered by the Language Training and Testing Centre) in Becky's GEPT-related course, as reported by herself (Extract 4.18). There were the mid-term and final test papers in University A which were modelled on GEPT test item types, developed by the local Taiwanese publishers (Table 4.3, Extract 4.17). However, as distinct from several previous studies (Alderson and Wall, 1993; Cheng, 1997, 2005; Stecher et al., 2004), there was little evidence of curriculum narrowing, focusing on skills tested, or change in teaching activities as intended effects from test innovation, even in the GEPT-related remedial course in University B (4.3.1). The analysis of the language focus and language skills targeted by the activities in the observed lessons (Figure 4.9 and 4.10) revealed individual differences among the teachers teaching the same courses. Teachers' different responses to English learning, test taking or the graduation requirement were more highlighted than the influences of the GEPT or other tests accepted by the requirement

on teaching. Similarly, teaching methods were another aspect that showed little evidence of test influences in this study. Since none of the English proficiency tests accepted by the graduation requirement was intended to bring change to teaching methods in the university English classrooms, it was difficult to link the different teaching methods that each teacher adopted to any particular test.

Lastly, individual differences between the teachers were not only evident in the classroom observations data, but also in what they revealed in the interviews, reflecting ‘washback variability’ (Green, 2006, 2.2.4) among teachers. The teachers’ lessons reflected different degrees of washback. Their perceptions of the GEPT washback and the impact of the graduation requirement impact also varied. The findings of the current study also provided evidence for how teacher factors might explain the presence or absence of washback on teaching (Watanabe, 1996, 2004). The teacher factors that mediated or prevented washback could be seen from the following findings: 1) teachers’ perceptions of how effective the graduation requirement could be in urging students to take English proficiency tests and in boosting their English proficiency (see Extract 4.25), 2) their willingness to comply with what the requirement demands (see Extract 4.21), 3) their preference of one test over the others (see Extract 4.22), and 4) their beliefs in what they should teach (see, for example, Extracts 4.9, 4.12, 4.19).

6.2.2 What are the effects of the English requirement for graduation on the non-English majors in Taiwanese universities? (RQ2)

To answer the second research question, the individual interviews of the 18 non-English majors (9 in each case university) were analysed inductively in terms of the influence of the GEPT or other English tests on them and their learning, and triangulated with related parts of the teacher interviews (5.2). The findings emerged

from the analysis are summarised below.

First, the GEPT was considered as the most important test among the tests accepted by the graduation requirement. Students aligned the requirement with passing the test, and the majority of the interviewees from University B had either taken the test or had planned to take it soon. Students had insufficient knowledge of other English tests and what scores were set as benchmarks for graduation. Most of them also took it for granted that the requirement was equivalent to a GEPT requirement. Such findings indicated a strong presence of the influence of GEPT on these students (see Extracts 5.1, 5.2). This also applied to the cohort of students in the case university without the requirement (University A). Students there mentioned GEPT most frequently when asked about what they thought the graduation requirement was (5.5). Although there were also references to TOEFL and TOEIC when the learners considered which English proficiency test to take during their university years, they mainly associated GEPT to the graduation requirement. TOEFL and TOEIC were associated more with the students' own academic plans in the future (see Extract 5.7).

The analysis of both the learner interview data and the related teacher interview data revealed differences in two main aspects: differences between the perspectives of the teachers and learners, and the individual differences among the learners themselves. First, similar to the findings of the few previous washback studies with a focus on the learners (2.3.2.1, 2.3.2.2), the findings of this study suggested that the learners viewed the impact of the requirement on their English learning from a different angle from their teachers. The teachers held a more negative view of the requirement (5.6.1). They were sceptical of its intended effect in promoting the students' English learning and their motivation to learn. The teachers were concerned with the possible adverse effects the

requirement might bring (see Extract 5.11). They were also concerned that the loopholes of the regulations for the graduation requirement also had some loopholes for students to avoid taking any external English language test (see Extracts 5.13, 5.14). Similarly, they were concerned about the difficulty in fully implementing the requirement (see Extracts 5.12, 5.15). Nevertheless, what concerned the teachers was not necessarily what the students cared about. The learners were more concerned about the relative easiness or difficulty in meeting the benchmark set by the graduation requirement and the role of the requirement in their English learning during their university years (5.6.2).

The learner interview data shed further light on the individual differences on their perceptions of the graduation requirement, its impact on them and on their English learning (5.6.2, 5.6.6). Their attitudes about the implementation the graduation requirement varied according to how they perceived the compulsory elements and also the stakes entailed in the requirement. Several of the students acknowledged the need and the benefits for such regulations to compel them to learn more English (see Extract 5.16-5.20); yet, others questioned the necessity for making it compulsory for their graduation (see Extracts 5.22, 5.23). The perceived stakes of fulfilling the requirement could also affect the learner's attitudes towards the implementation of the graduation requirement. Those who believed that their English was not good enough for them to reach the benchmark were reluctant in accepting the implementation (see Extract 5.24), whereas those who considered it easy to fulfil did not take it too seriously (see Extract 5.25).

The learners' perceptions of the impact of the graduation requirement also revealed some individual differences. There were students who perceived little impact (5.6.6.1)

but there were also students who associated stress and anxiety with the compulsory requirement in order to obtain their degrees (5.6.6.2). Those students who considered themselves to be little influenced were more concerned with their opportunities to learn more English than with being required to provide official proof of English proficiency (see Extracts 5.29, 5.30). For them, their motivation to learn English would not be influenced much by the implementation of the graduation requirement as they had their own learning goals, and the requirement would only make them work harder for this particular high-stakes purpose (see Extracts 5.26, 5.27, 5.31). Yet there was one exception that suggested that students were also likely not to be influenced by the graduation requirement because of their little interest in English and limited motivation to learn the language (Extract 5.33).

The rest of the students, on the other hand, associated stress and anxiety with the graduation requirement (5.6.5.2). However, they were still different in how they chose to face the implementation. Some viewed the pressure accompanying the requirement as a positive force on their English learning and thus, welcomed the implementation (see Extracts 5.35, 5.36). There were those who did not support the implementation as they disliked the 'side effects' the requirement might bring along (see Extracts 5.37, 5.38). Others were not supportive either, not so much because they had negative feelings towards the implementation, but because they feared that they would not be able to meet the requirement, because of their poor proficiency and their past failures in test-taking (see Extracts, 5.39-5.41).

6.2.3 To what extent students' learning power explain their perceptions of the washback on learners and their learning? (RQ3)

To answer this research question, the student interviewees' ELLI profiles were analysed

by relating their dimensions of learning power to their interview data concerning their perceptions of the graduation requirement and its impact on them and their learning (5.6).

The findings suggested that the overall strength of the learners' ELLI profiles on the seven dimensions could, in some way, be indicative of how much impact the learners considered the requirement would bring on them (5.6.6.1, 5.6.6.2. see 6.4.3 for more elaboration). As the ELLI instrument is used to assess individual learners' learning power (2.5), their profiles can show whether they consider themselves as effective learners who are active, creative, take control of their own learning and know how to advance in learning. By relating the interviewees' profiles to their interview data, it was found that those who had stronger profiles were more likely to perceive less impact from the graduation requirement, as they were more likely to take charge of their own learning (5.6.6.1). This could be evidenced from their efforts in English learning and test taking beyond the fulfilment of the requirement and their concerns for English courses. They did not perceive the benchmark for their graduation as high stakes or difficult to reach, and were thus motivated to learn English for less instrumental purposes (see discussion in 6.4.4 below). On the other hand, those who reported that they were influenced or would be influenced by the graduation requirement were mostly the learners with the weaker learning power profiles (5.6.6.2). Their English learning and test taking were more driven by the compulsory English requirement than their own learning goals and thus, mainly for instrumental benefits (6.4.4).

In addition to the strength of the overall profile, 'resilience', one of the seven dimensions, was found to be indicative of the extent to which the learners experienced anxiety towards the requirement (5.6.6.1, 5.6.6.2, see 6.3.3, 6.3.4 for further

elaboration). Those who rated themselves low on this dimension tended to be more anxious about having to take external English tests as they lacked confidence in their English proficiency and were more likely to be discouraged by their past failures in test taking (see extracts, 5.39, 5.41). It was also likely that they might feel overstressed, which resulted in a decrease in their motivation to learn English (see extract 5.37, 5.38). On the contrary, the learners who had higher scores on 'resilience' were less likely to be intimidated by the requirement. They welcomed the challenge of taking external English tests, and were not afraid of making mistakes and failures. Even with those who experienced anxiety, they tended to see the requirement as a necessary evil (see extract 5.34, 5.35; also see discussions in 6.4.4).

In short, the overall picture of an individual learner's learning power, as defined by the seven dimensions, could shed some light on the extent to which the learner perceived impact from the graduation requirement. In addition, 'resilience', could also inform the extent to which the learner experienced anxiety towards test taking in order to graduate.

6.3 The Power of the GEPT

Unlike most previous washback and impact studies that explored the ways a designated test was tailored to a curriculum can affect teaching and learning (2.3), the fact that the graduation requirement accepted a number of English proficiency tests external to the universities English curriculum makes the research context more complex. The analyses in Chapter 4 and 5 revealed that the GEPT exerted more influences on teaching and learning in Taiwanese university English education than any other English proficiency tests accepted by the requirement. However, it seemed that the majority of the GEPT washback effects manifested in the two universities in fact resulted from the social impact of the test. Its social impact was evidenced on the following: 1) the choice

of the commercial English self-learning magazines which publish GEPT mock items and related contents (4.3), 2) the choice of the local publisher representing international teaching materials to develop tests modelled on GEPT item types (4.6), 3) the parents' preference on the GEPT (5.5), and 4) students' assumption of the GEPT as necessary in future job hunting (5.5). The findings suggest that using test results for high stake purposes without designating a specific test, washback will be most eminent from the test that is perceived as the most important in the local society. The following section discusses the strong impact of the GEPT, the locally developed English proficiency test, on the Taiwanese society (6.3.1). Section 6.3.2 goes on to discuss the role of the English degree requirement in reinforcing its social impact in the universities, with evidence of the GEPT washback on English teaching and learning that can be linked to its social impact. In the final section (6.3.3), I will discuss how the GEPT impact was reinforced by the graduation requirement, from Shohamy's (2001, 2003, 2007) concept of the discrepancy between language educational policies and *de facto* policies.

6.3.1 The social impact of the GEPT

The social impact of the GEPT was evidenced in this study through different types of stakeholders, beyond or within the university system. External stakeholders included the parents, the publishers of monthly-issued English learning magazines for all citizens, and also the publishers that represented the international EFL teaching material. Stakeholders within the university system, like non-English majors, might also take the GEPT for purposes other than fulfilling the graduation requirement.

The most explicit evidence of the GEPT social impact is within the community of local publishers who develop English learning magazines for lifelong learning (Studio Classroom: <http://www.studioclassroom.com.tw/sc/index.php>, English Digest:

<http://www.english4u.net/web/products/digest.aspx>). An important feature that stems from the commercial nature of those monthly-issued magazines is to cater to the current needs of the potential buyers in order to promote sales. With the popularity of the GEPT in society, the magazines that were once not test-oriented have been changed into materials that can prepare readers for the GEPT (4.3). The inclusion of the GEPT-related contents and practice items that explicitly refer to the test in the issues since the test came into play is thus a reflection of how big the influence the test is on Taiwanese society. In addition to the evidence provided in Figure 4.9, the websites of the publishers also reflect the importance of the test. We can see that on the English Digest website (<http://www.english4u.net/web/products/digest.aspx>), the test is clearly stated in the 'quick menu' at the right side of the page. On the webpage, the description of the magazine includes the proficiency level it intends to develop, which is the intermediate to high intermediate level of the GEPT. Furthermore, the description of the contents in the CD-Rom attached to the magazine also includes GEPT mock tests. In this study, Ben used the magazines as the teaching materials in his lessons, but denied having the intention to prepare his students for the test (4.3), which was evidenced in his lessons with no signs of test preparation, practice of mock test items or any reference to the test. Thus, the GEPT washback on teaching materials as teachers use the magazines in their classrooms without the purpose of test preparation is in fact a product of the GEPT social impact on the magazines.

Similarly, local publishers who represent international English teaching materials that are designed for institutional use can also come under the strong social impact of the GEPT. The publishers here are different from those mentioned above, in that they import international teaching materials and promote the sales of those materials in local educational settings instead of developing teaching materials. The findings of the

present study revealed that evidence for GEPT washback did not lie in the teaching material itself, so much as in the mid-term and final test developed by the local publisher, with item types modelled on those in the GEPT (4.6). The following is a comparison of the test instructions for the 'question and statement response' section (translation in bracket) and an item of the listening test developed by the local publisher and the equivalent in the practice GEPT test provided by the official website: (<http://www.ltc.ntu.edu.tw/geptpracticee.htm>).

Listening	University A test paper	GEPT elementary level
Question and statement response: test instruction	每題請聽錄音機播出一英語問句或直述句之後，從試題冊上 A、B、C 三個回答或回應中，找出一個最適合的作答。每題只播出一遍。(For each item, please listen to a question or a statement from the audio recorder. Choose the most appropriate answer from the three answers or responses in a, b, and c. Each item is played only once.)	請聽錄音機播出一個問句或直述句後，從下面 A、B、C 三個回答或回應中，找出一個最適合的作答。每題只播出一遍。(Please listen to a question or a statement from the audio recorder. Choose the most appropriate answer from the three answers or responses stated in a, b, and c. Each item is played only once.)
Sample test item	1. (Audio: How often do you clean the house?) A. Yes. My house is very clean. B. Twice a week. C. We usually clean the house on Sunday.	1.(Audio: Who's that tall handsome man?) A. He's studying. B. He's my cousin. C. He's not very happy.

Table 6.1 Comparison between University A test item and the GEPT elementary level

The above table shows that the test instructions in both tests are only slightly different, and the test item in the test paper used in University A is exactly the same as that in the GEPT. As revealed by the teachers, the test papers were developed by the local publisher based on the contents of Top Notch, the non-GEPT-related, international

teaching material (4.3.4). Thus, the local publisher's attempt to shape a small part of the test for the GEPT, while preserving the internationally-recognised material contents, suggested how the local publishers retouched the global materials to meet the locals' needs. The local publisher's action is certainly an evidence of social impact of the GEPT, and when the deliberately designed test is used in universities, there will be GEPT washback on internal testing.

The social impact of GEPT is also evidenced by the parents, who can influence the learners on the selection of which English proficiency test to take. Findings from the student interviews revealed that parental influence was why some students prioritised the GEPT over other English tests (5.5). Although they also expressed their intention to take TOEFL or TOEIC test in the near future to fulfil important academic goals, their decisions to take the GEPT first was actually persuaded by their parents. Since the graduation requirement did not favour the results of one test over other tests, it was more likely that the parents' eagerness for their children to pass the GEPT was because they perceived the test to be more important than other English proficiency tests in society. Thus, parental influence in the context of this study is a manifestation of the GEPT social impact on the parents, which in turn shaped washback to the learners.

Stakeholders within the university system such as the learners may also reflect the social impact of the GEPT. Some students in University A revealed that they took the GEPT instead of other English proficiency tests, mainly because they considered a certificate of English proficiency test as essential for job searching (5.5). Students' assumption of the GEPT as being a test that will be accepted or requested by future employers illustrates a strong impact on the society.

The above findings suggested that those stakeholders were influenced more by the GEPT in society than the test within the context of the graduation requirement and hence, their actions probably resulted more from the impact of GEPT on general society than its impact on the graduation requirement, or the graduation requirement itself. What I argue here is that the policy made by the Ministry of Education (1.3.1), and the accompanying measures the universities have taken to boost university students' English proficiency, reinforce the social impact of the GEPT on teaching and learning in the universities.

6.3.2 The reinforcement of the graduation requirement

The findings reported in Chapter 4 and 5 suggest that the implementation of the English graduation requirement reinforces the GEPT social impact in two ways. First, for the majority of the non-English majors, the need to provide a proof of English proficiency for graduation is an imperative or the need to pass the GEPT. In other words, the students are compelled to prepare themselves for a test among the many English proficiency tests available for the purpose of graduation and they choose the GEPT because of its perceived status in the society. It is the same for the English teachers in the universities as they speculate that the GEPT is the test most likely to be taken by their students. The direct alignment of passing the GEPT with fulfilling the graduation requirement by both English teachers and the students in the current study is evidence of the strong social impact that the GEPT has, being reinforced in the university system.

Secondly, through the hands of the curriculum designers, who are usually the English departments, the English curriculum may reflect the social impact of the GEPT. The GEPT social impact is reinforced in English teaching and learning in the universities

when the curriculum designers shape the direction of the preparation courses or remedial courses, the intended curriculum change for the graduation requirement, to focus particularly on the GEPT but not other tests. An example is the remedial course in University B (4.3.1). Despite having a generic course title that did not suggest any link to any test, teachers were instructed by the English department to incorporate GEPT contents and related teaching materials. The course aim was narrowed by the department from helping students who failed to fulfil the graduation requirement to reach the level of proficiency as required, to assisting them especially to pass the intermediate level of the GEPT. In this case, GEPT washback manifested in those courses is thus mediated by the curriculum designers who are influenced by the power of the GEPT in the society.

The above findings offer some explanations why one particular test can exert the strongest influence when test takers can choose from a number of tests. The conditions or purposes (e.g., admission, promotion, placement or graduation) of a test, as Madaus (1990) pointed can determine whether a test is high stakes or not. However, this cannot be fully applied in determining the stakes that any tests receive in the context of the current study. When students are given the liberty to choose any test stated in the graduation requirement, each test can be considered as high stakes if its result is used for graduation. However, the findings in this study suggest that the importance of a test in the society as perceived by the stakeholders is another consideration needs to be taken alongside with test use. What makes the GEPT stand out from the internationally recognised and certified English proficiency tests such as TOEFL or TOEIC is its wide recognition among Taiwanese citizens. This may be the reason why Taiwanese parents will prefer their children to take the test as it is very likely that the GEPT is the test they know most about. In addition, unlike TOEFL, TOEIC, or IELTS, which are usually

used for a certain purpose like further studies abroad or business English, the GEPT test scores can be used in a wide range of areas in the Taiwanese society such as university admission, job application or governmental scholarship application. The GEPT is thus perceived by the majority of stakeholders as more important than other English proficiency tests in the society, therefore its social impact is stronger than that of the other tests. The implementation of the graduation requirement, which uses English test result for high stakes purpose, further reinforces the already strong social impact of the GEPT in the university system via its impact on teaching and learning. The advantage of test scores for multiple uses and the status of the test in the society are similar to Gates' (1995) ideas of test 'utility' and 'monopoly' that determines the extent of washback intensity. Nevertheless, both ideas refer to the social context in which a test is used and how important a test is in the society. Therefore, in the high stakes context where stakeholders are given multiple choices, I argue that, it is the test stakeholders perceived as the most important in the society that would bring the most eminent washback.

6.3.3 GEPT and the language education policy

Another issue that surfaces from the findings of this study is the complex and often controversial relationship between language tests and educational policy. In 2.2.2 I discussed such relationship. What needs to be emphasised is the social and political agenda a language test can entail. These less obvious agenda may result in language tests being used as powerful mechanisms in the implementation of covert, *de facto* policies, which often deviate greatly from the original policies (Shohamy, 2007):

“Even while policies which are expressed in official documents provide relatively transparent information about specific decisions regarding languages, much of language policy is realised through a variety of indirect actions and practices that serve as *de facto* policies that can override

and contradict existing policies and create alternative policy realities.” (p.120).

Although the Ministry of Education (MOE) in Taiwan partly sponsored the Language Testing and Training Centre (<http://www.lttc.ntu.edu.tw/TST.htm>) in the development of the GEPT, the test has rarely been officially announced or included in language educational policies. However, the strong social impact of the GEPT clearly demonstrated the discrepancy between the *de facto* and original policies. The policies of the MOE regarding tertiary English education were more concerned about the general issues of the establishment of common English standard, the mapping of the available English proficiency tests to the levels of the CEFR and the promotion of university students' English proficiency (1.3.1).

In addition, as I mentioned in 1.3.1, there has not been any official MOE publication or announcement of what should be included in the graduation requirement. Thus, each university implements the graduation requirement for English proficiency with different regulations concerning curriculum change or compensation measures for students who cannot pass external English proficiency tests. The original policies are more generic; and the *de facto* policies of each university, in the form of the regulations of the graduation requirement, are institution specific.

What really makes the impact of the requirement turn out to be different from the intended impact of the original policies is the GEPT. Shih's (2010) study of the GEPT washback on university policies provided similar findings. The graduation requirement of one case university in his study included GEPT mock test as compensation for students who were not able to pass the required level of the GEPT. As the test is perceived by most stakeholders as being one of the most important in society, the power of the GEPT influences the *de facto* policies carried out in the

universities to accentuate the test itself. The original intention of setting generic levels of English proficiency as benchmark and standards for university students to meet is thus contradicted by the unintended impact of one specific test on teaching and learning. Thus, in the end, the impact of the *de facto* policies is far different from the intended impact of the original policies. Similarly, it may be instructive to map scores of different English proficiency tests in the graduation requirement to a level in the CEFR, for universal recognition of students' English proficiency in terms of generic rating scales. Nevertheless, without considering the local language learning contexts and also the social agenda of the GEPT, the good intention seems to be challenged by teachers and students' alignment of the graduation requirement to the GEPT instead of generic proficiency levels. In conclusion, the findings of this study provide a different exemplification of Shohamy's (2004, 2006, 2007) argument about how language tests can affect actual language educational policy, in the condition that the language test is not deliberately developed and implemented for the purpose of control.

6.3.4 Summary

The GEPT social impact was found to be shaped by local publishers, parents and also the learners themselves (6.3.1). The local publishers included those that developed and published monthly commercial English learning magazines and those who imported and promoted the sales of international EFL teaching materials. The social impact of the GEPT was evident in the magazine publishers' inclusion of GEPT-related contents and mock GEPT test items in their magazines and the attached CD-rom and also in their development of test papers for institutional use by modelling on GEPT item types. Both findings revealed how the two types of local publishers were influenced by the GEPT in the social sphere as their actions, based on the purpose of promoting sales, reflected what they perceived their customers (the Taiwanese citizens, including teachers and

university students) wanted the most. Secondly, parental influence on the non-English majors' prioritising the GEPT over other English proficiency tests (5.5) could be also considered as evidence of the social impact of the GEPT. It was likely that the parents' perceptions of the GEPT being more important than other English tests in society made them ask their children to prioritise the test, even though the test was not viewed as superior by the related policies or the graduation requirement. Lastly, the learners who assumed that the GEPT would be most useful for future employment also reflected the social impact of the test.

In addition, it was revealed that the social impact of GEPT was further reinforced by the implementation of the English graduation requirement in the universities where teachers and learners aligned this test with the requirement and the curriculum designers directed the remedial courses in University B towards preparing for GEPT tests (6.3.2). First, the compulsory graduation requirement pushed university students and teachers to focus on one particular test, which in this case, the GEPT, for the students to be able to graduate. Since no test is favoured by the graduation requirement, the focus on the GEPT was due to its perceived status in the society. The implementation of the graduation requirement brought the social impact of the test into the university system. Similarly, the curriculum designer, the English department in University B, shaped the remedial course with a generic course title into something directly related to GEPT, by requiring teachers to teach GEPT contents using GEPT materials (4.3.1). Such action reflected how GEPT washback on teaching was a result of its social impact being reinforced by the graduation requirement.

The findings suggest for a similar context with multiple tests that are not tailored to the curriculum, the test that is likely to bring the most influences will be the one that is

perceived by the stakeholders the most important in the society. In this study, the strong social impact of the GEPT challenged the good intentions of the original policies and the mapping of the graduation requirement to the CEFR. Therefore, such social agenda embedded in a language test should be taken into consideration when the test is included in language educational policies.

The power of the GEPT test as shown in the social, and educational contexts of this study has been discussed above. In the next section I will discuss the washback mechanism on learners themselves.

6.4 Factors that mediate washback on learners and learning

A few recent studies on washback have suggested different factors that may mediate washback to the learners (2.3.2). For example, Green (2006) argues about the role of teachers and courses in shaping washback to the learners. He points out that there is a discrepancy between learners' perceptions of tests, test preparation and the perceptions of teachers. Scott (2005) argues the need to be cautious in reporting test scores to the parents as they may have limited information on the tests their children have to take, and their perceptions of the tests may shape test impact on learners, especially young learners. On the other hand, Gosa (2004) emphasizes learners' individual differences, and argues for the consideration of student variables in studying washback. Tsagari (2006)'s study explores more in depth about the student variables, and reveals the role of learners' attitudes and feelings, and motivation that shape washback. Shih (2007) lists all the possible factors that may shape washback on learners and learning: extrinsic factors, intrinsic factors and test factors.

In this section, I will discuss the findings of this study with reference to Bailey's (1996)

washback model and some of the conceptualisations of washback to the learners from the abovementioned studies, including Green's (2006b) washback model (Figure 2.5) and Shih's (2007) washback model of learning (Figure 2.6). Green's framework encapsulates factors that explain both washback variability and washback intensity. Likewise, the current study has revealed findings that reflected both issues and related factors. Shih's model is of great relevance to my study, because the educational and research contexts in which our studies were conducted were similar.

The following sections provide a further operationalisation on the following areas: the role of publishers, the intrinsic factors that shape washback to the learners, the ways learning power can inform the washback mechanism on the learners themselves and lastly, washback on learner motivation.

Bailey's model serves as the theoretical basis of this study (2.2.6, Figure 2.4). This study further confirms the roles of stakeholders external (material writers) and internal (curriculum designers and teachers) to the university setting in shaping the influences of GEPT on students. The findings of this study, however, reflect more than what Bailey's model has covered. Therefore, I will use Bailey's model as the basis of the following discussions, but at the same time incorporate Green's and Shih's models, particularly Green's discussion (2006b) on the relationship between washback, stakes and difficulty and the intrinsic factors highlighted by Shih model.

6.4.1 The role of publishers

Publishers have played some significant roles in shaping the GEPT washback on teaching and learning in the universities. The role of the material publishers/designers as important stakeholders in the washback process is considered in Bailey's (1996)

model. Most previous washback studies found that publishers exerted test influence through the teaching or learning materials they designed/published (2.3.1.2). Since most studies reported teachers' reliance on test-oriented textbooks in their classes, what the learners learn were largely dependent on what the publishers considered as contents that could prepare learners for the test. However, the role the publishers played in this study was different from that of publishers played in other studies.

In the context of this study, a broader definition of 'publishers' is used. In many EFL contexts similar to Taiwan, publishers of EFL teaching and learning materials include not only those who publish materials they develop themselves but also those local publishers that import and represent EFL materials designed by well-known international publishers. However, few previous washback studies have pointed out the differentiation between the two types of publishers and few have discussed how the local publishers representing international materials could contribute to the washback process. An unexpected finding of this study suggests that the local publishers can also play a significant role in mediating GEPT washback (4.6, 6.2.1). It was not so much through teaching materials that the international publishers can exert GEPT washback but through the the mid-term and final tests that the local publishers designed for universities, according to the GEPT item types. In addition, the classroom observation data revealed that none of the teachers mentioned the GEPT item types in their final exams. Not only could the material shape washback on teaching and learning, but the local publishers representing international teaching materials were also likely to exert influences of a test.

6.4.2 Intrinsic factors: Personal perceptions of test, test stakes and test difficulty

There are three factors contained in the category of intrinsic factors in Shih's model,

these being: individual differences, personal characteristics, and personal perceptions of the test. As mentioned above, based on the findings of this study, I do not think the factors are explained with enough clarity in Shih's model. Furthermore, I think that there are likely to be more than three intrinsic factors that can mediate washback on the learners and their learning.

I would like to expand the factor 'personal perception of the test' and integrate it with Green's conceptualisation of washback intensity (2.2.5, Figure 2.2). Shih argues that the assumption of tests with higher stakes always bring stronger test impact (e.g., Madaus, 1990; Shohamy et al., 1996) may not be true and that the mechanism of washback on learning in his context, like the research context of this present study, is much more complex. He thus adds two factors as test factors in his model: 'immediate importance of the test' and 'relative difficulty of the test in relation to students' proficiency' (Shih, 2007, p.151). Shih considers these two test factors as indicating why GEPT, as a high stakes test, has had little washback on the learning of his student participants. Although I agree with Shih on two additional test factors, the findings of my study suggest that the impact of the requirement on the learners may be related less to the actual stakes and difficulty of the test, and more to the learners' perceptions of test stakes and difficulty. Therefore, I argue that the phrase, 'personal perceptions of the test' is too broad, and that washback on learners and their learning should encompass personal perceptions of a test, and the complex relationship between perceptions of test stakes, perceived difficulty of the test (Watanabe 2001), and washback intensity (Green 2006b).

To begin with, in the context of the present study, it is problematic to associate stakes directly with the importance of a test. Even though the GEPT is perceived as the most

important test in the society and the majority of the students in this study confirmed that the test was important to them, not all of them perceived passing the test as high-stakes. This is because the GEPT is a test with five levels and the threshold level the university requires the learners to meet can be too high, too low or just appropriate for any student. Some learners in this study were nonchalant about the graduation requirement, considering that it would have little impact on them because it was not difficult for them to meet the benchmark. Thus, they did not perceive the first stage of the intermediate GEPT as high stakes (5.6.6.1), not because they did not consider the test important, but rather that they did not find passing the required level difficult. On the other hand, some learners feared that the GEPT intermediate level would be too difficult for them to pass, which might result in their failure to graduate from the universities. As a result the test was perceived by them as high stakes. This resonates what Watanabe (2001) refers to as learner's perception of difficulty. Shih's 'relative difficulty of the test in relation to the students' proficiency' already challenges the concept of objective difficulty and highlights the relationship between test difficulty and students' proficiency. However, without actual measurement of students' English proficiency; it is difficult to say whether the test difficulty they perceive will correspond to how they will actually perform in the test. For example, despite his lack of experience in taking the GEPT test, Bryan thought it would be very difficult for him to reach the standard stated in the graduation requirement, considering his poor English results in high school and in the University entrance examination. For him, fulfilling the requirement in order to graduate was perceived as high stakes because his perception of the difficulty to pass the GEPT intermediate level was high.

In conclusion, the above discussions show that the dichotomy of high and low stakes can be problematic, particularly in the context where learners are required to reach a

designated level of English tests for high stake purposes. In fact, students' perceptions of the test stakes were not only related to the importance of the test but also to their perceptions of the level of difficulty to pass the test.

Nevertheless, the operationalisation of the interaction between test stakes, test importance and test difficulty above is not sufficient enough to explain washback on the learners and their learning in the complex context of the current study. Watanabe argues that appropriate level of difficulty and importance as perceived by the learners are the key motivators for learners to prepare for the test. However, the relationship between perceived difficulty, perceived stakes and the test impact to the learners was more complex in this present study than in Watanabe's research context where the test investigated was of unarguably high stakes. In the next section, I will discuss how the Effective Lifelong Learning Inventories (ELLI) can help to provide more insights into the intrinsic factors that mediate washback.

6.4.3 Intrinsic factors: Learning power

ELLI is a tool that assesses a person's learning power: the capacity and desire to learn and to go on learning throughout life (2.5). Although the learning here refers to a general concept of learning and is not specifically used to indicate English learning or learning of any languages, the different learning profiles of the 18 student interviewees are helpful to explore the different manifestation of test impact on the learners. ELLI profiles consist of seven learning dimensions and draw on a wide range of variables that can impact on individual learners; capacity and their motivation to learn. The learning profiles are useful to understand how motivation to learn can be affected by required language tests in two ways.

First, the overall learning profile as represented by ELLI can be a useful indicator of a student's motivation to learn. The stronger the learners rate themselves on each of the seven dimensions, they are more likely to be learners who take more active actions in learning and who take charge of their own learning. This may be evidenced by the following findings. The majority of those who had the stronger profiles among the 18 students in this study had gone beyond what the requirement demanded and had their own plans of learning English and taking English proficiency tests. The graduation requirement for them was just one goal they could achieve along the learning path they have arranged for themselves. Quite differently, those with the weaker learning profiles tended to be less active learners and were pushed by the graduation requirement to learn English and to take a test solely for the purpose of receiving their degrees.

Secondly, one of the seven learning dimensions – resilience – was found to have the capacity to explain the students' anxiety towards test taking and fulfilling the graduation requirement. Learners who consider themselves as resilient are those who can face challenge, not afraid of mistakes and failure, and would embrace risk in learning even though they may feel frustrated and anxious. On the contrary, learners who score low on this dimension are more likely to feel more anxious and discouraged when they make mistakes. In this study, what the students expressed about their feelings towards the requirement corresponded to some extent with their resilience score in ELLI. Below are two illustrations, with the higher scorer on resilience in University A and the lower scorer on resilience in University B.

Archer had one of the strongest overall learning profiles in University A. Bryan had the weakest overall learning profile in University B and also had a relatively low score on the 'resilience' dimension. Both said they did not perform well in previous English tests. Archer failed to pass the second stage of the GEPT intermediate level, which consisted

of speaking and writing tests, however, this failure did not stop him from learning English nor from taking more tests. Acknowledging that the second stage was more demanding for him and the TOEIC had the advantage of international certification, he moved on to take TOEIC preparation course. On the other hand, Bryan was deeply discouraged by his repeatedly low scores in English in mock tests and the university entrance exam. His use of strong negative words like 'doomed', 'hopeless' and 'despair' as he expressed his feelings towards the requirement, showed that he did not believe in himself that he would be able to meet the requirement. His lack of confidence echoed with the significantly low scores he rated himself on the 'resilience' dimension. In accordance with Rea-Dickins et al.'s (2007) discussions on learners' learning power and their learning process, the abovementioned findings also showed that most learners experienced anxiety when facing challenges in relation to preparing for external English tests. The learners with strong ELLI profiles were able to deal with their anxiety effectively and take charge of their own learning regardless of what they received in their university English classes. However, those with weaker ELLI profiles were more likely to be overwhelmed by anxiety and feel intimidated by the implementation of the English proficiency requirement for graduation.

Although the students' English proficiency test results and their academic grades in the EAP curriculum were not collected, the analysis of their interviews along with their ELLI profiles still suggested the positive link between their learning power and English test scores as evidenced in Rea-Dickins et al's study (*ibid.*). Most students with the very strong ELLI profiles have met the stated requirement for graduation, the first stage of the intermediate GEPT or equivalent in the case of University B. The students who had overall weak learning power in both universities were likely to reveal their unsuccessful experiences of English tests (e.g. Ben had very low score for

university entrance examination; Aaron and Alex failed elementary GEPT). It would be too assertive to say students' learning power may be indicative of their scores in English tests based on the findings of this study. Yet learners with higher learning power are indeed more likely to perform better in language tests than those with weaker learning power.

This section has provided a discussion of how learning power can inform learner's perceptions of the English graduation requirement and test washback on their learning. In the next section, I will discuss washback on learner motivation.

6.4.4 Washback on learner motivation

The studies of both Watanabe (2001) and Tsagari (2006) are useful in understanding what test-related factors motivate students. However, my findings in Chapter Five suggest that the test-related factors they have discussed are not extensive enough to encompass the complexity of the context in this study. Watanabe, as previously mentioned, attributes a positive effect on learner's motivation for test preparation to the appropriate level of difficulty perceived by the learners themselves. Tsagari, on the other hand, highlights students' perceptions of the degree of importance and the status a test has in mediating washback on learner motivation. Tsagari further warns about two pitfalls that may accompany washback to the learners: long-term anxiety and stress, and the overpowering instrumental benefits that tests bring to the learners.

The findings of this study echo Tsagari's concerns to some extent. For those learners who perceived reaching the required level of a test as important, difficult and high stakes, and who have weaker learning profiles, it is likely that they will suffer anxiety and huge stress from the implementation of the graduation requirement. This is the case

with Bryan and some students in University A. Judging from their past failures in GEPT or other English tests, they regarded the graduation requirement as a huge burden. There was high likelihood that they would not be able to graduate even if they had passed all of the courses and received credits for their subjects related to their majors. There was also the potential threat that they had to stay in the universities more than four years just to fulfil the graduation requirement. The high anxiety they had might de-motivate some of the learners to give up on English learning and test preparation. However, the majority of the learners still expressed their willingness to study more English to fulfil the graduation requirement even if they were afraid that they might not be able to perform well enough. These findings are similar to Paris et al. (1991)'s study, in which they found that the pressure of not being able to do well on tests applied to students of all achievement levels. However, as lower-achievers perceived themselves more negatively, they might give up on making efforts and using appropriate testing strategies when facing a high-stakes test. In this study, the students who expressed greater frustrations towards the graduation requirements were indeed those who had less positive self-perceptions (as seen in their ELLI profiles). Although they did not reveal that they would give up on test taking, it was clear that they were very reluctant to have their graduation pending on whether or not they had a certificate of English proficiency.

It is also true that instrumental motivation is more powerful when the learners are required to provide test scores for high stake purposes, especially in the case that they may not be able to receive their diploma. Several of the learners admitted the requirement was a necessary evil that could compel them to spend more time on learning English and on preparing for the GEPT or any other tests. Nevertheless, for the few learners who did not perceive the threshold level as high stakes nor perceive

reaching the level as difficult, they could be motivated by less instrumental benefits. For example, Blair in this study, expressed her preference of English courses with clear objectives and teaching plans for all of the four years in the university and said that such courses, rather than the graduation requirement, might motivate her more to learn English (see 5.6.6.1 for more examples). The ELLI profiles can provide an explanation complementing Tsagari's arguments. Learners who regard themselves as effective learners may be influenced less by the need to provide test scores for high stake purposes as they take charge of their own learning and are more likely to be motivated learners. On the other hand, it may be understood that learners with weaker overall learning power profiles or significantly lower 'resilience' are less motivated learners and can suffer more from the anxiety. These findings are consistent with several studies in the general education field on high-stakes testing and motivation (Benmansour, 1999; Paris et al., 1999; Pollard et al., 2000). Studies like Paris et al. (*ibid.*) and Pollard et al. (*ibid.*) have discovered that the high achieving students were less likely to be affected by high-stakes assessment than the lower achieving ones. Although there was no instruments or actual learning performances in this study that could differentiate students between high or low achievers, the students who were nonchalant towards the implementation of the graduation requirement were indeed those who have already passed the required proficiency standard. In addition, similar to Benmansour's (1999) findings, those students' self-efficacy, as evidenced by their high learning power, was not only related to lower anxiety towards the graduation requirement, but also higher intrinsic goal orientations (self-control on English learning and test taking, studying abroad).

Similarly, learning power of a student as an important factor to understand washback can also fit well into Green's (2006b) model of washback intensity. In other words,

washback intensity may be highest if:

- 1) a learner perceive passing a test as important,
- 2) a learner is an effective learner with high degree of learning power.
- 3) a learner perceive the difficulty of the test or reaching the required standard as challenging, neither too easy or unachievable

It is important to note here that the possibility of the second in predicting the third condition, especially the dimension of 'resilience' in considering whether the learner is likely to be up for a challenge or be too intimidated by the challenge.

6.4.5 Summary

The above sections have demonstrated that Shih's model is not sufficient to understand fully the mechanism of washback on the learners and their learning in the context of this study. One additional extrinsic factor, 'the publishers', should be included in the extrinsic factors. Although local publishers who represent internationally-designed materials are common in EFL contexts, the roles that the local publishers can play in the washback process have not been explored elsewhere, to the best of my knowledge. The intrinsic factors in Shih's model are not as elaborated as other categories. The findings of this study suggest that all of the following intrinsic factors need to be put into consideration: a student's perceptions of test stakes, test difficulty and test importance, and the student's learning power. The interactions of the intrinsic factors can help to explain how washback on the learners and their learning is mediated.

6.5 Summary of the chapter

In this chapter, I first summarised the key findings of this study in the order the three main research questions. I then discussed one significant finding that permeates the whole study in 6.2, i.e., the strong social impact of the GEPT and the role of the

graduation requirement in reinforcing the impact of the test in English teaching and learning in the universities. The discussions were made with reference to Shohamy's framework of the power of tests in influencing language educational policies. I then discussed washback on the learners and their learning, by relating the findings of the present study to Bailey's model of washback, Green's washback model that encompasses test construct, test characteristics, washback variability, washback intensity and also, Shih's model of washback on learning that shares the same context as this study (6.3). One focus is on the operationalisation of the complex relationship between stakes and impact, and the other focus is on the role of learning power in explaining the washback on learners and their learning.

CHAPTER 7 Conclusion and Implications

7.1 Overview

The present study has explored the impact of the graduation requirement for English proficiency on the non-English majors and on their EAP curriculum in Taiwanese universities. Two case universities were chosen as research sites, one with the graduation requirement implemented, University B, and one without, University A. The main data collected included video/audio recordings of classroom observations (4 teachers and 7 classes in University A and 3 teachers and 3 classes in B), individual teacher and learner interviews (4 teachers and 9 learners in A and 3 teachers, 9 learners in B), and ELLI learning power profiles of the students (216 in A, 238 in B).

This research has attempted to better understand the consequences of the implementation of the graduation requirement, particularly on the learners. It has sought to ascertain which test accepted by the graduation requirement has exerted the most influence, and in what ways the non-English majors, who were the main target of the graduation requirement, have been affected. The focus on the learners was the attempt to draw a fuller picture of washback to learners and their learning. This study has further explored how a student's learning power profile could be used to explain washback variability among individual learners.

In the final chapter of the dissertation, I shall reflect on the strengths and limitations of this study (7.2), and suggest ways forward for future studies (7.3). In 7.4, I will discuss the implications of the research findings for different stakeholders in the process of the impact from the graduation requirement. Finally, in 7.5, linking the significant research findings to the reason why I wanted to conduct this study, I conclude by arguing for the

need for future washback and impact studies not to neglect the complex agenda of language tests, and to consider the impact from the learners' eyes.

7.2 Strengths and limitations of the study

7.2.1 Research strengths

The strengths of the present study are evident in three areas: 1) the exploration of the relationship between the learning power of individual learners and their perceptions of test impact; 2) the discussion of the test as perceived not only in the institutions but also in the society; 3) the focus on non-English majors, the major learners that come under the influence of the graduation requirement.

First, the current study has attempted to further explore the learner variables in washback variability among learners, which has been an under-researched (2.3.2.2). The few studies that have focused on learners did not explore the intrinsic variables related to the learners themselves; nor did they go further in discussing the relationship between the learner variables and the learners' perceptions of test influence. This study is significant in its systematic and quantitative measurement of one particular intrinsic learner variable, learning power, among all the learner participants through the ELLI instrument. The significance lies in the fact that few of the previous studies have made this attempt. Furthermore, by relating learning power assessed by ELLI to their interviews, this study has also explored how their learning power might explain their varied perceptions of the impact of the graduation requirement (5.6). Although learning power can only be considered as one among numerous learning variables that can affect test influences, the findings of this study have nevertheless shown that the kind of learners the learners view themselves as can explain the different extent of impact of the graduation requirement they perceive.

In terms of learner participants, this study focused on major stakeholders of the English proficiency requirement for graduation in Taiwanese universities. Previously related studies were limited, either to including only English majors or to conducting research in universities of technology (Shih, 2006, 2007; Tsai and Tsou, 2009). Shih (2006, 2007) admitted that this was a key limitation in his study, which only involved students who were English majors, because it was not possible to differentiate test preparation from their regular English training and learning throughout their four years in university. The key stakeholders are, in fact, non-English majors who usually receive one or two years of English training, and are required to demonstrate their competence in a subject that they once were not obliged to follow in order to graduate. In addition, students in universities of technology are not representative of all university students as there are more general universities than universities on the vocational path in Taiwan. This study is thus important in targeting learners who are at a majority in coming under the influence of the graduation requirement, namely the non-English majors in general universities.

A major strength of the current study is that it has taken into consideration not only how the tests were perceived among stakeholders in the institutions but also how the tests were socially perceived. Related studies with similar contexts mainly focused on the test influences on teaching and learning in the institutions (Shih, 2006, 2007, 2008, 2010; Wu and Chin, 2006; see 2.4), assuming GEPT as the most influential test. In the context where learners are given the freedom to choose from multiple English proficiency tests not related to their English curriculum, such presumption shows an implication of the high social status the GEPT is perceived, despite the reluctance of acknowledgement from the researchers. Since contextual factors are essential in understanding test washback, it makes less sense to exclude contextual factors related

to the society that can be linked to washback in the institutions. This study has examined the test effects induced by the graduation requirement in the universities within a bigger picture of viewing the tests from the social perspective (6.3). Without the macro perspective, it would not be easy to link the GEPT washback on internal testing materials to the GEPT social impact on the local publishers that imported international EFL/ESL materials. Local publishers seem to have played complex roles in the washback process. The findings have also explained why the GEPT was the most influential test among other English proficiency tests accepted by the graduation requirement and also how the requirement has enhanced the GEPT washback. The study showed that the power of a test is determined by its importance as perceived by the public. It is this power that makes an external English proficiency test exert greater influences on teaching and learning in institutions than other external English proficiency tests.

7.2.2 Research limitations

First of all, the limited time and accessibility to conduct classroom observations in University B created several related limitations. The limitations were the unbalanced number of lessons observed between two case universities, and the limited number of sessions on the course directly related to the graduation requirement in University B. Although there was little difference in the length of the research duration in the two universities, the fixed timetable for all 'English II' courses (3.7.3) and the approaching to the end of the semester during the time I conducted research in University B caused some hindrance to the collection of the observation data. Timetable constraints made it impossible for me to observe as many teachers in a week in University A. In addition, with the semester coming to an end, there was no opportunity to extend the research duration, and the observable classes were cut short when some sessions had to be used

for revisions, self study and final exams. As a result, the numbers of the lessons observed in the two universities, i.e. 5 in University B and 17 in University A, were substantially unbalanced. These constraints also meant that there was only one lesson in the GEPT-related course that I could observe as a follow-up observation for Becky's remedial class, 'English Reading and Writing'. Therefore, although little evidence except for the direct reference to the GEPT preparation material was observed in Becky's one and only lesson (4.3), the possibility for more manifestation of GEPT washback in her other lessons throughout the semester when she did use the GEPT material could not be ruled out.

The above constraints also resulted in the restricted selection of teacher participants and their classes for observation. The limited time and accessibility became a hurdle for exploring whether there was 'washback variability' (Green, 2006, p.39, 2.2.4) as there was only one remedial class available for observation at the time of the research. The absence of Becky's reference to the GEPT material might result from her personal dissatisfaction with the material or her own teaching beliefs (4.3) whereas this might not be the case with other teachers teaching the same GEPT-related remedial course (see teachers' difference responses to test preparation in Alderson & Hamp-Lyons, 1997); Burrows, 2004; Watanabe, 2004; Green, 2006, 2007). Furthermore, the limited time available for research made it extremely difficult to compensate the loss of one ideal teacher participant who taught both courses, one related and the other not related to the graduation requirement (3.7.1). Consequently, the comparisons of different teachers teaching the same course and the comparisons of the same teachers teaching exam prep and non-exam courses, suggested by Watanabe (2004) as the ideal design in exploring washback, could not be achieved, although this was planned

Another limitation is that the present study is not a longitudinal one. Longitudinal studies have been particularly fruitful in the exploration of the washback phenomenon. First, the appearance of extensive test preparation as the test approaches compared to the usual English training can be captured. Secondly, observer effect can be greatly diminished as the familiarity between the researcher and teacher and learner participants increases. Such familiarity also helps in eliciting participants' attitudes, beliefs and perceptions more in depth. Thirdly, learning outcomes which usually involve tests or assessments at the end of courses are as important as learner perceptions in exploring washback on learning, and are also essential in understanding the mechanism of washback. It is more likely for longitudinal studies to collect learner performances before and after preparation for a test. In terms of the context of the current study, the short duration of research (4 – 5 weeks in each university) has indeed brought about limitations on the limited evidence of extensive test preparation (Chapter 4), observer effect (4.4 on Ben) and the lack of learning outcomes (Chapter 5).

Lastly, the number of cases under study may also be a limitation in understanding the impact of the graduation requirement for English proficiency. As detailed in 1.3.1, the autonomy of institutions in higher education allows the universities to tailor the regulations of the graduation requirement to their needs, which created different contextual factors that may affect how the requirement impacts on English teaching and learning. Although the fundamental difference between the two cases was evidenced by the different degrees of washback and impact, another research site with regulations different from University B was likely to be able to shed more light on the impact of the graduation requirement.

7.3 Directions for further research

Three directions are suggested here for further research in relation to 1) expanding the length of the study and involving more teacher participants, following Watanabe (2004)'s suggestions, 2) expanding the size of the study in including more case universities and other related stakeholders, and 3) tracking learning in a longitudinal study and expanding the use of ELLI in relation to extended qualitative accounts of the learners.

7.3.1 Expanding the length of the study and involving more teacher participants for classroom observation

This study was conducted for a period of four to five weeks in each case university, involving only one teacher teaching a class related to the graduation requirement. For future research, more teachers teaching both courses related and not related to the graduation requirement and their classes should be involved to allow the comparison between the two types of courses (see Watanabe's suggestions, 2004). The involvement of more teacher participants will also allow the exploration of washback variability among the teachers. Such an exploration will further enable researchers to see the role of the teachers in determining which test among the English proficiency tests they should prepare students for. Research over a longer timescale such as a whole academic year or at least an entire semester, will be helpful in several ways. The extension of research duration will be useful in reducing the observer effect and the reluctance in sharing thoughts with researchers, especially for ethnographic studies that allow the complexity of the washback phenomenon to emerge. It will also be useful to observe whether the extent of test preparation would differ at different periods in the semester, the academic year, or when the deadline of test taking approaches. Although for the context of the current study, the 'real' deadline of presenting an official proof of

English proficiency is the submission for receiving the undergraduate degree, it is still interesting to know whether teachers include more elements of test preparation if they know their students are going to take the GEPT soon.

7.3.2 Expanding the size of the study in including other related stakeholders and more case universities

Equally important as the length of the study and number of teacher participants is the scope of the study. It could have involved a wider range of stakeholders and case universities. The findings of the current study have suggested that other than teachers and the learners, the curriculum designers (a.k.a. English department), the publishers and the parents also play an essential role in shaping GEPT washback on the requirement-related courses, on the internal testing system and also on the learners and their learning. Bailey's washback model (1996) and previous studies that involve other stakeholders, in addition to teachers and learners, have been fruitful in mapping out the ecology of how different stakeholders interact and contribute to the washback process. It is important for future studies to involve those stakeholders, to better understand the impact of the implementation of the graduation requirement. The various stakeholders that are involved in the process from deciding the regulations, deciding the teaching contents and materials of requirement-related courses, to actual teaching and learning within and beyond the classrooms, should all be taken into consideration. In particular, future investigations should focus on the factors that make those stakeholders prioritise the GEPT over other English proficiency tests.

Likewise, future research of a larger scale should also expand the study to include more cases that are similar or different in their regulations for the graduation requirement concerning related curriculum changes, or means of compensation or substitution for

English proficiency test scores. The findings of this study have shown that the implementation of the graduation requirement in one university has indeed created a context different from that of a university without the requirement. Only by taking into account the above mentioned stakeholders will their perceptions, their decisions, their actions and behaviours in the context of a particular institution will the investigation be more complete. It will be interesting to see whether the strong social impact of the GEPT reinforced by the graduation requirement in the case institutions will be reflected in other universities with similar or different regulations.

7.3.3 Tracking learning in longitudinal studies and expanding ELLI along with extended qualitative accounts of learners

Another important issue to consider that the findings and limitations of this study has brought up is the need to track learning for a longer timescale and also to make use of the ELLI along with extended qualitative accounts of learners in a larger scale over a longer period of time. First, from the perspective of washback and impact, I believe it will be more fruitful for future investigations into the impact of the graduation requirement to track non-English majors' English learning longitudinally. Ideally, their learning should be tracked from their first year to the time they graduate, by including their performances in their compulsory EAP courses or any optional English courses, their efforts in test preparation (extracurricular lessons or self-learning), and their attempts at taking English proficiency tests. Since one important purpose of implementing the graduation requirement is to promote university students' English learning and improving their English proficiency, a long-term observation of their learning throughout their four years will be ideal in examining whether the purpose has been realised.

The ELLI can be used as one instrument for the tracking of the students' learning. The findings of this study have suggested how individual learners' learning power may explain the varied extent of impact from the graduation requirement the students perceived. However, the length of the individual student interviews in this study was limited to 5 to 10 minutes, and the contents were mainly related to the impact of the graduation requirement and English proficiency tests but not on what they rated themselves as a learner on the ELLI questionnaire. In order to have a further understanding of the relationship between students' ELLI learning profiles and their perceptions of test washback, it will be important for future research to incorporate longer qualitative accounts from the learners on their perceptions of the requirement impact, their attitudes towards preparation for the graduation requirement, their experience in taking this assessment of learning how to learn, their thoughts on the learner profiles as feedbacks, and also stories on their own learning. These extended qualitative accounts of the learners can thus contribute to deeper understanding of the university students' learning throughout the four years, under the impact of the graduation requirement.

Green's (2007) study, which has related teacher and learner perceptions to measurable learning outcomes, has demonstrated that 'the response of the individual learner to the demands of the test and to other features of the learning context appear to influence outcomes to a greater extent than their choice of course and the content of their classes' (p.314). Therefore, in order to draw a fuller picture of test influences on the learners and their learning, future research may incorporate instruments like ELLI in showing intrinsic learner variables systematically, along with interviews, and learning outcomes on a longer timescale.

7.4 Implications for stakeholders

Based on the findings of the current study, implications for different stakeholders of the graduation requirement for English proficiency can be made, including implications 1) for policy makers, 2) for curriculum designers and decision makers concerning the regulations for the graduation requirement, 3) for teachers and learners.

7.4.1 Implications for policy makers

The research context detailed in 1.3.1 has pointed out that the major purpose of the policies concerning university students' English proficiency is to increase their international competitiveness by raising their English proficiency. That is to say, the intended effects of the graduation requirement of language proficiency, the reactive action of the universities to the policies, are to motivate the university students in English learning, to raise their English proficiency by propelling them to take external English proficiency tests and to pass the designated level. Yet the findings of this study cast doubts on the effectiveness of implementing the graduation requirement in the universities. The effectiveness of the policy is challenged by: the narrowing of the requirement that does not favour any English proficiency tests to simply one test – GEPT, and the limitations of the graduation requirement alone in affecting the students and their English learning.

First, findings on both the impact of the graduation requirement on the EAP curriculum for non-English majors (Chapter 4) and its impact on non-English majors (Chapter 5), albeit limited in extent, were related largely to one particular test. In other words, the impact of the graduation requirement on teaching and learning was almost equivalent to the washback of the GEPT on teaching and learning. Since none of the English proficiency tests were tailored to the English curriculum in the universities, it might be

problematic that what non-English majors learned and were tested in their classes and what they did in their self-learning were geared towards the GEPT alone.

Secondly, the limited effect the graduation requirement had on the non-English majors and their English learning challenged the substantial purpose of the requirement. For those students who took charge of their own learning, they were already highly motivated in English learning and they would prepare themselves for external English proficiency tests even without the requirement. The requirement might impel some of the students to work harder in order for them to graduate but there might still be the possibility that students just chose to give in to the high degree of anxiety and take the easy way out. In addition, much of the non-English majors' experience in test preparation and test taking, or lack of it, was shaped more by the extrinsic factors such as their parents, future studies they would take or credentials for future work applications than by the requirement. Even though the learning outcomes of the non-English majors were not measured and included in the present study, it is still debatable how much the intended effect of implementing the graduation requirement has been realised.

Therefore, the abovementioned issues provide the following implications for the policy makers: 1) the importance of taking into consideration how different English proficiency tests are perceived in the society, 2) the need to think beforehand about the consequences of the policies when brought into practice. The findings of this study has demonstrated that among various English proficiency tests not related to the English curriculum, the test that is perceived as the most important in a wider context will exert the greatest influences on English teaching and learning. In this case, the test is GEPT. As mentioned in 1.3.1, the initial intention of the original policies was to establish a

common index of English proficiency for university students while at the same time relating the index to the CEFR for international recognition. Without considering the different status of the different English proficiency tests available, the policy makers failed to foresee the strong social impact of the GEPT. The power of the GEPT resulted in the discrepancy between the *de facto* policies as realised by the graduation requirement and the original policies. Thus, I believe it is very important for the policy makers to take into account the complex agendas a language test can entail, as argued by Shohamy (1993, 1996, 1998, 2001) before announcing any policies related to language tests. In fact, what is the most important is that the policy makers should not only focus on determining what policies to announce and execute but also on the probable consequences the policies may bring about. Echoing the findings of previous washback and impact studies on curriculum innovation, the results of this study have demonstrated that the real consequences are likely to be different from the intended and expected. Therefore, even though the consequences are not likely to be fully predictable, I believe it is still the policy makers' duty to study the consequences and impact of the implementation of the policies may bring.

7.4.2 Implications for stakeholders in charge of determining the graduation requirement and related curriculum change

Although the graduation requirement for English proficiency may be a reactive action towards the policies related to raising university students' English proficiency, the English departments, the curriculum designers, and the school authorities, instead of the policy makers, are the stakeholders who decide what to be included in the requirement. In the current study, the stakeholders in charge were not only responsible for adding the remedial course in the students' third year for those who couldn't meet the requirement, but also for determining the purpose of the remedial course. It is

important for these stakeholders who hold a predominant position in this affair to understand the significance of the decisions they make to the learners. As discussed in 6.3.2, the English department in University B was essential in shaping the GEPT washback in the remedial course by requiring the teachers to teach reading and writing towards GEPT preparation. The orientation of the course will vary if another test is prioritised or if the English department allows the teachers to decide for themselves and their students what should be learned in that course. The question is, however, on what grounds the English department made the decision in directing the remedial course to be GEPT-oriented and how the grounds can be soundly justified.

The stakeholders who devise the regulations of the graduation requirement should also be aware of the possible deficiencies found in the regulations that can threaten the original purpose of such implementation. Similar to Shih's (2006, 2010) study, the teachers in this study have pointed out some deficiencies such as the impossibility for full execution of the graduation requirement and the possibility to substitute language proficiency certificates with remedial courses or any similar compensation plans (See Extracts 5.12~5.15). These deficiencies seem difficult to avoid, either in universities with high achieving students or those with students that have poor performance in English proficiency. There are likely to be problems afterwards if the stakeholders in charge implement the graduation requirement, overlooking or without being aware of the loopholes.

The issues discussed above suggest that the curriculum designers, the English departments and the school authorities should acknowledge the significance of their decisions on the graduation requirement and should be able to justify how their decisions are made. Moreover, they should not be too optimistic and be aware of the

deficiencies that may challenge the purpose of implementing the requirement. One important thing that they should do is not to neglect the voices of the teachers and especially the learners, but to involve them in the process of decision making. This study has provided evidence of how teacher factors explain the presence or absence of washback on teaching despite being required to teach certain contents or using the same assigned materials (Chapter 4, Watanabe, 1996, 2004). Likewise, the findings have also shown that variability among individual learners is 'central to the understanding of the complex process of washback' (Green, 2007, p.314) and impact of the graduation requirement. Since it is impossible for the different groups of stakeholders to view the graduation requirement and its implementation from the same angle, it is more likely to develop mutual understandings and establish common grounds by involving all of them in the process of decision making.

7.4.3 Implications for teachers and learners

Following the same argument, one implication for teachers and university students is to hold discussions and to reach a consensus with regard to teachers' perceptions of what learners want from the lessons and what the learners actually want to learn. Echoing Green's (2007) argument, the findings of this study have also shown that 'the nature and extent of washback to learners does not bear a transparent relationship to washback to teachers' (p.314). The non-English majors did not always consider what they received from the lessons was good for their English learning nor did they see the graduation requirement the same way as their teachers did. However, there was little evidence showing that such issues had been discussed between them and their teachers. I believe that as the field of ESL/EFL and also the language testing field are now valuing more of the learners' voices, teachers should provide more opportunities for the learners to express their thoughts, their attitudes towards EAP curriculum for them and

the graduation requirement.

Another implication for teachers and learners is the need to pay more attention to students' learning power. The ELLI questionnaire was used in a restricted manner in this study, i.e., only for the purpose of measuring learning power at the individual level. However the learner profiles, as the developers have proposed, are only the starting point in the whole process of using ELLI as a tool for formative assessment (Deakin Crick et al., 2004). They propose that the learner profiles should be distributed back to the learners, because the profiles can 'reflect back to the learners what they say about themselves in relation to their personal power to learn' so that the learners can understand their own learning better (Deakin Crick, 2007, p.150). Deakin Crick (*ibid.*) has also reported the use of the ELLI in classroom-based research for pedagogical purposes. The individual and the collective profiles of learning power of the students can not only be used to diagnose what the learners need to move forward in learning, but also help teachers to devise teaching and learning strategies accordingly to help strengthen their students on the dimensions of learning power. Therefore, it will be fruitful for teachers and learners to recognise and make use of the reflective, diagnostic, scaffolding functions of the ELLI.

7.5 Reflection and concluding remarks

The current study, conducted out of concerns for the 'after' scene, following the announcement of policies and reactive actions from the institutions, has findings that shed light on test influences within and beyond the institutions and the interaction between the two. For test influences within the institutions, consistent with findings of previous washback studies, the findings of this study have demonstrated that variability among teachers and learners at the individual level seem to be more evident than impact

resulted from the graduation requirement. The adoption of the ELLI instrument in presenting individual student's learning power has presented one facet of intrinsic factors that may explain the variability among learners.

On the other hand, the most valuable findings, I believe, are those which provide insights into the impact of the graduation requirement within a big picture that encompass the contextual factors within and beyond the institutions. The presumption that aligns the GEPT with the requirement in previous related studies limits the exploration to studying only the GEPT washback. However, the findings of analysing the classroom observations, teacher and student interviews in this study have pointed out that linking all the evidence of the influences of GEPT directly to the washback of GEPT is a little too arbitrary. Context is crucial in the exploration of test influences (Alderson and Wall, 1993; Watanabe, 2004). Without full recognition and acknowledgement of the complexity of the real contexts underlying the graduation requirement, only a partial picture can be showed. In fact, several instances of GEPT influences in this study are well beyond the control of the institutions. They are brought into the institutions by some stakeholders other than teachers and learners. In order to differentiate between evidence linked to GEPT washback, the social impact of the GEPT (see definitions of washback and impact in 2.2.1) and the interaction between the two, both the institutional and social contexts in which the test is used should be taken into account. It is through this fuller picture that the following argument in the present study can be developed: in the context where various tests are available but none of them are attached to the curriculum, it is the test that is perceived as the most important in the society that exerts the greatest influence.

As a teacher myself, and also once involved in the decision making process for the

implementation of the graduation requirement for English proficiency and the development of supplementary measures, I have always been concerned about the consequences our decisions could bring to our students. On the other hand, I have also realised that this area has received limited interest from the policy makers, the university authorities, the English curriculum designers, or even the fellow teachers. I remembered my students' expressions of anxiety and their opinions when they heard rumours about the implementation. I also felt their relief when they knew the graduation requirement was implemented not for them, but for new-coming students. I then realised that viewing the graduation requirement from only the angle of the decision makers was not enough. What we perceive as best for our students may not be what really is best for them. It should be our responsibility to not only make careful decisions while listening to our students' voices, but also to keep track of the aftermath of policy implementation and evaluate the outcomes of our decisions.

The findings reported here further suggest that we should keep in mind what McNamara and Roever (2006) remind us of:

language testing has a real impact on real people's lives, and we cannot cease our theoretical analyses at the point where the test score is computed. Just like language use, language testing is and has always been a social practice; the very power of tests has a mesmerizing effect on consciousness of their social character.'

(p.8)

The social agenda language tests may entail in the exploration of the test influences and of policies involving language tests should not be neglected in future research.

REFERENCES

- Alderson, J.C., 2004. Foreword. In L. Cheng, Y. Watanabe, & A. Curtis, eds. *Washback in language testing: Research contexts and methods*. Mahwah, New Jersey: Lawrence Erlbaum, pp. ix-xii.
- Alderson, J. & Banerjee, J., 2001. Language testing and assessment: state of the art review, part one. *Language Teaching*, 34(4), 213–236.
- 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.
- Andrews, S., 2004. Washback and curriculum innovation. In L. Cheng, Y. Watanabe, & A. Curtis, eds. *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 37–50.
- Andrews, S., Fullilove, J. & Wong, Y., 2002. Targeting washback--a case-study. *System*, 30(2), 207-223.
- Bachman, L.& Palmer, A., 1996. *Language testing in practice*, Oxford: Oxford University Press.
- Bailey, K., Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Bailey, K.M., 1999. *Washback in language testing*. Princeton, New Jersey: Educational Testing Service. [online] Available at: <http://www.ets.org/Media/Research/pdf/RM-99-04.pdf> [Accessed 03 October, 2010].
- Bassey, M., 1999. *Case study research in educational settings*. Buckingham: Open University Press.
- Benmansour, N. 1999. Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal of Educational Studies*, 4, 1-15.
- Blaxter, L., Hughes, C. & Tight, M., 2006. *How to research*. 3rd ed. Buckingham: Open University Press.
- Bogdan, R.C. & Biklen, S.K., 1998b. *Qualitative research in education: An introduction to theory and methods*. 3rd ed. Needham Heights, MA: Allyn & Bacon.
- British Association for Applied Linguistics (BAAL), 2006. *Revised Recommendations on Good Practice in Applied Linguistics*. [online] Available at: http://www.baal.org.uk/about_goodpractice_full.pdf [Accessed 03 October 2010].

- British Educational Research Association (BERA), 2004. *Revised Ethical Guidelines for Educational Research*. [online]. Available at <<http://www.bera.ac.uk/files/guidelines/ethical1.pdf>>. [Accessed 03 October 2010]
- Broadfoot, P.M., 2005. Dark alleys and blind bends: testing the language of learning. *Language Testing*, 22(2), 123-141.
- Broadfoot, P., Abbot, D., Croll, P., Osborn, M., Pollard, A. & Towler, L., 1991. Implementing National Assessment: issues for primary teachers. *Cambridge Journal of Education*, 21(2), 153-168.
- Brown, J., 2000. University entrance examinations: strategies for creating positive washback on English language teaching in Japan. *Shiken JALT Testing and Evaluation SIG Newsletter*, 3(2), 7-30.
- Burrows, C., 2004. Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In L. Cheng, Y. Watanabe, & A. Curtis, eds. *Washback in language testing: research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 113-128.
- Carr, M. & Claxton, G., 2002. Tracking the Development of Learning Dispositions. *Assessment in Education: Principles, Policy & Practice*, 9(1), 9-37.
- Chan, S. & Pollard, D.E., 2001. *An Encyclopaedia of Translation: Chinese-English, English-Chinese*. Hong Kong: Chinese University Press.
- Chapman, D.W. & Snyder, C.W., 2000. Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development*, 20(6), 457-474.
- Chaudron, C., 1988. *Second language classrooms: Research on teaching and learning*, Cambridge: Cambridge University Press.
- Chen, L. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the Basic Competence Test in English*. Unpublished PhD thesis, College of Education, The Ohio State University, Ohio, United States.
- Cheng, L., 1997. How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11(1), 38-54.
- Cheng, L., 1998. Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.
- Cheng, L., 1999. Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, 15(3), 253-271.
- Cheng, L., 2005. *Changing language teaching through language testing: A washback study*, Cambridge University Press.
- Cheng, L., Watanabe, Y. & Curtis, A., 2004. *Washback in language testing: Research contexts and methods*, Mahwah, NJ: Lawrence Erlbaum.
- Coffield, F., Moseley, D., & Hall, E. 2004. *Learning Styles and Pedagogy in Post-16 Learning: A systematic and critical review*. London: Learning and Skills Research Centre.

- Cohen, L., Manion, L. & Morrison, K., 2007. *Research Methods in Education*. 6th ed. London: Routledge.
- Council of Europe, 2001. *Common European Framework of Reference for Languages: learning, teaching and assessment*. Cambridge: Cambridge University Press.
- Crooks, T.J., 1988. The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Davies, A., 1985. Follow my leader: Is that what language tests do. In Y. Lee et al., ed. *New Directions in Language Testing*. Oxford: Pergamon, pp. 3-13.
- Davies, A., 1990. *Principles of language testing*, Oxford: Blackwell.
- Deakin Crick, R., 2007. Learning how to learn: the dynamic assessment of learning power. *Curriculum Journal*, 18(2), 135-153.
- Deakin Crick, R., Broadfoot, P. & Claxton, G., 2004. Developing an Effective Lifelong Learning Inventory: the ELLI - Project. *Assessment in Education: Principles, Policy & Practice*, 11(3), 247-272.
- Deakin Crick, R. & Yu, G., 2008. Assessing learning dispositions: is the Effective lifelong learning inventory valid and reliable as a measurement tool? *Educational Research*, 50(4), 387-402.
- Delanty, G., 2005. *Social science: philosophical and methodological foundations*, Buckingham: Open University Press.
- Derrida, J. 1982. *Margins of Philosophy*. Chicago: University of Chicago Press.
- Dornyei, Z., 2007. *Research methods in applied linguistics*, Oxford: Oxford University Press.
- Entwistle, N. (1988). Motivational factors in students: Approaches to learning. In R.R. Schmeck, ed. *Learning Strategies and Learning Styles*. New York and London: Plenum, pp.21-49.
- Entwistle, N., McCune, V. & Walker, P. (2001). Conceptions, styles and approaches within higher education: Analytic abstractions and everyday experience. In R.J. Sternberg and L.F. Zhang, eds. *Perspectives on Thinking, Learning and Cognitive Styles*. Mahwah, New Jersey: Lawrence Erlbaum, pp.103-136.
- Evans, B.A. & Hornberger, N.H., 2005. No Child Left Behind: Repealing and unpeeling federal language education policy in the United States. *Language Policy*, 4(1), 87-106.
- Fassnacht, C. & Woods, D., 2005. Transana v2.0x. Madison, Wisconsin: The Board of Regents of the University of Wisconsin System.
- Ferman, I., 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 context and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 191-210.
- Foucault, M., 1982. The subject and power. *Critical Inquiry*, 8(4), 777-795.

- Fox, J. & Cheng, L., 2007. Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education: Principles, Policy & Practice*, 14(1), 9-26.
- Frederiksen, J.R. & Collins, A., 1989. A systems approach to educational testing. *Educational Researcher*, 18(9), 27-32.
- Fröhlich, M., Spada, N. & Allen, P., 1985. Differences in the communicative orientation of L2 classrooms. *TESOL Quarterly*, 19(1), 27-57.
- Gates, S., 1995. Exploiting washback from standardized tests. In J. Brown & S. Yamashita, eds. *Language testing in Japan*. Tokyo: Japan Association for Language Teaching, pp. 101-106.
- Gibson, C. & Swan, A., 2008. *The significance of sociolinguistic backgrounds of teachers of IELTS Test preparation courses in selected Malaysian institutions.*, unpublished IELTS research report.
- Gipps, C., 1994. *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Gosa, C., 2004. *Investigating washback: A case study using students' diaries*. Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, United Kingdom.
- Green, A., 2006a. Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113-134.
- Green, A., 2006b. Watching for washback: Observing the influence of the International English Language Testing System academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333-368.
- Green, A., 2007a. Washback to learning outcomes: a comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education: Principles, Policy & Practice*, 14(1), 75-97.
- Green, A., 2007b. *IELTS Washback in Context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Guba, E.G. & Lincoln, Y.S., 1994. Competing paradigms in qualitative research. In E. Guba & Y. Lincoln, eds. *Handbook of qualitative research*. Thousand Oaks: Sage Publications, pp. 105-117.
- Haladyna, T.M., 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.
- Hammersley, M. & Atkinson, P., 2007. *Ethnography: Principles in Practice*. 3rd ed. London: Routledge.
- Hamp-Lyons, L., 1997. Washback, impact and validity: ethical concerns. *Language Testing*, 14(3), 295-303.
- Hargreaves, E., 1997. The diploma disease in Egypt: learning, teaching and the monster of the secondary leaving certificate. *Assessment in Education: Principles, Policy & Practice*, 4(1), 161-176.

- Harlen, W. & Deakin Crick, R., 2003. Testing and Motivation for Learning. *Assessment in Education: Principles, Policy & Practice*, 10(2), 169.
- Hawkey, R., 2006. *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*, Cambridge: Cambridge University Press.
- Hayes, B. & Read, J., 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 context and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 97-112.
- Herman, J.L. & Golan, S., 1993. The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12(4), 20–25.
- Heron, J. & Reason, P., 1997. A participatory inquiry paradigm. *Qualitative inquiry*, 3(3), 274.
- Holliday, A., 2007. *Doing and writing qualitative research*. 2nd ed. Thousand Oaks: Sage Publications.
- Hughes, A., 1993. *Backwash and TOEFL 2000*. Unpublished manuscript, University of Reading.
- Hughes, A., 2003. *Testing for language teachers*. 2nd ed., Cambridge: Cambridge University Press.
- International Language Testing Association, 2000. *Code of Ethics*. [online]. Available at: http://www.iltaonline.com/images/pdfs/ILTA_Code.pdf [Accessed 03 October 2010].
- Johnson, D.M., 1992. *Approaches to research in second language learning*, New York: Longman.
- Jones, B.D., 2007. The unintended outcomes of high-stakes testing. *Journal of Applied School Psychology*, 23(2), 65–86.
- Jones, M.G. & Hargrove, T.Y., 2003. *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield Pub Inc.
- Katz, L.G., 1985. Dispositions in early childhood education. *ERIC/EECE Bulletin*, 18(2), 1–3.
- Kunnan, A.J., 2009. Testing for citizenship: The US Naturalization Test. *Language Assessment Quarterly*, 6(1), 89–97.
- Kunnan, A.J., 2009. Politics and legislation in citizenship testing in the United States. *Annual Review of Applied Linguistics*, 29(1), 37-48.
- LeCompte, M.D. & Preissle, J., 1993. *Ethnography and qualitative design in educational research*. 2nd ed. New York: Academic Press.
- Language Training and Testing Centre. *General English Proficiency Test*. [online] Available at: <http://www.lttc.ntu.edu.tw/geptpracticee.htm> [Accessed 03 October, 2010]

- Liu, Y., 2007. An exploration of the assessment of learning power with 13 and 14 year olds in a Chinese language training centre. Unpublished master thesis, Graduate School of Education, University of Bristol, Bristol, United Kingdom.
- LTTC Annual Report, 2008. [online] Available at: <<http://www.lttc.ntu.edu.tw/Annualreportfiles/2008annualreport.pdf>>. [Accessed 03 October, 2010].
- LTTC College Student English Proficiency Test Statistical Report, 2008. [online]. Available at: <<http://www.lttc.ntu.edu.tw/academics/CSEPT/九十七年大學校院英語能力測驗成績統計分析.pdf>> [Accessed 30 Dec. 2010].
- Lyotard, J.F., 1984. *The Postmodern Condition: A Report on Knowledge*. Minnesota: University of Minnesota Press.
- Lynch, B.K., 2001. Rethinking assessment from a critical perspective. *Language Testing*, 18(4), 351-372.
- Lynch, B. & Hamp-Lyons, L., 1999. Perspectives on research paradigm and validity: Tales from the Language Testing Research Colloquium. *Melbourne Papers in Language Testing*, 8, 57-93.
- Madaus, G.F. & Kellaghan, T., 1992. Curriculum evaluation and assessment. In P. Jackson, ed. *Handbook of research on curriculum*, pp. 119-154.
- Madaus, G., 1988. The influence of testing on the curriculum. In L. Tanner, ed. *Critical issues in curriculum*. Chicago, Illinois: Chicago University Press, pp. 83-121.
- Markee, N., 1997. *Managing curricular innovation*, Cambridge: Cambridge University Press.
- McCallum, B., Gipps, C., McAlister, S. & Brown, M., 1995. National curriculum assessment: Emerging models of teacher assessment in the classroom. In H. Torrance, ed. *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment*. Buckingham: Open University Press, pp. 57-87.
- McNamara, T.F., 2000. *Language testing*, Oxford: Oxford University Press, USA.
- McNamara, T.F., 2001. Language assessment as social practice: Challenges for research. *Language Testing*, 18(4), 333-349.
- McNamara, T.F., 2005. 21st century shibboleth: Language tests, identity and intergroup conflict. *Language Policy*, 4(4), 351-370.
- McNamara, T.F., 2006. Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly*, 3(1), 31-51.
- McNamara, T.F. & Roever, C., 2006. *Language testing: The social dimension*, Malden, MA and Oxford: Wiley-Blackwell.
- Mehrens, W.A., 1998. Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6(13). [online] Available at: <<http://epaa.asu.edu/ojs/article/viewFile/580/703>>. [Accessed 03 October 2010].

- Mehrens, W.A., 2002. Consequences of assessment: What is the evidence? In G. Tindal & T. Haladyna, eds. *Large-scale assessment programs for all students: validity, technical adequacy, and implementation*. Mahwah, New Jersey: Lawrence Erlbaum, pp. 149–177.
- Menken, K., 2006. Teaching to the test: How No Child Left Behind impacts language policy, curriculum and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521–546.
- Menken, K., 2008. *English learners left behind: Standardized testing as language policy*. Clevedon, England: Multilingual Matters.
- Menken, K., 2009. No Child Left Behind and its effects on language policy. *Annual Review of Applied Linguistics*, 29, 103–117.
- Messick, S., 1989. Validity. In R.L. Linn, ed, *Educational Measurement*. 3rd ed. New York: ACE and Macmillan, pp. 13-103
- Messick, S., 1996. Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Mickan, P. & Motteram, J., 2008. *An ethnographic study of classroom instruction in an IELTS preparation program*, unpublished IELTS research report.
- Miles, M.B. & Huberman, A.M., 1994. *Qualitative data analysis: An expanded sourcebook*, Thousand Oaks: SAGE publications.
- Mitchell, R., Parkinson, B. & Johnstone, R., 1981. The foreign language classroom An observational study. Stirling Educational Monographs no.9. Stirling, Department of Education, University of Stirling.
- Morrison, K. & Tang, F.H., 2002. Testing to destruction: a problem in a small state. *Assessment in Education: Principles, Policy & Practice*, 9(3), 289–317.
- Morrow, K., 1986. The evaluation of tests of communicative performance. In M. Portal, ed. *Innovations in Language Testing*. Windsor: NFER-Nelson, pp. 1-13.
- National Taiwan University, 2002. *Guo li Taiwan da xue jin jie ying yu ke cheng shi xing ban fa*. [National Taiwan University Guidelines for Advanced English Study]. [online] Available at: <<http://curri.aca.ntu.edu.tw/cur/english.htm>>. [Accessed 03 October 2010].
- Newsfield, T. (2005). TOEIC washback effects on teachers: A pilot study at one university faculty. *Toyo University Keizai Ronshu*, 31(1), 83-106. [online]. Available at: <<http://www.tnewfields.info/Articles/PDF/Newfields-Washback.pdf>>. [Accessed 30 Dec. 2010].
- Nunan, D., 1992. *Research methods in language learning*, Cambridge: Cambridge University Press.
- Ochs, E., 1979. Transcription as theory. In E. Ochs & B. Schieffelin, eds. *Developmental Pragmatics*. New York: Academic Press, pp. 43-72.
- Oxenham, J., 1984. *Education versus qualifications?* London: Allen & Unwin.

- Pan, Y.C., 2009. The impact of test design on teaching. *The International Journal of Educational and Psychological Assessment*, 3, 94-103.
- Paris, S.G., Lawton, T., Turner, J. & Roth, J., 1991. A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20.
- Patton, M.Q., 2002. *Qualitative research and evaluation methods*. Thousand Oaks: Sage Publications, Inc.
- Pearson, I., 1988. Tests as levers for change. In D. Chamberlain & R. Baumgardner, eds. *ESP in the classroom: practice and evaluation*. London: Modern English Publications, pp. 98-107.
- Phillips, H., 1960. Problems of translation and meaning in fieldwork. In R. Adams & J. Preiss, eds. *Human organisation research: Field relations and techniques*. Homewood, IL: Dorsey, pp. 290-307
- Pollard, A., Triggs, P., Broadfoot, P., McNess, E. & Osborn, M. (2000). *What Pupils Say: changing policy and practice in primary education*. London: Continuum.
- Popham, W.J., 1987. The merits of measurement-driven instruction. *Phi Delta Kappan*, 68(9), 679-682.
- 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 context and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 171-190.
- Qi, L., 2005. Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142-173.
- Qi, L., 2007. Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice*, 14(1), 51-74.
- Rea-Dickins, P. & Scott, C., 2007. Washback from language tests on teaching, learning and policy: evidence from diverse settings. *Assessment in Education: Principles, Policy and Practice*, 14(1), 1-7.
- Rea-Dickins, P., Kiely, R., & Yu, G., 2007. Student identity, learning and progression: the affective and academic impact of IELTS on 'successful' candidates, in McGovern, P and Walsh, S (Eds.), *IELTS Research Reports Volume 7*, (pp. 59-136), British Council and IELTS.
- Robb, T. & Ercanbrack, J. 1999. A study of the effects of direct test preparation on the TOEIC scores of Japanese university students. *TESL-EJ*, 3 (4) A-2. (p. 1-22). [online]. Available at: <<http://www.tesl-ej.org/wordpress/issues/volume3/ej12/ej12a2/>> [Accessed 30 Dec. 2010].
- Saif, S., 2006. Aiming for positive washback: a case study of international teaching assistants. *Language Testing*, 23(1), 1-34.
- Scott, C., 2005. *Washback in the UK primary context with EAL learners: exploratory case studies*. Unpublished PhD thesis, Graduate School of Education, University of Bristol, Bristol, England.

- Scott, C., 2007. Stakeholder perceptions of test impact. *Assessment in Education: Principles, Policy & Practice*, 14(1), 27-49.
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 4, 453-472.
- Scouller, K. & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education*, 19, 267-279.
- Shepard, L.A., 1993. Evaluating test validity. *Review of Research in Education*, 19(1), 405-484.
- Shih, C.M., 2006. *Perceptions of the General English Proficiency Test and its washback: A case study at two Taiwan technological institutes*. Unpublished PhD thesis, Department of Curriculum, Teaching and Learning, Ontario Institute for Studies in Education, University of Toronto, Toronto, Canada.
- Shih, C.M., 2007. A new washback model of students' learning. *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes*, 64(1), 135-161.
- Shih, C.M., 2008. The General English Proficiency Test. *Language Assessment Quarterly*, 5(1), 63-76.
- Shih, C.M., 2009. How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8(2), 189-206.
- Shih, C.M., 2010. The washback of the General English Proficiency Test on university policies: a Taiwan case study. *Language Assessment Quarterly*, 7(3), 234-254.
- Shohamy, E., 1992. Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. *Modern Language Journal*, 76(4), 513-521.
- Shohamy, E., 1993. *The power of tests: The impact of language tests on teaching and learning (NFLC Occasional Papers)*, National Foreign Language Centre: Washington D.C.).
- Shohamy, E., 1998. Critical language testing and beyond. *Studies in Educational Evaluation*, 24(4), 331-345.
- Shohamy, E., 2001. *The power of tests: A critical perspective on the uses of language tests*. 2nd ed. London: Pearson Education.
- Shohamy, E., 2004. Assessment in multicultural societies: Applying democratic principles and practices to language testing. In B. Norton & K. Toohey, eds. *Critical pedagogies and language learning*. New York/London: Cambridge University Press, pp. 72-93.
- Shohamy, E., 2006. *Language policy: Hidden agendas and new approaches*, London: Routledge.
- Shohamy, E., 2007. Language tests as language policy tools. *Assessment in Education: Principles, Policy & Practice*, 14(1), 117-130.

- Shohamy, E., 2008. Language Policy and Language Assessment: The Relationship. *Current Issues in Language Planning*, 9(3), 363–373.
- Shohamy, E., Donitsa-Schmidt, S. & Ferman, I., 1996. Test impact revisited: washback effect over time. *Language Testing*, 13(3), 298-317.
- Smith, M.L., 1991a. Meanings of test preparation. *American Educational Research Journal*, 28(3), 521-542.
- Smith, M.L., 1991b. Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Smith, M.J., 1998. *Social science in question*, Thousand Oaks: Sage Publications.
- Spada, N. & Fröhlich, M., 1995. *COLT: communicative orientation of language teaching observation scheme: coding conventions and applications*, National Centre for English Language Teaching and Research, Macquarie University, Sydney.
- Spolsky, B., 1994. The examination-classroom backwash cycle: Some historical cases. In D. Nunan, V. Berry, & R. Berry, eds. *Bringing about change in language education*. Hong Kong: University of Hong Kong, Dept. of Curriculum Studies, pp. 55–66.
- Spolsky, B., 1995. *Measured words: The development of objective language testing*, Oxford: Oxford University Press.
- Spratt, M., 2005. Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
- Stake, R.E., 1995. *The art of case study research*. Thousand Oaks: Sage Publications.
- Stake, R., 2005. Qualitative case studies. In N. Denzin & Y. Lincoln, eds. *Handbook of Qualitative Research*. Thousand Oaks: Sage Publications, pp.433-466.
- 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 context and methods*. Mahwah, New Jersey: Lawrence Erlbaum, pp. 53-72.
- Taiwan Ministry of Education. Projection 2005-2008. [online] Available at : <http://www.moe.gov.edu.tw> >. [Accessed 03 October, 2010].
- Thomson, K., and Falchikov, N. (1998). Full on until the sun comes out: The effects of assessment on student approaches to studying. *Assessment and Evaluation in Higher Education*, 23, 379-390.
- Temple, B. & Young, A., 2004. Qualitative research and translation dilemmas. *Qualitative Research*, 4(2), 161-178.
- TOEIC Taiwan Scores Statistical Report, 2008. [online]. Available at: http://www.toEIC.com.tw/pdf/2008-1-12_TOEIC.pdf>. [Accessed 30 Dec. 2010].

- Tsagari, C., 2006. *Investigating the washback effect of a high-stakes EFL exam in the Greek context: participants' perceptions, material design and classroom applications*. Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, United Kingdom.
- Tsagari, D., 2007. Review of washback in language testing: How has been done? What more needs doing? ERIC [online] Available at: <http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED497709> .> [Accessed 03 October 2010].
- Tsagari, D., 2009. *The complexity of test washback: An empirical study*, Frankfurt am Main: Peter Lang Publishing.
- Tsai, Y. & Tsou, C.H., 2009. A standardised English Language Proficiency test as the graduation benchmark: student perspectives on its application in higher education. *Assessment in Education: Principles, Policy & Practice*, 16(3), 319–330.
- University of Bristol, 2008. Graduate School of Education Ethics procedure. [online] Available at: <http://www.bristol.ac.uk/education/research/networks/ethicnet/procedures> [Accessed 03 October 2010].
- Van Lier, L., 1988. *The classroom and the language learner*. London: Longman.
- Vernon, P.E., 1956. *The measurement of abilities*. London: University of London Press.
- Vongpumivitch, V., 2010. The General English Proficiency Test. In L. Cheng, & A. Curtis, eds. *English language assessment and the Chinese learner*. New York: Routledge, pp. 158-172.
- 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., 1997. Impact and washback in language testing. In C. Clapham & D. Corson, eds. *Language Testing and Assessment. Encyclopedia of language and education*. Dordrecht; Boston: Kluwer, pp.291-302
- Wall, D., 2000. The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System*, 28(4), 499–509.
- Wall, D., 2005. *The impact of high-stakes examinations 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), 41-69.
- Wall, D. & Horak., 2006. The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the Baseline study. TOEFL monograph series. Report number: RR-06-18, TOEFL-MS-34. Princeton, NJ: Educational Testing Service. [Accessed 15 May 2010].
- Wall, D. & Horak, T., 2007. Using baseline studies in the investigation of test impact. *Assessment in Education: Principles, Policy & Practice*, 14(1), 99-116.

- Watanabe, Y., 1996. Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318-333.
- Watanabe, Y., 1997. *The washback effects of the Japanese university entrance examinations of the English-classroom-based research*. Unpublished PhD thesis, Department of Linguistics and Modern English Language, Lancaster University, Lancaster, England.
- Watanabe, Y. 2001. Does the university entrance examination motivate learners? A case study of learner interviews. In A. Murakami, ed., Akita Association of English Studies., *Trans-Equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*. Akita, Japan: Akita University, pp. 100-110.
- Watanabe, Y. 2004a. Methodology in washback studies. In L. Cheng, Y. Watanabe & A. Curtis, eds. *Washback in language testing: Research context and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 19-38.
- Watanabe, Y. 2004b. Teacher factors mediating washback. In L. Cheng, Y. Watanabe & A. Curtis, eds. *Washback in language testing: Research context and methods*. Mahwah, NJ: Lawrence Erlbaum, pp. 129-146.
- Wiseman, S., 1961. *Examinations and English education*, Manchester: Manchester University Press.
- Wood, D. & Fassnacht, C., 2005. Transana software. *Wisconsin Center for Education Research, University of Wisconsin-Madison*. (Accessed May, 15, 2010).
- Wu, R., & Chin, J., 2006. An impact study of the Intermediate-Level GEPT. *Proceedings of the Ninth International Conference on English Language Testing in Asia, Taipei*, 41-65.
- Wu, J., & Wu, R., 2007. Using the CEFR in Taiwan: The perspective of a local examination board. Paper presented at the Fourth European Association for Language Testing and Assessment (EALTA) Conference, Sitges, Spain.
- Yin, R.K., 2003. *Applications of case study research*, Thousand Oaks: Sage Publications.
- Yin, R.K., 2008. *Case study research: Design and methods*, Thousand Oaks: SAGE publications.
- http://www.toeic.com.tw/toeic_news_31.jsp. [Accessed 30 Dec. 2010].

Appendix A GEPT Test Construct

Intermediate Level (GEPT)

General Level Description

An examinee who passes this level can use basic English to communicate about topics in daily life. His/her English ability is roughly equivalent to that of a high school graduate in Taiwan. Intermediate level English ability is recommended for: administrative, marketing, and sales personnel; technicians; nurses; hotel reception personnel; switchboard operators; police officers; tourism industry workers.

Skill-Area Descriptions

Listening

An examinee who passes this level can understand general conversation in daily life situations and grasp the general meaning of public announcements, weather forecasts, and advertisements. At work, he/she can understand simple product introductions and operating instructions. He/she can catch the general meaning of native English speakers' conversations and inquiries.

Reading

An examinee who passes this level can read short essays, short stories, personal letters, advertisements, leaflets, brochures, and instruction manuals. At work, he/she can read job-related information, company notices and operation manuals, as well as routine documents, faxes, telegrams and e-mail messages.

Writing

An examinee who passes this level can write simple messages and narratives. He/she can write about things he/she has learned and use simple English to write about his/her own experiences or about topics with which he/she is familiar.

Speaking

An examinee who passes this level can use simple English to discuss or describe general daily life topics, introduce his/her daily life, work, family, and experiences, and state his/her outlook on general topics. At work, he/she can ask and answer basic questions and can carry on basic conversations with native English speakers in social settings.

< http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_i.htm > [Accessed 17 March 2007]

Test Formats and Structures

Stage	Component	Part	Item Type	Time(mins.)
First	Listening	1	Picture Description	30 (approx.)
		2	Question or Statement Response	
		3	Short Conversation	
	Reading	1	Vocabulary & Structure	45
		2	Cloze	
		3	Reading Comprehension	
Second	Writing	1	Translation	40
		2	Guided Writing	
	Speaking	1	Reading Aloud	15 (approx.)
		2	Answering Questions	
		3	Picture Description	

High-Intermediate Level

General Level Description

An examinee who passes this level has a generally effective command of English; he/she is able to handle a broader range of topics, and although he/she makes mistakes, these do not significantly hinder his/her ability to communicate. His/her English ability is roughly equivalent to that of a university graduate in Taiwan whose major was not English.

High-Intermediate level English ability is recommended for: business professionals; secretaries; engineers; research assistants; airline flight attendants; airline pilots; air traffic controllers; customs officials; tour guides; foreign affairs police; news media personnel; information management personnel.

Skill-Area Descriptions

Listening

An examinee who passes this level can understand conversations in social settings and grasp the general meaning of lectures, news reports, and TV/radio programs. At work, he/she can understand brief reports, discussions, product introductions, and operating instructions.

Reading

An examinee who passes this level can read written messages, instruction manuals, newspapers, and magazines. At work, he/she can read general documents, abstracts, meeting minutes, and reports.

Writing

An examinee who passes this level can write general work-related reports and messages. In addition to topics related to daily life, he/she can write about current events and more complex or abstract subjects.

Speaking

An examinee who passes this level can fluently express his/her opinion about issues he/she is interested in. At work, he/she can socialize with native English speakers; explain the contents of a task or project; conduct business transactions and participate in discussions; give brief reports or express his/her opinions in a meeting.

< http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_hi.htm > [Accessed 17 March 2007]

Test Formats and Structures

Stage	Component	Part	Item Type	Time(mins.)
First	Listening	1	Question or Statement Response	35(approx.)
		2	Short Conversation	
		3	Short Talk	
	Reading	1	Vocabulary & Structure	50
		2	Cloze	
		3	Reading Comprehension	
Second	Writing	1	Translation	50
		2	Guided Writing	
	Speaking	1	Answering Questions	20 (approx.)
		2	Picture Description	
		3	Discussion	

**Appendix B Effective Lifelong Learning Inventory: validated English version
and ELLI items corresponding to dimensions of learning power**

1	I like it when I have to try really hard to understand something.
2	When I am really interested in something I find it easy to learn.
3	There is at least one person at home who is an important guide for me in my learning.
4	I like to question the things that I am learning.
5	I prefer to work on a problem on my own.
6	Getting to the bottom of things is more important to me than getting a good mark.
7	When I am learning something new I try to think of other things that I have already learned
8	I sometimes do something different when I am learning just to see what will happen.
9	If I get stuck with a learning task I can usually think of something to do to get round the problem.
10	When learning is hard, it's usually because I didn't have enough help.
11	Stories help me in my learning
12	I often have a good idea of how long something is going to take me to learn.
13	When I'm stuck I don't usually know what to do about it.
14	I prefer an interesting question to an easy answer.
15	I tend to avoid trying to learn new things because I don't like feeling confused and uncertain.
16	I like to think things out logically and carefully when I'm learning.
17	When I'm not able to master something, it's usually because I don't know how to go about it.
18	I like it when I can make connections between new things I am learning and things I already know.
19	I like to try out new learning in different ways.
20	When I find learning boring I can usually find a way to make it interesting.
21	I like to have a good reason to learn something.
22	I often use my imagination when I'm learning.
23	If I wait quietly, good ideas sometimes just come to me.
24	When I don't understand something I tend to struggle with it for a while.
25	When learning is hard, I tend to find it interesting.
26	I like working on problems with other people.
27	I like learning new things when I can see how they make sense for me in my life.
28	When I have to struggle to learn something I think it's probably because I'm not very clever.

29	I like to learn about things that really matter to me.
30	I feel that my family is an important source of learning for me.
31	When I learn things well it's usually because I had plenty of help.
32	I'm continually improving as a learner.
33	I get my best ideas when I just let my mind float free.
34	I can feel myself improving as a learner.
35	I don't like to accept an answer till I've worked it out for myself.
36	I like facing new challenges.
37	It is important to respect other people's views even if they are different from mine.
38	I usually learn best on my own.
39	Talking things through with my friends helps me to learn.
40	I feel it's OK to try different things out in my learning.
41	I enjoy discussing difficult problems with my friends.
42	I expect to go on learning for a long time.
43	My learning ability stays pretty much the same.
44	There is at least one person in my community who is an important guide for me in my learning.
45	When I have trouble learning something I tend to get upset.
46	If I do get upset when I'm learning, I'm quite good at making myself feel better.
47	Sometimes when I start a task I don't know what I am going to do until I see my friends getting on with it.
48	I like to be able to improve the way I do things.
49	I often change the way I do things as a result of what I have learned.
50	I find it difficult when my teacher criticises my work.
51	I can usually work well with people I have just met
52	I know that if something is important I can find a way to learn it.
53	I know I can learn in my own way, even if my friends think it's a waste of time.
54	I find it difficult when I am not given clear instructions about how to do something.
55	There is at least one person at home who has helped me to learn.
56	I like to try out new ways of doing things even if there is very little time.
57	I know it's easy to learn if all my friends are learning the same as me.
58	I learn equally well on my own and with others.
59	I have ways of making myself learn if I don't feel like learning.
60	I think about everything that I will need before I begin a task.
61	I need others to be able to learn.
62	I make connections between what I am learning and what I have learned before.
63	I always do the same thing when I am learning.

64	I know I can find a way of solving a problem if I have enough time to think.
65	I like to find my own ways of doing things.
66	Sometimes good ideas just come into my head.
67	I like to imagine how other people might feel and think about things.
68	I find learning hard when I am not told how to go about it.
69	I like to find my own ways of doing things even if everybody else is doing it a different way.
70	I always begin a task without thinking about it first.
71	I often look back and think about what I have learned.
72	I am able to use other people's suggestions in my work.

● **ELLI Items corresponding to the seven dimensions of learning power**

Changing and learning

32	I'm continually improving as a learner.
34	I can feel myself improving as a learner.
42	I expect to go on learning for a long time.
48	I like to be able to improve the way I do things.

Critical curiosity

1	I like it when I have to try really hard to understand something.
2	When I am really interested in something I find it easy to learn.
4	I like to question the things that I am learning.
6	Getting to the bottom of things is more important to me than getting a good mark.
14	I prefer an interesting question to an easy answer.
16	I like to think things out logically and carefully when I'm learning.
25	When learning is hard, I tend to find it interesting.
35	I don't like to accept an answer till I've worked it out for myself.
36	I like facing new challenges.

Meaning making

7	When I am learning something new I try to think of other things that I have already learned
18	I like it when I can make connections between new things I am learning and things I already know.
21	I like to have a good reason to learn something.
27	I like learning new things when I can see how they make sense for me in my life.

29	I like to learn about things that really matter to me.
37	It is important to respect other people's views even if they are different from mine.
62	I make connections between what I am learning and what I have learned before.

Fragility and dependence

10	When learning is hard, it's usually because I didn't have enough help.
13	When I'm stuck I don't usually know what to do about it.
15	I tend to avoid trying to learn new things because I don't like feeling confused and uncertain.
17	When I'm not able to master something, it's usually because I don't know how to go about it.
24	When I don't understand something I tend to struggle with it for a while.
28	When I have to struggle to learn something I think it's probably because I'm not very clever.
31	When I learn things well it's usually because I had plenty of help.
43	My learning ability stays pretty much the same.
45	When I have trouble learning something I tend to get upset.
47	Sometimes when I start a task I don't know what I am going to do until I see my friends getting on with it.
50	I find it difficult when my teacher criticises my work.
54	I find it difficult when I am not given clear instructions about how to do something.
57	I know it's easy to learn if all my friends are learning the same as me.
61	I need others to be able to learn.
63	I always do the same thing when I am learning.
68	I find learning hard when I am not told how to go about it.
70	I always begin a task without thinking about it first.

Creativity

8	I sometimes do something different when I am learning just to see what will happen.
11	Stories help me in my learning
19	I like to try out new learning in different ways.
22	I often use my imagination when I'm learning.
23	If I wait quietly, good ideas sometimes just come to me.
33	I get my best ideas when I just let my mind float free.
40	I feel it's OK to try different things out in my learning.
56	I like to try out new ways of doing things even if there is very little time.
66	Sometimes good ideas just come into my head.

67	I like to imagine how other people might feel and think about things.
----	---

Learning Relationships

3	There is at least one person at home who is an important guide for me in my learning.
5	I prefer to work on a problem on my own.
26	I like working on problems with other people.
30	I feel that my family is an important source of learning for me.
38	I usually learn best on my own.
39	Talking things through with my friends helps me to learn.
41	I enjoy discussing difficult problems with my friends.
44	There is at least one person in my community who is an important guide for me in my learning.
51	I can usually work well with people I have just met
55	There is at least one person at home who has helped me to learn.
58	I learn equally well on my own and with others.
72	I am able to use other people's suggestions in my work.

Strategic Awareness

9	If I get stuck with a learning task I can usually think of something to do to get round the problem.
12	I often have a good idea of how long something is going to take me to learn.
20	When I find learning boring I can usually find a way to make it interesting.
46	If I do get upset when I'm learning, I'm quite good at making myself feel better.
49	I often change the way I do things as a result of what I have learned.
52	I know that if something is important I can find a way to learn it.
53	I know I can learn in my own way, even if my friends think it's a waste of time.
59	I have ways of making myself learn if I don't feel like learning.
60	I think about everything that I will need before I begin a task.
64	I know I can find a way of solving a problem if I have enough time to think.
65	I like to find my own ways of doing things.
69	I like to find my own ways of doing things even if everybody else is doing it a different way.
71	I often look back and think about what I have learned.

Appendix C Field note for Amy's EIST II class, lesson on 07.05.2008

Data Log

The data log is a list of data collected during each classroom observation, and includes data to be collected by the end of the week if not available at the time.

Lesson: 英文(二)大二下學期 (trans: EIST II, second semester of Year II) lower level of proficiency

Date: <u>7 May 2008</u>	Time: 13:10~15:00	
CT (Class Teacher): <u>Amy</u>	Data Collected:	Collected?
	Audio files	N/A
Total students in Class: <u>30-40</u>	Video files	Yes
Physical conditions (noise, temperature, light, interruptions): Language labs with no monitors in front of each student (B1406). Air-conditioned. With audio/video equipment. Computer. Overhead. Instructors are equipped with microphones.	Worksheets/textbooks	- Topnotch 1 Unit 9 - Handouts from Tactics for Listening
	CT's lesson plans	N/A
	Blackboard input	- Grammar, vocabulary, - Usage - Answers to textbook exercises.
Notes - Ss hated the video recorder placed at the front of the classroom, stating that 'how scary!' or 'such a pressure!' - Place to put video recorder were negotiated between Ss and me.	CT taped interview data	Yes
	Students taped interview data	N/A
	Other	

--	--	--

Observed Lesson Summary Sheet

Overt Washback:
N/A
Teaching content:
<p>Part 1</p> <ul style="list-style-type: none"> - T went through grammar point on p.104 (e.g. be going to, will, future tense). This section is about day, date, month and related vocabulary words. T went through all this and had Ss practice how to pronounce dates. - T then distributed supplementary materials she photocopied from another textbook (Tactics for Listening) and had students practice listening comprehension on dates. <p>Part 2</p> <ul style="list-style-type: none"> - T went through p.104, 105 on the textbook, on car rental, hotel reservation, related grammar points and grammar exercises. T gave answers on whiteboard. T assigned one exercise as homework to be checked next time. - T then played a video clip, related to the theme of travelling and asked Ss to pay specific attention on the dates. T asked questions about the dates the ppl in the video would start their trip and would return to their country. - T explained the video content as the content was much more difficult than the textbook content and Ss did not have the transcripts. <p>Part 3</p> <ul style="list-style-type: none"> - Lastly, T went through the vocabulary words on vehicles on p.109.
Teaching format:
Materials:
<p>Topnotch 1.</p> <ul style="list-style-type: none"> - T has a focus on function and thus, she supplemented the textbook content with similar contents from textbooks/materials she has used in the past.
Teaching method:
<p>Grammar-Translation for vocabulary words.</p> <p>Communicative language teaching, 'cuz she emphasises on language functions.</p>
Students' motivation:
<ul style="list-style-type: none"> - Ss were too talkative and T's voice could not be heard clearly.

- SS would respond to T but in a passive way. They answered T's questions when called on one by one.
- Ss with better proficiency answered more to T's questions.

Others:

- As I put the video recorder at the back of the classroom, I could only record T and the backs of Ss. Some Ss who sat near the video recorder still found the video recorder irritating and intimidating while some played with the recorder.
- I sat near the recorder, at the back of the classroom.

Appendix D ELLI traditional Chinese version

ELLI 問卷

以下問題，請依您個人經驗選擇合適的答案，在 1-4 數字框畫叉。	1 完全 不像 我	2 只有一 點像 我	3 還蠻 像我	4 非常 像我
1. 我喜歡那種不得不非常努力才能理解一件事情的感覺。	1	2	3	4
2. 當我真正對某件事物感興趣的時候，我覺得它比較容易學。	1	2	3	4
3. 家裡至少有一個人對我的學習提供重要的指導作用。	1	2	3	4
4. 我喜歡對我正在學習的事物提出問題。	1	2	3	4
5. 我更喜歡獨自解決問題。	1	2	3	4
6. 了解事物的本質對我來說比獲得好成績更為重要。	1	2	3	4
7. 當我學習新事物的時候，我嘗試聯想一些我已經學過的東西。	1	2	3	4
8. 當我學習的時候，我有時會用一些不同的方法只是為了看看這樣做會怎麼樣。	1	2	3	4
9. 如果我在完成一項學習任務中遇上了困難，我通常能想辦法來解決難題。	1	2	3	4
10. 當我覺得學習困難的時候，通常是因為我沒有得到足夠的幫助。	1	2	3	4
11. 把學問轉化成故事的方式來理解，這對我的學習幫助很大。	1	2	3	4
12. 我總是很清楚的明瞭學習某樣東西要花多久時間。	1	2	3	4
13. 在我遇上困難時，我通常不知道該怎麼辦。	1	2	3	4
14. 我喜歡一個有趣的問題甚於一個簡單的問題。	1	2	3	4
15. 我往往逃避學習新東西，因為我不喜歡感覺困惑或不確定。	1	2	3	4
16. 學習的時候，我喜歡詳細周密的思考，並做出	1	2	3	4

	合情合理的解釋。				
17.	當我不能掌握某樣東西的時候，通常是因為我不知道該如何著手做這件事。	1	2	3	4
18.	我很喜歡把新知識跟以前學過的知識 做連結 ， 幫助我學習新知識 。	1	2	3	4
19.	我喜歡用不同的方法學習新知識。	1	2	3	4
20.	當我覺得學習枯燥的時候，我通常能找到使 學習變 有趣的方法。	1	2	3	4
21.	我喜歡因為一個好的理由而去學習某樣東西。	1	2	3	4
22.	學習的時候，我常常會運用我的想像力。	1	2	3	4
23.	如果我靜靜等待，有時自然而然地會想到好主意。	1	2	3	4
24.	當不懂某樣事物的時候，我通常會 花 一段時間努力思考。	1	2	3	4
25.	當學習有難度的時候，我常常發現它很有趣。	1	2	3	4
26.	我喜歡和別人一起解決問題。	1	2	3	4
27.	當我知道所學的新東西對我有何意義的時候，我學起來會更有動力。	1	2	3	4
28.	當我在學習某樣東西的過程中 遇到不順遂 時，我想這可能是因為我不夠聰明。	1	2	3	4
29.	我喜歡學習我真正關心的東西。	1	2	3	4
30.	我覺得我的家人對我的學習有很大幫助。	1	2	3	4
31.	當我學習 順利 的時候通常是因為我獲得了足夠的幫助。	1	2	3	4
32.	做為一個學習者，我在不斷地進步。	1	2	3	4
33.	我最好的 想法產生 通常都是在 思緒 自由翱翔的時候得到的。	1	2	3	4
34.	做為一個學習者，我能感覺到自己的進步。	1	2	3	4
35.	我不喜歡直接接受一個答案，除非我自己設法弄懂它。	1	2	3	4
36.	我喜歡接受新的挑戰。	1	2	3	4
37.	尊重別人的觀點很重要，即使這些觀點和我自己的觀點不一樣。	1	2	3	4
38.	我自己一個人學習的時候通常 學習 得最好。	1	2	3	4

39 ·	和朋友詳細討論可以幫助我學習。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
40 ·	我覺得在學習的過程中嘗試不同的東西還不錯。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
41 ·	我喜歡和朋友們一起討論很難的問題。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
42 ·	我期望長期地持續學習。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
43 ·	基本上我的學習能力保持一致，沒有變過。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
44 ·	在我周圍的人當中，至少有一個人是我學習的重要引導人。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
45 ·	在學習中遇到困難的時候，我通常會變得很沮喪。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
46 ·	如果我在學習中變得很沮喪，我很善於調節自己，使自己能擺脫沮喪。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
47 ·	當我開始完成一個學習任務的時候，有時直到看到周圍的朋友取得進展，我才知道要怎麼做。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
48 ·	我希望能健全我做事情的方式。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
49 ·	我經常用我所學到的知識來改變我做事情的方式。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
50 ·	當老師批評我的作業的時候，我覺得很難受。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
51 ·	我通常可以和剛剛認識的人合作得很好。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
52 ·	如果我知道某樣東西很重要，我就能想出辦法來學會它。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
53 ·	我知道我能堅持用自己的方式來學習，即使我的朋友們覺得那只是浪費時間。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
54 ·	當我沒有收到明確的指示去做某事的時候，我會覺得很困難。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
55 ·	家裡至少有一個人幫助過我的學習。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
56 ·	只有一點時間，我也喜歡嘗試新的做事的方法。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
57 ·	我知道，如果我的朋友們都和我學習一樣的東西的話，我學習起來就比較容易。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
58 ·	不管我自己一個人學習或與別人一起學習都可學得一樣好。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
59 ·	如果我不喜歡學習的話，我會有方法迫使自己學習。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
60 ·	我開始一項學習任務前總會想想在學習過程中所有需要的東西。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

61 ·	我需要和其他人一起才能夠學習。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
62 ·	我把正在學習的東西和已經學習的東西 做聯結 。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
63 ·	我總是用同樣的方法和步驟 來學習 。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
64 ·	我知道如果有足夠的時間去思考，我就能找到一個解決問題的辦法。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
65 ·	我喜歡尋找自己的方式來做事。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
66 ·	有時候好主意自然就出現在我的腦海中。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
67 ·	我喜歡去想像其他人是怎麼樣考慮和看待問題的。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
68 ·	當沒有人告訴我該怎麼去做的時候，我發現學習是很困難。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
69 ·	即使所有人都在用另一種方式學習，我也喜歡尋找我自己做事的方式。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
70 ·	我開始一項學習任務前，從不會先思考一下這個學習任務。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
71 ·	我經常溫習我學過的東西。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
72 ·	我能夠在學習中採用別人的建議。	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

Note: The highlighted areas indicate changes of words or phrases from simplified Chinese to traditional Chinese, for the purpose of better reading (see 3.8 for explanation).

Appendix E Research ethics form for ethical considerations in this study

Name(s): Shwu-Wen Lin

Proposed research project:

A washback study of the degree requirement for English proficiency and the General English Proficiency Test on non-English majors in Taiwan: Case studies of two universities

Proposed funder(s): Self

Discussant for the ethics meeting: Shwu-Wen Lin, Guoxing Yu and Pauline Rea-Dickins

Please include an outline of the project or append a short (1 page) summary:

Please see the Appendix I for the summary of the project.

Ethical issues discussed and decisions taken (see list of prompts overleaf):

The British Educational Research Association (BERA), the British Association for Applied Linguistics (BAAL) and International Language Testing Association (ITLA, Codes of Ethics) ethical guidelines serve as the references for the discussions below.

1) Researcher access/exit:

- Access to be obtained only if this ethics form, the research proposal and the consent of the teacher and student participants are received by the case school.
- Exit time as negotiated between the researcher, the case universities and the participants to allow for flexibility while maintaining respect for the case universities and the participants.

2) Information given to the teacher and student participants:

- Basic information of the proposed study, e.g. background information and participants involved.
- Details on how, when and for how long they will be involved. Ask the participants willingness for return or further involvement in the research.

3) Participants' right of withdrawal:

- All participants have the right not to participate in or withdraw from the research at any time prior to the publication of the research results.

4) Informed consent:

- Voluntary informed consent will be received from the participants before

conducting classroom observation complaints procedure:

- If participants have any complaints during the research process, the researcher should deal with the complaints at first place. If after negotiation, there are still no solutions to the complaints, then the researcher should consult the Research Committee of the GSoE.

5) Anonymity/confidentiality:

- All participants have the right to anonymity and confidentiality. However, as no confidentiality is absolute, in the proposed research, it is difficult to achieve confidentiality in forms of video data. Thus, the informed consent should ask for participants' willingness to reveal their identities in videos.

6) Data collection:

- All of the procedures at the stage of data collection should aim at collecting data without bringing harm to any of the participants.

7) Data analysis:

- Data analysis should conform to the rule of anonymity and confidentiality concerning the identities of the participants and the university.

8) Data storage:

- All of the data will be stored in an anonymised form and in a locked file in the researcher's private laptop and portable hard disk and it is ensured that the data will be retrieved by the researcher alone through a set of password only known by herself.
- The Fifth Data Protection Principle in the Data Protection Act 1998 indicates that data processed for any purpose shall not be kept for longer than is necessary for that purpose. However, this principle does not apply to data processed only for research purposes. If all research activities in the proposed research meet the requirement of not causing damage or distress to the individual or as a result of the research itself, identify any data subject, the data collected from the research activities can be kept as long as they are used for research purposes only.

9) Data protection act:

- The storage and the use of the data collected should follow the provisions of the Data Protection Act (1998). According to the Act, all personal data should be used exclusively for research only. The personal data should not be used to support measures or decisions relating to any identifiable living individual. The personal data should also not be used in a way that will cause, or is likely to cause, substantial damage or substantial distress to any data subject.

10) Feedback:

- To ensure the validity of the researcher's interpretation of the data/findings,

the participants will be invited to provide feedback and comments on transcriptions, translations and analyses made by the researcher.

11) Responsibilities to the academic community:

- As BAAL ethical guidelines indicate, the responsibilities of applied linguists are to ‘maintain the integrity of applied linguistic enquiry, the freedom to research and study, and the freedom to publish and disseminate the results of their research’ (p.12).
- The codes of ethics by the ITLA indicate that language testers ‘should adhere to all relevant ethical principles embodied in national and international guidelines when undertaking any trial, experiment, treatment or other research activity’ (p. 3).

12) Reporting of research:

- ‘Publication of research results shall be truthful and accurate’ and ‘shall not permit identification of the subjects who have been involved’(Codes of Ethics, ITLA, p. 3).
- The participants have the right to read the reports of the research.
- The results of the research should be disseminated.
- The opinions expressed in the reports are those of the researcher, not of the University of Bristol.

If you feel you need to discuss any issue further, or to highlight difficulties, please contact the GSoE’s ethics co-ordinator who will suggest possible ways forward.

Signed: (Researcher) Signed: (Discussant)

Date: 12th May 2008

Appendix F Consent form for teachers

Dear Teacher,

Thank you for agreeing to participate in my research for my doctoral study in the University of Bristol. My research is to investigate the washback effects of the General English Proficiency Test to the learners and the impact of the graduation requirement on learner motivation. Your participation in this research will involve you in the following activities:

- 1) two individual interviews with me
- 2) observation of your two classes while you teach for four weeks
- 3) opportunities to comment and provide feedback on the translation of my data and on the data analysis.

The interviews will be audio-taped and the observations will be both video- and audio-taped.

As recommended by the ethical guidelines provided by the UoB/GSoE, BERA and BAAL, I would like to have a formal consent from you. Anonymity of all data will be ensured and your data will be used solely for this research. Your participation in this study is entirely voluntary. You may refuse to participate in this research. Such refusal will not have any negative consequences for you. If you begin to participate in the research, you may at any time, for any reason, discontinue your participation without any negative consequences. I would be very grateful if you could tick the boxes below and sign your name and date.

- I agree to participate in Shwu-Wen Lin's research, cooperating with her, and allowing her to be an observer in my classrooms under the circumstances that she will ensure anonymity of all participants and will not interfere in the normal activities in my classrooms and will also not interfere in the decisions about the classrooms and the lessons.

- I agree to anonymised use of all the data I will provide for Shwu-Wen Lin's research in conference presentations and academic publications.

Name:

Signature:

Date:

If you have any queries about the research, please contact me at :

Shwu-Wen Lin (S.W.Lin@bristol.ac.uk)

Best wishes,

Shwu-Wen Lin

Appendix G Consent form for the students

Dear Participant,

Thank you for agreeing to participate in my research for my doctoral study in the University of Bristol. My research is to investigate the washback effects of the General English Proficiency Test to the learners and the impact of the graduation requirement on learner motivation. Your participation in this research will involve you in the following activities:

- 1) an individual interview with me
- 2) observation of your class for 4 weeks
- 3) opportunities to comment and provide feedback on the data analysis.

The interviews will be audio-taped and the observations will be both video- and audio-taped.

As recommended by the ethical guidelines provided by the UoB/GSoE, BERA and BAAL, I would like to have a formal consent from you. Anonymity of all data will be ensured and your data will be used solely for this research. Your participation in this study is entirely voluntary. You may refuse to participate in this research. Such refusal will not have any negative consequences for you. If you begin to participate in the research, you may at any time, for any reason, discontinue your participation without any negative consequences. I would be very grateful if you could tick the boxes below and sign your name and date.

- I agree to participate in Shwu-Wen Lin's research, cooperating with her, and allowing her to be an observer in my classrooms under the circumstances that she will ensure my anonymity.

- I agree to anonymised use of all the data I will provide for Shwu-Wen Lin's research in conference presentations and academic publications.

Name:

Signature:

Date:

If you have any queries about the research, please contact me at :

Shwu-Wen Lin (S.W.Lin@bristol.ac.uk)

Best wishes,

Shwu-Wen Lin

Appendix H Sample mid-term and final test papers that modelled on the item types of the GEPT elementary level

Test 1

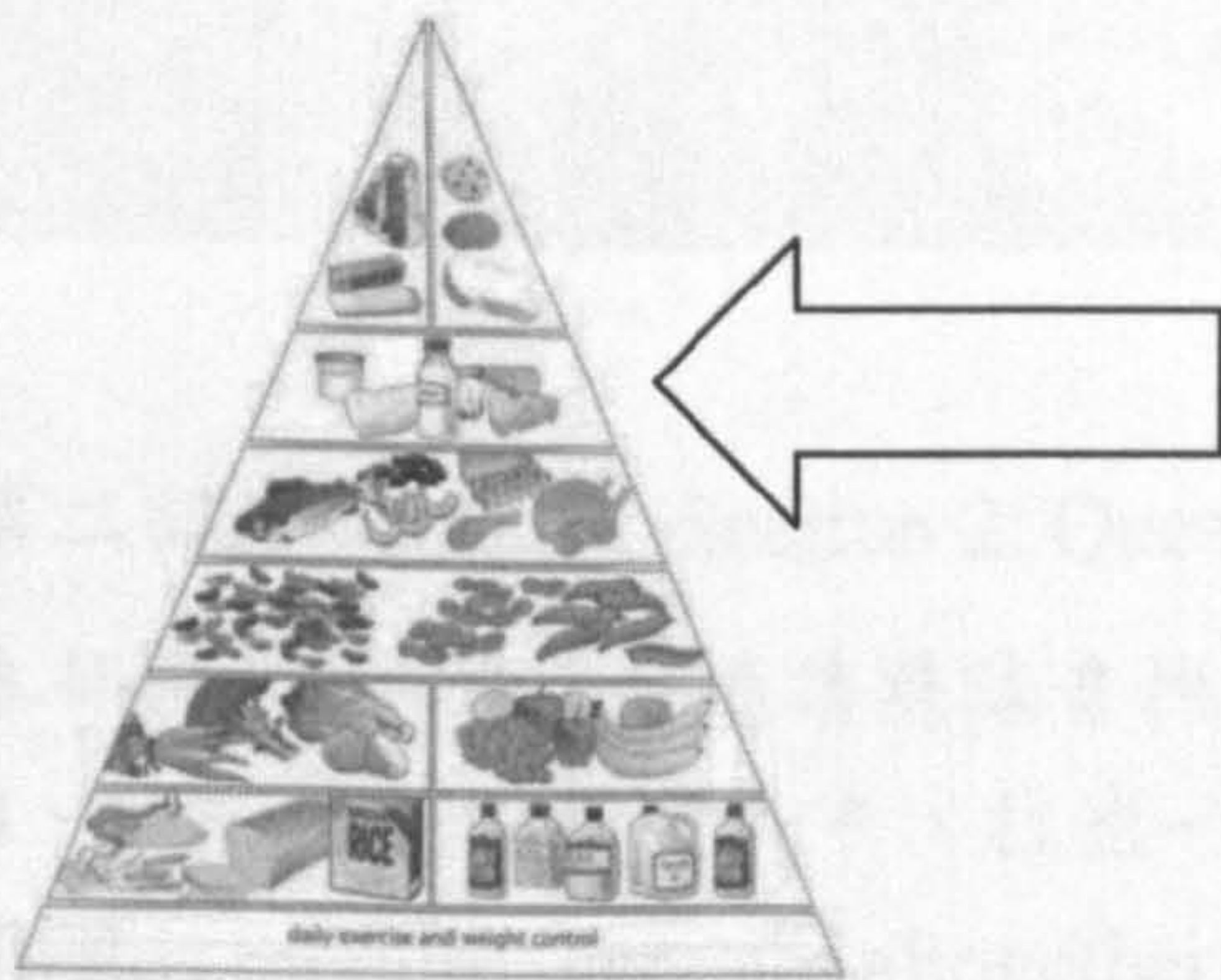
聽力測驗 (Listening comprehension)

第一部份：看圖辨義 (Section 1: Picture description)

本部份共4題，試題冊上有數幅圖畫，每一圖畫有1個描述該圖之題目，每題請聽錄音機播出題目以及A、B、C三個英語敘述之後，選出與所看到的圖畫最相符的答案，每題只播出一遍。

(In this section, there are altogether 4 items. There are several pictures on the test papers. Each picture corresponds to one item that describes the picture. For each item, please listen to a question and three statements a, b, c, and select the most appropriate answer for the corresponding picture. Each item is played only once.)

1. Question 1 _____



Picture 1

2. Question 2 _____



Picture 2

3. Question 3 _____



Picture 3

4. Question 4 _____



Picture 4

第二部份：問答 (Section 2: Question and statement response)

本部份共 10 題，每題請聽錄音機播出一英語問句或直述句之後，從試題冊上 A、B、C 三個回答或回應中，找出一個最適合的作答。每題只播出一遍。

(In this section, there are altogether 10 items. Please listen to a question or a statement and choose the most appropriate answer or response from a, b, c. Each item is played only once.)

1. _____
 - a. Do you want to go to a movie?
 - b. I know what you mean.
 - c. I'm so tired of the same old grind.

2. _____
 - a. Let's see. I'll paint it red.
 - b. It's depressing.
 - c. You'd better paint it quickly.

3. _____
 - a. Nothing I can put my figure on.
 - b. Blue is a good color. Let's do it!

c. I'll cheer you up.

4. _____

- a. It's terrific. I can't stand it.
- b. Definitely. I'm crazy about coffee.
- c. No. I'm a coffee lover.

5. _____

- a. Why? Don't you like it?
- b. Not really. It was fantastic.
- c. Certainly. I am avoiding it.

6. _____

- a. Yes. I am allergic to chocolate.
- b. Very nice. Thanks.
- c. Actually, no. I'm a vegetarian.

7. _____

- a. I'm so sorry. Want to try some?
- b. I used to be. Not anymore.
- c. Do you mind not doing that?

8. _____

- a. Don't worry about it. It's not a problem.
- b. I'd like to. But I think I'll pass.
- c. Not really. It's too spicy for me.

9. _____

- a. Sorry. That's a lot of work.
- b. Yes. You can do that again.
- c. Really? What color would you choose?

10. _____

- a. I'll have to take this back.
- b. Good idea. Let's go.
- c. I don't think he'll agree with it.

第三部份：簡短對話 (Section 3: Short conversation)

本部份共 4 題，每題請聽錄音機播出一段對話及一個相關的問題後，從試題冊上 A、B、C 三個備選答案中找出一個最適合的回答。每段對話及問題只播出一遍。

(In this section, there are altogether 4 items. Please listen to a short conversation and a related question and select the most appropriate answer from a, b, c. Each conversation and question is played only once.)

1. _____
A: Hard to understand.
B: Introverted.
C: Active.
2. _____
A: Her sibling.
B: Her boyfriend.
C: Her brother-in-law.
3. _____
A: She gets dumped.
B: She doesn't know, either.
C: She has a cut on her finger.
4. _____
A: He doesn't care for broccoli.
B: He's a vegetarian.
C: Broccoli doesn't agree with him.

閱讀測驗 (Reading Comprehension)

第一部分：詞彙和結構 (Section 1: Vocabulary and Structure)

本部份共 10 題，每個題目裏有一個空格。請就試題冊上 A、B、C、D 四個選項中找出一個最適合題意的字或詞作答。(In this section, there are altogether 10 items. There is a blank in each item. Please select the most appropriate word or phrase from a, b, c, d, to fill in the blank.)

1. If you are a middle child, you're probably the silent rebel _____ the family's values.
 - a. of
 - b. to
 - c. from
 - d. against

2. If you are the youngest child, you're probably creative _____ art, music, and other ways.
 - a. in
 - b. on
 - c. for
 - d. at
3. The first child has to _____ very fast.
 - a. clown up
 - b. grow up
 - c. sign up
 - d. rule up
4. The word "emotion" is similar in meaning _____ the word "feeling."
 - a. for
 - b. with
 - c. to
 - d. of
5. The fried grasshoppers tasted _____. I think I like it a little bit.
 - a. crazy
 - b. soft
 - c. terrible
 - d. crunchy
6. When someone is overweight, he or she is _____.
 - a. sedentary
 - b. obese
 - c. diet
 - d. diabetes
7. Some parents just _____ their children whatever they want.
 - a. feed
 - b. process
 - c. blame
 - d. saturate
8. Heart attacks, stroke, and hypertension are all _____.
 - a. decades
 - b. organizations
 - c. diseases
 - d. intakes
9. I used to have sweets a lot. But now I've been _____.
 - a. feeding with

- b. going on
 - c. getting along
 - d. cutting back
10. I can't eat meat. So I think I'll pass _____ the chicken.
- a. through
 - b. on
 - c. to
 - d. above

第二部分：段落填空 (Section 2: Cloze)

本部份共 15 題，包含三個段落 5 個空格。請就試題冊上 A、B、C、D 四個項中選出最合題意的字或詞作答。

(In this section, there are altogether 15 items, including 3 paragraphs with 5 blanks each. Please select the most appropriate word or phrase to fill in the blanks.)

Questions 11-15

When you take a look 11 the health-eating pyramid that suggests daily eating habits to 12 heart diseases, you would find that basic things like daily exercise and weight 13 are even more important than watching out 14 what you eat. So, other than going on a strict diet, we need to try not to live a 15 lifestyle.

- _____ 11. a. on
 b. in
 c. with
 d. at
- _____ 12. a. avoid
 b. gain
 c. respond
 d. lose
- _____ 13.** a. connect
 b. contact
 c. control
 d. comfort
- _____ 14.** a. at
 b. with
 c. to
 d. as

- _____ 15. a. annual
b. sedentary
c. compare
d. grain

Questions 16-20

Like it or not, the truth is that the best friend in your life should be yourself. When you think back, ___16___ you are a popular person ___17___ lots of friends, there is so much time you need to face yourself. And only you know how you really feel ___18___ every single moment. So, cherish yourself. And take good care ___19___ your own body. ___20___, you'll need to spend most of your life time with it.

- _____ 16. a. as
b. if
c. even if
d. so as to
- _____ 17. a. with
b. on
c. have
d. into
- _____ 18. a. of
b. at
c. on
d. about
- _____ 19. a. on
b. to
c. as
d. of
- _____ 20. a. However
b. Other than
c. After all
d. Although

Questions 21-25

I'd like to introduce some personality traits to you today. A person's

personality includes his/her usual behavior, thoughts, and emotions. In other words, that's the person's pattern of behavior. Personality comes both from nature and nurture. Nature means when someone is born, something in his genes has decided part of him/her. It's genetic. And nurture means as the time goes by, the environment that the person is in also has some influence to form part of him/her.

- _____ 21. What title best suits this passage?
- The influence of our environment.
 - The decision from genes.
 - The formation of personality traits.
 - Personal behavior.
- _____ 22. What kind of a passage do you think it is?
- A lecture.
 - A part of a TV show.
 - A conversation.
 - A phone call.
- _____ 23. Which one decides a person's personality traits?
- Nature.
 - Nurture.
 - Neither.
 - Both.
- _____ 24. Which of the followings is not included in personality traits?
- Emotions.
 - Thoughts.
 - Achievements.
 - Behavior.
- _____ 25. Which of the followings is close to the meaning "the person's pattern of behavior"?
- Who this person is.
 - How this person usually acts.
 - Where this person works.
 - When this person usually eats.

寫作測驗(Writing)**第一部分：單句寫作 (Section 1: Sentence Writing)**

請將每題作答的句子完整地寫出來，題目不需抄寫。評分重點包括內容、文法、用字、標點符號及大小寫。

(Please write down the complete sentence for each item. There is no need to rewrite the item. The criteria for scoring include contents, grammar, use of vocabulary, punctuation and capital and small letters.)

Question 1-5 句子改寫 (Sentence Rewriting)

1. The first child is always pushed to be successful.
→ Parents always _____.
2. Personality traits can be generalized by birth order
→ We can _____ about personally traits from birth order.
3. Spending so much time in front of a computer makes her angry.
→ She is _____.
4. Flying doesn't make her afraid.
→ She isn't _____.
5. I am addicted to chocolate.
→ I am a _____.

Questions 6-10 句子合併 (Sentence Combining)

6. The World Health Organization has a report.
The report says that 6 out of 10 deaths in the region are due to obesity.
→ _____.
7. Dr. Litonjua blames it.
And he calls it "malling."
→ Dr. Litonjua blames _____.
8. Processed foods and fast foods are often the most available and the cheapest.
Processed foods and fast foods are rich with sugar and saturated fats.
→ Processed foods and fast foods _____.

9. There is another word for genetics in discussing personality.

Nature is the word.

→ _____.

10. I got back from vacation.

And after that, I have been feeling a little out of sorts.

→ _____.

Questions 11-15 重組 (Sentence Reconstruction)

11. lead / students / active / how many / non-sedentary / an / lifestyle

→ _____?

12. go to / but / use to / I / now / I / do / a gym / don't

→ _____.

13. much / not / a / I'm / eater / pizza / of

→ _____.

14. needs / sleep / lots / she / get / to / of

→ _____.

15. she / object to / herself / talking about / from time / unlike Ted / doesn't / to time

→ _____.

第二部分：段落寫作 (Section 2: Paragraph Writing)

請根據提示，寫一篇約 50 字的文章描寫描寫以下的資料。

(Please write a paragraph around 50 words according to the prompts below.)

依照下圖，請寫出一篇約五十字的短文來描述陳先生和陳太太的粉刷房屋計畫。
(Write a paragraph around 50 words to describe Mr. and Mrs. Chen's plan for refurbishing their house according to the picture below.)

Describe the personalities of the children.



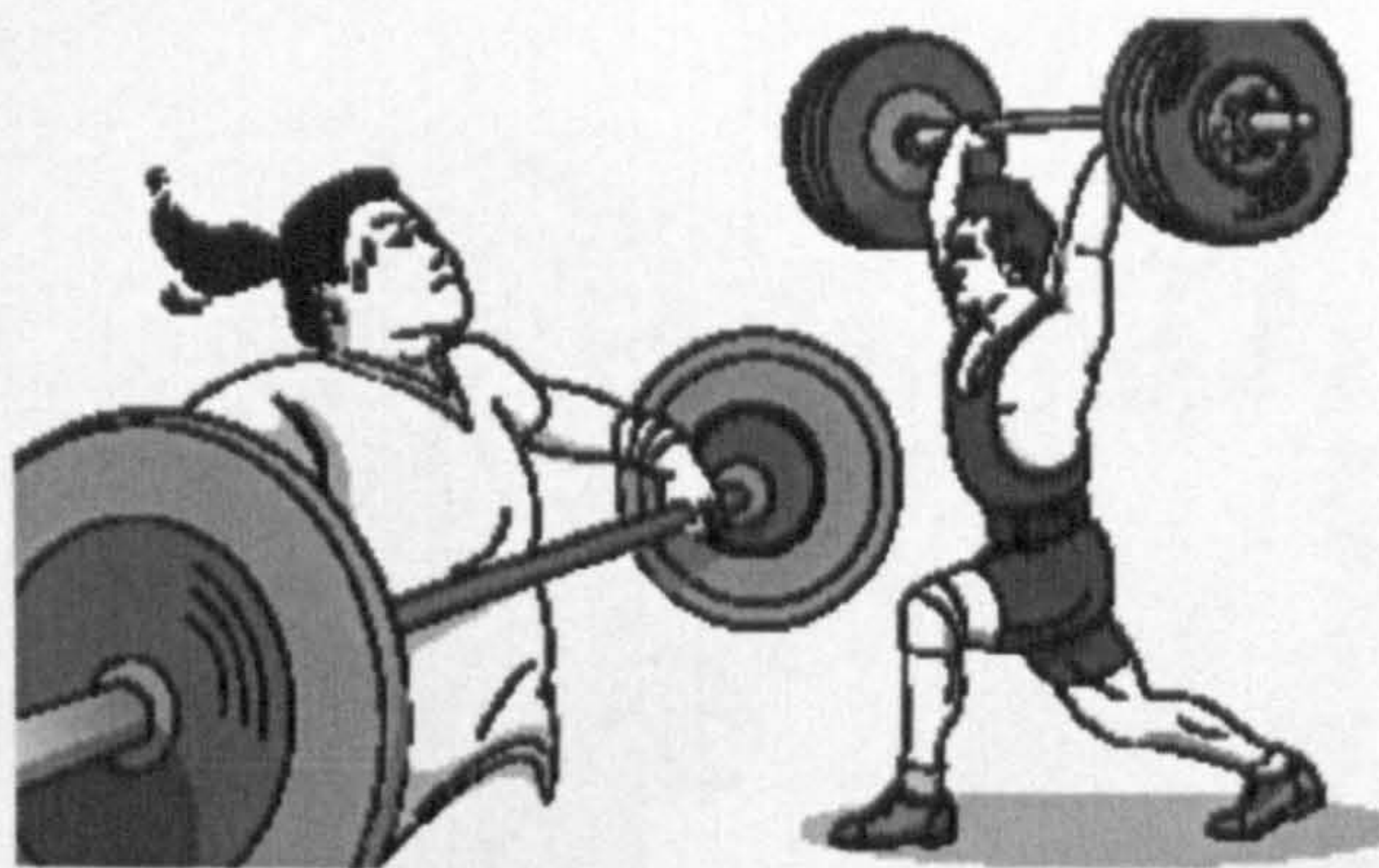
Test 2**聽力測驗 (Listening comprehension)**

第一部份：看圖辨義 (Section 1: Picture description)

本部份共4題，試題冊上有數幅圖畫，每一圖畫有1個描述該圖之題目，每題請聽錄音機播出題目以及A、B、C三個英語敘述之後，選出與所看到的圖畫最相符的答案，每題只播出一遍。




(In this section, there are altogether 4 items. There are several pictures on the test papers. Each picture corresponds to one item that describes the picture. For each item, please listen to a question and three statements a, b, c, and select the most appropriate answer for the corresponding picture. Each item is played only once.)

1. Question 1 _____



Picture A

2. Question 2 _____

Go walking		250 calories / hour
Go bike riding		500 calories / hour
Go running		572 calories / hour

Picture B

3. Question 3 _____



Picture C

4. Question 4 _____



Picture D

第二部份：問答 (Section 2: Question and statement response)

本部份共 10 題，每題請聽錄音機播出一英語問句或直述句之後，從試題冊上 A、B、C 三個回答或回應中，找出一個最適合的作答。每題只播出一遍。

(In this section, there are altogether 10 items. Please listen to a question or a statement and choose the most appropriate answer or response from a, b, c. Each item is played only once.)

1. _____

- A. Yes. My house is very clean.
- B. Twice a week.
- C. We usually clean the house on Sunday.

2. _____

- A. No, but I can play the piano.
- B. Lisa can play the guitar very well.
- C. I learned how to play the guitar two years ago.

3. _____

- A. Yes. He likes potatoes very much.
- B. Yes. He is surely a couch potato.
- C. No, but he likes fruits and vegetables.

4. _____

- A. Sure. That would be great.
- B. I sometimes go dancing with my friends.
- C. I usually go dancing twice a month.

5. _____

- A. Really? That would be fun.
- B. That's too bad.
- C. No, I am too busy.

6. _____

- A. Sorry. They have been sold out.
- B. Yes. Here is a smaller size.
- C. Sure. How would you like to pay for it?

7. _____

- A. My sister gave it to me a few months ago.
- B. It looks great, and it's inexpensive as well.
- C. He's buying a bag for his wife.

8. _____

- A. It's too cold in winter without gloves.
- B. What about these? See if they are better.
- C. Yes. They look great, and are warm enough.

9. _____

- A. It must mean a lot to you.
- B. That watch is too expensive. Let's see if there's a cheaper one.
- C. Where did you get that?

10. _____

- A. It's on the seventh floor.
- B. Sorry, we don't sell shoes but coats.
- C. Pumps are on the ground floor.

第三部份：簡短對話 (Section 3: Short conversation)

本部份共 4 題，每題請聽錄音機播出一段對話及一個相關的問題後，從試題冊上 A、B、C 三個備選答案中找出一個最適合的回答。每段對話及問題只播出一遍。

(In this section, there are altogether 4 items. Please listen to a short conversation and a related question and select the most appropriate answer from a, b, c. Each conversation and question is played only once.)

1. _____

- A. Tomorrow morning.
- B. Tomorrow afternoon.
- C. Sunday morning.

2. _____

- A. She is too heavy.
- B. She is very sick.
- C. She exercises too much.

3. _____

- A. At the restaurant.
- B. At the department store.
- C. At the supermarket.

4. _____

- A. A teller and a customer.
- B. A manager and a secretary.
- C. A clerk and a shopper.

閱讀測驗 (Reading)

第一部分：詞彙和結構 (Section 1: Vocabulary and Structure)

本部份共 10 題，每個題目裏有一個空格。請就試題冊上 A、B、C、D 四個選

項中找出一個最適合題意的字或詞作答。(In this section, there are altogether 10 items. There is a blank in each item. Please select the most appropriate word or phrase from a, b, c, d, to fill in the blank.)

_____ 01. My brother is _____ because he works out six days a week and often walks to work.

- (A) a couch potato (B) in great shape (C) out of shape
(D) a sweet tooth.

_____ 02. A: What physical activities do you do to _____ a lot of calories?

B: I go swimming and running.

- (A) share (B) lift (C) burn (D) walk

_____ 03. The school _____ is used for a lot of different sports. Students play football and soccer in the fall and baseball in the spring.

- (A) athletic field (B) pool (C) track (D) gym

_____ 04. I always take aerobics classes or run on the treadmill at the _____ on weekends.

- (A) athletic field (B) pool (C) track (D) gym

_____ 05. This beach resort hotel has a tennis court and an 18-hole golf _____.

- (A) pool (B) track (C) course (D) court

_____ 06. Excuse me, I want to buy my wife a nightgown. Can you tell me where the _____ is?

- (A) lingerie (B) hosiery (C) accessories (D) electronic

_____ 07. Could you gift _____ this dress? It's my sister's birthday.

- (A) turn (B) tie (C) wrap (D) charge

_____ 08. The shoe department is upstairs, _____ the third floor.

- (A) in (B) on (C) to (D) beyond

_____ 09. The food court is downstairs, _____ the basement.

- (A) in (B) on (C) to (D) up

_____ 10. The children's department is upstairs. Take the _____ to the second floor and walk to it.

(A) ground (B) briefs (C) windbreaker (D) escalator

第二部分：段落填空 (Section 2: Cloze)

本部份共 15 題，包含三個段落 5 個空格。請就試題冊上 A、B、C、D 四個項中選出最合題意的字或詞作答。

(In this section, there are altogether 15 items, including 3 paragraphs with 5 blanks each. Please select the most appropriate word or phrase to fill in the blanks.)

Questions 11-15

Brooke Ellison is the first quadriplegic to graduate from Harvard University. When she was 11 years old, a terrible accident changed her life. She was hit by a car 11 her way home from school, leaving her paralyzed from the neck down. She 12 move her arms or legs and has to breathe through a very special machine. She has to 13 all her time in a wheelchair. And yet, she was not 14 by that. She stays active every day. Now, She's a doctoral candidate at the State University of New York at Stony Brook. She gives speeches to encourage those 15 face adversity of any kind. She tells them about her life and teaches them to always be active.

- _____ 11. (A) in (B) to (C) for (D) on
- _____ 12. (A) can't (B) has to (C) should (D) will not
- _____ 13. (A) cost (B) take (C) spend (D) make
- _____ 14. (A) discouraged (B) encouraged (C) couraged (D) uncouraged
- _____ 15. (A) when (B) who (C) where (D) to

Questions 16 - 20

Forty-year-old Joan Chen started to worry about her health after she found out she could not fit in most of her pants any longer. She 16 5 kg within 6 months and those pants became too tight. In her free time, she enjoys 17 TV and eating junk food like potato chips and soda. She knew she needed to change her lifestyle. The first thing she tried to change was her 18. She tried to 19 fatty foods and sweets and eat more healthy food like fruit and vegetables. She also started to work out at least 30 minutes at a time, threes times a week. After all, healthy diet and 20 exercise are the key to a healthy life.

- _____ 16. (A) put on (B) lost (C) put off (D) took on
- _____ 17. (A) watch (B) to watch (C) watched (D) watching
- _____ 18. (A) look (B) diet (C) aerobics (D) athlete
- _____ 19. (A) spend (B) take (C) avoid (D) cook
- _____ 20. (A) regular (B) oily (C) many (D) often

Questions 21-25

If you want to go shopping, there are things you need to consider. If you are looking for a bargain, 21 sure you go when there's a big sale. One problem with a sale is that there might not be a lot of selection. And if you are not satisfied 22 the product after you have bought it, it's sometimes hard to exchange. Many stores also refuse to give a 23 . It's very likely that you can't get your money back. If you're looking for clothes, don't forget to 24 . It's always a good idea to bring enough 25 if you're shopping at some small shops. They probably don't accept credit cards or checks.

- _____ 21. (A) make (B) making (C) to make (D) having made
- _____ 22. (A) in (B) with (C) to (D) into
- _____ 23. (A) return (B) repeat (C) refund (D) review
- _____ 24. (A) try them on (B) take them off (C) wrap them up (D) put them down
- _____ 25. (A) charge (B) check (C) cash (D) tips

寫作測驗(Writing)

第一部分：單句寫作 (Section 1: Sentence Writing)

請將每題作答的句子完整地寫出來，題目不需抄寫。評分重點包括內容、文法、用字、標點符號及大小寫。

(Please write down the complete sentence for each item. There is no need to rewrite the item. The criteria for scoring include contents, grammar, use of vocabulary, punctuation and capital and small letters.)

Question 1-5 句子改寫 (Sentence Rewriting)

1. Mary always goes to the gym on Saturday morning.

→ Mary _____ tonight.

2. I go mountain climbing every Monday, Wednesday and Friday.

→ I _____ a week.

3. I play the guitar every day.

→ How _____ ?

4. Do you want to play soccer sometime?

→ Why _____ ?

5. I am going to the park.

→ I am on _____ to the park.

Questions 6-10 句子合併 (Sentence Combining)

6. Tim buys a beautiful purse.

It's for his wife.

→ _____ a beautiful purse.

7. The children's department is downstairs.

The children's department is in the basement.

→ The _____.

8. Do you have these loafers?

Do you have a larger size?

→ Do _____ ?

9. These pajamas aren't comfortable to wear

This nightgown is comfortable to wear.

→ The nightgown is _____.

10. You can't wear open shoes at a temple in Thailand.

It is disrespectful.

→ It _____ in Thailand.

Questions 11-15 重組 (Sentence Reconstruction)

11. Let _____.

see / find / I / something / you / better / me / if / can

.

12. Could _____ ?

wrap / for / you / them / gift / me

13. We _____.

size/ have / yellow / smaller / in / a

14. Please _____.

to / this / give / your / windbreaker / brother

15. How _____?

pay / like / that / would / for / you / to

第二部分：段落寫作 (Section 2: Paragraph Writing)

請根據提示，寫一篇約 50 字的文章描寫描寫以下的資料。

(Please write a paragraph around 50 words according to the prompts below.)

提示：以下的圖片是 Sam 平常的一天。

(Prompt: The following picture shows a typical day of Sam's life.)

