

**Investigating the Washback Effect of the MUET as
a University Entry Test on Students in Malaysia**

Najihah Mahmud

PhD

University of York

Education

December 2018

Abstract

The influence that a test has on teaching and learning is commonly known as washback. Existing literature has acknowledged the complexity and multi-directionality of the phenomenon. This study seeks to explore the washback effect of the Malaysian University English Test (MUET), a high-stakes compulsory university entry test. Focusing on students as the main participants in this research, the issue of how Malaysian university students perceive the MUET and their own self-efficacy in relation to the test is examined. The study also investigates the washback length of the MUET, i.e. the continuation of the influence of the test even after the students have sat it. Therefore, this study aims systematically to re-examine beliefs concerning washback by investigating the relationship between the students' perceptions of the MUET in terms of its importance and difficulty, their own self-efficacy and the language learning strategies they employ in preparing for the test. Within a mixed methods approach, a student questionnaire, student interviews and classroom observation were employed to elicit data. Data were also collected using a teacher questionnaire for triangulation purposes. In general, the findings suggest that the students' perceptions play a major role in mediating the washback effect of the MUET, especially with regard to perceived test importance and self-efficacy. It was found that the students' perceptions of the test shaped their goals and consequently stimulated their use of language learning strategies when preparing for the test. The findings also revealed that measuring washback length was as complicated as expected due to the difficulty of controlling for other intervening variables. It is hoped that this study will inspire more research on washback and contribute more knowledge with regard to the scarcely explored area of longer term washback.

Table of Contents

<i>Abstract</i>	2
<i>Table of Contents</i>	3
<i>List of Tables</i>	8
<i>List of Figures</i>	11
<i>Acknowledgements</i>	12
<i>Declaration</i>	14
CHAPTER 1: INTRODUCTION	15
1.1 Background to the study	16
1.2 Statement of the problem	19
1.3 Context and purpose of the study	22
1.4 Aims and research questions	29
CHAPTER 2: LITERATURE REVIEW	30
2.1 Introduction	30
2.2 The concept of washback	30
2.2.1 Defining and describing washback	31
2.2.2 Defining “washback” in the present study	33
2.3 Mechanisms of washback	34
2.3.1 Alderson and Wall’s (1993) 15 washback hypotheses	34
2.3.2 Hughes (1993) trichotomy of backwash model	37
2.3.3 Bailey’s (1996) basic model of washback	38
2.3.4 Watanabe’s (2004) washback dimensions	40

2.3.5	Green's (2007a) model of the direction, intensity and length of washback	42
2.3.6	Shih's (2007) model of washback on students' learning	44
2.3.7	Zhan's (2009) model of washback on the learning process.....	48
2.4	Dimensions of washback	51
2.4.1	Direction: positive and negative washback.....	52
2.4.2	Extent, strength or intensity of washback	62
2.4.3	Length of washback	63
2.5	Complexity of the washback phenomenon	65
2.6	The role of students' perceptions in influencing washback.....	66
2.6.1	Perceptions of tests	67
2.6.2	Perception of self-efficacy	71
2.7	Measuring washback in the study.....	75
2.8	Conceptualizing washback in the study.....	82
2.9	Assessment context in Malaysia	84
2.10	Malaysian University English Test (MUET).....	86
2.10.1	Previous studies on the MUET	90
2.11	Summary	97
CHAPTER 3: METHODOLOGY.....		99
3.1	Introduction	99
3.2	Aim and research questions.....	99
3.3	Philosophical assumptions	101
3.4	Context of the study.....	103
3.4.1	Research site and criteria for selection	104

3.4.2	Study sample.....	105
3.5	Research design.....	109
3.6	Research approach	113
3.6.1	Quantitative research approach.....	113
3.6.2	Qualitative research approach.....	114
3.6.3	Mixed methods approach.....	115
3.6.4	Methodological justification for the approach based on prior washback studies	116
3.7	Research instruments	118
3.7.1	Quantitative data collection	118
3.7.2	Qualitative data collection	132
3.8	Data analysis	139
3.8.1	Quantitative data analysis	139
3.8.2	Qualitative data analysis	140
CHAPTER 4: RESULTS AND ANALYSIS		142
SECTION I: WASHBACK FROM THE MUET ON THE STUDENTS.....		142
4.1	Students' perceptions of the MUET.....	143
4.1.1	Perceived importance of the MUET	144
4.1.2	Perceived difficulty of the MUET	151
4.1.3	Students' self-efficacy in relation to the MUET.....	155
4.1.4	Students' perceptions of their MUET teachers.....	157
4.2	Students' perceptions and language learning strategies	165
4.2.1	Relationship between perceived test importance and language learning strategies	176
4.2.2	Relationship between perceived test difficulty and language learning strategies.....	178
4.2.3	Relationship between self-efficacy and language learning strategies.....	179
4.3	Washback length and intensity of the MUET.....	181

4.3.1	Washback length.....	181
4.3.2	Washback intensity	191
SECTION II: MUET WASHBACK FROM THE TEACHERS' PERSPECTIVE		194
4.4	Teachers' perceptions of the MUET	195
4.5	Teachers' attitudes towards aspects of learning	202
4.5.1	Recommended learning strategies	203
4.5.2	Proposed teaching activities.....	205
4.5.3	Strategies to motivate students to learn	206
4.5.4	Perceptions of the use of mock tests	207
4.6	Teachers' medium of instruction	208
4.6.1	Lesson preparation	208
4.6.2	Teaching activities	209
4.6.3	Teaching materials and resources	213
4.7	Teachers' difficulties in teaching the MUET	214
4.8	Classroom observation	215
4.8.1	Participant organization	216
4.8.2	Activities conducted in the MUET preparation classrooms	219
4.8.3	Teaching materials	220
4.8.4	MUET preparation classroom observation	221
4.9	Summary	234
CHAPTER 5: DISCUSSION		235
5.1	Introduction	235
5.2	RQ1: What are students' perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?	235

5.3	RQ2: How does washback of the MUET operate and to what extent do students’ perceptions seem to have a washback effect on their language learning strategies?	246
5.3.1	English language learning inside the classroom	247
5.3.2	English language learning outside the classroom	253
5.3.3	Students’ perceptions of the MUET and their language learning strategies	255
5.4	RQ3: What is the intensity of the washback effect of the MUET and what is its length? How do these appear to influence washback on the learners?	257
CHAPTER 6: CONCLUSION		261
6.1	Introduction	261
6.2	Summary of the main findings	261
6.3	Limitations of the study	265
6.4	Implications of the study	268
6.5	Suggestions for future research.....	271
Appendices.....		274
Appendix A – Student Questionnaire (Group A)		274
Appendix B – Student Questionnaire (Group B).....		278
Appendix C – Teacher Questionnaire.....		282
Appendix D – Student Interview Questions		287
Appendix E – Communicative Orientation of Language Teaching Observation Scheme (COLT, Spada & Frohlich, 1995)		288
Appendix F – MUET Sample Paper		289
References.....		321

List of Tables

Table 2.1 Characteristics of deep and surface approaches to learning	80
Table 2.2 MUET test scores for each component.....	87
Table 2.3 The MUET scoring guide	88
Table 3.1 Form 6 term duration and course outline.....	106
Table 3.2 Demographic information of the teachers	109
Table 3.3 Overall structure and sources of the student questionnaire	123
Table 3.4 Student interview participants	137
Table 4.1 Perceived importance of the MUET	144
Table 4.2 Perceived importance of each component tested in the MUET.....	149
Table 4.3 Students' perceived self-efficacy related to the MUET.....	156
Table 4.4 Teachers' role in encouraging students in MUET preparation class as perceived by students	158
Table 4.5 Teachers' teaching activities in MUET preparation class as perceived by students ..	160
Table 4.6 Students' reported language learning strategy use according to frequency.....	166
Table 4.7 Students' language learning strategies in the top 10 ranked according to the mean scores	167
Table 4.8 Students' language learning strategies in the bottom 10 ranked according to the mean scores.....	173
Table 4.9 Differences in learning strategies by gender.....	175
Table 4.10 Differences in learning strategies by proficiency level	176

Table 4.11 Correlations between perceived test importance and language learning strategies..	177
Table 4.12 Correlations between perceived test difficulty and language learning strategies.....	179
Table 4.13 Correlations between self-efficacy and language learning strategies.....	180
Table 4.14 Students' language learning strategies before and after the MUET (memory)	182
Table 4.15 Students' language learning strategies before and after the MUET (cognitive).....	184
Table 4.16 Students' language learning strategies before and after the MUET (compensatory)	186
Table 4.17 Students' language learning strategies before and after the MUET (metacognitive)	187
Table 4.18 Students' language learning strategies before and after the MUET (affective)	189
Table 4.19 Students' language learning strategies before and after the MUET (social)	190
Table 4.20 Regression analysis.....	193
Table 4.21 Teachers' recommended learning strategies (Q3)	204
Table 4.22 Teachers' proposed teaching activities (Q4)	205
Table 4.23 Teachers' methods for motivating students to learn (Q5)	206
Table 4.24 Teachers' perceptions of mock tests (Q6)	207
Table 4.25 Teachers' lesson preparation (Q1).....	209
Table 4.26 Teaching activities (Q2).....	210
Table 4.27 Teaching materials and resources (Q3)	213
Table 4.28 Teachers' difficulties in teaching the MUET (Q2).....	214
Table 4.29 Participant organization	216
Table 4.30 Activities conducted in the MUET preparation classrooms	219
Table 4.31 Teacher 1's observation scheme data	222
Table 4.32 Extract from classroom observation	223
Table 4.33 Extract from classroom observation	223

Table 4.34 Teacher 2's observation scheme data	226
Table 4.35 Extracts from classroom observation.....	227
Table 4.36 Extracts from classroom observation.....	228
Table 4.37 Teacher 3's observation scheme data	229
Table 4.38 Extract from classroom observation	230

List of Figures

Figure 2.1 Basic model of Washback (Bailey, 1996, p. 264)	39
Figure 2.2 A model of washback direction, variability and intensity (Green, 2007a, p. 24).....	43
Figure 2.3 A washback model of students' learning (Shih, 2007, p. 151).....	46
Figure 2.4 Mechanism of the effect of CET-4 on the learning process (Zhan, 2009, p. 278)	49
Figure 2.5 Key components of the instruction process (Mertler, 2007, p. 42)	61
Figure 3.1 Distribution of students in Group A and Group B according to gender.....	107
Figure 3.2 Distribution of students in Group A and Group B according to language proficiency.	108
Figure 4.1 Students' perceptions of the difficulty of the MUET	151
Figure 4.2 Students' perceptions of the difficulty of the MUET according to gender	154
Figure 4.3 Students' perceptions of the difficulty of the MUET according to English proficiency	154
Figure 4.4 Seating arrangement from classroom observation	217
Figure 4.5 Seating arrangement from classroom observation	218
Figure 4.6 Screenshot from classroom observation.....	224
Figure 4.7 Screenshot from classroom observation.....	224
Figure 4.8 Screenshot from classroom observation.....	225
Figure 4.9 Screenshot from classroom observation.....	232

Acknowledgements

I would like to take this opportunity to extend my greatest gratitude and appreciation to everyone involved in this journey of a lifetime of mine. First and foremost, I would like to thank my supervisor, Dr. Khaled El Ebyary, for his endless guidance, support and encouragement for me to complete a PhD under his supervision. It was a hell of a ride but it was worth it. I would also like to thank my Thesis Advisory Panel, Dr. Bill Soden for his invaluable input, detailed comments and ongoing support throughout my study years in the University of York. I could not wish for better supervision and support than these people provided.

The conduct of this research would not have been possible without the generous help from the student participants, teachers, who have contributed to this research by completing questionnaires, taking part in interviews and classroom observations. Without them, the completion of this research project would not have been possible. I am forever thankful both for their cooperation and kindness for assisting me with my data collection. In addition, the dissertation would not have shaped up so well without the help of my proofreader, Dr. Catriona Scott.

Most important of all, I would like to thank my dearest family and friends, both in the UK and Malaysia, who made this journey bearable at times when I was at my lowest and who were always there to celebrate my accomplishments, although I was thousand miles away from home. I would also like to extend a big thank you to the people in the Research Centre for Social Sciences for your kindness and friendship, and countless cups of Yorkshire Tea.

Last but not least, the funders of this studentship, Universiti Malaysia Kelantan, and the Ministry of Higher Education Malaysia, are gratefully acknowledged for the financial support and opportunities. Thank you.

Declaration

I hereby declare that no portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning. I further declare that this thesis is my own original work, except where reference is made in the text of the thesis to the work of others.

CHAPTER 1: INTRODUCTION

Language assessment is becoming one of the key areas of study in the field of educational research due to its major influence on teaching and learning. It is used to obtain evidence and information concerning students' knowledge, skills and abilities in the tested language. Candidates' performance in carrying out language tasks in the test will be used as an evidence of their mastery of the language. Unlike specific tests that measure expertise in a specific skill, language assessment is something that most people have to undergo regardless of their background. The results obtained from a language assessment or test, especially a high-stakes language test, will to a certain extent affect someone's life as such tests are used to make many crucial decisions, for example determining who should have access to important opportunities in areas including employment, further education and immigration. In other words, the results are often used as a gatekeeping device to filter and exclude those who are deemed to be unqualified.

In the field of education, it is well known that tests, especially high-stakes tests, have a considerable influence on teaching and learning. Specifically, in language education, the influence that tests exert over teaching and learning is known as "washback" or "backwash". These two terms are interchangeable, but as "washback" is commonly used in the field of language assessment, it will be used throughout the thesis (for a detailed discussion, see 2.2.1).

Tests, especially those with important consequences, are viewed as effective tools for exercising change. They are used as a lever of change to influence teaching and learning in the classroom. Shohamy (1993) discussed the power of tests in great detail in one of her seminal articles examining the impact of language tests on teaching and learning. According to Shohamy, a test is deemed powerful to the extent that the curriculum comes second after the test when it

comes to preparing and conducting lessons in the classroom. Educational innovations and newly designed curricula have been communicated to teachers through tests by authorities. If used correctly, such a powerful educational device can be immensely beneficial to policymakers, testers, teachers and students (Shohamy, 1993).

Based on extensive background research and reading on washback, it is apparent that washback has primarily been associated with the negative consequences of tests. However, this phenomenon was not empirically investigated until the 1980s. Since then, researchers in the field of language assessment have examined washback based on different perspectives and lenses as the findings of research on washback have continued to reveal the complexity of the phenomenon, an issue discussed further in the literature review (see 2.5).

This chapter situates the study within educational debate and the exploration of washback from a broader context (world-wide) to an immediate context (Malaysia). The purpose of this chapter is to provide the reader with a general idea of the focus of this study and why it was carried out. Against this background, the position of this study is identified and explained and the three research questions used to guide this study are delineated.

1.1 Background to the study

Previous research in countries such as Korea (Cho, 2004), Taiwan (Pan, 2014; Pan & Newfields, 2012), China (Qi, 2005) and Hong Kong (Qian, 2014), to name but a few, indicate that summative assessment can be one of the major stimuli for teaching and course design. High-stakes tests, usually referred to as “large-scale standardised tests or public examinations” (Shih, 2013, p. 13), are believed to have important consequences for classroom instruction and practice, as well as for syllabus and curriculum planning. In the field of assessment research, there is general consensus

that tests can influence the teaching and learning process, a phenomenon that has come to be known as the “washback” effect (Alderson & Wall, 1993).

According to Pearson (1988), the term washback originated from the backward direction of a test, as it often came at the end of a course. Washback is closely associated with learners and teachers, as the two main stakeholders of a test, besides test designers and policymakers (for more detail on the definition of washback and how it is defined in the present study, see 2.2.1 and 2.2.2.) Instead of focusing on the reliability and validity of tests, studies on the washback effect have centred on aspects that are related to the stakeholders of tests, such as the school curriculum, the behaviours of teachers and learners inside and outside the classroom, their perceptions of the test and how the test scores are used (Cheng, Watanabe, & Curtis, 2004).

Alderson and Wall (1993, p. 120), in their pioneering study on washback in Sri Lanka, hypothesized that “Tests that have important consequences will have washback” and conversely “Tests that do not have important consequences will have no washback”. In other words, the higher the stakes of a test, the greater the impact it will have on the teaching and learning process. According to Qi (2007), authorities tend to be tempted to resort to manipulating high-stakes testing in the name of “curriculum innovation” to attain immediate outcomes, as this is claimed to be the “quick and most cost-effective way to improve education” (p. 52). Studies on the washback of high-stakes testing have reported that tests can change students’ learning behaviour when preparing for the test if the stakes are sufficiently high, for example motivating them to put more effort into learning (Cho, 2004; Pan & Newfields, 2012; Thomas, 2005) and promoting learner autonomy (Pan, 2014; Stecher, 2002). However, the influence of washback is not limited to students. It has also been reported that high-stakes testing influences teachers’ teaching styles (Ferman, 2004; Green, 2014; Qi, 2005), as teachers align their classroom instruction to the test or

“[teach] to the test” (Pan & Newfields, 2011, p. 267); this can be both beneficial and detrimental for students.

From the literature on washback considered in this thesis thus far, it is apparent that there seems to be a negative tone or connotation associated with the term washback. This could be due to pressure exerted by the top-down management of the learning institution, in turn potentially influenced by a higher authority, such as the Ministry of Education (MoE). For instance, in Malaysia, the MoE aims to improve the quality of education by measuring students’ test scores and thus some teachers will resort to teaching to the test, narrowing the focus of their lessons to produce as high scores as possible. Therefore, it is common within the Malaysian educational system to see English language learners memorizing by rote rather than trying to understand a concept, because what matters to them is passing the test and what matters for the teachers is being able to present a good report of their students’ achievement to the board members of the institution.

On the surface, this might not appear harmful, but Tsagari (2006) cautions against the drawbacks that may accompany washback to the learners, particularly long-term anxiety and stress, even if there are considerable instrumental benefits that tests might bring to the learners. Ryan and Brown (2005) posit that when students are under great pressure to attain a certain score or a set minimum requirement on a high-stakes test, they will not be learning in a supportive environment. Hence, supporting students’ autonomy through a high-stakes testing policy might not be possible.

Although washback has been studied for almost three decades, research continues to reveal new aspects that need to be taken into consideration when investigating washback, further highlighting the complexity of the phenomenon. It has been established that test designers alone cannot engineer desirable change (Cheng et al., 2004); a well-designed test does not necessarily

bring about the anticipated outcomes in terms of improvements in language learning and language use. As pointed out by Alderson (2004), teacher-related factors play a crucial role in shaping the washback effect, as their beliefs and understanding of the nature and the rationale for the test will to a certain extent influence the way in which they prepare their students for the test. Similarly, students' beliefs and perceptions concerning the test itself are equally crucial and can critically influence the washback effect of the test. However, despite Bailey's (1999) call long ago for washback researchers to carry out more research on students' perceptions of a test, to date relatively few studies have been published from students' perspectives.

1.2 Statement of the problem

The connection between (1) testing and (2) teaching and learning practices has commonly been explored in the burgeoning research in this field (Cheng, Andrews, & Yu, 2011; Gebril & Brown, 2014; Luong-Phan & Effeney, 2015), sparked by Alderson and Wall's research into washback in the late 1980s. Alderson and Wall's (1993) seminal publication on washback raised the notion of its complexity and the need for more in-depth research, not only to describe what washback looks like, but also to account for what occurs (Alderson & Wall, 1993). With the increasing research undertaken on washback, it appeared that understanding of tests and test design were developing and the use of tests by different education bodies grew. It was thought that the mechanisms of washback were uniquely connected to "the test" and the context. However, as pointed out by Alderson and Wall (1993), although it is widely known that there is a relationship between testing and teaching and learning practices, the complexity of the washback concept makes it difficult to identify how tests affect teaching and learning practices without taking into account other mediators or variables that may or may not contribute to the influence. Similarly, Stoneman (2006)

observed that the washback phenomenon has yet to be clearly explained and explored in depth despite abundant literature in language assessment focusing on the impact of testing on teaching and learning.

To date, as noted above, research on the washback effect, particularly in foreign language learning, has tended to focus on teachers' perspectives as they are viewed as among the most important stakeholders in assessment, whereas there is a dearth of washback studies on learners (Pan, 2014). According to Cheng (2008) and Spratt (2005), as test-takers, learners' points of view need to be taken into consideration as they are the ones directly affected by these tests. It is unfair to learners and has proven rather problematic to employ a "one-size-fits-all" approach (Chu, 2009) in making judgments concerning the washback effect of high-stakes tests. Stoneman (2006) pointed out that not much is known about how learners and their learning are affected by tests or how washback works for them, as such aspects are often missing from research or not explicitly elaborated; this is still the case 12 years on. It is important for stakeholders to bear in mind that different learners might or might not react differently to the same tests, as articulated in one of Alderson and Wall's (1993) washback hypotheses, namely that "Tests will have washback effects for some learners and some teachers, but not for others" (p. 120). Even though this was proposed a quarter of a century ago, it is still the case that this rather fundamental aspect of washback has not adequately been addressed.

It has been reported that less proficient learners tend to be more worried about tests compared to highly proficient learners, but do not prepare for the test until the last minute (Chu, 2009; Stoneman, 2006; Watanabe, 2001), a phenomenon that could be attributed to test anxiety (Chen & Hsieh, 2011). However, Ferman (2004) found that less proficient learners adopted intense learning for the test in order to improve their scores, which contradicts the aforementioned studies.

Hughes (1993), as one of the earliest scholars to give due attention to washback, developed a preliminary model of washback in which the “participants” are one of the three main components, together with “processes” and “products” (see 2.3.2). Based on Hughes’ (1993) tri-partite model, positioning learners as the principal stakeholders in the washback phenomenon, it is important to take into consideration the variety of effects that tests can have upon learners (Pan, 2014) and also other variables. For example, Tsai and Tsou (2009) found that learners’ negative opinions of a test will lead to a decrease in motivation to learn. It would be interesting to explore the relationship between learners’ perceptions of the test and how these perceptions influence their course of action, for instance in terms of their language learning strategies when preparing for the test. Washback on learners is associated with how students react towards a test, for example by amplifying or reducing their effort to learn because of their perceptions of the test. The current study thus seeks to explore how learners’ perceptions determine the different extent of washback they experience by looking at the language learning strategies they adopt when preparing for a high-stakes English language test.

Concerns about the lack of empirical studies aimed at determining the relationship between how the two most important stakeholders of testing, namely the teachers and learners, are affected have been raised in the washback literature (Cheng & Curtis, 2004; Wall, 2005). Thus, this study was prompted by the scarcity of washback studies focusing predominantly on the learners, but at the same time considering the perspective of the teachers as interlinked. Moreover, as noted by Pan and Newfields (2012), most of the studies available on learner washback have not adopted an experimental or quasi-experimental research design with control and experimental groups. Comparing two or more groups could help to see clearly the washback effect of the high-stakes test under investigation. Hence, in the present study, the washback of a high-stakes language test

was explored not only from the perspective of the students who were preparing for the test, but also the students who had already taken the test. This focus on the effects before and after the same high-stakes test will help to validate the findings (see 3.4.2).

In addition, also with reference to research methodology, most studies on washback from high-stakes testing have employed self-report data, whether from learners or teachers. As noted by Shih (2013), such an approach does not yield rich data and there can be a mismatch between self-reported information and actual behaviours, as well as a risk of expectancy bias (Yu, 2018). In view of this, Pan and Newfields (2012, p. 119) proposed that in order to get a “more accurate and dynamic picture” of the washback effect on different stakeholders, future research should include more classroom observation data, which will be addressed in this study. The target examination chosen for investigation in this study was the Malaysian University English Test (MUET), a high-stakes language test in Malaysia. The following section discusses the MUET and the context and purpose of this study in detail.

1.3 Context and purpose of the study

In the context of learning English as a second language (L2), the ultimate goal of learning is the ability to communicate effectively using the language. As the predominant international language of the world, the importance of mastering English is often emphasised by authorities, especially in countries where English is not the first language. English also holds a unique role in many of the countries colonised by the British in the past. As a result of their history, many of these colonised countries have adopted a social system of communication in which English is widely used, especially among those more highly educated. Malaysia was among the countries that experienced a period of colonisation, long enough to have English embedded in present use. During the

colonisation era, English was not only the language of administration and other government and economic matters, but was also the medium of instruction in highly prestigious schools in Malaysia, making it a symbol of power and a sign of the educated.

Fast forward to the present time, English still plays a major role in Malaysia. This is due to the fact that English language is a lingua franca used widely – indeed globally – in a variety of social contexts, such as political, business, education, industry and media (Kassim & Ali, 2010; Menon & Patel, 2012). Thus, in order to compete in the international arena, a strong workforce with high proficiency in English is needed. A good command of English will boost students' marketability and competitiveness in securing a job. As Hanapiah (2004) contends, alongside the skills in their respective fields, graduates who are proficient in English are more likely to be employed. In Malaysia, it is common for English proficiency to be listed as a requirement for potential candidates in many job advertisements. Mastery of English has been cited as one of the most important factors for graduate employment (Sirat, Buang, et al., 2004; Sirat, Bakar, Lim, & Katib, 2004). Those who are proficient in English will have more opportunities; graduates are expected to have a sufficient level of English language proficiency to meet the anticipated needs of the job market, both locally and internationally (Pan & Newfields, 2012).

As mentioned earlier, test results are often used as a gatekeeping filter for candidates in gaining access to important opportunities in areas such as employment, further education and immigration. One of the most popular utilizations of high-stakes tests is as a university entrance exam. Two of the high-stakes language tests most commonly used worldwide in relation to further education are the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL). These two tests are commonly used to help make critical decisions concerning admission to institutions for academic training (Chalhoub-Deville & Turner,

2000), as candidates are expected to possess a certain level of English language proficiency in order to meet the linguistic demands of their respective course of training. Similarly, various countries have their own specific high-stakes language tests for university entrance, for example the National Matriculation English Test (NMET) in China (Qi, 2005, 2007), the Entrance Exam of the Universities (EEU) in Iran (Yunus & Salehi, 2012) and the EFL National Oral Matriculation Test in Israel (Ferman, 2004), to name but a few. Likewise, in Malaysia, the MUET is a compulsory component for candidates who are applying to study in Malaysian public universities.

The MUET was first introduced in 2000 with the aim of bridging the gap in English language needs between secondary and tertiary education (Malaysian Examination Council, 2001) and consolidating and enhancing the English proficiency of students preparing to enter Malaysian public universities (Lee, 2004). The four language skills, reading, writing, listening and speaking, are tested in the MUET and performance on the test is reported in terms of an aggregated score with respect to six levels of achievement, referred to as Bands 1–6, Band 1 being the lowest and Band 6 the highest (see Chapter 2 for more detail). As a criterion-referenced test, namely one that is designed to assess students' academic performance against pre-determined standards or criteria, monitored by the Malaysian MoE, the MUET is used to gauge the overall English language proficiency of candidates applying for their first degree programme at tertiary level. It is designed and administered by the Malaysian Examination Council and is recognized in Malaysia and Singapore (Othman & Nordin, 2013). This test is significant for pre-degree students as it serves as an indicator of their English language proficiency, enabling them to enrol on their desired course (Kaur & Nordin, 2006). Students need to obtain a minimum band on the MUET, according to their desired field of study, for entry into public institutions of higher learning in Malaysia. Examples of the specific minimum requirements for courses according to field of study are as follows: (a)

Arts and Social Science – MUET Band 2; (b) Science, Technology, Engineering and Mathematics (STEM) – MUET Band 3; (c) Law and Medical – MUET Band 4.

Moreover, the introduction of the MUET was also intended to curb the worrying deficiency in the standard of English language proficiency among Malaysian students once they reach university. As noted by Omar (1992) 27 years ago, many students are unable to perform well once they are at university due to poor mastery of English. Hence, the MUET was introduced to help bridge the language gap between pre-university and university levels as the English language learning environment in Malaysia was deemed not conducive for English language development. Thang (2004), based on her 2001 doctoral dissertation, investigated the motivational elements of approaches to studying among Malaysian undergraduates. She distributed a total of 1,500 questionnaires to all first and second year students from three different faculties at the Universiti Kebangsaan Malaysia (UKM), a public university in Malaysia, to elicit data and 721 were returned. She found that most of the students expressed awareness of the importance of language, but primarily focused on the extrinsic value rather than the intrinsic value of learning English. The students appeared to have some reservations with regard to learning English to a certain extent. Thang (2004) attributed this to the prejudice attached to those who attempted to use English as “showing off”. In my experience, both as a student and a lecturer in the Malaysian context, this issue, while it might appear extreme to outsiders, is actually common, especially among less proficient learners. The lack of a conducive environment hampers and demotivates many students in Malaysia from actually practising English in real life.

Fast forward to 12 years later and Ganapathy and Ying’s (2016) study on factors that influence Malaysian students’ attitudes towards and motivation for learning English elicited similar findings. ‘Attitude’ refers to an outcome of chronologically cultivated values and beliefs

developed over time in a given socio-cultural context (Liu, 2014) whilst ‘motivation’ is denoted as a force that strengthens and guides behaviour towards a goal or to a particular action (Brown, 2000; Saville-Troike, 2006). Their study aimed to identify ways of helping students to learn English more effectively by tackling the problem from its roots. In their qualitative case study, involving 20 students participating in focus group interviews, they discovered that although most of the participants generally had positive attitudes towards learning English, several had developed negative attitudes. The reason for this lay in inadequacies in terms of teaching and resources and a lack of effort to impart awareness of the importance of mastering the language.

According to Philip and Koo (2006), it is common for students to remain silent in language classrooms in Malaysia. The issue of students seldom volunteering to answer questions or give opinions was identified by Hamid (2001), who noted that students try their best not to be called on by lecturers during lessons by avoiding eye contact. Drastic measures should be taken to “pull” them out of their comfort zone and start taking English language learning more seriously to prepare them for their future undertakings. Therefore, the MoE in Malaysia decided to introduce a standardized English test for university entry, the MUET.

In the context of the present study, the MUET is regarded as a very important test, indeed one of the most “fearsome” tests, by Malaysians. Aside from MUET preparation classes provided by institutions, there are also many private learning institutions that provide MUET preparation classes for those who are willing to spend extra to help them prepare for the test. Textbooks containing model answers and past questions can also easily be found at local bookstores in Malaysia. Typing “MUET Tips” in the Google search engine also yields more than one million hits for websites, blogs and vlogs sharing information on the MUET, for example “Find out how you can score Band 6 in MUET” and “MUET Reading Exam Guide and Tips by a MUET expert”,

to name but a couple. However, studies have shown that a majority of Malaysian undergraduates still have a poor command of English, despite having 11 years of formal education in primary and secondary school (Jalaluddin, Awal, & Bakar, 2008; Kaur & Nordin, 2006) on top of having to sit this high-stakes English language test.

However, there appeared to be an attempt to exploit the MUET itself by using it as a university exit test, on top of it being used as a university entry test in Malaysia. This caused a lot of pressure especially among the students as the test now comes with two important consequences not only for university entry, but also for university graduation purposes. As a lecturer myself, there was a period of time when a lot of students in the university that I was teaching were not able to graduate just because they did not achieve MUET Band 3, which was the minimum requirement set to graduate in most public universities in Malaysia at that time. There was a case of one student who retook the MUET 7 times but still did not manage to obtain Band 3 and was not able to graduate even though she had finished all her courses. With such a high pressure and such high stakes at hand, the MUET would surely cause intense washback to both the students and the teachers/lecturers. Due to this, a few years later, the MOE decided to discard the MUET graduation requirement altogether and universities were given the autonomy to manage university students' English language mastery by strengthening their EAP/ESP courses offered at the respective institutions.

Although the MUET was introduced to prepare students to cope with the demands of English at the tertiary level, Nambiar and Ransirini (2012) reported that many students continue to struggle at university as they do not possess a level of English language proficiency that would enable them to handle academic tasks. Similarly, Othman and Nordin (2013) highlighted that students need to possess a certain level of English proficiency to cope with the linguistic demands

in tertiary education, as most public universities in Malaysia use English as their medium of instruction. Most of the reference materials and resources used in teaching and learning are in English. Students embarking on tertiary education need to have at least a minimal mastery of the language of instruction in their courses, for example Band 4 in the MUET if they are applying for medicine or law courses. Introducing the MUET to bridge this gap has appeared to be rather unsuccessful. The most recent published report on MUET results, in 2014, showed that less than 1% of 182,457 candidates had achieved the highest score, i.e. Band 6, whereas more than 50% had only achieved the two lowest bands (Band 1 and Band 2). Moreover, Ali (2014) claimed that the MUET seems to have overridden classroom instruction and activities, which is in contradiction with its objectives.

Recently, there has been a call to the Ministry of Education Malaysia to completely abolish the MUET. It was reported in one of the local newspapers in Malaysia that some parties (politicians and scholars included) claimed that the MUET was “stressful and a financial burden on students” (Berita Harian, 23 February 2019). According to the report, the MUET was seen only as a university entry test and it should have been combined with the high school English language syllabus instead of having it as a separate test. One person even went to the extent of accusing the Malaysian Examination Council of making financial gain from the MUET as students are required to pay a certain fee every time they sit for this test. When news of this nature was brought up as headlines in local newspapers, it showed that many Malaysians in particular, were still not aware of the actual objectives of why the MUET was introduced in the first place, 20 years ago. This has given rise to grave concerns regarding the execution of the MUET and the need to explore the washback of the MUET as a university entry test. Hence, it is deemed timely for this research to be carried out in order to explore the washback impact of the MUET in greater detail.

1.4 Aims and research questions

The overarching aim of this study was to investigate the washback effect of the MUET on Malaysian university students. To do so, this study focused on examining (i) students' perceptions of the MUET and (ii) their language learning strategies. In addition, this research was also interested in understanding the length of the washback effect from the MUET on students. The research questions guiding this study were as follows:

1. What are students' perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?
2. How does washback of the MUET operate and to what extent do students' perceptions seem to have a washback effect on their language learning strategies?
3. What is the intensity of the washback effect of the MUET and what is its length? How do these appear to influence washback on the learners?

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter begins by providing a detailed explanation of the concept of washback, including its origin, debates concerning the definition of washback and the dimensions and mechanisms of washback. It provides a review of empirical washback studies in general education and in English language learning. The review of the washback literature helped to identify relevant baseline theories surrounding the washback phenomenon and to provide an extensive overview of previous research to help position the current research in relation to the literature and ultimately support the discussion of the study findings. Specifically, this study focused on the washback effect of the Malaysian University English Test (MUET), a high-stakes standardized English language test. This study was interested in exploring the relationship between students' perceptions of the test and their language learning strategies when preparing for the test. The second part of the chapter covers the theoretical and conceptual justifications for the study. Relevant washback models from previous studies are then described before presenting information on washback in language assessment specifically, which was the focus of this study. Before concluding the chapter, a review of the literature specific to the washback of the MUET in Malaysia is presented to help establish the context of this study.

2.2 The concept of washback

As noted by Alderson and Wall (1993, p. 1), "tests are held to be powerful determiners of what happens in the classroom". Thus, if one wishes to change how students learn or how teachers teach, one needs to change the methods of assessment. This shows how powerful a test can be in shaping

and influencing the teaching and learning process. This phenomenon is referred to as “washback” (Alderson & Wall, 1993; Bailey, 1996; Messick, 1996; Pearson, 1988) or “backwash” (Biggs, 1995; Hughes, 2002; Spolsky, 1994). Although the term “backwash” emerged first and can be found in several prominent dictionaries, such as Cambridge, Merriam-Webster and Oxford, this term is commonly used in the field of general education. In contrast, the term “washback” is more prevalent in applied linguistics and the language testing literature (Wall, 2012).

In general, the concept of washback is rooted in the notion that the teaching and learning process is highly influenced by tests or examinations (Cheng & Curtis, 2004). However, there is more to washback than simply the impact a test will have on teaching and learning. Wong and Chan (2009) went further and related tests, in particular language tests and their washback effects, to a form of “social engineering”, as the content of a language test constitutes texts that are constructed within the context of social practice. Such social engineering “may manifest in curriculum change, certification requirements and public demonstration of achievement as a valued outcome based on test results” (Wong & Chan, 2009, p. 253). The following sections discuss the terminology of washback in detail to establish the definition of the concept and the scope of washback for this study.

2.2.1 Defining and describing washback

A definition commonly used for washback is that it is “the influence of testing on teaching and learning” (Alderson & Wall, 1993, p. 115). Although washback and backwash carry the same meaning, the term washback is more prevalent in the language teaching and testing literature, while “backwash” is more commonly found in general education research. Some researchers consider only foreseen and intended effects of tests as washback, as for them the primary goal of a test is

the control of curricula (Spolsky, 1994). However, most educators seem to agree that any effects that a test have on teaching and learning, be they positive or negative, intended or unintended, can be referred to as washback (Alderson & Wall, 1993; Bachman & Palmer, 1996; Cheng, 2005; Cheng et al., 2004; Hughes, 2002; Hung, 2012). This is rooted in the notion that washback in general deals with any impact associated with tests, regardless of the nature of the impact.

In this vein, as thoroughly argued by Cheng (2005, p. 112), washback refers to “an intended or unintended (accidental) direction and function of curriculum change on aspects of teaching and learning by means of a change of public examinations”. Hence, any impact associated with the introduction of a targeted test for a specific purpose can be deemed to constitute washback. Other interchangeable terms for washback are “measurement-driven instruction” (Popham, 1987; Shohamy, 1992), “curriculum alignment” (Madaus, 1988; Resnick & Resnick, 1992; Shohamy, Donitsa-Schmidt, & Ferman, 1996), “test impact” (Andrews, 2004; Bachman & Palmer, 1996; Wall, 1997), “systemic validity” (Frederiksen & Collins, 1989), “consequential validity” (Messick, 1989, 1996) and “washback validity” (Morrow, 1986). As noted by Hsu (2010), although different terms have been preferred by different researchers, they all concern different aspects of the same phenomenon.

Another common term for washback and one that is used interchangeably by researchers in the field of assessment is “impact”. Although this may seem to carry the same meaning as washback, impact is normally used when referring to the effects of testing in general educational circles, as opposed to washback, which refers to the effect of testing specifically on teaching and learning (Booth, 2012). For Bachman and Palmer (1996), impact operates at two levels: (i) the micro level – a narrow view within the classroom (test effects on the teachers and learners); (ii) the macro level – a broader and more holistic view beyond the classroom (i.e. test effects on society

and the educational system). Rather than considering the two concepts separately, Wall (1997, p. 291) treated washback as a subset of impact by defining the term impact as “any of the effects that tests may have on individuals, policies or practices, within the classroom, the school, the education system, or society as a whole” and classified washback as the type of impact that appears in the classroom, as it is “more frequently used to refer to the effects of tests on teaching and learning”. Hamp-Lyons (1997, p. 299) adopted a similar view, suggesting that washback should be taken into account within the scope of impact, rather than being treated as a different concept altogether.

In this study, the term “backwash” will be retained in direct quotations and the term “impact” will be used in a non-technical sense, i.e. as an alternative when referring to anything associated with the effect, consequence, or influence of testing.

2.2.2 Defining “washback” in the present study

As argued by Alderson and Wall (1993), the terms washback and backwash can be used interchangeably since they refer to the same thing and carry the same meaning. For the purpose of this study, the term “washback” was chosen as it is commonly used in the field of language assessment. In this study, washback was used in the sense suggested by Hughes (2002), namely that it not only includes the effects of tests on teaching and learning, but also on the educational system and society as a whole. Bailey (1999) noted that “There are differing points of view about what the construct may encompass” (p. 9). As more research is being conducted on washback, the definition of washback keeps expanding, with researchers taking increasing numbers of factors into consideration when investigating washback. Starting from looking at student outcomes and test preparation strategies, then uncovering the layers of washback complexity, washback research has now expanded to consider individual differences in terms of the factors that could affect

washback. As suggested by Bachman and Palmer (1996), the washback of a test should be investigated with reference to the contextual variables of societal goals and values, the educational system in which the test is used and the potential outcomes of its use, as washback is more than simply the effect of a test on teaching and learning.

Having defined washback, to gain a better understanding of its nature it is timely to study how washback works by examining its various mechanisms. Therefore, the following section reviews the mechanisms of washback, examining the key washback models employed in this study. The review made it possible to conceptualize and exemplify the various aspects involved in the washback phenomenon relevant to this study.

2.3 Mechanisms of washback

This section starts with an overview of three seminal washback models, namely Alderson and Wall's (1993) 15 washback hypotheses, Hughes' (1993) trichotomy of backwash model and Bailey's (1996) basic model of backwash. These three models are often referred to when teasing out the complexities of the washback phenomenon and describing how washback functions. In addition to these models, four other more recently developed washback models relevant to the present study are discussed, namely Watanabe's (2004) washback dimensions, Green's (2007a) model of washback direction, variability, and intensity, Shih's (2007) washback model of students' learning and Zhan's (2009) washback on the learning process model.

2.3.1 Alderson and Wall's (1993) 15 washback hypotheses

In language testing, washback is no longer a foreign concept among researchers and educators. Perhaps the most prominent catalyst for research on washback arena stemmed from Alderson and

Wall's (1993) Sri Lankan study, which investigated how examinations might influence the teaching and learning. In their seminal article, entitled "Does Washback Exist?", they reviewed studies conducted in the Netherlands, Turkey, Nepal and Kenya, followed by their own large-scale study of washback in Sri Lanka (Wall & Alderson, 1993), investigating the implementation of a new test and its impact on updating curricula and teaching methodology. Based on their study, they proposed 15 hypotheses referring to areas of teaching and learning that are generally affected by washback, as follows:

1. A test will influence teaching.
2. A test will influence learning.
3. A test will influence what teachers teach.
4. A test will influence how teachers teach.
5. A test will influence what learners learn.
6. A test will influence how learners learn.
7. A test will influence the rate and sequence of teaching.
8. A test will influence the rate and sequence of learning.
9. A test will influence the degree and depth of teaching.
10. A test will influence the degree and depth of learning.
11. A test will influence attitudes to the content, method, etc. of teaching/learning.
12. Tests that have important consequences will have washback.
13. Tests that do not have important consequences will have no washback.
14. Tests will have washback on all learners and teachers.
15. Tests will have washback effects for some teachers and some learners, but not for others.

(Alderson & Wall, 1993, pp. 120–121)

Their argument centred on the need to define clearly various dependent variables in washback research to see their relationship. According to Green (2007a), the potential dependent variables of washback indicated in the 15 washback hypotheses, including the content, methods, rate,

sequence, degree and depth of teaching and learning, have been and are still being used to guide washback studies. Alderson and Wall (1993) concluded that more research on washback was needed and that such research should be based on a more strictly defined specification of washback, using a range of instruments and including classroom observation for data triangulation. They also called upon the researchers to consider the findings in the research literature in at least two areas: (i) motivation and performance and (ii) innovation and change in educational settings.

Three years later, Alderson and Hamp-Lyons (1996), in an attempt to explore and understand the nature of a high-stakes language test preparation classroom, revisited the 15 washback hypotheses and argued the need to refine them further as, based on the literature, some were rather general and too simplistic. They further suggested an expansion of the fifteenth washback hypothesis, proposing that the amount and type of washback will vary according to following:

1. the status of the test (the level of the stakes);
2. the extent to which the test is counter to current practice;
3. the extent to which teachers and textbook writers think about appropriate methods for test preparation;
4. the extent to which teachers and textbook writers are willing and able to innovate.

(Alderson & Hamp-Lyons, 1996, p. 296)

Alderson and Wall's (1993) hypotheses highlighted the potential complexity of washback, as opposed to the previous view that the relationship between tests and teaching and learning is linear. However, McNamara (2000) pointed out that the 15 washback hypotheses did not mention factors that contribute to how and why teachers and learners behave in a certain way in the classroom. To date, we know that washback does exist, but we do not know how or why. The key issue here is that the underlying factors that cause washback are manifold and need to be explored further. The

15 washback hypotheses might have helped washback researchers to look at washback differently and opened up more research avenues, but many aspects still need to be explored.

2.3.2 Hughes (1993) trichotomy of backwash model

Developing a basic model of washback, Hughes (1993) attempted to illustrate the mechanisms by which washback may work, categorizing the types of effects that might occur in relation to three main components, namely the “participants”, “processes” and “products” of an educational system. Participants refer to those who are directly involved in the test, such as teachers, students, test developers and policymakers. According to Hughes (1993), a test can first influence the perceptions and attitudes of participants towards their work, which will then affect their behaviours. The specific term used to label this route of action is “processes”. Processes refer to “any actions taken by participants which may contribute to the process of learning, such as materials development, syllabus design, changes in teaching methodology, the use of test-taking strategies”, undertaken by the participants to obtain their desired products, i.e. “what is learned and the quality of learning” (Hughes, 1993, p. 2).

Hughes (1993) also proposed five conditions that need to be met in an attempt to promote positive washback, as follows:

- i. Success on the test must be important to the learners.
- ii. Teachers must want their learners to succeed.
- iii. Participants must be familiar with the test and understand the implications of its nature and content.
- iv. Participants must have the expertise which is demanded by the test (including teaching methods, syllabus design, and materials writing expertise).
- v. The necessary resources for successful test preparation must be available.

(Hughes, 1993, pp. 2–3)

However, this model does not clearly explain why participants have various perceptions of tests and react differently. Individual differences were not taken into consideration in this model, despite Hughes' (1993) emphasis on the consequences that a test can have for learners. As mentioned earlier, the stakes of the test are not solely a property of their importance in terms of consequence, but also relate to how the stakeholders perceive them. Hence, there is a possibility that two students might have different perceptions of the same test and might react differently to it. Bearing in mind that it has been highlighted in numerous studies how complex washback is, Hughes' (1993) model needs to be developed further as it does not denote other factors which might affect the process of teaching and learning besides the test itself and focuses too much on learners' desire for success.

2.3.3 Bailey's (1996) basic model of washback

Based on Hughes' (1993) washback trichotomy and Alderson and Wall's (1993) 15 washback hypotheses, Bailey (1999) developed a washback model (Figure 2.1) to illustrate the inter-relationships of the mechanisms involved.

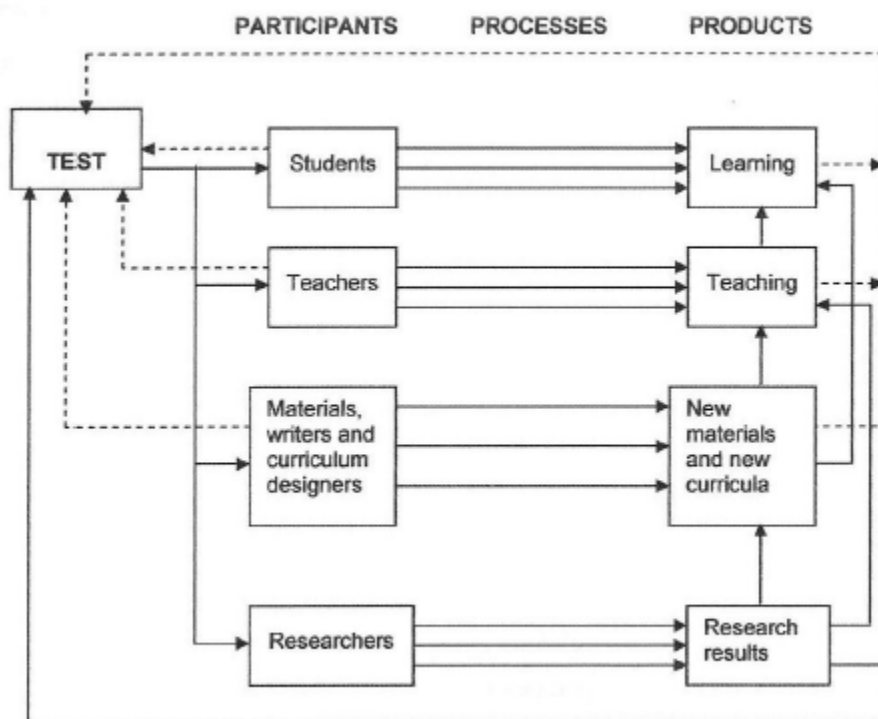


Figure 2.1 Basic model of Washback (Bailey, 1996, p. 264)

Bailey (1996) distinguished teachers from students and teaching from learning in her model. She also added the role that researchers play in the process of the washback of a test. Following the flow of Bailey's (1996) washback model from left to right, a test is predicted to affect the participants' (teachers and students) perceptions of the test and their endeavours. These perceptions and attitudes towards the test will then affect their behaviours, for example in the form of their teaching and learning strategies to prepare for the test. The unbroken lines represent the direction of impact that one would normally expect, for example the influence of a test on teaching and learning. One of the shortcomings of Bailey's (1996) model of washback, as pointed out by Hamp-Lyons (1997) and Wall (1997), is that she does not show precisely what the intermediate processes are and how they lead to the corresponding products. Taking into account that Bailey's

(1996) basic model of backwash was developed to capture how washback works, “processes” comprise a vital component of the model that need to be identified, as is a focus of this study.

Since the development of Bailey’s (1996) model and the earlier proposals of Alderson and Wall (1993) and Hughes (1993), many washback studies have been conducted, either to test them, or to contribute new discoveries to the field. The following sub-sections present reviews of three recent models of washback, all developed based on empirical evidence, that were used to guide this study.

2.3.4 Watanabe’s (2004) washback dimensions

One of the overarching aims of this study was to explore the length of the washback effect. To date, to the best of my knowledge, only one washback model has explicitly included washback length, namely that of Watanabe (2004). Watanabe (2004) conceptualized washback as comprising five dimensions: specificity, intensity, length, intentionality and value.

- ***Specificity***

Washback can be general or specific, depending on how broad or limited the scope of the test is. General washback refers to the “effect that may be produced by any test” (Watanabe, 2004, p. 20). In contrast, specific washback denotes only one particular aspect of a test or test type. A useful example of specific washback is when test designers introduce a new component into a test in the hope that the teachers and learners will emphasize this particular aspect in teaching and learning.

- ***Intensity***

The higher the stakes of a test, the more value it carries, hence making the washback effect stronger or more intensity; the converse holds for low-stakes tests. Thus, washback intensity can be “strong” or “weak”. Washback is considered to be strong when it has the power to determine everything that happens in the classroom and influence most, if not all, stakeholders, in particular inducing teachers and learners to react in the same way to prepare for a particular test. If the test affects only a part of classroom events, or some teachers and learners but not others, the intensity of the washback is considered weak. Cheng (1998) suggested that the intensity of the test is related to its stakes.

- ***Length***

Washback can have either short-term or long-term effects on stakeholders, especially the learners. If learners are seen to be increasing their efforts to learn and adopt certain learning strategies in preparing themselves for the test, but discard them once the test is over, the influence of the test is considered a short-term effect. If the influence endures even after the test ends, there is a long-term effect.

- ***Intentionality***

When test developers design a test, its objectives can be linked to intended washback, namely when a test does what it is meant to do. However, a test can produce intended or unintended washback, or both, depending on the objectives of the designers and those who implement it.

- *Value*

Despite the negative connotation of the term washback, a test can yield both negative and positive washback. However, there are no clear parameters for determining whether the washback of a test is positive or negative because the evaluation depends on who carries out the investigation within a particular context (Cheng & Curtis, 2004). Cheng (2005, p. 8) concluded that the value of a test will “largely depend on where and how it works and within which educational contexts it is situated”.

2.3.5 Green’s (2007a) model of the direction, intensity and length of washback

Figure 2.2 presents Green’s (2007a) basic model of the direction of washback. This model was developed and tested based on an extensive empirical study investigating whether test preparation classes were advantageous in helping learners to improve their writing scores on IELTS. This study involved 663 students participants which represented 50 nationalities from eight universities, two colleges of further education (FE) and seven private language colleges in the UK. In his study, Green (2007a) was concerned with the influence of the IELTS Academic Writing Module on preparation for academic study and the equivalence between IELTS test preparation and other forms of English for Academic purposes directed at university study. He utilised IELTS Academic Writing Module, tests of grammar and vocabulary, questionnaires, focus group interviews and also classroom observations at selected centres. Based on the outcome of his study, Green (2007a) asserted that there was adequate evidence to support his washback model which consists of three washback dimensions: direction, variability and intensity. When it comes to determining the direction of washback (see top part of the model) whether it is positive or negative, Green (2007a) proposed that the closer the characteristics of a test reflect the focal construct as understood by the

course providers and learners, i.e. the bigger the overlap, the higher the chance for positive washback to occur. Conversely, if the overlap is small, there is a high possibility that the test will produce negative washback. For example, taking the MUET into consideration, the MUET is intended to reflect the language or the target needs of prospective undergraduate students to survive the language demands at tertiary level in Malaysia. Hence, the better the MUET represents the language skills required in the university, the more likely it is to engender positive washback. If a test and curriculum are not in line with the focal construct of the test, then there is a high probability for negative washback to occur.

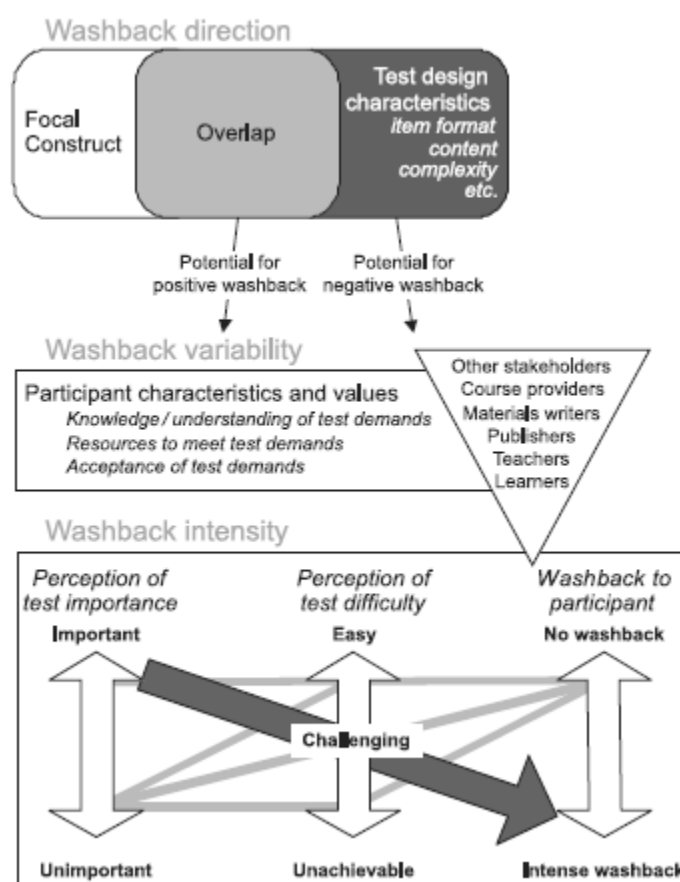


Figure 2.2 A model of washback direction, variability and intensity (Green, 2007a, p. 24)

In terms of variability, Green (2007a) emphasized the importance of individual differences when investigating washback, as different participant characteristics and values lead to different washback effects; no one learner is the same as another.

The third dimension covered in Green's (2007a) model is washback intensity. According to the model, perceptions of the importance and difficulty of a test will influence the intensity of washback for participants. For example, if the test is perceived to be important and sufficiently challenging, it will lead to intense washback. The relationship between perception and intensity is clearly illustrated in the model and seems straightforward, but it is not as simple as it might appear. For instance, if a test is regarded as important but easy, the washback intensity could be weak; indeed, according to Green (2007a), under such conditions there will be no washback. However, the extent to which this is the case requires empirical testing. It is thus crucial to investigate both the perceived importance and difficulty of tests as separate constructs to establish their relationship with washback intensity.

Moreover, Booth (2012) argued that Green's (2007a) model omits a wide range of mediating factors needed to demonstrate a more complex view of the directional nature of washback. In this regard, Shih's (2007) washback model of student learning might provide a complementary perspective on Green's (2007a) washback model.

2.3.6 Shih's (2007) model of washback on students' learning

Shih (2007) investigated the washback on learning of the General English Proficiency Test (GEPT) in Taiwan, exploring how the test is linked to the social and educational context in which the test is administered. His research was also driven by the belief that individual differences are among the factors influencing washback that require greater attention in language assessment research.

He collected data using interviews, observations, and reviews of relevant documents of department chairs, three English teachers, 15 students and three students' family members making sure that his participants reflected the diversity of faculty members and students in two universities in Taiwan. On top of that, he also carried out observation both in classrooms and in self-study centre at both universities over the period of 8 weeks. Shih (2007) found that the test itself was not the sole factor affecting students' test preparation strategies, suggesting the intricacy of washback and the need to take other factors into account in determining the potential washback of a test. Hence, Shih (2007) developed a tentative washback model of students' learning (Figure 2.3), distinguishing five different categories: content of learning, total time on learning, learning strategies, learning motivation and anxiety. Possible factors that could lead to washback were listed according to three groups, namely extrinsic factors, intrinsic factors and test factors.

A washback model of students' learning

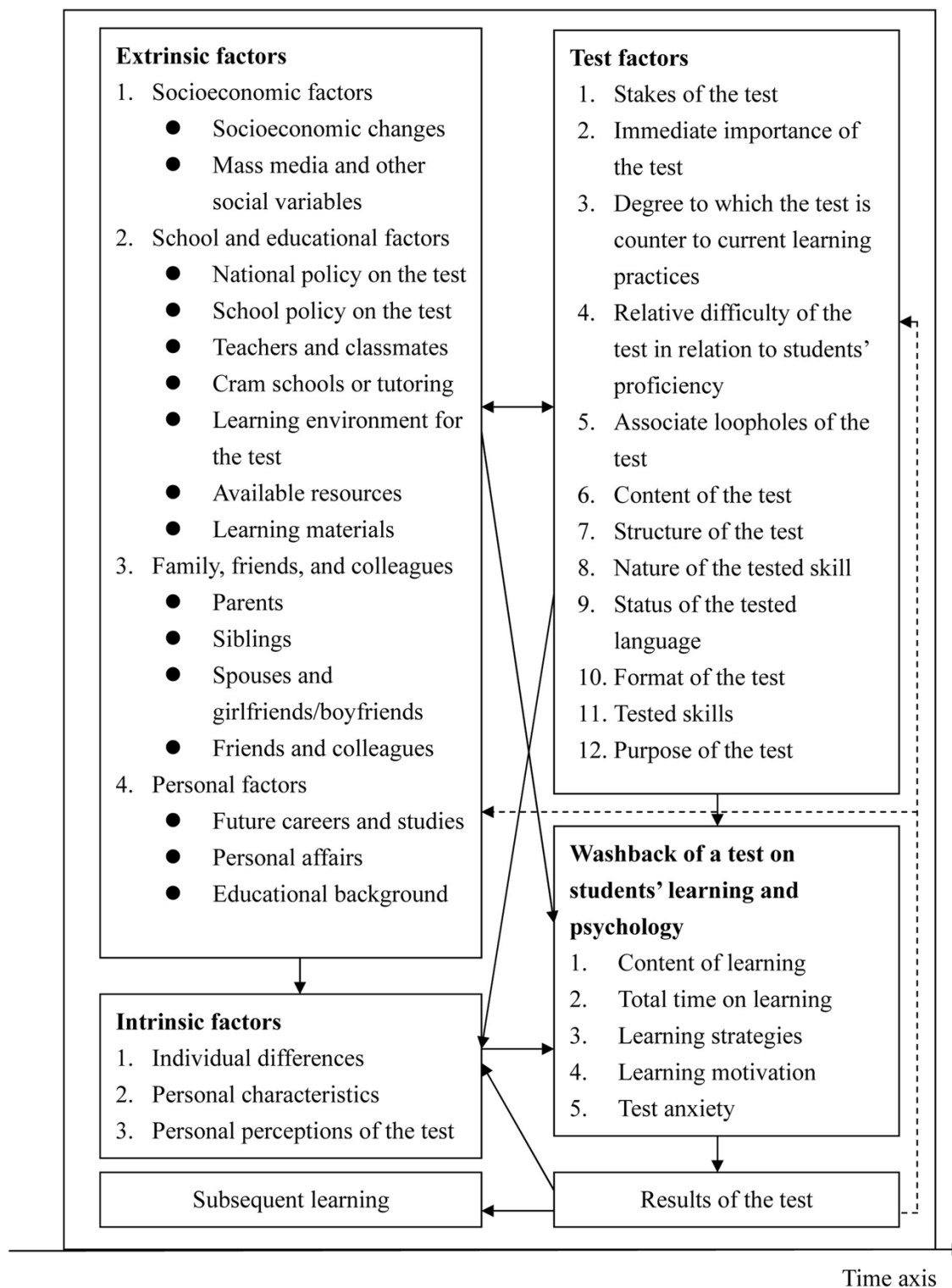


Figure 2.3 A washback model of students' learning (Shih, 2007, p. 151)

The arrows in the model indicate the relationships between these factors and how they influence and affect each other. These relationships were delineated based on the findings of previous washback studies, except for the relationships marked with dotted lines between the results of tests and extrinsic and test factors. These relationships were yet to be proven empirically, but were viewed as mostly likely to happen.

As mentioned previously, since washback effects are liable to change over time, Shih (2007) included the time factor in his model as one of the variables reflecting the impact of a test on students' subsequent learning. In this regard, when designing a test or educational policies, it is also important to investigate the aftermath of the test. Moreover, the process of preparing for a test might also affect students' learning after they sit the test, considered in greater depth in this study in terms of the dimension of washback length. As can be seen in Shih's model, test factors may affect extrinsic factors and vice versa. For example, if students' future studies (extrinsic factor) are determined by how well they perform in the university entrance exam (test factor), they might increase the total time spent on learning or use specific learning strategies to prepare for the test. Again, it has to be highlighted here that different learners operate differently, hence the need to take their perceptions into account, as addressed in Shih's (2007) washback model of students' learning. Shih (2007) explained that individual differences cover students' various reactions towards the test. Some students may be willing to prepare for the test while some students may think otherwise. Personal characteristics refer to the impact of students' personalities or other inherent characteristics, for instance fighting their own laziness when preparing themselves for the test. The third intrinsic factor listed in the model is personal perceptions of the test which could also affect students' test preparations based on the consequence(s) that comes with the test.

Another study that looked into individual differences in washback study in greater detail is Zhan's (2009) study on washback and possible selves as discussed in the next section.

2.3.7 Zhan's (2009) model of washback on the learning process

Zhan (2009) contributed significantly to recent advances in washback by developing a model addressing washback on the learning process based on his doctoral study. His aim was to define a comprehensive theoretical model of washback on the learning process utilising qualitative case study. Zhan (2009) attempted to systematically investigate how a group of non-English major students experienced washback from the 2006 revised College English Test Band 4 (CET-4) written test. The data for his study consisted of 4 focus group interviews, 20 classroom observations, 25 post-class interviews, 202 diary entries and 50 post-diary interviews, collected from five students in different time periods, from the day when they began their College English learning to the day when they took the CET-4 written test. Echoing previous researchers' conclusions on the complexity of the washback phenomenon, Zhan (2009) further found that the learners' various visions of their possible second language (L2) selves were closely related to the complexity of washback. He suggested that the possible "L2 self" could be regarded as a guide to the mechanism of washback affecting the learning process. This was vital in this study as one of the overarching aims was to explore the processes involved in washback. Figure 2.4 illustrates the process (mechanism) of washback on learning depicted by Zhan (2009).

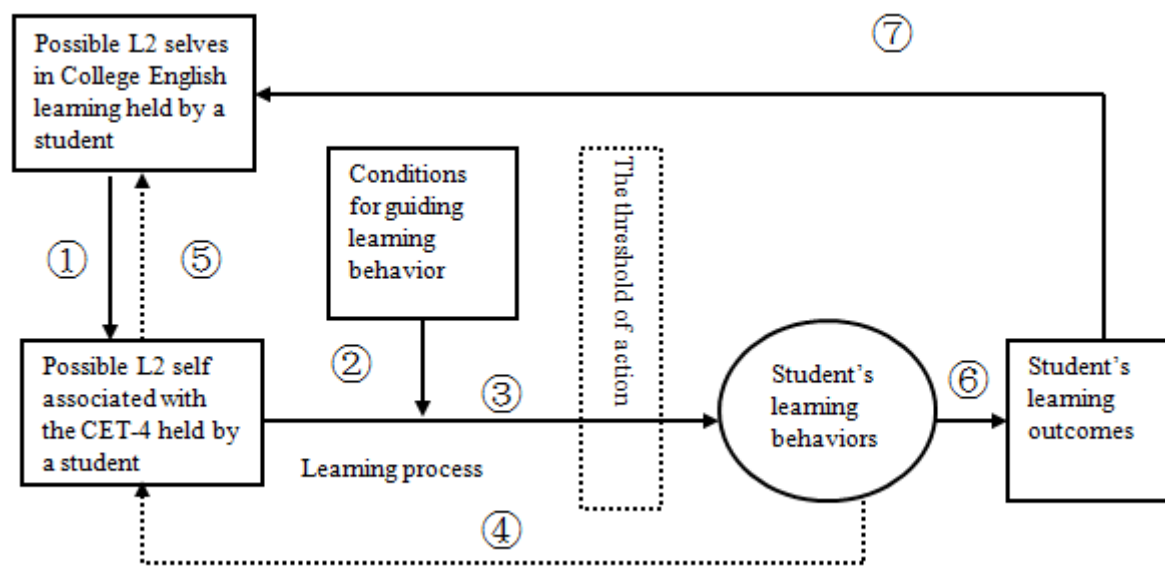


Figure 2.4 Mechanism of the effect of CET-4 on the learning process (Zhan, 2009, p. 278)

As shown in Figure 2.4, each solid and dotted line marked by a number represents a part of the whole process. In phase ①, students may generate multiple possible L2 selves associated with the test that they have to take in their language class. For example, students may envision themselves as someone who can speak fluent English or who will have English native-speaker friends in the future. Students will generally have more than one possible self in the beginning and will then decide which possible L2 self they want to pursue and eliminate those that they perceive to be irrelevant to their goals. However, Zhan (2009) cautioned that having a possible L2 self or selves in mind does not necessarily trigger students' motivation to learn. Certain conditions guiding the capacity of possible selves need to be met at point ② to exert a powerful motivational influence on learning behaviour. These conditions include: (i) "availability of an elaborate and vivid future self-image"; (ii) "perceived plausibility"; (iii) "harmony between the ideal and ought selves"; (iv) "necessary activation/priming"; (v) "accompanying procedural strategies"; (vi) "the offsetting impact of a feared self" (Dörnyei, 2009, pp. 18–22).

Next, at point ③, learners will cross the threshold of action and work towards their goals, a process that Zhan (2009) described as translating one's possible self in relation to the test into action. In this stage, the test begins to operate on the students, affecting their learning behaviour. Zhan (2009) noted that this process may be in a state of flux, such that it is possible for the students to return to their original possible L2 selves associated with the test via ④ if there is a need for them to refine their imagination of a possible L2 self. By doing so, the students will regain motivation and take appropriate actions according to their desired possible-self guides. As mentioned previously, the students will then eliminate the other possible self/selves that they think are not relevant to their goals. However, the elimination process may not be rapid as it is normal for students to change their minds ⑤ according to what they experience along the line. They may experience competition between the possible self that is associated with the test and the one that is not associated with the test. As an example, when the non-test possible self dominates, the possible self associated with the test may lose its motivational capacity and the influence on learning may disappear. However, this process can take a turn such that the possible self associated with the test takes control again; it may thus resume its motivational capacity and its influence on learning may reappear. This competition may occur over and over again, until the student's possible self is finally realized.

The students then sit the test at ⑥ and obtain their results. Once they know their test results, they will evaluate whether their possible self/selves have been realized or not ⑦. If the possible self is realized, they will "update their possible L2 selves and consider choosing another possible L2 self to pursue in the next step of learning" (Zhan, 2009, p. 280), but if not, they may re-evaluate their possible selves and pursue these again in their next step of learning, resuming the process of the test washback on learning.

According to Zhan (2009), there are two crucial phases in the development of a possible self associated with a test. Phase one involves the process of “constructing” the possible self, which is further divided into internal factors and external factors. Internal factors consist of the perceived value of the test, the perceived quality of past learning and test-taking experience and personal ambitions after study (e.g. university), while external factors include significant others, the immediate learning environment and the broader context.

Once students have constructed their possible selves, the second phase is the process of “realizing” these possible selves, which also involves internal and external factors. The internal factors include the perceived harmony of possible L2 selves with the outside world, perceived consequences of not acting, perceived behaviour control, knowledge of the test, self-knowledge, past test-taking experiences and perceived efficiency of strategies in preparing for language tests. The external factors entail the students’ significant others and the immediate learning environment. Zhan (2009) also highlighted that internal and external factors correlate with each other in every phase. Hence, both types of factors and their correlation need to be taken into consideration to encourage positive washback.

2.4 Dimensions of washback

As proposed by language assessment scholars (e.g. Alderson & Wall, 1993; Brown & Hudson, 2002; Buck, 1988; Hughes, 2002), washback is commonly categorized into two dimensions. The first is direction, which addresses whether the washback effect is positive or negative (2.4.1), or both. The second dimension of washback is intensity (Cheng, 2005), which is also referred to as the extent (Bachman & Palmer, 1996) or strength (Gates, 1995) of washback (see 2.4.2). Other than direction and intensity, washback can be measured in terms of its length (2.4.3), i.e. whether

it is long term or short term. This latter dimension has not yet been explored to any great extent in the literature on washback. The following sub-sections discuss in detail what each dimension entails and the research carried out in an attempt to measure them.

2.4.1 Direction: positive and negative washback

The concept of washback direction refers to the positive and negative effects a test could have on the process of teaching and learning. Bailey (1996) views the positive and negative directions of washback in terms of whether the test encourages or discourages learners from attaining their learning goals. At first glance, Bailey's (1996) definition suggests that there is positive washback for test-oriented learners if test familiarization and practice fill class time. However, when it comes to learning goals, it must be noted here that different stakeholders have different goals. A test can either be beneficial or detrimental to learners depending on what they deem appropriate and the educational goals they are trying to achieve (Hamp-Lyons, 1987; Hughes, 2002; Mehrens, 1998). Some learners might be more instrumentally motivated as far as their learning goals are concerned, as opposed to those who are more intrinsically motivated. For example, learners with instrumental goals will tend to prefer their teachers to teach to the test so they can obtain high marks on the test; in contrast, integrative learners will tend to want to learn for the sake of gaining knowledge. Competing goals between learners and stakeholders make it difficult to establish if washback is positive or negative, as it is determined by the attainment of these various goals, primarily those of learners. Thus, the direction of washback is rather an elusive concept that needs a clear and more elaborated definition to address the differences in goals held by different stakeholders. According to Booth (2012), attempting to outline what denotes positive and negative washback is

a contentious matter as there are many aspects involved in making such a judgement. Different stakeholders might have different ideas or views of what positive or negative goals look like.

Cheng and Curtis (2004) suggested that the context of the test, the timing and length of time from taking the test, the test objective(s) and the different approaches used by different stakeholders in light of the test need to be taken into consideration in deliberating whether a test is harmful or beneficial to the stakeholder(s). Bailey (1996, pp. 264–265) provided a list of possible processes that students preparing for a high-stakes test with important consequences might engage in:

1. practising items similar in format to those on the test;
2. studying vocabulary and grammar rules;
3. participating in interactive language practice (e.g. target language conversations);
4. reading widely in the target language;
5. listening to non-interactive language (radio, television, practice tapes, etc.);
6. applying test-taking strategies;
7. enrolling in test preparation courses;
8. requesting guidance in their study and feedback on their performance;
9. requesting or demanding unscheduled tutorials or test-preparation classes (in addition to or in lieu of other language classes);
10. skipping language classes to study for the test.

Referring to the list above, there is no definitive way or guideline for determining whether such behaviours constitute positive or negative washback. For example, “skipping language classes to study for the test” might be regarded as positive washback by test-oriented instrumental learners, but could be regarded otherwise by teachers or test designers.

- *Positive washback*

A test is designed and introduced to bring about positive washback, but what is positive washback? A “loose” definition of positive washback is when the introduction of a test brings about a beneficial effect in the attainment of stakeholders’ goals. However, there are no explicit statements of what positive washback entails as different stakeholders may have different perceptions based on their individual needs and goals. For example, as mentioned earlier, for instrumental learners, the goal is to attain good scores in the test and hence a lesson that is test oriented is deemed favourable because it will bring them closer to their goal. However, other stakeholders, such as teachers or test designers, may have different aims. They might perceive teaching to the test as negative washback as it results in narrowing of the syllabus. In a perfect world, positive washback would be when all the stakeholders’ goals are in line, which is rarely the case.

When a test promotes the attainment of educational goals for learners, teachers, or both, it is said to produce positive washback (Bailey, 1996). Moreover, as noted by Messick (1996, pp. 241–242), “for optimal positive washback there should be little, if any, difference between activities involved in learning the language and activities involved in preparing for the test”. A good test that will bring about positive washback will normally shape how teachers prepare lessons in the classroom in line with the test, but without necessitating the teacher neglecting any aspects of the learning of the target language.

A test is typically designed with a particular purpose in mind and positive washback is clearly intended by test designers. Nevertheless, a well-designed test does not necessarily lead to positive washback. There is much more to washback than simply relating it to tests that have a good design or are high stakes. Ren (2011) conducted a study of the washback effect of the College English Test Band 4 (CET-4), a high-stakes national English examination, on teaching and

learning English in five universities in Tianjin, China. In his study, involving 210 students, he found that introducing a test with important consequences like CET-4 was able to drive most students, whether they loved learning English or not, to work hard on doing so (Ren, 2011). The consequences or the stakes of the test might have triggered students' motivation to work hard, as the outcome of the test would affect their future undertakings. High-stakes tests are commonly used for recruitment purposes, such as international student admissions, graduation and employment. Hence, it may be that the design of the test does not guarantee positive washback, but rather it is the consequences of the test that do so.

According to Xie and Andrews (2012), higher task value is associated with greater engagement in test preparation activities. Commonly, high-stakes language tests encompass several, if not all, the language skills: listening, speaking, reading and writing. Some test designers assign different weighting to these components by allocating higher weighting to skills that are deemed more important or those they wish to see improved. Teachers and students have been found to focus more on tasks that they consider to be imperative for the outcome of the test. Indeed, Nambiar and Ransirini (2012) point out that different washback effects depend on perceived task importance. No matter what test designers have in mind when designing the test, how students perceive task importance will determine the washback effect of the test. Ferman (2004) undertook a rigorous study examining whether washback occurred after the introduction of a new high-stakes national EFL oral matriculation test in Israel in terms of educational processes, the participants and the products of learning. She used four types of instrument to elicit data: document analysis and structured questionnaires, structured interviews and open interviews undertaken with teachers, students and EFL instructors. She found that introducing an oral component into the test with the intention of improving students' speaking skills increased the focus on oral language skills among

the teachers, students and also parents. This change in focus included an increase in the time allocated for the development of these skills, an accelerated pace of learning, employment of teaching and learning strategies geared towards test success and the promotion of learning oral skills.

- *Negative washback*

As mentioned earlier, the term washback has often had a negative connotation. This might be due to the fact that more findings reported in the literature have concerned negative washback than positive. According to Taylor (2005, p.154), negative washback happens when a “test’s content or format is based on a narrow definition of language ability, and so constrains the teaching/learning context”. For instance, if the students are allowed to memorise texts or scripts for their speaking test, then there is great pressure to practise memorising rather than to practise the skill of speaking itself. The washback effect of a test is considered to be negative when too much focus is placed on passing the test as opposed to mastery of the subject being taught. Among the undesirable consequences of a test are narrowing of the curriculum, test drilling and rote memorization. Negative washback is commonly reported in washback studies. For example, Hayes and Read (2004) in their study on the impact of the IELTS test on the way international students prepare for academic study in New Zealand conducted a survey in 96 language schools throughout New Zealand. They then followed up the questionnaire by interviewing 23 teachers engaged in preparing students for the IELTS Academic Module about preparation courses. This was done in Phase 1 of the study. In Phase 2, Hayes and Read (2004) conducted a classroom study to compare two IELTS preparation courses. Data was elicited through classroom observations, teacher interviews, teacher and student questionnaire, and pre and post testing of the students for

triangulation purpose. From their comprehensive study, Hayes and Read (2004) found that both teachers and students focused on practising test tasks rather than developing academic language proficiency. Also, Damankesh and Babaii (2015) conducted a study on the washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. Their study consisted of data gathered from 80 Iranian male students who were selected from six classes in four high schools in the Northern Guilan Province, the cities of Siahkal and Shaft. These students were asked to answer final examination questions designed following the typical format of high school final examinations in Iran. Out of the 80 student participants, 30 were randomly selected for think-aloud protocol where they were asked to vocalize their mental processes and operations while completing the final examination. From the findings of this study, Damankesh and Babaii (2015) concluded that the tests exerted a predominantly negative influence, with the students employing strategies such as guessing blindly, cheating, memorizing and rote learning, which could have detrimental effects on their creativity and hinder meaningful learning; however, the influence was not entirely negative, with some of the strategies fostering students' abilities. Similarly, Ferman (2004) reported that the students in her study had a high tendency to memorize the material rather than to acquire and develop language skills. This phenomenon could explain why, in some cases, even though the students' test scores are high, when asked to demonstrate their skills, they are unable to do so.

Qi (2005) surveyed 986 students undertaking the NMET in China – the university entrance test in English for the entire country – and found that the students believed studying for the test is education and getting high scores is evidence of educational excellence. In the long term, such beliefs could be harmful for the educational process, leading to learners defining their ability in terms of test scores rather than what they are actually capable of doing, as poor test scores would

hamper their self-efficacy in language learning. For example, a good student might perform poorly in a writing test if the topic given is not within his or her knowledge base at all. The student might perform better if a different topic was given. Hence, care must be taken when it comes to interpreting test result as test validity plays an important role in reflecting students' performance through test scores. This is supported by Ren's (2011) study on the CET-4, which showed that the test was likely responsible for students' inability to use English for real-life purposes, or going beyond answering questions in English language tests. Specifically, Ren (2011) found that the test encouraged rote memorization of linguistic forms, which hindered communicative growth. The teachers were also found to restrict the curriculum and the teaching objectives to cater to the test. Linked to this, as the students had limited time before the test, their focus was solely on passing and hence they paid particular attention to the skills tests, as also reported in other studies (e.g. Akpinar & Cakildere, 2013; Zhan & Andrews, 2014).

In a recent study of the effect of mock tests on Iranian EFL learners' final test scores conducted by Khodabakhshzadeh, Zardkanloo, and Alipoor (2017) with 51 IELTS students at the Mahan Language Institute in Birjand, Iran, it was revealed that using mock tests in IELTS preparation classes could have a positive effect on overall scores in the actual IELTS test. Khodabakhshzadeh et al. (2017) adopted a quasi-experimental study, comparing two groups of students preparing for the IELTS, Group 1 with mock testing and Group 2 without mock testing; they found that the Group 1 students outperformed the Group 2 students. Their study also showed that when it comes to high-stakes tests, such as IELTS, practising test-taking strategies appears to be more effective than teaching course content, especially in the Iranian context in which many students lack such test-taking strategies. When students' performance on the test is a higher priority than their mastery of the target language, the test is deemed to cause negative washback. This is

not only the case for students, but also teachers. Nambiar and Ransirini's (2012) study reported that teachers involved in teaching test preparation classes were under a lot of pressure and felt disempowered. Moreover, lack of time and understanding of the test itself can cause teachers to feel trapped and restricted by the test. In their study, the teachers were overwhelmed by the test and failed to see it as only a format to test language proficiency (Nambiar & Ransirini, 2012).

Jianrattanapong (2011) discussed the washback from Thai university entrance examinations in terms of English language components. The university entrance examination in Thailand does not include a direct test of writing; rather, the test consists of multiple choice items, which are suitable for diagnostic or progress tests, but do not measure writing ability. This resulted in negative washback, as the teachers did not ask their students to practise writing because it was not tested. Jianrattanapong (2011) also highlighted the issue of students having the ability to memorize complex grammatical features and attain very high scores on the test, but lacking the ability to use English in real life. Her argument centred on the fact that given the importance of writing, to make full use of the washback effect, it should be tested in the Thai University Entrance Examination. The main reason given for not testing writing is that it is costly in terms of time and money to train teachers to score writing papers. Because of this, students perform poorly in writing once they are at university.

A further significant effect of tests with strong consequences for teachers lies in performance evaluation. In some contexts, teachers' performance is gauged by how well their students do on tests. As mentioned by Buck (1988, p. 17), there is:

...a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success.

Hence, success in tests not only benefits the students, but also the teachers and the learning institution indirectly. Green (2014) cautions that an excessive focus on test results will lead to teachers neglecting the official curriculum and directing their teaching more towards what is tested rather than ensuring learners' mastery of content knowledge. Learning objectives that are not covered in the test may be overlooked and ignored, as more attention is paid to "training" and "prepping" the students to perform well on the tests. This clearly depicts "teaching to the test", which is commonly associated with the negative effects of tests or negative washback. This is an enduring issue and the same findings keep being reported, despite the problem being pointed out by Swain (1985, p. 43) 30 years ago, namely that "teachers will teach to a test: that is, if they know the content of a test and/or the format of a test, they will teach their students accordingly".

However, in the face of many washback researchers interpreting teaching to the test as one of the negative effects of testing, Kober (2002) claimed that there is often confusion about what the term really means and it is often defined only at the surface level. To help define "teaching to the test" with greater clarity, Kober (2002) distinguished three forms: extremely unethical, common and extremely ethical. In its more common form, teaching to the test refers to "direct preparation for a particular test", such as administering model questions, familiarizing the students with the answer sheets, or focusing instruction on a limited number of skills, while in its extreme unethical forms, it simply means "cheating". Giving students actual questions for a secure version of a standardized test falls under this description. The extreme ethical form of teaching to the test, in contrast, is when instruction is focused on the most important knowledge and skills as outlined in the content curriculum standards (Kober, 2002), i.e. positive washback. As long as lessons are planned according to the content curriculum standards, narrowing the curriculum is unlikely to

happen. Hence, when teaching to the test is reported in washback studies, care must be taken in interpreting the findings, as this may not simply constitute negative washback.

Unfortunately, over the years, the practice of teaching to the test has developed a considerably negative connotation among researchers and educators. Most interpret teaching to the test in its unethical form, which Mertler (2007, p. 43) described as the act of “narrowing or limiting the scope of instruction to only that content that is specifically covered on any assessment”, also commonly known as “curriculum alignment” in the field of washback. Mertler (2007) questioned whether teaching to the test is actually a bad practice. His argument centred on the notion that the three key components of the instructional process – curriculum, instruction and assessment – should always be aligned with one another. These components are interdependent and inform each other, as depicted in Figure 2.5.



Figure 2.5 Key components of the instruction process (Mertler, 2007, p. 42)

When designing a test for a particular course (assessment), test developers have to consider both the specific content of the course (curriculum) and how it is taught (instruction) in the classroom. For appropriate instruction and subsequent learning to take place, these three components must be aligned (Mertler, 2007) in terms of “constructive alignment” as opposed to “curriculum alignment”. Constructive alignment is an outcomes-based teaching approach in which “what it is intended students should learn and how they should express their learning is clearly stated before

teaching takes place” (Biggs, 2014, p. 5). In this case, learning objectives are spelled out first, before teaching and assessment methods are developed accordingly. Constructive alignment thus focuses on learning first, as opposed to centring learning around tests and catering to learners based on what is going to be tested. In this instance, teaching to the test is not as negative as commonly depicted in previous studies.

However, the “narrowing or limiting [of] the scope of instruction to only that content that is specifically covered on any assessment” (Mertler, 2007, p. 43) is a concern for many washback researchers in the field of assessment. As pointed out by Hughes (2002, p. 1), “If a test is regarded as important, then preparation for it can come to dominate all teaching and learning activities”, which can potentially lead to negative washback and also directly affect the intensity of the washback, a measure of washback in addition to its direction, as discussed in the next sub-section.

2.4.2 Extent, strength or intensity of washback

A key dimension of washback is its strength, also referred to as its extent or intensity. This refers to how strong or how weak the washback effect of the test is, evidenced by the extent to which the learners conform to the test demands (Cheng, 2005). In this study, for ease of understanding and coherence, the term intensity is used, as employed both in Watanabe’s (2004) and Green’s (2007a) washback models (see 2.3). Washback intensity is often associated with the stakes of a test; the higher the stakes, the stronger the washback effect (Green, 2007a). When it comes to determining the major factors in washback intensity, the constructivist critics of high-stakes testing consider test use and the associated consequences to be more dominant factors than test design (Crooks, 1988; Gipps, 2002; Shohamy, 1992). Tests with higher stakes will trigger stronger washback as they carry greater consequences for the stakeholders.

Moreover, as pointed out by Gates (1995), the prestige of a testing organization and its domination in the educational market will also contribute to the degree of washback intensity. For example, in the case of gaining entry to a university for which applicants are required to take and obtain a certain score on a particular language test, the washback intensity of the test will most likely be strong. However, Madaus (1988) asserted that it is how the stakeholders perceive the test rather than the reality that will affect their behaviour towards the test. If the applicants feel they have other choices or alternatives to the test, the intensity of washback for them may be weak (for further discussion, see 2.6.1). Hence, the more the teachers or students feel the weight of the consequences of the test, the more likely it is they will adjust their behaviour, doing things they would not normally do for the test (Alderson & Wall, 1993). This clearly indicates the complexity of washback and that the relationship between testing and teaching and learning is not simply linear (see 2.5).

2.4.3 Length of washback

If washback is found to exist, how long does it actually last? This is one of the washback dimensions as yet underexplored in language assessment research. Termed “length” by Watanabe (2004), it was listed together with specificity, intensity, intentionality and value in outlining the complexity of the washback phenomenon. According to Watanabe (2004), washback can have either short-term or long-term effects on the stakeholders, especially the learners. If learners are seen to increase their efforts to learn and adopt certain learning strategies in preparing themselves for the test but discard them once the test is over, the influence of the test is a short-term effect. If the influence remains even after the test ends, the washback effect can be considered long term (Watanabe, 2004). However, washback length remains on a conceptual level as no attempt has

been made to empirically measure it. Shohamy et al. (1996) investigated the impact of two established national tests in Israel which are Arabic as a second language (ASL) and English as a foreign language (EFL). They employed student questionnaires with 62 ASL and 50 EFL student participants and structured interviews with both teachers (9 ASL, 16 EFL) and inspectors (2 ASL, 4 EFL). Analysis of documents concerning the tests issued by the Ministry of Education Inspectorate, teaching materials and new courseware were also carried out in this study. According to Shohamy et al. (1996), washback evolves over time. Their work shows that the introduction of a test may initially have a considerable impact on teaching and learning, but this may wear off, particularly if the test is later shown to have little power. Previous literature on washback has shown that the closer the exam date, the more intense the washback. However, to establish whether a test has a long-term or short-term washback effect, it is necessary to investigate students' learning behaviour not only before the test, but also after they have sat it.

To date, there has been a lack of work empirically testing washback length as such research requires a longitudinal study design to elicit data (Hoque, 2011; Scaramucci & Kobayashi, 2013). Only tentative comments and claims have been made regarding washback length and no concrete conclusions have been reached as the focus of the majority of washback studies has been on short-term washback effects (Hoque, 2011). Scaramucci and Kobayashi (2013), investigating the washback effect of the Cambridge English Test in Brazil employing interviews, classroom observation and a review of school documents, could only assume that "the influence of one exam will probably continue until the next one" (p. 19), but they provided no empirical evidence to support this assumption.

It is thus not known what kind of test will have long-term or short-term washback, or what factors contribute to or hinder the long-term washback of a test. Among these factors, it is not clear

to what extent students' motivation in learning might simply be to obtain a good score on the test or beyond. According to Booth (2012), it is generally accepted that tests, especially high-stakes tests, can influence students' motivation to learn. However, the intensity of the washback may decrease over time as the test score is no longer needed. More research on the long-term effects of washback is therefore required as highlighted by Tsagari (2007), who pointed out that studies of "how high-stakes exams motivate students to learn and whether they can help sustain students' motivation for learning after the exam" (p. 56) would make a valuable contribution to the field of washback studies.

2.5 Complexity of the washback phenomenon

To quote McEwen (1995, p. 42), "what is assessed becomes what is valued, which becomes what is taught". Washback has the power to compel "teachers and learners to do things they would not necessarily otherwise do because of the test" (Alderson & Wall, 1993, p. 117). In other words, when preparing for a test, using Hughes' (1993) washback trichotomy as a point of reference, participants (e.g. teachers, learners) do things they would not normally do in the "process" to obtain desired "products". Thus, as argued by Yang (2013), in its traditional form washback is viewed as a linear stimulus-response mode, which suggests that if a test is properly designed, it will have positive outcomes.

While the definition of washback may suggest simple and linear relationship between a test and outcomes, it has been found to be far more complex. The apparent simplicity of this concept has been challenged over and over again with the discovery of positive and negative washback effects from evidence-based studies demonstrating that there are factors other than the test which influence both teachers' and learners' behaviours in preparing for the test. It does not take much

reading in the language assessment field to see how complex washback is. Alderson (2004) noted that a number of washback researchers were surprised by their own discoveries when investigating this phenomenon, as they discovered that the introduction of a new test did not necessarily induce the intended change(s) and any difference tended not to be significant or less than expected. This provided the motivation for this research in terms of re-examining beliefs concerning washback and taking into consideration other factors that may mediate the process of teaching and learning beyond washback, particularly focusing on the role of students' perceptions.

2.6 The role of students' perceptions in influencing washback

Ensuring that a test is valid is arguably one of the most important concerns in designing a test. A test is considered valid when it measures what it claims to measure. Creating a valid test of an individual's language skills is a difficult task, requiring many steps. The values and social meaning of a test are two main concerns in test validity as introduced by Messick (1989). He argued that social values will play an important part in the intended or unintended outcomes of test use. Assessing the social consequences of a test (also referred to as consequential validity) for learners is also necessary in washback studies to establish whether the test scores do indeed mean what the test designers intend them to mean. Moreover, as emphasized by McNamara (2006), test validation should be an ongoing process, addressing new issues as they arise, especially in the context of language testing research.

In the Malaysian context, Bidin, Jusoff, Aziz, Salleh, and Tajudin (2009) investigated motivations and attitudes related to learning English among diploma students in one of the tertiary institutions in Malaysia, the University Teknologi MARA (UiTM). A total of 620 students from three different UiTM campuses were asked to complete a questionnaire pertaining to their personal

characteristics, motivations and attitudes. Bidin et al. (2009) found that the students were more extrinsically motivated than intrinsically motivated in terms of learning English. They attributed this finding to the students' uncertainty concerning the purpose of learning English as it was one of their compulsory courses necessary to graduate. It seemed rather clear that not knowing the true purpose of undertaking the course affected their perceptions and indirectly affected their behaviour. Hence, this study focused on the perceptions of learners concerning the consequences of tests, tapping into their views of the test itself in terms of perceived difficulty and importance and also their perceptions of their own ability when it came to performing on the MUET.

2.6.1 Perceptions of tests

Previous research has investigated the potential effect of students' behaviour on outcomes in the form of test results or performance. However, few washback studies have explored the cause(s) of such behaviour, in this case, the students' perceptions of the test (Zhan, 2009). This study looked at two main factors that have the potential to affect washback as posited by Green (2007a) in his washback model, namely perceived test importance and perceived test difficulty. According to Green (2007a), at an appropriate level of importance and difficulty perceived by students, a test will lead to intense washback. As past literature has suggested, it is the social aspect of the test that makes it powerful. No matter how important the test is supposed to be, if those who deal directly with the test, i.e. teachers and students, do not share this view, the test will not have strong washback. Students perception of the value of a test is associated with greater or lesser engagement in preparation, regardless of their level of language proficiency.

Xie and Andrews (2012) surveyed 870 test takers' perceptions in Hong Kong concerning the influence of test design on test preparation and found that the value of the test is associated

with greater engagement in test preparation, alongside students' self-efficacy. They reported that positive endorsement of intended test demand was related to a higher evaluation of test importance. The higher the stakes of the test, the more effort the students will put in, as high-stakes tests generally have greater consequences or constitute a potential threat to their future pathways and undertakings. Similarly, perceived test difficulty plays a vital role in influencing how students prepare for a test, as concluded by Watanabe (2001) based on interviews with Japanese students with regard to their test preparation practices for university entrance exams. In his study, he found that perceived test difficulty could potentially explain the complex relationship between students' test preparation and their motivation; namely a test at the appropriate level of difficulty could have a positive effect on students' test preparation. Hence, Watanabe (2001) called for more research on students' perceptions, as the causes of washback do not lie in the test alone, but also in students' perceptions of the difficulty of the test, a rather neglected area of research in washback studies.

Allen (2016a) investigated washback to the learner from the IELTS in the Japanese tertiary context. In his study, 190 students sat IELTS twice over a period of 12 months. The participants were given two half-day workshops focusing on the productive skills components. Other than that, the participants prepared for the tests independently. Allen (2016a) focused on the students' test preparation strategies and score gains from the test at time 1 to the test at time 2. He also conducted interviews with 19 students to explore the possible mediating factors of washback. The study found a significant increase in speaking ability, especially among those who prepared intensely for the test. It appeared that after the first test, the students started to focus significantly more on speaking and writing as opposed to reading. From his interviews with the students, Allen (2016a) discovered that learner perceptions and their access to resources were prominent factors in shaping washback to the learners. To achieve positive washback, these mediating factors have to be addressed, but

have often been overlooked when investigating washback. However, his study only involved high academic achievers. The participants adopted a test-focused approach when preparing for the test, paying more attention to the tasks that were going to be tested as commonly observed in many washback studies. The students were found to have changed their learning strategies when they were preparing for the second test, focusing more on productive skills as opposed to receptive skills. The reasons given for changing or not changing their preparation strategies were perceived difficulty (they tended to perceive the unfamiliar parts of the test as difficult, so they focused more on those aspects because they considered the other skills had already been covered in school), perceived efficiency and effectiveness, knowledge of how to study and improve and assistance from others.

Conversely, Allen (2016b) found that those who perceived the test as not being important to them tended to study little for it. As commonly reported in other washback studies, the students tend to change what they learn rather than how they learn for a test. However, in Allen's (2016b) study, the students appeared to be willing to adopt new ways of studying English, but it seemed difficult for them to achieve this due to the limited time and support they had in preparing for the test. He concluded that the sociocultural and educational context is among the numerous factors that need to be considered when it comes to introducing a test to achieve positive washback or to ensure the consequential validity of the test. It is not just a matter of designing a test properly that determines the direction of the washback of the test.

Sato and Ikeda (2015) looked at test takers' perceptions of the skill being measured by items (face validity) in high-stakes tests to see if their perceptions matched intentions of the test developers. Face validity is one of the crucial elements in achieving and maximizing positive washback, as the test developers' intentions should be in line with what the students perceive the

test is measuring. Sato and Ikeda (2015) argued that the gap between students' perceptions of the skills being measured and the test developers' intentions was one of the main factors hindering positive washback on learning. Participants from two EFL countries, South Korea (n = 98) and Japan (n = 80), were given a set of past university entrance examinations for their respective countries and were asked to indicate what skill they thought each item was measuring. The data were then compared with the test developers' intentions. It was found that the overall agreement was 59.1% for the Korean students and 71.8% for the Japanese students. There were some items with significantly low face validity, for example items aiming to measure the students' ability to read between the lines, which were perceived by the students as measuring the ability to understand the content objectively. The majority of the students also perceived items created to measure writing skills as tapping into reading skills instead. Based on their findings, Sato and Ikeda (2015) argued that to achieve positive washback, the gap between test takers' perceptions and the test developers' intentions needs to be addressed, as the test takers' perceptions of the test appeared to have some effects on the content of their learning. However, the relationship between test takers' perceptions and their actual learning has not yet been empirically investigated to any great extent.

As contended by Shohamy (2014), a test is deemed powerful when it is given social and political functions by the authorities, not because of its technical strength. In countries with exam-oriented educational systems, such as Korea and China, society as a whole seems to believe that getting good scores in high-stakes tests will open up more opportunities for success in today's competitive world (Cho, 2004). Due to this, Xie and Andrews (2012) concluded that negative washback should be attributed to the misuse and abuse of test results, not the test design, no matter how good or how bad it is. Therefore, apart from the pedagogical impacts of a test, the social impact that the test engenders cannot be ignored when researching washback (Cho, 2004).

Gebril and Eid (2017) stated that when researching washback, the social context in which test preparation takes place has to be taken into consideration because within this social context, a complex network of variables mediates test washback (for a more in-depth discussion, see 2.5). Thus, it is crucial to examine other variables that may interfere with or contribute towards the positive washback of a test. Several variables have been identified in the literature as strong washback factors, among which students' self-efficacy or how they perceive the test in relation to their ability (Embse & Hasson, 2012; Roderick & Engel, 2001; Xie & Andrews, 2012) has received some interest in the literature and was also addressed in this study. The next section discusses students' perception of their self-efficacy in relation to tests.

2.6.2 Perception of self-efficacy

Self-efficacy is one of the crucial mechanisms that needs to be considered in trying to understand the washback from high-stakes tests on students' learning. It is defined as individuals' beliefs in their abilities to perform a task (Bandura, 1986) and/or "the belief of one's capabilities to organize and execute courses of action required to produce given attainments" (Bandura, 1997, p. 3). In other words, it refers to certain beliefs about one's own ability to learn and to achieve goal(s). These beliefs are very specific to a particular situation in a particular context. Based on their belief system, learners will perform in certain ways. A basic principle is that the higher a person's self-efficacy, the more he/she will believe in his/her capability to accomplish a task; conversely, the lower self-efficacy, the less capable the person will consider himself/herself. People are less likely to put their full effort into attempting something if their self-efficacy is lower concerning that task. In the language assessment literature, self-efficacy has been identified as one of the principal variables predicting students' engagement in test preparation (Gosa, 2004; Xie & Andrews, 2012)

and the kinds of learning strategies employed (Damankesh & Babaii, 2015; Shih, 2013). As opposed to investigating students' perceived self-efficacy in relation to a test, most studies have focused more on students' actual proficiency level when investigating the washback of a test.

Studies have shown that students' proficiency level plays a major role in determining how learners react to a test. Some studies investigating high-stakes tests have found that due to the stakes and consequences of the test, lower proficiency learners in particular believe that by increasing their effort when preparing for a test, they will be able to perform well and get good grades. These learners will resort to cramming and engage in intense learning for the sake of the test (Ferman, 2004; Shohamy et al., 1996). However, cases have also been identified of low-proficiency learners not preparing at all because they were too overwhelmed by the stakes of the test and perceived that no matter how much effort they invested in the task, they would not be able to do well (Chu, 2009; Watanabe, 2001). Thus, a test that is too challenging for the students' level may induce high test anxiety, hampering the motivation to learn (Chen & Hsieh, 2011).

Generally, it has been argued that low-proficiency learners tend to worry more about the stakes of the test compared to high-proficiency students. This notion was highlighted by Shih (2007), who concluded that higher proficiency students might view the test as too easy for them and put lower effort into preparation compared to less proficient students. Indeed, according to Watanabe (2006), a test at the appropriate level of difficulty for the students will possibly motivate them to prepare for the test due to test anxiety. However, Pan (2014) argued that high-proficiency students already tend to be eager to learn, even without having to sit a high-stakes test.

A very interesting exploratory study on test anxiety, heart rate and performance in A-level French oral mock exams conducted by Daly, Chamberlain, and Spalding (2011) used heart-rate monitors and questionnaires to measure test anxiety. They found that the measure of heart rate was

closely related to physiological arousal rather than test anxiety. Their findings showed that it is actually the increase in arousal, not anxiety, that will lead to higher performance, suggesting a positive effect of arousal on students' performance. However, it has to be noted here that mock exams could be considered less demanding compared to high-stakes tests. Hence, the results might differ if data were collected for the actual test.

Pan's (2014) study reported that students' proficiency levels gave rise to significant differences in the amount of effort expended by learners in preparing for the test, such that higher proficiency learners used a greater variety of resources and types of practice compared to less proficient learners and they appeared to believe that preparing for the exit tests would help enhance their language skills and improve their intrinsic motivation. Pan and Newfields (2012) also found learners' proficiency level to be one significant factor that could contribute to students' overall motivation to prepare for a test. Their study found that lower proficiency learners appeared to be less motivated compared to their higher proficiency counterparts.

However, Fan, Ji, and Song's (2014) study revealed otherwise. They conducted a study investigating the washback effect of the Fudan English Test (FET) on students' learning of English, focusing specifically on the role of gender and English language proficiency in shaping washback. The FET is a university-based English language test, developed and used by Fudan University for the purpose of measuring the students' English language abilities and inducing more positive washback on English language learning and teaching in the university. They collected data from 335 students through a questionnaire and conducted semi-structured follow-up interviews with 13 students. The students mentioned that the FET did not have much impact on their motivation as it was a newly developed test used only within the university. Fan, Ji, and Song (2014) found that neither gender nor English language ability exerted a significant impact in

shaping the washback of the FET on the students' learning practices. Hence, there seemed to be a missing link between the students' language proficiency and their course of action, which they identified as students' self-efficacy. They proposed that the washback phenomenon could be better explained by self-efficacy, which is in part based on students' interpretation of their actual performance (Schunk & Pajares, 2009, p. 36).

As far as students' self-efficacy was concerned, Wong and Chan (2009) investigated the washback of the MUET through student questionnaire and found that the students' beliefs that they had a weak command of language skills created a tension between perception and the actual case. In their study, Wong and Chan (2009) were referring to the tension between students' beliefs about what they were capable of and what the results of the tests suggest they were really capable of. However, they did not systematically measure the students' actual performance and their study was solely based on self-reported data which necessitated the need for this study to be carried out (for more in-depth discussion, see 2.10.1). Thus, efforts are needed to resolve this tension to help students unlock their full potential, including more washback research focusing on self-efficacy. This will also help to overcome the issue highlighted by Shohamy (2014) concerning the lack of attention paid to how a test is used, its importance for the lives of students and the place of a test in society.

As discussed earlier, self-efficacy in high-stakes testing relates to students' perceptions of the difficulty or ease of a test and their current perceived ability. If the level of the test is perceived not to be within the students' capacity, their motivation to learn will be negatively affected (Gosa, 2004). That is to say, if the test is perceived to be too difficult, students will tend to put less effort into learning as they will be overwhelmed; they will consider that no matter how hard they study,

they will not be able to perform well and it was an unrealistic endeavour to begin with (Chu, 2009; Stoneman, 2006; Watanabe, 2001).

Students' self-efficacy might thus be a more significant factor in driving their learning intensity than their actual ability. Roderick and Engel (2001) found that even low-proficiency students could hold positive attitudes of a test as long as they viewed the goal as attainable. A study by Li, Zhong, and Suen (2012) that looked into students' perceptions of the impact of the College English Test (CET), a high-stakes test in China, specifically focused on self-efficacy as one of the washback impacts of the test. It was reported that the students felt more confident about their reading and listening ability as a result of preparing for the test. According to the authors, the increase in students' self-efficacy in relation to these two skills, might have been caused by the higher weighting given to them compared to writing and speaking. This encouraged the students to spend more time on developing their reading and listening skills, hence making them more confident about these two skills as a result of preparing for the test.

However, Li, Zhong, and Suen's (2012) study did not look at the students' behaviour as a result of having higher or lower self-efficacy when preparing for the test. This is crucial to determine if students' increase in self-efficacy is actually caused by spending more time on certain skills. It would be helpful to collect other types of evidence, such as through observations and interviews, to gain a more in-depth understanding of how and why students hold particular perceptions of the impact of a test, making it timely and relevant to undertake this study.

2.7 Measuring washback in the study

High-stakes tests are generally believed to have the power to bring about change in course design or classroom practices (Pan & Newfields, 2012). Such tests are commonly used by authorities,

especially when new educational policies are being implemented, due to their alleged effectiveness in delivering fast outcomes (Shohamy, 2014). Furthermore, high-stakes tests are more cost-effective (Qi, 2007) than other interventions or innovations used to improve the quality of education within a short amount of time, such as hiring more competent teachers or conducting more classes with the aim of encouraging and motivating the students to learn. However, studies on the use of high-stakes tests to promote positive washback have reported ambiguous results to date (Cho, 2004; Pan, 2014; Pan & Newfields, 2011, 2012; Qi, 2007; Thomas, 2005). It is unclear whether students are motivated or demotivated by high-stakes tests, as the literature presents mixed findings.

Cho (2004) investigated the use of the Test of English as a Foreign Language (TOEFL) as one of the university graduation requirements in South Korea. Realizing the status and the power of English as an international language, several universities in South Korea started to employ well-established high-stakes English language tests, for example the TOEFL, the Test of English for International Communication (TOEIC) and the Test of English Proficiency (TEPS), as a graduation requirement. Findings from telephone interviews with eight English teaching faculty members revealed that these high-stakes language tests drove students to work for the tests, or at least develop some interest in learning English. This was then confirmed by the students themselves, who reported positive attitudes towards the tests. The findings revealed that these students felt that the tests increased their motivation to learn English. However, there was no evidence or indication that the findings were significant, as there were inadequate data to generalize the findings of the study (Cho, 2004).

In China, more recent studies by Shih (2013) and Pan (2014) on high-stakes testing echoed Cho's (2004) findings concerning students' motivation, but Shih (2013) noted that although

students may display an increased level of motivation, the effort taken to learn could be rather minimal. They would tend to put less effort into learning English if they perceived that they could not achieve the test score cut-offs. Similarly, if the goal was deemed easy to achieve, they would not study as hard as they might. As explained by Liu (2014) in her conceptual paper on motivation and attitude to arouse students' potentialities in learning English, it is on the basis of the students' perceptions, be it about teachers, class, curriculum etc., that form the students' attitudes towards English language. In Shih's (2013) case, it was the students' perceptions of the test difficulty and their self-efficacy that to an extent shaped their attitude and how they approach the test.

Pan and Newfields (2011), in a large-scale washback study in Taiwan, compared two groups of teachers and students to see how their pedagogy and learning strategies might have been influenced by a specific test-driven policy. Similar to South Korea, there is a trend in Taiwan to employ high-stakes language tests as an additional exit requirement to enhance students' English proficiency and this policy has been supported by Taiwan's Ministry of Education. Using a quasi-experimental design, two phases of data collection, involving both students and teachers, were carried out: (i) phase 1 – questionnaire; (ii) phase 2 – interview. Pan and Newfields (2011) found that the initial aim of using the test to enhance students' motivation for English language learning seemed to have failed, as the findings showed a gap between the teachers' and students' classroom goals. The teachers in their study preferred “regular teaching” and decided not to include a great amount of test preparation in class. In contrast, the students, especially those with lower English proficiency, wanted explicit test preparation in class. They wanted to learn the tips and tricks to do well in the English language test just to enable them to graduate. Another report on the same study published in the following year revealed four slight changes brought about by the test, i.e. increased motivation for English study, more time allocated to studying English, more variation in the

methods employed to study English and more test-related practice. Pan and Newfields (2011, 2012) found a slight increase in the amount of time students spent studying for the test, but the effects of the test on the students' motivation level appeared to be partial at best.

The evidence reported thus far does not adequately explain the relationship between high-stakes testing and washback in terms of students' motivation: the term "motivation" in these studies was not clearly defined and was used rather loosely. Using motivation as a measurement construct and reporting that students' motivation appeared to increase following the introduction of a test, signifying positive washback, is problematic as motivation is a broad concept. Hence, to address this issue, the current study used data on the students' language learning strategies as indicative of the washback process, aiming to explain the washback effect of the MUET on the students.

There is no single specific scale that can be used to measure washback. "Process" in this study refers to "any actions taken by participants which may contribute to the process of learning, such as materials development, syllabus design, changes in teaching methodology, the use of test-taking strategies" (Hughes, 1993, p. 2). These processes are undertaken by the participants to obtain the desired products in terms of "what is learned and the quality of learning" (Hughes, 1993, p. 2). Most washback studies that have focused on the product or the outcome of a test in the form of students' test scores have been unable to establish if an increase or decrease in test scores is an indication of washback. Thus, the previous literature has suggested that washback can be measured by observing how students react to the test under investigation through the kinds of actions they take to prepare for it. Studying the processes involved when the students are preparing for a test should yield more reliable findings when investigating washback than only relying on the students' test scores. Measuring student outcomes in terms of their scores is deemed problematic as other

variables need to be taken into consideration, for example their proficiency and learning background. For instance, if a bad test is introduced and a good student sits it and does well, it cannot be concluded that the test has positive washback just by looking at the score obtained. Hence, this study focuses more on the process than the product of a test.

As the main stakeholders in a test, learners have been reported to use multiple language learning strategies to prepare for the test. According to Pan (2014), high-stakes tests promote students' autonomous learning. Exploring the various washback effects on students' learning under the influence of different policies, with or without an English language graduation requirement, Pan (2014) employed two sets of student questionnaires and used baseline data as a point of comparison to see if these tests had any effect on students' motivation, learning activities and test performance. He reported that the students who sat the tests appeared to be more independent in their own learning and take matters into their own hands if there were a requirement to pass for graduation compared to those without such a requirement. The former used in-class test-related materials and school resources to prepare themselves for the test. However, the strategies used by these learners were mostly geared towards performing well in the test, not focused on in-depth learning to really acquire language skills. There was no statistically significant difference in terms of the language learning activities employed by the participants in this study. Most of the students reported having frequently used traditional language learning activities, such as reading textbooks, memorizing vocabulary and idioms and practising sentence patterns, to name but a few.

In the same vein, Shih (2013) in a very comprehensive study for his doctoral dissertation, used both quantitative (questionnaires) and qualitative (semi-structured individual interviews with students and teachers) approaches to elicit data. He reported that most of the students in his study

seemed to employ more surface strategies in their English language learning process, rather than deep strategies. As conceptualized by Marton and Saljo (1976), a behaviourist view of learning closely reflects “surface-level” processing, including increasing knowledge, memorizing and acquiring facts or procedures to be used at a later date. “Deep-level” processing, in contrast, includes abstracting meaning and interpreting to understand reality. Shih (2013) referred to Lublin’s (2003) characteristics of deep and surface approaches to learning to help him determine which approaches the students employed, as depicted in Table 2.1.

Table 2.1

Characteristics of deep and surface approaches to learning

Deep Approaches	Surface Approaches
<ul style="list-style-type: none"> • Actively seek to understand the material/the subject • Interact vigorously with the content • Make use of evidence, inquiry and evaluation • Take a broad view and relate ideas to one another • Motivated by interest • Relate new ideas to previous knowledge • Relate concepts to everyday experience • Tend to read • Study beyond the course requirements 	<ul style="list-style-type: none"> • Try to learn in order to repeat what they have learned • Memorize information needed for assessments • Make use of rote learning • Take a narrow view and concentrate on detail • Fail to distinguish principles from examples • Tend to stick closely to the course requirements • Motivated by fear of failure

Source: Lublin (2003, pp. 3–4)

Pan and Newfields (2011) found that the learners in their study allocated more time to studying English because of the test and adopted more test-related practices with more variation in the methods used. The washback literature suggests that learners are most likely to resort to traditional methods rather than more communicatively oriented methods when preparing for a test (Pan, 2014;

Zhan & Andrews, 2014). Zhan and Andrews (2014) also found that students were more willing to change “what” they learned (content) than “how” (method) they learned. This type of washback is described as “superficial” (Cheng, 1998) or “quantitative” (Andrews, 1994). As the test draws closer, “desperate” learners rely on their old test preparation methods and use language learning strategies (e.g. rote memorizing) that they know will work best for them, even though this has only a short-term effect.

Akpinar and Cakildere (2013), investigating two high-stakes language tests in Turkey, namely the Kamu Personeli Dil Sınavı (KPDS) and the Üniversiteler Arası Dil Sınavı (ÜDS), found that most learners focused more on passing the exam than improving skills not included in the test. These two tests only brought about positive washback for reading, which was the only skill tested. The learners in their study reportedly neglected the other three language skills – speaking, listening and writing – as they were not tested. However, Akpinar and Cakildere (2013) did not clearly identify the objective(s) of the two tests, making it difficult to determine if the washback effect on the skills that were not tested was as negative as they claimed. In China, Ren’s (2011) findings were similar: the students had little incentive to learn anything that was not tested and put very little effort into doing so as their primary motive was to pass the test.

A review of past studies shows that the most popular language learning strategies used tend to be conducting test analysis by studying model questions and practising test-taking skills and drilling what is tested (Xie & Andrews, 2012; Zhan & Andrews, 2014). Learners have also been reported to place more emphasis on test tasks or components with higher value so they can perform well in the test.

2.8 Conceptualizing washback in the study

Alderson and Wall (1993) suggested that findings from the field of motivation could shed light on possible mediating variables when investigating the washback phenomenon. More than a decade later, Watanabe (2006) highlighted the importance of theories of motivation in understanding washback on learning. He believed that “the process of washback being generated to the learners also seems to be mediated by some psychological factors much as the case of washback to the teacher” (Watanabe, 2006, p. 2). In the same year, Tsagari (2006) expressed her concern regarding the small number of studies dealing with the issue of learners’ motivation and its relation to tests, although a number of researchers had claimed that tests can be used to motivate learners (Hayes & Read, 2004; Wall, 2000). Due to the complexity of the relationship between washback and motivation, Tsagari (2006) called on more researchers to look into how high-stakes tests motivate students to learn and whether they can help sustain students’ motivation for learning after the exam.

In terms of Alderson and Wall’s (1993) and Watanabe’s (2006) views concerning the need to examine the role of motivation in the process of washback, it is believed that if learners wish to pass a test, they will be more likely to be influenced by the examination in their learning. Motivation is a very broad concept and its theories are very wide ranging. In the field of washback, most researchers who have addressed motivation have tended to view at it at the surface level, treating it as a general concept. For instance, Pan (2014) examined the relationship between motivation and high-stakes tests but articulated no specific theory of motivation.

However, several washback researchers have started to narrow their focus when researching motivation in washback. For example, Shih (2013) used self-determination theory to explore learners’ motivational regulation types. Another study on motivation and washback

conducted by Xie and Andrews (2012) used the expectancy–value theory of motivation to explain the washback phenomenon. They found that positive endorsement of intended test demand was related to higher evaluation of test importance. Gosa (2004) used the major theme “expectations” in the same theory to explain why the students in her study experienced little washback in class, but she did not establish any link between expectations and the diverse amounts and types of washback the students experienced. Zhan’s (2009) study, using a qualitative interpretative approach, revealed another theory of motivation that can be used to explain one of the uncharted areas in the cycle of washback, namely washback on the learning process. Through cyclical data analysis, Zhan (2009) reported that possible L2 selves emerged as an overarching theme. Possible selves refer to the “vivid portrayal of one’s self in future states, including thoughts, images and senses” (Zhan & Andrews, 2014, p. 74). In other words, they relate to how one views or pictures oneself in the future. In this study, which was interested in exploring the relationship between students’ perceptions of a test, their English language ability and the washback of the test, Bandura’s (1986) self-efficacy theory was chosen as a guide as it deals with how individuals perceive their own ability (as discussed in 2.62).

Based on the previous discussion and the washback models presented in 2.3, germane to this study, it was assumed that students’ perceptions not only of the test (test importance, test difficulty), but also of themselves (self-efficacy), might influence (or not) to a certain extent their behaviours (language learning strategies) when preparing for a test. Hence, the following two research questions were formulated to guide the study:

1. What are students’ perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?
2. How does washback of the MUET operate and to what extent do students’ perceptions seem to have a washback effect on their language learning strategies?

The third research question was formulated to explore two of the dimensions of washback that have rarely been investigated, particularly together, namely washback intensity and length. Pertaining to washback length, washback has a short-term effect if the influence of the test on students' learning dissipates once the test ends. However, the factors that determine or influence the length of washback have not previously been explored. Hence, the third research question was formulated to also investigate washback length:

3. What is the intensity of the washback effect of the MUET and what is its length? How do these appear to influence washback on the learners?

Having established the theoretical perspective and the research questions of the study, the following section covers the high-stakes language test under investigation, i.e. the MUET.

2.9 Assessment context in Malaysia

The Malaysian education system is divided into three levels: primary, secondary and pre-university. Primary school consists of six years, with students starting in Year 1 at the age of seven. Students in Malaysian public schools study English from Year 1. Malaysia having been a British colony in the past, English has a high status among Malaysians. It is considered the country's second language and is used widely for commerce, education and management. From Years 1 to 3 (lower primary), English is taught for 240 minutes per week, followed by 210 minutes per week as a compulsory subject in Years 4 to 6 (upper primary). At the end of Year 6, students sit their first national examination, the Ujian Penilaian Sekolah Rendah (UPSR; Primary School Assessment Test).

Next, they proceed to secondary school for another five years of learning, in which English is taught for 200 minutes per week. Students sit a national test at the end of Form 3 (the Penilaian

Menengah Rendah [PMR]; Lower Secondary Assessment) and again at the end of Form 5 (the Sijil Pelajaran Malaysia [SPM]; Malaysian Certification of Education). The SPM is considered one of the most important national examinations as the certificate is needed to apply for further education and employment opportunities and can be regarded as the local version of the General Certificate of Education ([GCSE]; Rethinasamy & Chuah, 2011). To be awarded the certificate, the students need at least to pass the English language paper. In all three national examinations, only reading and writing are tested in the English language paper, not listening or speaking, which explains why the English language learning syllabus in Malaysian schools does not focus on the latter to any great extent.

The pre-university level consists of Form 6 (Lower 6 and Upper 6) and matriculation (Semester 1 and Semester 2). The final school examination, Sijil Tinggi Pelajaran Menengah ([STPM]; High School Certificate) is taken at the end of Form 6. Matriculation students have to sit final exams at the end of each semester and their cumulative grade point average (CGPA) will be calculated at the end of their studies. Both of these pre-university level examinations are considered for entry into tertiary education. In the Malaysian context, gaining entry to universities, especially public universities, is considered a highly important, both by the students and their parents, as it is a symbol of success. With university education, it is assumed that one will have a bright future ahead. Unlike the three national examinations, UPSR, PMR, and SPM, in which English is a compulsory subject, it was historically not tested officially at the pre-university level. This made it difficult for university administrative officers to gauge applicants' English language proficiency and they had to rely on students' English language grades on the SPM for university entry purposes. Moreover, prior to 2000, there was a two-year gap in which English was not taught at all, raising many concerns, as this clearly put the students at a disadvantage in terms of preparing

for tertiary education. In Malaysian public universities English tends to be the medium of instruction. Students are expected to have a certain level of proficiency in English as lectures, reading materials, references and assessments are predominantly in English. This “educational dilemma”, as Wong and Chan (2009) put it, led to representatives from Malaysian universities discussing a solution and the birth of a new English test, the MUET, in 1999. This was a major change in language test policy in Malaysia.

2.10 Malaysian University English Test (MUET)

This section examines the MUET in detail to provide background on the test under investigation. This is then followed by a review of relevant studies specifically on the MUET to establish what has been done and what needs to be done in research on the washback of the MUET.

The MUET was first introduced in 2000 with the aim of (1) bridging the gap in English language needs between secondary and tertiary education and (2) consolidating and enhancing the English language proficiency of students preparing to enter Malaysian public universities. It has become one of the entry requirements for tertiary level education, specifically for first degree studies, in all public universities in Malaysia. This move was seen as necessary and timely, although some had reservations about using the MUET scores for university entry. This test is valid in Malaysia and some parts of Singapore. The MUET is a high-stakes standardized English language test that is administered and monitored by the Malaysian Examination Council (MEC). The MEC (2001, p. 11) stated that the MUET syllabus “seeks to consolidate the English language ability of pre-university students to enable them to perform effectively in their academic pursuits at tertiary level, in line with the aspirations of the National Education Philosophy”.

Nambiar and Ransirini (2012) described the MUET as a criterion-referenced proficiency test, similar to IELTS and TOEFL. As shown in Table 2.2, the test includes all four language skills – reading, writing, listening and speaking – with reading given the highest weighting of 45% of the overall band, followed by writing (25%), speaking (15%) and listening (15%).

Table 2.2

MUET test scores for each component

Test Component	Time	Maximum possible score	Weighting (%)
Listening	1/2 hour	45	15
Speaking	1/2 hour	45	15
Reading Comprehension	2 hours	135	45
Writing	1 1/2 hours	75	25
Total	4 1/2 hours	300	100

Performance on the test is reported in terms of an aggregated score with respect to six levels of achievement, referred to as Bands 1–6, Band 1 being the lowest and Band 6 the highest as depicted in Table 2.3.

Table 2.3

The MUET scoring guide

Aggregated score	Band	User	Communicative Ability	Comprehension	Task performance
260–300	6	Highly proficient user	Very fluent; highly appropriate use of language, hardly any grammatical errors	Very good understanding of language and context	Very high ability to function in the language
220–259	5	Proficient user	Fluent; appropriate use of language; few grammatical errors	Good understanding of language and context	High ability to function in the language
180–219	4	Satisfactory user	Generally fluent; generally appropriate use of language; some grammatical errors	Satisfactory understanding of language and context	Satisfactory ability to function in the language
140–179	3	Modest user	Fairly fluent; fairly appropriate use of language; many grammatical errors	Fair understanding of language and context	Fair ability to function in the language
100–139	2	Limited user	Not fluent; inappropriate use of language; very frequent grammatical errors	Limited understanding of language and context	Limited ability to function in the language
Below 100	1	Very minimal	Hardly able to use the language	Very limited understanding of language and context	Very limited ability to function in the language

The MUET is designed to prepare students to meet the demands of English language at tertiary level. The different weighting given to the different skills in the MUET depict the test designers' view of the demands of tertiary education in terms of language skills; reading has the highest weighting at 45% of the overall band score, reflecting the fact that students have to do a great deal

of reading at university and most of the resources are written in English. According to Alderson, Clapham, and Wall (1995, p. 149), weighting is given based on the extra value assigned to a language component believed to be central to the curriculum or to the concept of proficiency. As most universities in Malaysia use English as their medium of instruction, a minimum level on the MUET is required for certain courses. For example, for critical courses that are considered particularly demanding in terms of English language ability, such as medicine, law and TESL, applicants are required to obtain at least Band 4 or Band 5 in the MUET.

The practice of using high-stakes English tests to bridge a potential language gap is also common in other countries. For example, in the UK, university applicants whose first language is not English must obtain a certain minimum IELTS score, typically 6.5, to enrol on university courses. This is to ensure that the students are able to meet the linguistic demands of the course, especially in the Anglophone context.

Yang and Badger (2015) explored the extent to which IELTS preparation classes help international students meet the admission requirements and the demands of A-level Economics courses in terms of their English language needs in a college in the UK. They collected four sets of data, comprising field notes from preliminary observations in the first two weeks of lessons, student interviews, teacher interviews and teaching materials. Rather than focusing only on certain language skills or all language skills in general, Yang and Badger (2015) undertook a comprehensive analysis of the washback of IELTS preparation classes for each of the language skills: reading, listening, writing and speaking. Their analysis found that the IELTS preparation classes did help the students in terms of their language skills to a certain extent, but failed to cater to the demands of the Economics course. There were gaps between the IELTS course and the Economics course that needed to be addressed. However, they deemed it unfitting and unfair to

expect an English language course to prepare international students not only to meet the language demands of the Economics course, but also other demands, namely enabling the students to understand colloquial language or be able to engage in independent study or research. In the case of the MUET, as mentioned earlier in this section, the main purpose of the test is to cater to the students' English language needs in tertiary education. The following section discusses in detail previous studies conducted on the MUET thus far.

2.10.1 Previous studies on the MUET

Singh, Ismail, & Safinas (2012) conducted a study on the MUET to investigate if there were any relationship between students' informal exposure to English and their achievement in the test. They measured students' informal exposure to English by looking at the average hours spent by the respondents on studying the four skills. They used a questionnaire adapted from a past study to gather data from 50 undergraduates in a public university in Malaysia. They found a positive relationship between the degree of informal exposure to English and achievement in the MUET. However, it should be noted that the respondents took the MUET before their informal exposure to English was recorded, which suggests the findings are rather back to front.

Othman and Nordin (2013) investigated the predictability of MUET scores in relation to CGPA with students on a BEd TESL programme. The reason for investigating this field of study was that the MUET result was used as one of the main criteria for getting onto the desired course at the desired university. As already noted, for high-demand courses, such as medicine, law and TESL, the applicants are required to attain a minimum Band 4 in the MUET. They collected 111 students' MUET results and their CGPA to examine the significance of any relationship. Unlike IELTS, each component tested in the MUET has a different weighting. Aside from collecting

students' overall MUET band level, they also collected the scores that the students had achieved for each component. They found that the MUET components could actually predict the students' CGPA, with listening and reading emerging as the strongest predictors of students' academic achievement. They concluded that for students to do well academically, they would need to attain good competence in reading and listening skills. However, they did not tap into other mediating variables, such as motivation, scholastic aptitude, attitude, exposure to English medium instruction or previous academic performance, which could also contribute to students' academic achievement. While their findings seem enlightening, they are also limited to BEd TESL students, taking courses related to English. Hence, it was not a particular surprise that the MUET scores could predict students' CGPA to a certain extent. Research with students not taking English language courses might yield different findings in terms of using MUET results to predict current achievement.

In contrast, Thomas and Noordin (2013) looked into the relationship between students' emotional intelligence (EI) and their achievement in the MUET based on their MUET score. Basing their work on Goleman's (1996) claim that an individual's success depends 25% on their IQ and 80% on the effectiveness of how they manage their emotions in life, they wanted to explore the potential relationship between students' EI, consisting of well-being, self-control, emotionality, sociability and global traits, with their performance in the MUET. They used the Trait Emotional Intelligence Questionnaire Short Form (TEIQue-SF) developed by Petrides et al. (2007) to collect data. From their findings, it appeared that there was a positive relationship between three out of four of the EI traits – well-being, sociability and emotionality – and achievement in the MUET, but the relationship was either low or negligible and thus there was also a positive relationship between the global trait and achievement in the MUET. This suggests

that an increase in EI traits could improve students' attainment in the MUET. It is interesting to note that one of the traits of EI, as defined and introduced by Salovey and Mayer (1990) 28 years ago, is the ability to perceive emotions in oneself and others. In other words, there is a possibility that students' perceptions play a vital role in determining their actions and indirectly affect their achievements. This provided a motivation to explore the relationship between students' perceptions and their language learning strategies in greater depth in this study, focusing on the MUET.

Nopiah et al. (2011) investigated the relationship between students' English proficiency, measured using their MUET score, and their academic achievement, measured using their current CGPA. Unlike Othman and Nordin's (2013) study, which took TESL students as the sample, Nopiah et al. (2011) looked at engineering students. They also wanted to see if loading hours would affect students' CGPA or not. Loading hours refer to the number of credit hours taken by the students per semester. They postulated a hypothesis based on Lahmers and Zulauf's (2000) work, which found that an increase in students' loading hours could increase their CGPA. Nopiah et al.'s (2011) study involved 266 engineering students in various specialties, including chemical engineering, civil engineering, etc. Information on the students' MUET scores, loading hours and CGPA were obtained from the dean's office. Based on correlational statistical analyses, they found a positive but weak correlation between students' MUET scores and their CGPA and the relationship between these two constructs was significant. They also found that an increase in loading hours did not help students to increase their CGPA; rather, their CGPA declined. This could be attributed to the students' heavy study load, which might have caused them to be overwhelmed. Nonetheless, it was interesting to see that there was correlation between English

proficiency and students' achievement even though they were not English major students. This shows the significance of English in the Malaysian context.

A more recent study on the MUET conducted by Rashid (2014) focused on the washback effects of the test on teachers' perceptions of their classroom teaching and teaching materials. Rashid (2014) based her research questions on Wall and Alderson's (1993) 15 hypotheses and narrowed them down to focusing on how the MUET affected teaching in terms of what the teachers teach and the degree and depth of teaching. Her study involved seven teachers who were teaching preparation classes and she used a self-report survey adapted from the work of Cheng (2005), Qi (2005) and Tsagari (2007) to elicit data. The questionnaire contained 46 items that examined the teachers' perceptions of the washback effects of the MUET on their teaching, teaching materials and depth of teaching.

Based on her observation of the current trend in language testing in Malaysia, she found that the MUET seemed to have replaced classroom instruction in that students were trained to sit for the test rather than to master the language. This is in contradiction of the MUET's main objective, which is to bridge the gap in English language needs between secondary and tertiary education. In her study, although the teachers disagreed that their classes were MUET-driven, they admitted to reminding their students constantly of the importance of the test for their future, alongside the importance of mastering and applying the English language skills that were taught in the classroom. In relation to classroom activities, they stated that although they did not quote any MUET questions from past years, the activities they conducted in the classroom were relevant to the MUET. Additional teaching and learning materials outside those set in the syllabus were also included in the lesson to help students perform well in the test and master language skills at the same time. Despite the teachers' efforts to vary their teaching and learning materials, most of

them agreed that the students' test scores could be raised by teaching to the test. This was reflected in the findings, as the teachers admitted to having taught the students test-taking strategies, such as tips on how to do well in the test even though they did not possess the required skills. Although the teachers believed students' mastery of language skills to be more important than a high test score, they had to prioritize their students' immediate needs first, namely to excel in the test.

However, Rashid (2014) did not consider the students' perspectives in her study and hence failed to confirm whether the teachers' concerns were baseless or not. Furthermore, her findings were based on only seven respondents and richer data could have been obtained had she interviewed the teachers, rather than asking them to complete a questionnaire. Positive washback was reported when it came to the text book and course materials employed in the classroom. According to the teachers, using a text book based on the MUET specification outlined by the MoE helped the students to familiarize themselves with the test format and the level of difficulty of the actual test. Rashid's (2014) findings echoed Cheng's (1997) and Tsagari's (2007) claim that an established high-stakes test will have positive washback in terms of the text book used. This has further been confirmed by the rapidly increasing numbers of MUET text books being made available by different publishers. This study concluded that the MUET had both positive and negative washback in relation to teaching, specifically what teachers teach and the degree and depth of teaching.

To date, Nambiar and Ransirini's (2012) washback study is one of the few on the MUET to take into account both teachers' and students' perspectives. The scarcity of washback studies in Malaysia and the grave concern over the decline of Malaysian undergraduates' English language proficiency prompted them to conduct this study. Multiple data gathering instruments, including questionnaires, interviews and classroom observation, were used to gather data from the two key

stakeholders of the MUET: teachers and students. Nambiar and Ransirini (2012) focused on students and the teachers at the pre-university level (Form 6) in two secondary schools and a matriculation centre. Teacher and student questionnaires were utilised to elicit data from 108 students and 9 teachers altogether. On top of the questionnaires, Nambiar and Ransirini (2012) also conducted interviews with the 9 teachers and non-participant classroom observations with 8 classes for 17 hours in total. From the findings, they found that the students were of the opinion that the MUET had improved their English language proficiency, but it was not enough to prepare them for university. However, when this study was conducted, the students were not yet enrolled in university. They would not be able accurately to gauge the demands in terms of English at the tertiary level. Hence, this study aimed to address this gap by including both students preparing for the MUET and those who had already taken the exam.

Nambiar and Ransirini's (2012) study also revealed that the teachers and the students were not entirely aware of the objectives of the MUET, which might hinder washback. The MUET was specifically designed to prepare students to undertake courses at the tertiary level, hence the different weighting of components according to the importance of the various skills. The test designers assigned reading the highest weighting as they viewed this as the skill most used at university. However, the findings showed that both students and teachers perceived speaking skills to be those most needed at university. The students were of the opinion that of the four language skills, their speaking skills had improved the most. Moreover, the teachers revealed that they liked the MUET because of its speaking component, as this skill had not previously been tested in any other national high-stakes test in Malaysia. The increased attention paid to speaking skills could certainly be considered positive washback, as long as it does not undermine other skills perceived as less important, such as listening. The findings showed that the teachers neglected listening skills

altogether when asked to rate the skills in order of usefulness at university. The findings also showed that there was a mismatch between what the teachers thought they focused on in their lessons and focus reported by the students. Both the teachers and the students tended to minimize the importance of listening skills, which can be considered negative washback of the MUET. The students in this study expressed their willingness to engage in meaningful learning activities that were not directly related to the MUET. The teachers expressed a similar view, but cited time constraints a major obstacle as they only had just enough time to prepare the students for the test.

Wong and Chan (2009), in a critical examination of the impact of the MUET, distributed questionnaires to 200 undergraduate students at a public university in Malaysia who had already sat the exam. Their study reported that the MUET did not seem to have had any dramatic effects on language improvement in terms of the four skills. They attributed this to the short period of time – the two-year pre-university gap – in which students could improve their English language mastery and any significant improvement seemed rather ambitious. However, they found that the MUET could be considered useful in aiding the students in their educational endeavours at university in terms of preparing them for the workplace, as mastery of English equates with better work prospects.

Moreover, Wong and Chan's (2009) study showed that the educational environment and social practices in the Malaysian context could be the reason why the English proficiency level of the undergraduates was below average. A learning atmosphere that was neither conducive to nor supportive of enhancing language proficiency made it difficult for the students to find the opportunity or motivation to practise and use English, despite knowing how important it was. In particular, Wong and Chan (2009) found that the students were not inclined to engage with reading materials in English. They also believed English to be difficult to learn and master, signifying low

self-efficacy. In terms of the specific language components tested in the MUET, the inclusion of speaking did not appear to be very successful in encouraging these students to communicate more in English. Listening, on the other hand, was considered easy by the students, as reflected in their high listening scores on the MUET. These students claimed that they were “highly skilled in recognizing speakers’ attitudes, roles and relationships” (Wong & Chan, 2009, p. 271).

2.11 Summary

In the first part of this chapter, the definitions of the key terms in the present study and the concept of washback was discussed. It was then followed by discussions of several key issues of washback which includes mechanisms of how washback operates, dimensions of washback, complexity of the washback phenomenon and conceptualisations of washback in the current literature. Next, one of the main focus of this study, which is on the role of students’ perceptions in influencing washback, specifically perceptions of tests and perceptions of self-efficacy, was elaborated in great detail before moving deeper into a review of the assessment scenario specific to the Malaysian context of this study,.

The literature reviewed in this chapter has shown that washback is a complex phenomenon, that despite some notable studies into the impact of testing, the need for more empirical studies into how washback works, or does not work, still remains. 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 further explored. Although we now know that tests exert washback differently on different groups of stakeholders, we know little about how washback operates for learners. This lack of understanding of learners as participants in the washback cycle provided the theoretical motivation

for this research. In addition, washback on learning should be studied in relation to the students' perceptions of the test under investigation itself in order to see to what extent their language learning strategies when preparing for the test were/were not influenced by their perceptions.

The second issue is related to the methodological concerns for washback studies to provide a more holistic view of the phenomenon. Even though the focus of this study is on the learners, data from the teachers were collected as well for triangulation in order to compensate what was lacking from the students' data. The literature review has also revealed that little information is known or published on one of the underexplored washback dimensions, known as washback length. One possible reason this washback dimension is still underexplored is due to the difficulty in measuring the length of the washback itself. Hence, this study attempted to also collect empirical data on the length of the washback of the MUET and hopefully contribute to a wider understanding of this phenomenon.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter begins by restating the three research questions for this study, investigating the washback effect of the Malaysian University English Test (MUET) as a university entry requirement for students in Malaysia. The research paradigm underpinning this study is then outlined. The chapter goes on to describe how the research was planned and designed to address the aims and objectives. It sets out the design and implementation of the research methods and how the data were analysed. The final section of this chapter explores the research procedures, the analysis of data and changes made after piloting the instruments.

3.2 Aim and research questions

The overarching aim of this study was to investigate the washback effect of the MUET on Malaysian university students. In doing so, it focused on examining (i) students' perceptions of the MUET and (ii) their language learning strategies related to the test. In addition, this research was interested in understanding the intensity and length of washback from the MUET on the students. The research questions guiding this study were as follows:

1. What are students' perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?
2. How does washback of the MUET operate and to what extent do students' perceptions seem to have a washback effect on their language learning strategies?
3. What is the intensity of the washback effect of the MUET and what is its length? How do these appear to influence the washback on the learners?

An understanding of the research designs used in previous washback studies was important to seek answers to the above research questions and gain a better understanding of the washback phenomenon. Identifying an appropriate research design from the washback literature is essential. Parahoo (1997, p. 396) defines a research question as “the broad question which is set at the start of a study”. In order to answer research questions, an overall strategy to integrate the different components of the study in a coherent and logical way, also known as research design, needs to be developed carefully and is tightly related to the research questions (De Vaus, 2001). Research designs are associated with the practical arrangements of obtaining answers to the research questions. According to De Vaus (2001, p.9), the function of a research design is to “ensure that the evidence obtained enables us to answer the initial question as unambiguously as possible”. The relationship between the research questions and the design of the research is fundamental to the research process as a whole, as it determines the quality of the overall research.

In developing the research questions, the type of knowledge that the researcher wishes to obtain, whether descriptive, explanatory or predictive, first needs to be established. This study was interested in explaining the relationship between students’ perceptions of the test, their self-efficacy and the language learning strategies they adopted when preparing for the MUET. As discussed in the literature review, some knowledge concerning the variables in the study is already available; hence a new research design was developed to explore the relationships between the various components in greater depth.

To determine what design to employ in research, the underpinning philosophical assumptions need to be established first. Identifying the philosophical stance and ontological and epistemological perspectives underpinning a study helps the researcher shape the process of the research (Creswell & Clark, 2011) and decide which data collection method(s) to employ (Cohen,

Manion, & Morrison, 2007). A more detailed discussion of the philosophical assumptions underpinning this study is provided in the following section.

3.3 Philosophical assumptions

A research paradigm or “worldview” (Creswell & Clark, 2011) is defined as “a way of looking at the world” (Mertens, 2010, p. 7) through “a basic set of beliefs that guide actions” (Guba, 1990, p. 17). These relate to a researcher’s views of how research should be carried out, i.e. the judgement of how best to study a certain phenomenon, before deciding on which methods to employ. According to Guba and Lincoln (1994), there are three guiding principles for a paradigm: ontology, epistemology and methodology.

Ontology deals with the nature of reality being studied, which relates to “the study of being, that is, the nature of existence” (Gray, 2004, p. 16), or more simply, “a specification of what exists” (Newby, 2014, p. 35). Essentially, the ontological consideration of a phenomenon helps the researcher seek answers to whether something exists as it is believed to or independently of the researcher’s belief(s). There are two contrasting perspectives associated with ontology: (i) that held by realists, who contend that a phenomenon can exist independently of human beliefs and judge research data as facts to be assembled into an understanding; (ii) that held by relativists or constructivists, who interpret evidence through moral, political, economic and cultural perspectives and accept that beliefs about reality are subjective and personal, as different individuals interpret a phenomenon in different ways (Guba & Lincoln, 1994; Newby, 2014).

Epistemology refers to beliefs about the way in which knowledge is interpreted, particularly in terms of how the researcher acquires knowledge. This is one of the main branches of thinking in philosophy that concerns itself with “the study or the theory of what constitutes

knowledge” (Atkins & Wallace, 2012, p. 24). Epistemology seeks an answer to the question “How can we be sure?”, for example when it comes to evaluating if someone’s interpretation of a certain phenomenon is correct. According to Newby (2014), the only way of doing so is by studying the evidence and assessing how it was collected. Epistemologically, there are two opposing extremes: positivist (realism) and interpretivist (constructivism). Positivists view the world as an objective reality. Hence, they use rigorous scientific methods to uncover knowledge that is highly objective and empirically verifiable. When it comes to making sense of the information gathered, researchers operating within a positivist paradigm detach themselves from their data and analyse them in a manner that ensures the data are value-free. Hence, they are not likely to accept evidence presented by interpretivists, who acknowledge “some degree of subjectivity in the researcher and other participants” when “[seeking] to throw light on a particular case or situation” (Atkins & Wallace, 2012, p. 22).

This stance in this study is neither strongly positivist nor radically interpretivist, but is rather pragmatic, i.e. accepting elements of the two paradigms, as it involves both inductive and deductive thinking. This perspective is consistent with Onaiba’s (2013) comprehensive study undertaken for his doctoral dissertation on the washback effect of a revised EFL public examination on teachers’ instructional practices, materials and the curriculum in Libya, which reflected aspects of the two paradigms. His justification was that a more positivist orientation enabled him to gain generalizable data, while an interpretivist orientation helped him to collect deep and rich data (Onaiba, 2013). As washback is often deemed complex and multidimensional (Alderson & Wall, 1993), orienting to a single paradigm will not suffice. In washback research, it is the norm to collect data from various sources, thus triangulating the data to understand the phenomenon (e.g. Munoz & Alvarez, 2010; Pan & Newfields, 2011, 2012; Qi, 2007).

Although the notion of integrating aspects of two different paradigms has sparked intense debate among researchers, namely that the contradictory nature of the two paradigms will nullify each other (Guba & Lincoln, 1994; Smith, 1983 as cited in Onaiba, 2013), many washback researchers have started to acknowledge that “within the past decade, the borders and boundary lines separating these paradigms and perspectives have begun to blur” (Denzin & Lincoln, 2003, p. 246). Likewise, Creswell and Clark (2011, p. 45) noted that “more than one worldview might be used in a mixed methods study”. This means that a researcher may employ a quantitative methodological strategy in one phase of the study, followed by a qualitative approach in the next phase, although these two approaches belong to different paradigms.

Due to the complexity of washback, I could not afford to lose potentially valuable data by confining the study to a limited, highly-structured, philosophical stance. This study started with a questionnaire, primarily situated within a post-positivist paradigm, beginning with empirical measures of specific variables under investigation. At the same time, qualitative methodological strategies, namely interviews and classroom observation, were then employed to follow up and further explain the quantitative data obtained earlier. This reflects a more constructivist perspective, making it possible to get the best of both worlds (Creswell & Clark, 2011).

3.4 Context of the study

This section introduces the stakeholders of the MUET in this study – the students and the teachers – explaining the reason for selecting them and how they were recruited for this study. Given the focus of the study on a specific high-stakes language proficiency test, the MUET, the research context was Malaysia. The MUET is one of the requirements for university entrance in Malaysia. Hence, students at the pre-university level (Group A) were deemed suitable participants. However,

the study was also interested in exploring the length of washback in terms of lasting effects after students had taken the test, making it necessary to include those currently doing their first degree at the university level (Group B). Moreover, although the main focus of this study was on the students, data from teachers were also collected for triangulation purposes. As pointed out by Alderson (2004), teacher-related factors play a crucial role in shaping the washback effect of a test, as their beliefs and understanding of the nature and the rationale for the test will to a certain extent influence the ways in which they prepare their students for the test and how students respond.

3.4.1 Research site and criteria for selection

There are three main educational levels in Malaysia: school level (primary and secondary), pre-university level and university level. After students have completed secondary school and before pursuing study at the university level, they are required to undertake pre-university education, depending on their performance at school and also their preferences in terms of specialization. In Malaysia, there are three options for students wishing to undertake pre-university education: diploma courses (6 semesters), Form 6 (2 semesters) or A-level/matriculation (2 semesters). For all these three options, students are provided with MUET preparation courses to prepare them for sitting the test.

It was not possible to investigate all three types of pre-university institution in this study as they provide MUET preparation classes at different times. For example, matriculation colleges hold MUET preparation classes in the first semester, whereas Form 6 schools do so in both semesters one and two. Therefore, for this study, it was decided to focus on one pre-university institution, a Form 6 school. The schedule for Form 6 schools is less packed compared to that of

matriculation colleges and there are also more contact hours for MUET preparation classes, increasing the feasibility of collecting data.

This study was also interested in exploring the washback length of the MUET. Taking into consideration the time constraints on data collection, despite the need to consider certain validity and reliability issues, it was decided to include students who had already sat the MUET rather than conducting a longitudinal study with the same participants. Therefore, this study also included a group of first-year students from one of the public universities in Malaysia, who had recently sat the test, in the hope that they would still be able to recall their test preparation experience from 9–12 months previously. The public university chosen for this study was the University Malaysia Kelantan (UMK). Located on the eastern side of the peninsula, its students come from all over the country. The next section discusses the sample of the study in detail, consisting of the students in Group A and Group B.

3.4.2 Study sample

As noted above, two groups of students were recruited for this study. A combination of convenience and purposive sampling was used to recruit the participants for this study. The participants were selected because they were willing and available to be studied. In this case however, I cannot say with confidence that the individuals are representative of the population. However, this sample can provide useful information for answering questions and hypotheses. Group A comprised 137 pre-university students from three high schools offering Form 6 education. Students wishing to undertake Form 6 courses have to sit the Malaysian High School Certificate (STPM), a high-stakes standardized test, to gain entry to tertiary education. Two streams are offered in Form 6 education – Social Sciences and Science – and the MUET is a

compulsory subject for all students.. Form 6 courses run for three terms, as detailed in Table 3.1. The three schools were chosen because they were available at the time of data collection for this study. Furthermore, I had the permission from the schools' principals and the consent from the Form Six students and teachers to collect data at their schools during the English language lessons.

Table 3.1

Form 6 term duration and course outline

Term	Duration	Course outline
1	May–November (26 weeks)	<ul style="list-style-type: none"> • Teaching and learning • Coursework (for certain subjects) • Term 1 Examination (P1)
2	January–May (20 weeks)	<ul style="list-style-type: none"> • Teaching and learning • Coursework (for certain subjects) • Term 2 Examination (P2)
3	May–November (26 weeks)	<ul style="list-style-type: none"> • Teaching and learning • Coursework (for certain subjects) • Term 3 Examination (P3), Repeat Examination 1 (U1) and Repeat Examination 2 (U2)

To explore the washback length of the MUET, 238 students from UMK who had already taken the test were also recruited to form Group B. Although it appears somewhat ambitious to use a cross-sectional approach rather than carrying out a longitudinal study to investigate washback length, one of the aims of this study is to explore this aspect by involving students who have sat for the same test in the past (see 3.7.1).

Data collection for the study commenced approximately 9 to 12 months after they had sat the MUET. These students' insights were deemed valuable for the study as they could reflect on and compare their experiences when they were preparing for the MUET with their current English language learning experience and strategies. This was not a comparative study and the data from

Group B were thus used solely to explore the length of washback from the MUET, not to compare and contrast the findings from Group A.

Demographic data were collected from the respondents in both Group A and Group B. As indicated in Figure 3.1, there were more female students than male students, which was to be expected as the female population in Malaysia in general is higher, especially in higher learning institutions. In 2017, the Planning, Research and Policy Coordination Division, Ministry of Higher Education (MoHE) Malaysia, reported a male-to-female ratio of students enrolled at public higher education institutes of 1:1.6, with 37.94% male and 62.06% female (Malaysia Educational Statistics, 2017).

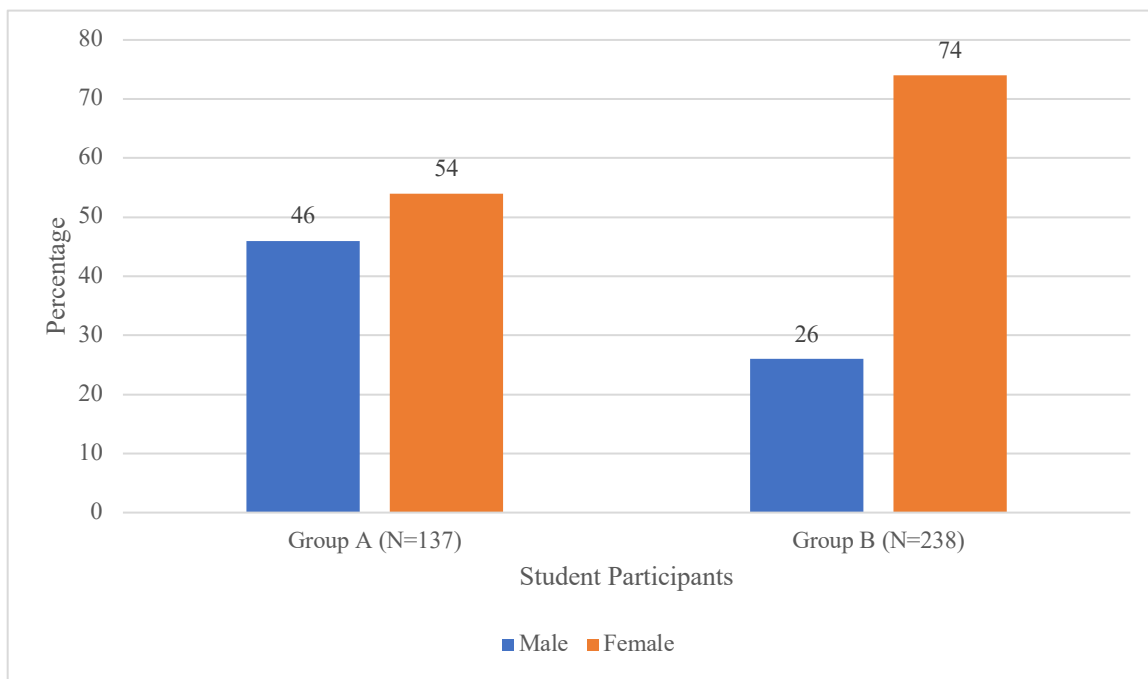


Figure 3.1 Distribution of students in Group A and Group B according to gender.

The students were categorized into two different English language proficiency groups based on the grades they had obtained on their English language test in the Malaysia Certificate Examination

(MCE), a national public examination equivalent to GCSE in the UK. The grading for the MCE is as follows: A+, A, A-, B+, B, C+, C, D, E and G. For this study, those who obtained grades A+, A, A-, B+ and B were categorized as high proficiency, while those who obtained grades C+, C, D, E and G were categorized as low proficiency. The decision for grouping the proficiency level into two levels (high and low) as opposed to three levels (high, intermediate, and low) was due to extreme imbalance in the number of participants for the high proficiency group (n=6) if three levels were used. Hence, the students were only categorized into high and low proficiency levels. As shown in Figure 3.2, Group A comprised only 24.1% high-proficiency students, whereas Group B comprised 48.3% high-proficiency students. This could be attributed to the uneven number of participants in the two groups, which was a limitation of this study.

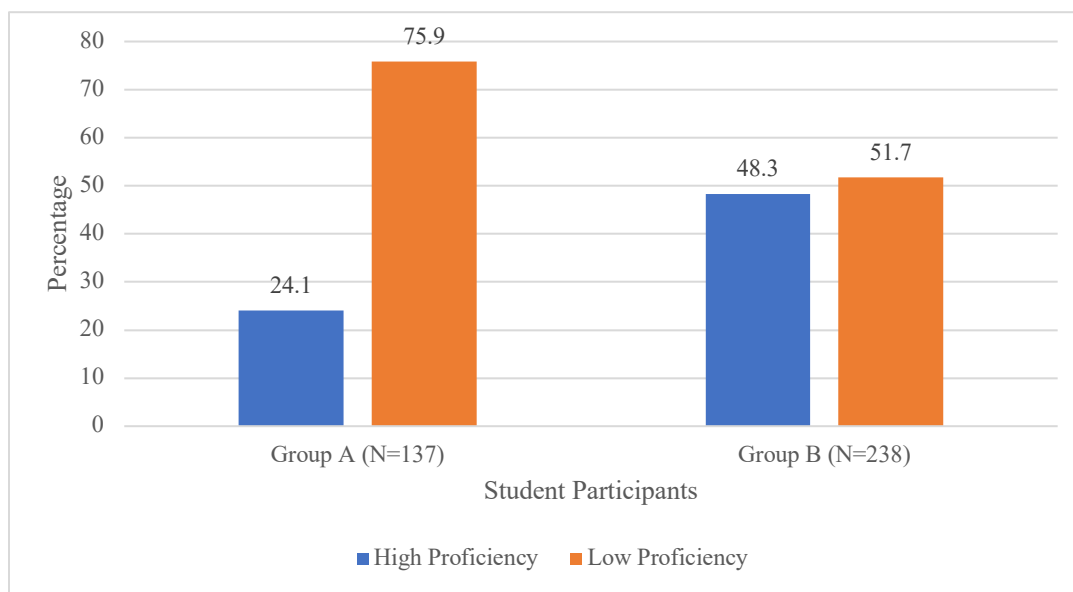


Figure 3.2 Distribution of students in Group A and Group B according to language proficiency.

In terms of the teacher participants, 55 were contacted and 36 responded to the online questionnaire, representing a 65.5% return rate. Table 3.2 presents the demographic information for the teachers who responded to the online questionnaire.

Table 3.2

Demographic information of the teachers

Variables	Categories	Sum	Percentage
Gender	Male	5	13.9
	Female	31	86.1
Teaching experience (years)	1–3	3	8.3
	4–6	2	5.6
	7–9	13	36.1
	> 10	18	50.0
Number of English teaching periods per week (hours)	< 9	4	11.1
	9–15	7	19.4
	16–21	20	55.6
	22–27	3	8.3
	> 27	2	5.6

Table 3.2 shows that 86.1% of the teacher respondents were female and only 13.9% were male. Most of them had more than 10 years of teaching experience. In terms of teaching load per week, most teachers taught English, not limited to the MUET preparation class, for 16–21 hours, except for 2 teachers who had to teach for more than 27 hours per week.

3.5 Research design

As the washback phenomenon is complex in nature, many researchers have noted that a simple causal relationship is inadequate to explain the concept (e.g. Lin, 2010). Hence, a combination of a causal-comparative and a correlational research design, underpinned by a mixed methods approach, was chosen for this study to allow for a more in-depth investigation of the participants'

behaviours and the meanings that they constructed. Drawing on the experience of earlier research on washback, washback researchers have either employed a comparative research design when making comparisons between groups (e.g. Green, 2006; Munoz & Alvarez, 2010; Pan & Newfields, 2012), or a correlational research design when exploring the relationship(s) between different variables (e.g. Green, 2007; Nazari & Nikoopour, 2011; Xie & Andrews, 2012). As the general aim of this study was to investigate the relationship between students' perceptions and their language learning strategies, a combination of comparative and correlational research design was deemed the ideal methodology.

According to Borg and Gall (1979, p. 445), causal-comparative research is commonly used to explore "possible causes for a behaviour pattern by comparing subjects in whom this pattern is present with similar subjects in whom it is absent or present to a lesser degree". This design is also known as *ex post facto* research, as "causes are studied after they have presumably exerted their effect on other variables" (Borg & Gall, 1979, p. 445). According to Cohen et al. (2011), the literal translation of *ex post facto* means "after the fact", which indicates "from what is done afterwards", "from after the event" or "from what has happened". In the context of educational research, *ex post facto* research design is commonly applied in "studies which investigate possible cause-and-effect relationships by observing an existing condition or state of affairs and searching back in time for plausible cause factors" (Cohen et al., 2011, p. 303), but in which experimental manipulation of the variables is not possible. This research design makes it possible to determine the cause-and-effect relationship between an independent variable and dependent variables.

As in experimental or quasi-experimental research, causal-comparative research is also a form of experiment, but without the strict controls of a true experiment, making any inferences of causations tentative (Cohen et al., 2011). Spector (1993, p. 42) went as far as to position causal-

comparative research as a form of quasi-experiment, as it involves a procedure that is intended to transform a non-experimental research design into a pseudo-experimental form. What distinguishes causal-comparative research from the experimental and quasi-experimental design is the control of the independent variables. In an experimental design, the researcher can at least manipulate one active variable and exercise control by randomization, assigning interventions or treatments to groups at random or assigning subjects randomly to groups (Cohen et al., 2011, p. 307). In causal-comparative research, the independent variable under investigation cannot be changed or manipulated as in true experiments, as the variable either already exists, has already happened, or it is unethical to do so (Cohen et al., 2011). In the case that experimental manipulation is not possible, a causal-comparative design can be used to test hypotheses regarding cause-and-effect relationships. For example, to investigate the effects of smoking during pregnancy on the growth of the embryo and child, ethically speaking it is immoral to have a group of pregnant mothers start smoking for the sake of the research. Instead, the researcher can have pregnant mothers who were already smokers participate in the study. The researcher must take things as they are, but can try to make use of selected procedures to disentangle them in order to attain an element of control and thus bridge the gap between the causal-comparative design and the experimental design.

With regard to educational research, when studies involving learners are carried out, strict ethical considerations need to be taken into account. In the context of this study, it was not possible to have students sit a test that they were not supposed to take. Furthermore, the aim was to study students who actually had to sit the test rather than giving them a mock or practice test. Under certain circumstances, Cohen et al. (2011) contend that a causal-comparative design is more suitable than the experimental method when it comes to preserving the authenticity of research

variables, as they are not manipulated or controlled but observed in their natural state. This avoids the risk of having artificial or unrealistic variables that could jeopardize normal interaction with other influential variables. When it comes to investigating the washback effect of a high-stakes English language test, the relationship between the test and the students' behaviours leading to various outcomes has been established in previous washback studies. However, further investigation is needed to identify the possible mediating variable(s) that cause the outcomes, aside from the test itself.

In causal-comparative research, the interpretation is rather limited compared to an experimental design as it is not possible to find out whether a particular variable is the cause or result of the behavioural pattern under investigation (Borg & Gall, 1979, p. 446). Hence, to bridge this gap, this study also adopted a correlational research design, which can be used to determine the relationship between variables through the use of correlation coefficients (Glatthorn & Joyner, 2005). According to Mertens (2010), although both causal-comparative and correlational research are generally employed to investigate phenomena encompassing the inherent characteristics of participants, causal-comparative research focuses on making group comparisons, whereas correlational research aims to provide an estimate of the magnitude of the relationship between variables. Glatthorn and Joyner (2005) caution that correlation is not causation and hence causation cannot be proved. However, by incorporating both correlational and causal-comparative methods in the research design, the hope was that it would be possible to draw conclusions in terms of the relationship both across groups and across variables.

Mertens (2010) added that advancements in the tools of statistical analysis have allowed researchers to combine both designs in their studies. Such researchers start by asking causal-comparative questions and then proceed by examining competing explanations beyond the initial

categorization using more complex correlational analysis (Mertens, 2010, p. 153). In the context of this study, the aim was to determine the washback of the MUET by exploring the relationships between students' perceptions of the test, their self-efficacy and the language learning strategies they adopted when preparing for the test. In addition, the research intended to investigate washback intensity and length by comparing the language learning strategies adopted before and after the students had sat the test.

3.6 Research approach

A brief explanation of the concepts of quantitative and qualitative research approaches is provided before introducing the approach adopted for this study, i.e. a combination of the two approaches, or a mixed methods approach. Methodological justifications based on previous washback studies are then set out.

3.6.1 Quantitative research approach

Creswell (2014, p. 4) defines quantitative research as “an approach for testing objective theories by examining the relationship among variables”. Using quantitative research allows researchers to test theories deductively, avoid bias when interpreting the data, control alternative explanations and disentangle the complexity of a situation or a phenomenon (Creswell, 2014). According to Talib (2013), quantitative research employs objective measures to generate numerical data, which are analysed and explained using statistics. This approach focuses on measuring and searching for the relationship between variables. Quantitative data are used to look at the overall tendency of responses from individuals and to note how this tendency varies among them (Creswell, 2008). The literature on past washback studies (see Akpınar & Cakildere, 2013; Green, 2006, 2007b; Pan,

2014; Ren, 2011; Xie & Andrews, 2012) shows that the findings for washback on learning have mainly been derived based on a quantitative research approach. However, Zhan (2009) argued that questionnaires restrict students' from expressing their perceptions due to their structured nature. He claimed that when it comes to investigating the washback effect, especially on learners, a more flexible approach employing qualitative methods that allow new details to emerge from the data should be deemed more appropriate.

3.6.2 Qualitative research approach

As in other washback studies, the complexity of the phenomenon under investigation requires a detailed and comprehensive understanding of the students' and teachers' perceptions generated from their beliefs, views and experiences. Zhan (2009), for example, employed semi-structured interviews and diaries to investigate systematically how a small group of non-English major students experienced washback from a written test revised in 2006 as these methods enabled him to collect in-depth data and gain insider perspectives.

Collecting qualitative data can be time consuming, but the data gathered are valuable and rich. According to Denzin and Lincoln (2003, p. 3), qualitative research involves an interpretive and naturalistic approach: "This means that qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them". As noted by Erickson (1991) and Keeves and Sowden (1992), a qualitative approach is useful for revealing complexities and a realistic picture of reality, deemed necessary to investigate a complex phenomenon such as washback. As far as the previous literature on washback is concerned, qualitative approaches (see Damankesh & Babaii, 2015; Zhan, 2009; Zhan & Andrews, 2014) seem to be used less commonly than quantitative or mixed methods approaches

due to its limitations in terms of the feasibility of gathering data and dealing with a large number of participants.

3.6.3 Mixed methods approach

The primary aim of this study was to gather data on the washback effect of the MUET and its relation to students' perceptions and language learning strategies among Malaysian pre-university ESL learners, as well as those in the early stages of university study. Given the complex nature of the washback phenomenon, it was necessary to examine both the teaching and learning of English language courses in the research context. Creswell (2014, p. 4) defined mixed methods as “an approach to inquiry involving collecting both quantitative and qualitative data, integrating the two forms of data [to provide] a more complete understanding of a research problem than either approach alone”. The use of a mixed methods approach makes the findings more valid and reliable (Mertens, 2010). According to Tsagari (2006), to enhance the validity of washback research, it is preferable to employ more than one method. A mixed methods approach not only enables the researcher to gain a more comprehensive picture of test washback, but also provides a richer and more rigorous data set than afforded by either quantitative or qualitative methods alone. Due to this, a growing number of washback studies have exploited a mixed methods approach to collect data (see Cho, 2004; Munoz & Alvarez, 2010; Pan & Newfields, 2011, 2012; Qi, 2005, 2007).

As suggested by Denscombe (2014), a mixed methods approach enables the triangulation of data to minimize the biases and drawbacks inherent in each data collection instrument; thus, the data will complement each other and provide a more complete picture of the washback phenomenon. Similarly, Creswell (2014) noted that since all methods have limitations, using mixed methods will enable the biases inherent in one method to be neutralized or cancelled by the

biases of another method. Combining quantitative and qualitative approaches potentially allows the quantitative data to be validated and proven in context by the qualitative data obtained, for example, from in-depth interviews or classroom observation. Moreover, students' opinions and perceptions, together with teachers' teaching behaviours, are better understood through a qualitative approach. Although it can be argued that qualitative data lack standard procedures for data collection and interpretation and are more prone to subjectivity and bias, complementing these with quantitative data that are based on large samples and statistical significance make it possible not only to capture a global picture of the phenomenon under investigation, but also to attain a grasp of participants' depth of feelings and thoughts (Mertens, 2010). Furthermore, this approach allows the researcher to look at a given phenomenon from different angles (Denscombe, 2014). Stronger conclusions can be drawn at the end of the study if data from various sources point to the same findings (Cohen et al., 2011).

3.6.4 Methodological justification for the approach based on prior washback studies

The washback effect is multi-faceted, involving the interaction of several independent and intervening variables beside the test itself. Such variables include the stakeholders (i.e. teachers, students, test designers, etc.), the stakes of the test, classroom conditions and curriculum resources, the management of practices in the institutions and the socio-political context in which the test is put to use (Tzagari, 2009, p. 6). As outlined in the literature, a wide range of research methods has been used to elicit data concerning washback on the process of teaching and learning involving various stakeholders. Questionnaires, interviews, classroom observations, diaries and document analyses are among the research tools commonly used in washback studies. Wall and Alderson (1993) highlighted the importance of using multiple research tools to collect data, as they allow

researchers to paint a complete picture of the washback phenomenon. In agreement with this, Cheng (2005, p. 67) argued that “survey data alone are useful but insufficient for understanding washback”. Including observational data can help researchers validate the responses obtained from the respondents through questionnaires and interviews, also referred to as data triangulation. This enables the researchers to make sense of or “be sufficiently critical of the answer they are given” (Wall & Alderson, 1993, p. 65).

Alderson and Wall (1993) classified the methods used in washback studies into two categories: (i) direct methods, such as interviews, observations and document analysis; (ii) indirect methods, such as questionnaires, tests and diaries. The findings from washback studies that have only employed indirect methods to elicit data (see Gebril & Brown, 2014; Jager, Maag Merki, Oerke, & Holmeier, 2012; Pan, 2014; Putwain, 2008; Ren, 2011) are deemed questionable, taking into consideration the drawbacks of self-reported data. Thus, to obtain a clearer picture of the washback phenomenon that might be revealed using only indirect methods, more recent studies have used both indirect and direct methods simultaneously (see Allen, 2016a,b; Fan et al., 2014; Khodabakhshzadeh, Zardkanloo, & Alipoor, 2017; Luong-Phan & Effeney, 2015).

This study employed a quantitative approach, making use of numerical quantification and statistical procedures to assess the participants’ overall perceptions of the impact of the high-stakes English language test on their language learning strategies. Bearing in mind the limitations of using a questionnaire in terms of the oversimplification of the findings and poor ecological validity (Cheng, 2005), interviews with the learners and classroom observation were also conducted for data triangulation purposes. Based on the above justification, the instruments developed for this study are discussed in the following section.

3.7 Research instruments

The use of both quantitative and qualitative data gathering techniques can assist researchers in shedding light on different aspects of the same issue and provide a more complete picture (Denscombe, 2014). Furthermore, data triangulation can increase the overall reliability of the research process, since more information is obtained. As this study aimed to generate data on washback length, data from students who had already sat the MUET were also collected. To provide multiple perspectives, data were gathered using a student questionnaire, individual telephone interviews with students, a teacher questionnaire and classroom observation. The use of the various data collection methods in a process of data triangulation can increase the truth value of the findings if each method yields similar results. In what follows, detailed accounts of the instruments developed in this study are presented.

3.7.1 Quantitative data collection

A questionnaire was used as the instrument to collect the quantitative data for this study. According to Langdrige (2004, p. 67), questionnaires are “useful if you want to know something about the opinions, beliefs or attitudes of large numbers or groups of people”. In the field of washback, questionnaires have been and are still very popular as a means of data collection among researchers (see Akpınar & Cakildere, 2013; Green, 2006, 2007b; Pan, 2014; Ren, 2011; Xie & Andrews, 2012). Thus, for this study, two versions of student questionnaires and one teacher questionnaire were developed. The items and constructs from relevant established questionnaires were adapted for the student questionnaires and the teacher questionnaire, as these ready-made questionnaires have been tested for validity and reliability, making it less time consuming to construct appropriate questionnaires. The following sub-sections discuss the design of these questionnaires.

- **Student questionnaires**

The aim of the student questionnaires was to explore perceptions of the MUET and potential learning strategies pattern that could be linked to the washback effect of the test. The data were collected from the two groups of participants, Group A, preparing to take the test for the first time, and Group B, who had already taken the test (see 3.4.2). Thus, to collect the data, two student questionnaires were prepared: Student Questionnaire A for Group A and Student Questionnaire B for Group B. For clarity, Group A comprised the main participants in the study. Group B was included for the sole purpose of understanding the length of the washback from the MUET.

Student Questionnaire A

Student Questionnaire A (see Appendix A) consisted of two main sections. The first section covered demographic questions, specifically gender and English proficiency levels. Section 2 comprised three main sub-sections, dealing with students' perception of (1) the test (MUET), (2) teaching activities and (3) language learning strategies.

All the items in Section 2 were assessed using a Likert-type scale, commonly used to measure perceptions, attitudes, preferences, level of agreement, etc., by asking participants to respond by selecting an option within a given range to a particular question or statement (Cohen et al., 2007; Sullivan & Artino, 2013). The more response options provided to the respondents, the greater the information that can be obtained concerning a specific item. For instance, a dichotomous item is only able to indicate the direction of the respondent's attitude (yes/no), whereas a three-point scale with a middle option enables respondents to indicate neutrality in their response.

As noted by Cummins and Gullone (2000), increasing the scale further increases its sensitivity and thus a five-point Likert scale or more tends to be preferred. Revilla, Saris, and Krosnick (2014) conducted a study to see if a five-point Likert scale or more (7–11) would be better in terms of agree-disagree (AD) rating scales and concluded that “despite what information theory states, there is no gain in information when an AD scale with more than five categories is used. There is, instead, a loss of quality” (p. 90). Therefore, a five-point Likert scale was chosen where deemed appropriate for the questionnaires in this study. A three-point scale was also used in the questionnaire, but only on one item for which the students were required to indicate the level of importance.

The Likert-scale categories used in this study were predominantly as follows: 1 = strongly disagree, 2 = disagree, 3 = undecided, 4 = agree and 5 = strongly agree. The “undecided” response option is typically included in Likert-type scales to minimize cases of missing data. This covers respondents who are unsure, or have no idea what the item is about, enabling them to choose this option as opposed to leaving the item blank. Kent (2015) categorized mid-point or neutral answers as a “valid indicator of the absence of attitudes, beliefs, opinions, or knowledge” or “inaccurate reflections of existing cognitive states” (p.57). Although researchers have argued that using neutral midpoint option in a questionnaire could tempt respondents to choose this category, Borgers, Hox, and Sikkel (2004) found a positive effect of introducing a mid-point option in questionnaire scales as it produces a larger relative difference. According to Schuman and Presser (1996, pp. 113–114), “To virtually measure any attitude, opinion, or belief question in a survey, a possible reply is ‘I don’t know’ [and therefore] respondents should be allowed, perhaps even encouraged, to see DK (don’t know) as a legitimate response”. In the case of this study, the possibility that students might not have an answer to the items could not be dismissed. Hence, it was decided to provide a neutral

mid-point option for most of the items. Where possible, qualitative data were used to provide further explanation in the case of non-committal findings from the questionnaire.

Once the format of the student questionnaire had been determined, the items were developed according to the pre-established variables in the study, namely students' perceptions of the MUET, self-efficacy and language learning strategies employed in preparing for the test. The items were selected and developed for each construct based on Czaja's (1998) three basic guidelines for developing a questionnaire: (i) whether respondents understand the words and terms used; (ii) whether the items developed are commonly understood by all respondents; (iii) whether the questions are sufficiently interesting for the respondents to answer. For this study, the development of the student questionnaire was generally based on Green's (2007a) study on the washback of the IELTS writing test as his was one of the main washback models used in this study, together with items adopted and adapted from other studies presented later in this section addressing individual parts of the student questionnaire.

Based on Bachman and Palmer's (1996) argument, this study specifically considered students' perceived consequences of the test as one of the possible contributory factors concerning the washback effect of the MUET. In this study, the main focus of investigation was on the learners. Thus, washback in this study mainly referred to the influence of the test on the learners and their language learning strategies, positive and/or negative, weak and/or strong, short-term and/or long-term, specific to the MUET. (These washback dimensions are discussed in greater detail in section 2.4.) The student questionnaire began with items pertaining to students' perception, comprising three subscales: (i) perception of test importance, (ii) perception of self-efficacy and (iii) perception of test difficulty. The perceived test importance scale asked students to evaluate the MUET in relation to their attainment, its instrumental value and its intrinsic value.

There were four items ($\alpha = .833$), two of which were drawn from the previous study conducted by Xie and Andrews (2012) on the relationship between test design and test uses concerning students' test preparation in Hong Kong. The overall structure of Student Questionnaire A is provided in Table 3.3, together with sources of items.

Table 3.3

Overall structure and sources of the student questionnaire

Section	Source
Section 1	
Background information	
Gender	
English language proficiency	
Part 1 : Perception of the MUET	
Importance – 4 items:	Xie and Andrews (2012)
Item 1 – The MUET is an important test for me.	
Item 2 – It is very important for my future undertakings that I do well in the MUET.	
Item 3 – If I do poorly in the MUET, my chance of getting into a top university will be affected.	
Item 4 – If I do poorly in the MUET, my chance of enrolling in my desired course will be affected.	
Self-efficacy – 4 items:	Pintrich et al. (1991)
Item 5 – I believe I will receive an excellent grade in the MUET.	
Item 6 – Taking into consideration its difficulty, I think I can perform very well in the MUET.	
Item 7 – Taking into consideration my ability, I think I can perform very well in the MUET.	
Item 8 – I'm confident I can do an excellent job on the assignment and tasks in the MUET class.	
Test difficulty – 1 item:	Ribeiro and Yarnal (2010)
Item 13 – On a scale of 1 to 5, how difficult do you think the MUET is for you?	
Perceived importance of skills – 4 items:	
Item 14a – Listening	
Item 14b – Speaking	
Item 14c – Reading	
Item 14d – Writing	
Part 2: Teaching activities – 4 items:	Cheng (2005)
Item 15 – Organize group work or discussion.	
Item 16 – Do mock exam like activities.	
Item 17 – Discuss textbook exercises.	
Item 18 – Organise real life language activities (e.g. mock interview, sketches, etc.)	
Part 3 : Language learning strategies – 50 items	Green (2007a)
Memory – Items 19–27	Oxford (1990)
Cognitive – Items 28–41	
Compensation – Items 42–47	
Meta-cognitive – Items 48–56	
Affective – Items 57–62	
Social – Items 63–68	

The Perceived Difficulty Assessment Questionnaire (PDAQ) developed by Ribeiro and Yarnal (2010) was adopted to assess students' views of the difficulty of the MUET. As this study sought to explore if self-efficacy plays an important role in determining the washback of the MUET, the relationship between self-efficacy and the language learning strategies employed by the students in preparing for the MUET was also investigated. Four items from the self-efficacy scale of the Motivated Strategy Learning Questionnaire (Pintrich et al., 1991) were modified to suit the context of the study ($\alpha = .869$), as presented in Table 3.3. All items were scored on a five-point Likert-type scale, except for those on the perceived importance of skills, which employed a three-point scale.

As this study also took into account teachers' views and perspectives for the purposes of data triangulation, it was considered necessary to collect data on the students' views of their teachers' behaviours in the class. According to Cheng (2005), to explore the type(s) of language learning activities that students undertake in the classroom, they need to be asked about the tasks teachers assign them. Learning opportunities in the classroom are typically provided by the teachers and students generally have minimal control over the lesson. Hence, asking students to identify what kinds of language learning activities are conducted in the classroom will provide beneficial data. In this study, the students were therefore asked about the kinds of activities that their teachers implemented in the English language classroom. For this aspect, four out of ten items ($\alpha = .830$) dealing with classroom teaching and learning activities derived from Cheng's (2005) student questionnaire, which was designed and validated specifically for a washback study in Hong Kong, were chosen based on their relevance to the study. A five-point Likert-type scale was used to measure the perceived frequency of activities (1 = never, 2 = rarely, 3 = sometimes, 4 = often and 5 = always). There was a concern, however, that terms such as 'often' and 'always' might be

interpreted in different ways by respondents. Since the questionnaire was fully translated into the Malay language, which is the native language of the respondents involved in this study, the issue with misinterpretation could be minimized. Furthermore, the students were briefed first on each section of the questionnaire before they started filling them in.

Since the focus of this study was on the washback process as opposed to washback product, it was deemed appropriate to explore students' use of language learning strategies rather than looking at their MUET scores. In terms of language learning strategies, Oxford's (1990) Strategy Inventory for Language Learning (SILL), a widely used and reliable research instrument (Oxford & Burry-Stock, 1995) was adopted. This scale comprises 50 items, divided into six categories:

- i. Memory strategies, such as grouping, imagery, rhyming and structured reviewing (9 items).
- ii. Cognitive strategies, such as reasoning, analysing, summarizing (all reflecting deep processing), as well as general practice (14 items).
- iii. Compensation strategies (to compensate for limited knowledge), such as guessing meanings from the context in reading and listening and using synonyms and gestures to convey meaning when the precise expression is not known (6 items).
- iv. Metacognitive strategies, such as paying attention, consciously searching for practice opportunities, planning language tasks, self-evaluating one's progress and monitoring errors (9 items).
- v. Social strategies, such as asking questions, cooperating with native speakers and becoming culturally aware (6 items).

(Oxford & Burry-Stock, 1995, p. 5)

When adopting a well-established instrument, Shih (2013) recommended ideally retaining all the items in the scale. Hence, for the main study, all 50 items from the SILL (Oxford, 1990) were employed to ensure reliability. The original SILL scale is rated on a five-point Likert-type scale (1 = never or almost never true of me, 2 = generally true of me, 3 = somewhat true of me, 4 = generally true of me, 5 = always or almost always true of me). However, Green (2007a) changed the wording of the scale in his washback study to indicate frequency rather than agreement as his intention was to identify how often the strategies were used. For this study, Green's (2007a) approach was adopted for the same reason (1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = always).

Student Questionnaire A was then piloted with 30 students preparing for the MUET, but who were not involved in the main study. The purpose of piloting the instrument was to test the validity and the reliability of the instrument, as the data from the questionnaire were used to measure the key variables in the study. From conducting the pilot study and analysing the data, some crucial changes were made to the student questionnaire. For the pilot study, most of the items in each construct were adopted from well-established questionnaires that have been validated and tested for reliability. However, in order to reduce the total number of items in the questionnaire and to reduce survey fatigue, only several items were selected from the scale as opposed to adopting the entire scale. This caused a problem when reliability analysis was undertaken for each construct as the Cronbach alpha value for some of the constructs were quite low. For the language learning strategies scale, which was adopted from the SILL (Oxford, 1990), the analysis for each sub-scale (six sub-scales altogether) of the language learning strategies was not carried out as well because the language learning strategies scale adopted in the pilot study was not taken as a whole

scale. Due to this, for the main study, the entire scale was adopted to enable a deeper and more thorough analysis on the language learning strategies and to ensure reliability of each scale.

Student Questionnaire B

For Group B participants, a different set questions (see Student Questionnaire B in Appendix B) was developed with a similar format to Student Questionnaire A. This comprised two sections covering the demographic characteristics of the participants (Section 1) and washback length (Section 2). Since the sole purpose of including Group B in this study was to measure the washback length of the MUET, data on their perceptions of the MUET were not collected. This major decision was made after the questionnaire was piloted and it was deemed unnecessary to collect data on Group B's perceptions of the test since it was only used to elicit data on the washback length of the MUET. To the best of my knowledge, there is no existing instrument, either quantitative or qualitative, for collecting data on this aspect of washback, perhaps because of the potential for mediation by other intervening variables and the difficulty of access to participants for longitudinal study.

Therefore, this study again employed the SILL (Oxford, 1990), but with a major change in the scaling of the items. As measuring washback length requires that data be collected both before and after the students sit the test to establish any changes over time, the participants from Group B was asked to evaluate the items in the SILL twice, first in terms of the strategies they used when they were preparing for the MUET and second concerning those they currently used for language learning. For Student Questionnaire B, the scale used was dichotomous, with 1 indicating "True" and 2 indicating "Not true". The reason for employing dichotomous items rather than a five-point Likert-type scale was to lessen the student's cognitive load, since the SILL comprises 50 items.

One additional open-ended question was added: *Is there any difference between your current English language learning strategies and the strategies that you used when you were preparing for the MUET in the past?* This open-ended question was added to supplement the quantitative data on washback length and provide more in-depth insight. Adding such a question made it possible to collect qualitative data from a considerable number of participants in a short amount of time. Similar to Student Questionnaire A, Student Questionnaire B was then piloted with 46 students who had already sat the MUET, but were not involved in the main study.

Translation procedures

Both the student questionnaires were in English, with items also translated into Malay to ensure that the respondents would be able to complete it. When piloting the instruments, the questionnaires were written in both English and Malay and at the end, the respondents were asked to choose their language preference. This helped determine the language to be used for the main study; it was not feasible to use both languages as it would make the questionnaire very long, potentially hampering respondents' motivation to respond and causing survey fatigue. Indeed, some of the respondents in the piloting for Student Questionnaire B commented that the questionnaire was rather too long and they were bored. Furthermore, using only one language minimized the need for translation, which might give rise to misinterpretation, particularly due to cultural differences. Hence, for the main study the questionnaires were only provided in Malay. Malay is the national language and the first language of the participants and thus using it in the questionnaire would reduce the cognitive load of the questions and overcome the drawbacks of an otherwise lengthy questionnaire. All items in the student questionnaires were composed in English first and then translated into Malay.

According to Sperber, Devellis, and Boehlecke (1994), many have taken translation in cross-cultural research for granted, especially when it comes to using previously validated instruments in another target culture and language. As pointed out by Sperber (2004), it is common practice for researchers to take a questionnaire translated by unqualified translators and use this version without validating it first. Researchers should be aware that when it comes to translation, there is a possibility that translators do not necessarily have knowledge of the specific content area of the instrument. Hence, there is a risk of inaccurate translation of items due to specific cultural gaps. This could then lead the respondents to misinterpret the items or the questions asked in the questionnaire, jeopardizing the overall findings of the study.

Thus, for this study, the back-translation method was employed, as suggested by Sperber (2004). This technique involves a translator rendering the questionnaire in the target language, then this being translated back into the source language by another translator blinded to the original questionnaire. The final step is to compare the two source language versions of the questionnaire. Once translated, the questionnaires were then e-mailed to peers, who were doctoral students from Malaysia and English language lecturers in Malaysia, for validation. Based on the comments and suggestions from this team of experts, several amendments in terms of the language structure and the format of the questionnaires were made before finalizing them for the main study.

Data collection procedures

The student questionnaires were administered to the two groups of students (Group A and Group B). As it was necessary first to get consent from the respective teachers and lecturers who were involved in this study, the exact number of questionnaires was distributed according to the number given by the teachers and the lecturers, making the return rate 100%. In total, Student

Questionnaire A was distributed to 137 students in Group A and Student Questionnaire B was distributed to 238 students in Group B. The medium of distribution was a paper-based questionnaire and the administration took place towards the end of the school term for Group A and halfway through the semester for Group B, which was after the students had taken the MUET and received their results. The students took 15 to 25 minutes to answer the questionnaires and they were collected on the same day. Data from the questionnaires were transferred into the Statistical Package for Social Sciences (SPSS) Version 23.0 for analysis.

- **Teacher questionnaire**

Although the main focus of this study was on the students, data from teachers were also collected for triangulation purposes. The aim of the teacher questionnaire was to obtain information on teachers' perceptions of the MUET and their attitudes towards aspects of learning in terms of preparing the students for the test. Data on teachers' medium of instruction were also collected, specifically concerning their lesson preparation, teaching activities and teaching materials and resources for the MUET preparation class. The teacher's questionnaire (see Appendix C) consisted of two sections: Section 1 relating to objective questions and Section 2 concerning subjective questions. All the items for Section 1 were adapted from Cheng's (2005) study on the phenomenon of the washback effect in Hong Kong, which were then modified to suit the context of this study. The slight modifications only involved changing the name of the test from the Hong Kong Certificate of Education Examination (HKCEE) to the MUET and certain questions, for example "What are the learning strategies you would recommend to your students in the context of the new 1996 HKCEE?" to "What are the learning strategies you would recommend to your students to

prepare for the MUET?” Other elements of the questionnaire, such as the format and the scales, were retained.

Similar to Cheng’s (2005) teacher questionnaire, Section 1 comprised three parts. The first part dealt with the demographic characteristics of the teachers, namely gender, number of years they had been teaching and the number of periods they taught English per week. Section 2 comprised 42 items within 6 categories addressing teacher’s perceptions of teaching, learning and assessment in relation to preparing students for the MUET. A five-point Likert-type scale was used (5 = strongly agree, 4 = agree, 3 = undecided, 2 = disagree, 1 = strongly disagree). In Section 3, three out of Cheng’s (2005) ten categories were adopted from the original questionnaire. These comprised 24 items altogether, which addressed aspects of classroom teaching and learning in the MUET preparation class, scored using a five-point Likert-type frequency scale (1 = never, 2 = seldom, 3 = sometimes, 4 = often, 5 = always).

As no interviews were conducted with teachers due to time constraints and the scope of the study, four open-ended questions were adopted from Takagi’s (2010) study on English language entrance examinations to Japanese universities. The open-ended questions were intended to enable the teachers to express and elaborate their views and ideas clearly with regard to the MUET. The teachers were given the option to submit the questionnaire with or without answering the open-ended questions. The four open-ended questions were as follows:

1. Does the MUET influence the way in which you teach your English classes? (If yes, how? If no, why not?)
2. Do you think the MUET assesses your students’ English ability appropriately? (If yes, how? If no, why not?)
3. Do you think that the MUET is necessary for university entrance? (If yes, how? If no, why not?)

4. Do you think that the MUET influences the future of Malaysian students? (If yes, how? If no, why not?)

Data from Question 1 were used to support the quantitative findings concerning teachers' medium of instruction in terms of lesson preparation and teaching activities, while the data from Questions 2, 3 and 4 provided qualitative findings concerning the teachers' perceptions of the MUET, not covered by the quantitative data. The teacher questionnaire was not piloted since it was adopted from well-established instruments from two comprehensive washback studies which are Cheng's (2005) and Takagi's (2010).

An online questionnaire was used to reach as many teachers as possible within a short amount of time. Qualtrics was used to distribute the online questionnaire. The online teacher questionnaire went live on 4 January 2017 and was taken down on 11 April 2017. A total of 55 teacher questionnaires were distributed and 36 were used for analysis due to missing data. The data from the open-ended questions could then be linked to the classroom observation data for validation purposes (see 3.7.2).

3.7.2 Qualitative data collection

Telephone interviews with the students and classroom observations were used to collect qualitative data for the study, as detailed in the following sub-sections.

- **Interviews**

Qualitative data were used as a "subsidiary counterbalance" in this investigation, providing a detailed and comprehensive understanding of the students' perceptions generated from their beliefs, views and experiences. This study used individual telephone interviews with students to

gather qualitative data. The aim of these was to elicit in-depth data from the students to complement the findings from the student questionnaire concerning their perceptions of the MUET, their self-efficacy and the language learning strategies they employed when preparing for the test. As pointed out by Atkins and Wallace (2012), interviews not only allow researchers to engage with the participants individually, but also enable them to collect various types of in-depth data, for example factual data, views and opinions, personal narratives and histories. When participants engage with researchers in interview sessions, there is the opportunity to probe and clarify responses obtained through other means of data collection. This can also be done after interview recordings have been transcribed. Respondents will then be given the opportunity to read the interview transcripts, giving their approval and agreement.

Although the initial plan was to use focus group interviews for the student participants, as this would enable them to listen to alternative points of view, disagree or agree and expand on their responses (Denscombe, 2014), individual telephone interviews were used instead due to several constraints. First, in terms of practicalities, researchers need to decide the time and venue for the interviews to take place (Runswick-Cole, 2011) and this was quite a challenge, especially with regard to the pre-university students (Form 6) as their timing was not very flexible; they were expected to abide by the rules of the school, which meant they were only available after school hours. Individual telephone interviews provided greater flexibility in terms of scheduling, thus not interrupting the students' learning. As far as conducting the individual interview via telephone is concerned, there was no technical difficulties with using two mobile phones as interview devices during the pilot study. The quality of the recorded phone calls was good and I could transcribe the recordings with ease even though those calls were made all the way from the United Kingdom to Malaysia.

Two-way communications with another human being entail potential risks associated with human interactions, for example embarrassment, anger, violation of privacy, misunderstandings and conflicts concerning opinions and values (May, 1991). Researchers need to be aware of these, but using telephone interviews might reduce such risks, enabling respondents to be more relaxed and able to share their opinions more openly taking into consideration the nature of students in the Malaysian context who are known to be more on the 'reserved' side. Although it cannot be denied that physically sitting with the respondents would foster a sense of trust and reduce the potential for misunderstanding, this is normally not the case when it comes to students in the Malaysian context. Anonymity is actually an advantage in this case.

The nature of focus group interviews gives the researcher/moderator less control over the data produced compared to individual or one-to-one interviews (Morgan, 1988). As the participants in focus group interviews are able to interact with each other when expressing opinions and doubts on the topic being discussed, there is a possibility that they will not express their own definitive individual views. Their responses might be influenced by other respondents' answers as they are speaking in a specific context, within a specific culture. Moreover, focus group interviews may discourage students who are not very confident or shy from taking part (Gibbs, 1997). Hence, for this study, individual interviews were deemed to be a more suitable approach.

Nevertheless, there are also disadvantages to interviews. The trustworthiness and reliability of this data collection method has been debated as there is no guarantee that the interviewees or respondents will tell the truth and not succumb to perceived social desirability, trying to provide the answers that they think the researcher would like to hear. One way of checking this is by having two questions that ask more or less the same thing but in different ways. The answers given to these questions can be used by the researcher to cross-check for consistency, enabling the

researcher to evaluate whether the data are trustworthy and reliable. The power relationship between the researchers and the interviewees can also be a threat to validity when it comes to interviews. For example, if interviewees feel intimidated or shy, they might not be able to communicate their opinions or thoughts freely. According to Atkins and Wallace (2012), this shortcoming when dealing with power relationships can be overcome by choosing an appropriate setting for the interview to take place. A less formal setting with a relaxed atmosphere will put the interviewees at ease, making them feel that they are participating in a conversation rather than a confrontation or an interrogation. Hence, structured individual telephone interviews were adopted as an appropriate method with the students for this study. This meant that the students could be in a place of their choosing. Also, as the interviewees were high-school students with limited flexibility in terms of time, making face-to-face interviews difficult to arrange. Moreover, the interviews could be conducted even when I was based in the UK. The telephone interviews took around 20–30 minutes each, which was shorter than expected.

For the student telephone interviews, eight open-ended questions with several probing questions on students' perceptions and experiences of preparing for the MUET were prepared (see Appendix D). The questions were prepared in line with the framework adopted for this study. The respondents were asked to describe their perceptions of the MUET pertaining to its importance, its difficulty and their self-efficacy. They were also asked to share their learning experiences and the language learning strategies they used when preparing for the MUET. Questions pertaining to their perceptions of the consequences of the MUET were also asked during the telephone interviews.

The interviews with students were conducted after the administration of the questionnaires, in which those who were interested in taking part in the telephone interview were asked to provide their contact details, including a pseudonym and telephone number. Although this meant some of

the questionnaires were not entirely anonymous and risked that the data would be traceable, the name given was only for the purpose of addressing the participants during the individual phone interview. The use of a pseudonym instead of their actual name aimed to reduce the risk of identifying participants. It was also clearly stated in the student questionnaire consent form that all information would be anonymous and treated in the strictest confidence for research purposes only. The data published in this study would not include information making it possible to identify anyone individually. According to Sudman and Bradburn (1974), maximizing subject anonymity can also help reduce the possibility of social desirability bias, such that participants are inclined to give answers that they consider socially acceptable as opposed to the truth as they see it. The student interviews were conducted with Group A only as they were the main focus of this study.

One of the reasons for conducting telephone interviews was to mitigate shyness or reluctance to respond. Moreover, the interviews were conducted in English and/or Malay, according to the participants' preferences. This was to enable them to express their thoughts and opinions fully as they were using the language in which they were comfortable. However, at least for some of the participants, the telephone interview method appeared not to be engaging. There were also some unavoidable technical issues with this method of interview, namely the difference in time zone (8 hours between Malaysia and the UK) and the poor quality of audio recordings from the phone conversations. The first problem was addressed by making arrangements in advance with the participants and good transcription software helped to improve the quality of the audio recordings. The interviewees' demographic information is presented in Table 3.4.

Table 3.4

Student interview participants

Pseudonym	Stage of Study	Total
Haslinda	Form 6	11
Husaini	Form 6	
Irdina	Form 6	
Shahirah	Form 6	
Umi	Form 6	
Maisarah	Form 6	
Azleen	Form 6	
Hidayah	Form 6	
Nazeerah	Form 6	
Syamimi	Form 6	
Marziana	Form 6	

According to Ary, Jacobs, Irvine and Walker (2013), there is no general rule in determining the number of participants for the purpose of collecting qualitative data. Lincoln and Guba (1985, p. 202) stated that when there is redundancy in the information gathered (saturation point), sampling should be terminated as no new information is forthcoming.

All the interviews were recorded with an audio recorder to ensure thorough data collection. The audio recordings were transcribed using the ExpressScribe software before exporting the data to NVivo 12.1 for analysis.

- **Classroom observation**

Washback researchers have highlighted the need for triangulation of data, gathered through both quantitative and qualitative methods from various data sources (Wall & Alderson, 1993; Watanabe, 2004). Combining the views of the main stakeholders involved in a test together with the researcher's interpretation supported by the literature is also suggested as part of data

triangulation. In particular, it has been recommended that data on stakeholders' perceptions be supported by direct observation of behaviour in the classroom.

Observation "offers an investigator the opportunity to gather 'live' data from naturally occurring social situations" (Cohen et al., 2011, p. 456). Alderson and Wall (1993) argued that although questionnaires and interviews provide insights into how participants believe they have been affected by a test, direct observation in the classroom can provide a corrective to potentially misleading questionnaire and interview data, in particular whether teachers' and students' responses in questionnaires and interviews are reflected in their behaviours. It can also contextualize otherwise incomprehensible responses, making the findings more valid and reliable. Ren (2011) added that classroom observation allows the researcher to capture the manners in which washback operates more accurately than relying on the information given by respondents by observing it directly. However, the review of the literature showed that few washback studies have elicited data through classroom observation, despite this being highlighted as one of the key points when researching washback.

The washback literature suggested from the start that there would be many intervening factors that interact in teaching and learning as a result of a test, especially a high-stakes test. Thus, this study employed the well-established Communicative Orientation of Language Teaching (COLT) observation scheme, developed by Spada and Frohlich (1995). COLT was chosen for this study as it focuses on teachers' and students' behaviour and interaction in the classroom (Allen, Frohlich, & Spada, 1983) and has been widely used in previous washback studies (Barnes, 2010; Burrows, 1998; Cheng, 1997, 2005; Green, 2007a; Read & Hayes, 2003). It consists of two parts, the first describing classroom events at the level of activity (Part A) and the second (Part B) addresses verbal exchanges between teachers and students or among students themselves as they

occur within each activity (Spada & Frohlich, 1995). Only Part A of the COLT (see Appendix E) was used, as the language used in the classroom was beyond the scope of this study. The observation scheme encompasses five main categories: time, participant organization, activity type, content and material used.

Five MUET preparation classes were observed at two high schools that provide Form 6 education. Structured classroom observation was carried out as it eases the task of observing and makes recording the data very much easier and more systematic than with no scheme. Consent from teachers to videotape their teaching sessions was also obtained. Bearing in mind the possibility that there might be a problem with the video recordings, audio-recording and field notes were employed alongside the observation scheme. For each classroom observation, a video recorder was set up at the back of the classroom to minimize any disruption and disturbance to the lessons, as well as to ensure that teacher and student interaction would be as natural as possible. In addition, I was not present in the classroom throughout the lesson, only at the beginning and at the end to set up the video camera and to give a short briefing to both the teachers and the students. Although there was a potential concern in terms of capturing the students' and the teachers' faces on the video, on viewing the recordings it was apparent that the visual was not sufficiently large to reveal the participants' identity. Having processed the recordings, I watched the video and recorded the data in the COLT observation scheme for further analysis.

3.8 Data analysis

3.8.1 Quantitative data analysis

The Statistical Package for Social Sciences (SPSS) Version 23.0 was used to analyse the quantitative data. Data from the questionnaires were input and analysed using descriptive and

inferential statistics. The analyses involved frequency distributions and item analysis for descriptive statistics and correlation analyses, as well as Kruskal–Wallis, Mann–Whitney U and McNemar tests and regressions for inferential statistics.

3.8.2 Qualitative data analysis

For the interview transcripts, content analysis was carried out based on the themes listed in the framework of this study (see 2.8), namely the students' perceptions of test importance, test difficulty and self-efficacy and their language learning strategies when preparing for the MUET. The first step taken in analysing the interview transcripts was to organize the data using manual coding in NVivo 12.1. This involved assigning the data category codes and labelling them. Relationships and patterns between the data were also recorded. Next, I conducted focused coding, which involved scanning the data for units of meaning – words, phrases, sentences, or respondents' ways of thinking – that were consistent with the aforementioned established themes. The numerous category codes assigned earlier were either eliminated, combined or subdivided. Repeating ideas were also addressed. The remaining codes were arranged into tables according to the themes assigned before the interviews. This arrangement allowed effective comparison of data between different participants and analysis of data both within and across categories. The qualitative data from the interview were then used to complement, support and explain the quantitative findings.

For the structured classroom observation, to enhance the reliability of the data, the recordings of the lessons, both audio and video, were recoded at least three months after the observation was carried out. This aimed to identify if there were any missing or unnecessary elements that needed to be added or discarded from the first manual coding. It also helped to ensure consistency in the coding. The data were then compared to ensure that the coding was consistent

with the criteria for analysis. A measure of inter-rater reliability was used where fellow doctoral colleagues were invited to be second coders, examining a sample of the classroom observations. For this purpose, two other coders, who were experienced researchers in the field of education were invited for their assistance. These independent coders were provided the relevant literature and briefed on the procedures and the instrument that was used to elicit the classroom observation data for this study. A copy of the SILL was given to each coder and they were requested to match participant behaviours from the video recording with the categories in the Communicative Orientation of Language Teaching (COLT) observation scheme. Once their observations were coded, the coders together with me compared our coding. In the event of a mismatch, the particular coding was scrutinized further until a common consensus was reached. All participant behaviours were coded in such a manner.

CHAPTER 4: RESULTS AND ANALYSIS

This chapter is divided into two main sections according to the participants, the students (Section I) and the teachers (Section II), to answer the three research questions, as follows:

1. What are students' perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?
2. How does washback of the MUET operate and to what extent do students' perceptions seem to have a washback effect on their language learning strategies?
3. What is the intensity of the washback effect of the MUET and what is its length? How do these appear to influence washback on the learners?

Both quantitative and qualitative research instruments were employed in data collection for triangulation purposes. The quantitative research instruments were the student questionnaire and teacher questionnaire; for qualitative data, classroom observation, student interviews and written responses from the open-ended questions in the teacher and student questionnaires were used. Two groups of students were involved, Group A (those who were preparing for the MUET) and Group B (those who had already sat the MUET). As previously noted in Chapter 3, this study was not comparative in nature. In what follows, the overall results from the questionnaire are reported first, followed by the data gathered through qualitative instruments.

SECTION I: WASHBACK FROM THE MUET ON THE STUDENTS

This section presents the findings from the students' data. The findings for Group A are addressed first, including students' perceptions of the MUET, followed by the relationships between these

perceptions and their English language learning strategies. This answers the first and second research questions. The findings for Group B students are then considered, exploring the washback intensity and length of the MUET, which covers the third research question.

4.1 Students' perceptions of the MUET

This section explores how the students perceived the MUET and the factors contributing to their perceptions. In this study, the students' perceptions of the MUET were examined in terms of perceived test importance, perceived test difficulty and perceived self-efficacy in relation to their performance in the MUET.

To analyse the quantitative data, descriptive analysis was used to report the frequency and percentage of the students' perceptions. The demographic aspects of the respondents considered, gender and English language proficiency (see 3.4.2), were cross-tabulated with their perceptions of the MUET before undertaking inferential analysis to see if there were any significant differences between the variables. Differences in the students' perceptions across gender and English language proficiency were tested for statistical significance using the Mann–Whitney U test, a non-parametric test, as the data were not normally distributed. A probability of less than 0.05 was taken as statistically significant.

Thematic analysis was used to analyse the qualitative data, based on the pre-established categories from the framework of this study and similar to the categories use for analysis of the quantitative data: (i) perceived test importance, (ii) perceived test difficulty and (iii) perceived self-efficacy. In discussing the interview data, relevant supporting evidence is presented in the form of excerpts of students' comments and examples based on their own experiences. The respondents'

pseudonyms are used to ensure anonymity. All interviews were conducted in Malay and translated verbatim into English to retain the meaning.

4.1.1 Perceived importance of the MUET

As suggested by Green (2007a), an appropriate level of perceived test importance could lead to intense washback for students under certain circumstances. Items 9 to 12 in Student Questionnaire A (see Appendix A) were concerned with students' perceptions of how important they thought the MUET was for them in general and for their future undertakings, rated on a five-point Likert-type scale, with 1 indicating "strongly disagree" and 5 indicating "strongly agree". In the questionnaire, students were required to choose from the five-point Likert scale (i.e. strongly agree, agree, undecided, disagree, strongly disagree). In order to mitigate some of the individual subjectivity in the students' interpretations of, for instance, the difference between strongly agree and agree, the Likert scale was reduced to a three-point scale (i.e. agree, undecided, disagree) for ease of data interpretation and presentation. This applies to all charts for this dataset. The results for the students' perceived test importance are presented in Table 4.1:

Table 4.1

Perceived importance of the MUET

Item	N	Disagree (%)	Undecided (%)	Agree (%)
The MUET is an important test for me	137	6.6	7.3	86.1
It is very important for my future undertakings that I do well in the MUET	137	5.1	8.0	86.9
If I do poorly in the MUET, my chance of getting into a top university will be affected	137	5.1	9.5	85.4
If I do poorly in the MUET, my chance of enrolling in my desired course will be affected	137	6.6	12.4	81.0

As depicted in Table 4.1, the high overall agreement for each item, ranging from 81% to 86%, indicates that the students regarded the MUET as a very important test. This finding was expected as the MUET is a high-stakes language test with important consequences. Next, a closer analysis was carried out on each item. The students were first asked to rate the importance of the MUET in general. Analysis of each item revealed that although 86.1% (n = 118) students agreed that the MUET was an important test for them, a few students disagreed. As shown in Table 4.1, 6.6% (9) students disagreed and 7.3% (10) students were undecided when it came to the perceived importance of the MUET in general: these findings were not expected. The qualitative data revealed that one possible explanation for some students not perceiving the MUET as important for them was that they could retake the test if they wanted to improve their scores, with 4 out of 11 respondents mentioning they had the option to do so. One respondent mentioned that:

I will retake the MUET until I get a better result.

(Husaini)

Since there is no limit on the number of times they can sit the MUET, some students may not think it was important to perform well the first time. Indeed, some students might take the MUET the first time to test the water. However, this test comes with a fee, so there is a financial cost. Moreover, there are time implications and students might not be able to take the test over and over again as they will have other study commitments they need to fulfil.

Most students (86.9%, n = 119) also agreed that performing well in the MUET was very important to their future undertakings. However, there appeared to be some students (5.1%, n = 7) who disagreed that doing well in the MUET would affect their future undertakings positively. The results also revealed that 8.0% (n = 11) were undecided. Although this percentage was small, it

raised the need to explore in greater depth these students' views. One possible explanation for this finding is that the MUET result is not usually required for job applications in Malaysia; rather, it is only important for university applications. Although getting employment is commonly related to graduating from a university, some of the students did not consider the MUET beyond its face value. By relating future undertakings to their future career, those who focused solely on the MUET score in this regard, not their English language mastery, might think that the result would be of no benefit to them in terms of applying for jobs in the future.

Another way of measuring perceived test importance is to look at how the students perceived the consequences of the test. In Table 4.1, 85.4% (n = 117) of students agreed with the statement *"If I do poorly in MUET, my chance of getting into a top university will be affected"*. This was expected to be 100%, as the main purpose of the MUET is to gain entry to university. However, Table 4.1 shows that 5.1% (n = 7) disagreed and 9.5% (n = 13) were undecided. To explore this matter further, questions related to consequences of the MUET were also included in the student interview.

During the interviews, the students were asked what would happen if they were not able to score well in the test and 5 out of 11 stated that they might have to retake the MUET for the following reasons:

If I do poorly in the MUET, maybe I cannot go to university.

(Husaini)

The problem is to further my study in the university because some universities require at least Band 3 and above.

(Hidayah)

The inability to pursue study at university was the students' main concern when it came to not getting the required band on the MUET. The test result would not only influence which university they applied to, but also which course or field of study some of them would take for their first degree. This was also supported by the findings for Item 12, which asked the students if the MUET result would affect the choice of their desired course. Similar to the above findings, the majority of the students (81%, n = 111) agreed that the MUET would influence their chances of enrolling in their desired courses, followed by undecided (12.4%, n = 17) and disagree (6.5%, n = 9). Nevertheless, it has to be highlighted here that 12.4% (n = 17) were undecided and 6.5% (n = 9) thought that performing poorly in the MUET would not affect their chances of enrolling in their desired course. This finding was completely unexpected, as it is common knowledge that students have to obtain a certain MUET band to be enrolled in certain courses. The qualitative findings revealed that getting accepted on the course that they applied for was not the main priority for 6 out of 11 students:

If I did not do well in the MUET, I would just have to change my study options and find an appropriate field that is on par with my MUET result.

(Marziana)

This appeared to indicate that for these students, as long as they were able to enrol in any course at the tertiary level, that was good enough for them. Two of the interview respondents mentioned that they did not mind changing the option for their field of study because they “believed” in what the MUET result would tell them in terms of their English language proficiency and choosing an appropriate course according to their MUET band would actually benefit them:

If I obtain a low score in the MUET, maybe I'll take a different course, because I'll be calculating my own ability because I cannot handle English that much, so I'll be taking another course that can match my own ability.

(Nazeerah)

Furthermore, some universities accept students with Band 1 in the MUET for certain courses. If a student does not wish to retake the test, he or she can just opt for a course with the lowest minimum requirement of the MUET, as explained by Husaini:

Even if I get Band 1 in the MUET, I can already further my study at tertiary level. It is just that I cannot be picky with what course I want to do or which university I want to go to.

(Husaini)

To determine possible factors that could influence such perceptions, comparisons were made in terms of gender and proficiency to see if there were any significant differences between these independent variables. The Mann–Whitney U test was run to determine if there were differences in perceived test importance scores between males and females. Those for females (mean rank = 81.22) were statistically significantly higher than for males (mean rank = 54.65): $U = 3235$, $z = 3.945$, $p = .000$. The female students appeared to be more concerned about the MUET than the male students. In terms of students' proficiency level, the perceived test importance scores for low-proficiency students (mean rank = 73.68) were statistically significantly higher than for high-proficiency students (mean rank = 57.65): $U = 2394$, $z = 2.172$, $p = .030$. This indicates that low-proficiency students appeared to consider the test as more important for them.

Since the MUET tests the four language skills, it was also important to explore which language skills students perceived as important and which language learning strategies they

focused on to help investigate the nature of washback from the MUET on each language skill. The students were asked to rate which of the four language skills – reading, writing, listening and speaking – they considered important for them to perform well in the MUET, rating them on a three-point scale of 1 for “not important”, 2 for “important” and 3 for “very important”. It has to be noted again that the weighting for each language component in the MUET is different; reading (45%), writing (25%), listening (15%) and speaking (15%).

Table 4.2

Perceived importance of each component tested in the MUET

Item	N	Not important (%)	Moderately important (%)	Very important (%)
Listening	137	0.7	39.4	59.9
Speaking	137	0.7	30.7	68.6
Reading	137	0.0	38.0	62.0
Writing	137	0.7	43.1	56.2

Table 4.2 shows high percentages for the moderately important (30.7–43.1%) and very important (56.2–68.6%) categories, indicating that most of the students perceived all the skills as very important for them to perform well in the MUET. One student (0.7%) perceived skills other than reading as “not important”. Speaking (68.6%, $n = 94$) and listening (59.9%, $n = 85$) have very high mean scores, indicating that these were the skills that the students considered to be some of the most important to perform well in the MUET. This is surprising, as these are the two skills that have the lowest weighting assigned (15%). Based on the qualitative data, it appears that this was because they had not previously been tested on these skills and were thus somewhat anxious. However, care must be taken in interpreting this finding, as the weighting of each language component also has to be taken into consideration, as discussed further in Chapter 5, together with

the findings obtained from the qualitative data. The interviews showed that 7 out of 11 students considered speaking to be their primary focus, over the other language skills:

Previously, when I was in the fifth form, the English test was mainly writing and answering objective questions. For the MUET, it is more challenging because now we have to speak and we have to study hard for that.

(Irdina)

One thing about the MUET that is troubling me is the speaking test because I was never tested on my speaking ability before and this is my first time in the MUET.

(Umi)

For the usual test like SPM they only focus on writing and comprehension solely, compared to the MUET which also tests your ability to speak and ability to listen attentively.

(Nazeera)

Reading, despite being ranked second highest (62%) in terms of being “very important” and the only skill that none of the students regarded as “not important”, was mentioned less by the participants than speaking and listening skills during the interviews. This could be attributed to the fact that they had prior knowledge of how to prepare for and answer reading questions in English, based on previous test experience and hence they focused more on the newly tested skills, namely speaking and listening. Although they did not mention reading as much as expected, the quantitative data clearly showed that the students were aware of the high weighting of the reading component in the MUET. As discussed later in this chapter, the qualitative data from the teachers revealed that the reading component in the MUET was the one that they were most concerned about (see 4.4).

4.1.2 Perceived difficulty of the MUET

Perceived test difficulty was measured on a continuous scale from 1 to 5, anchored at 1 “very easy” and 5 “very difficult”. On the scale, labels were only given to the values 1 and 5 to mark the extremes, but not to the values 2, 3 and 4. The students were asked to rate how difficult they thought the MUET was for them. Figure 4.1 presents findings for each level of perceived difficulty.

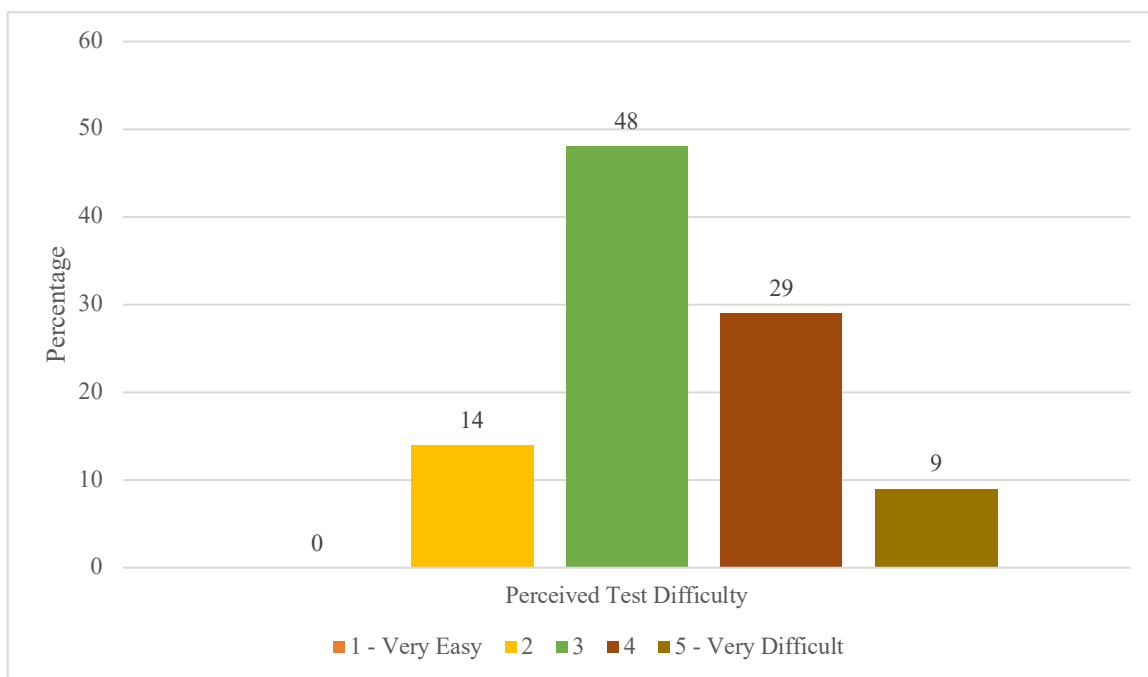


Figure 4.1 Students’ perceptions of the difficulty of the MUET

As depicted in Figure 4.1, 9% ($n = 12$) rated the MUET as very difficult and none of them rated it as very easy. Almost half of the students (48%, $n = 66$) rated the test difficulty at the mid-point of the scale, i.e. not too easy and not too difficult. Moreover, 29% ($n = 40$) viewed the difficulty of the MUET as being at level 4 and 14% ($n = 19$) at level 2. These findings indicates that, in general, the difficulty of the MUET is perceived as moderate to difficult.

The interview data were used to explore students' perceptions of the difficulty of the MUET in greater detail. Some reported viewing it as a difficult test because of the many language components being tested, i.e. reading, writing, speaking and listening. As mentioned by Husaini, the fact that the MUET includes all the language skills would require him to study harder to address each skill. Similarly, other students (5 out of 11) mentioned that the many components tested in the MUET made the test challenging for them, especially the speaking component:

The MUET is more challenging. Other English exams are alright but not the MUET. Speaking is hard.

(Maisarah)

It seems that some of the students would describe the MUET as “challenging” (translated from the word “*mencabar*” in Malay) rather than “difficult” and they treated this challenge as something positive to push them to work harder. Despite reporting the MUET as more challenging than any other English test that she had taken before, Haslinda stated that she loved studying for the MUET in class:

Not stressful at all because for me learning for the MUET in the classroom is enjoyable. Although it is a bit difficult, it is not stressful at all.

(Haslinda)

This challenge also encouraged the students to study harder, as confirmed by Azleen and Umi:

Preparing for the MUET is stressful, but it is a positive stress. It makes me want to learn the English language more.

(Azleen)

I think it is quite challenging, making me feel more motivated to do the exercises in the book. If before, I played a lot.

(Umi)

There were indications of pressure from having to sit for the MUET, but it was described as a positive kind of pressure by the students. None of the students mentioned being burdened by the test or portrayed the MUET in a negative manner. This shows that the difficulty of the MUET was viewed as within an acceptable range for them, even for the low-proficiency students. A test that is perceived as too difficult may hamper students' efforts and can be rather counterproductive. It seems that the difficulty of the MUET was not too overwhelming for the students, certainly not to the extent of discouraging them from learning.

A Mann–Whitney U test was run to determine if there were differences in test difficulty scores between male and female students. The distributions were dissimilar, as assessed by visual inspection. The test difficulty score for males (mean rank = 74.28) was higher than for females (mean rank = 64.51), but it was not statistically significantly different: $U = 1998.5$, $z = -1.548$, $p = .122$. Figure 4.2 shows that almost half of the male students ranked the MUET towards very difficult (46%) compared to the female students (31%). However, the difference was not significant, meaning that gender did not seem to be a great influence on how the students perceived the difficulty of the test in this study.

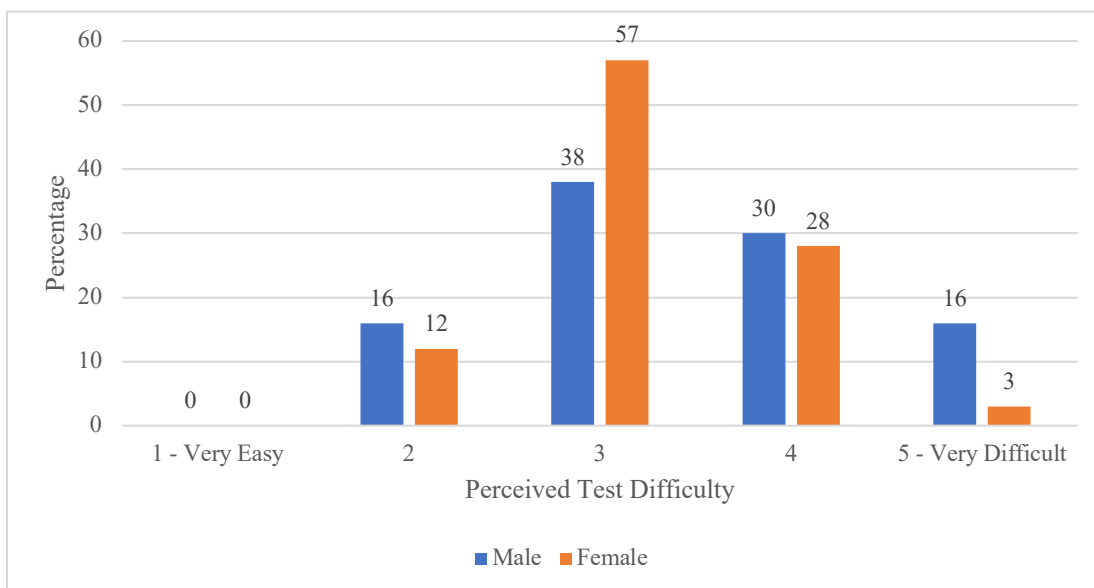


Figure 4.2 Students' perceptions of the difficulty of the MUET according to gender

Another Mann–Whitney U test was run to determine if there were differences in perceived test difficulty scores between high- and low-proficiency students, as shown in Figure 4.3.

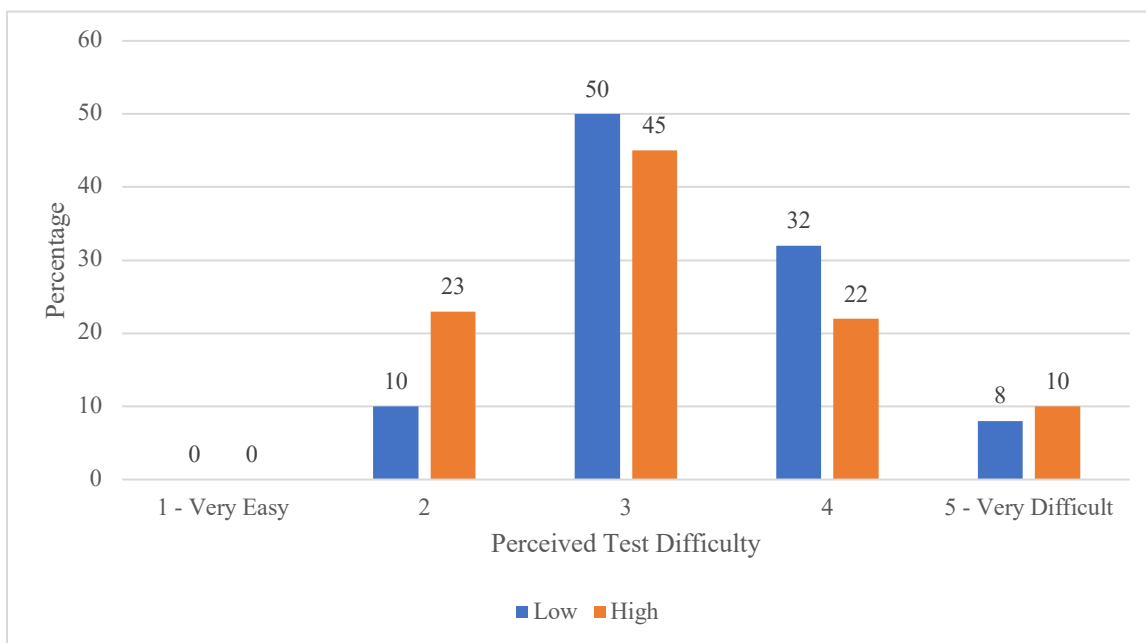


Figure 4.3 Students' perceptions of the difficulty of the MUET according to English proficiency

The students' proficiency was measured based on their English language result in the Malaysia Certificate Examination (MCE) (see 3.4.2). The perceived test difficulty scores for low-proficiency students (mean rank = 71.57) and high-proficiency students (mean rank = 62.76) were not statistically significantly different: $U = 2189.5$, $z = 1.274$, $p = .203$. Figure 4.3 shows that more low-proficiency students ranked the perceived difficulty level as 3 (50%, $n = 48$) and 4 (32%, $n = 31$) compared to high-proficiency students, with only 45% ($n = 18$) at level 3 and 22% ($n = 9$) at level 4. This finding was expected, as low-proficiency students might not be as confident about their English language proficiency compared to those of higher proficiency levels. However, it was interesting to note that 10% ($n = 4$) of the high-proficiency students rated the MUET as very difficult compared to low-proficiency students at 8% ($n = 8$). Although the difference was minimal, it is an interesting point to be discussed further in Chapter 5.

4.1.3 Students' self-efficacy in relation to the MUET

Four items in the questionnaire pertained to the students' perceived self-efficacy in relation to their performance in the MUET, as listed in Table 4.3. All four items were designed using a 5-point Likert-type scale and addressed students' perceptions of their current English language ability and how they thought they would perform in the MUET preparation class and in the actual MUET. These items concerned self-efficacy specifically in the context of the MUET, not students' self-efficacy in general.

Table 4.3

Students' perceived self-efficacy related to the MUET

Item Ranking	N	Disagree (%)	Undecided (%)	Agree (%)
I believe I will receive an excellent grade in the MUET (Q15)	137	5	28	67
Taking into consideration its difficulty, I think I can perform very well in the MUET (Q16)	137	6	21	73
Taking into consideration my ability, I think I can perform very well in the MUET (Q17)	137	3	24	73
I'm confident I can do an excellent job on the assignment and tasks in the MUET class (Q18)	137	5	25	70

Table 4.3 shows a very high percentage of agreement for all items measuring perceived self-efficacy in relation to performance in the MUET, ranging from 67% to 73%. This indicates that the students were generally confident in their ability to perform well in the MUET.

As shown in Table 4.3, the first three items concern how the students perceived their ability to perform well in the MUET. Although most of the students in this study indicated agreement (67–73%) with the statements, quite a number of students were not so sure (21–28%) and some disagreed with the statements (3–6%). Similar findings were found for item Q18, which deals with students' self-efficacy in relation to their performance in the MUET preparation classes.

To explore which independent variable under investigation might have influenced the students' perceived self-efficacy, Mann–Whitney U tests were run to determine if there were differences for gender and the students' English proficiency level. Visual inspection showed differences in the distributions of the self-efficacy scores for males and females. The self-efficacy scores for females (mean rank = 78.83) were statistically significantly higher than for males (mean rank = 57.45): $U = 3058.5$, $z = 3.206$, $p = .001$. There were also differences in the distributions for the high- and low-proficiency students, but the perceived self-efficacy scores for the high-

proficiency students (mean rank = 69.49) and the low-proficiency students (mean rank = 68.80) were not statistically significantly different: $U = 1920.5$, $z = -.094$, $p = .925$. This finding was expected, indicating that the students' proficiency level did not seem to affect their perceptions of their self-efficacy in terms of performing well in the MUET.

4.1.4 Students' perceptions of their MUET teachers

As mentioned earlier in the literature review in Chapter 2, tests with strong consequences not only affect the learners, but also the teachers. Teachers' and students' perceptions of the consequences of the test affect how teachers teach and how learners learn. For example, high-stakes tests often lead to teaching and learning to the test. This section presents the students' perceptions of the MUET in relation to their teachers' teaching activities in an attempt to provide a more holistic picture of the washback phenomenon. There were two sections in the student questionnaire that dealt with the students' perceptions of their teacher's role in encouraging them to practise their language skills (6 items) and their teacher's teaching activities (6 items). Table 4.4 reports the percentages for items pertaining to the students' perceptions of their teachers motivating and encouraging them to work in their MUET preparation classes compared to *before* they had to start preparing for the MUET.

Table 4.4

Teachers' role in encouraging students in MUET preparation class as perceived by students

Item ranking	N	Disagree (%)	Undecided (%)	Agree (%)
My MUET teacher makes me practise my writing skills more than before (Q3)	137	5	13	82
My MUET teacher makes me practise my reading skills more than before (Q4)	137	2	13	85
My MUET teacher makes me practise my speaking skills more than before (Q5)	137	3	7	90
My MUET teacher makes me practise my listening skills more than before (Q6)	137	2	12	86

All items show a high percentage of agreement, which indicates the students overall considered that their teachers made them practise their language skills more than “before” MUET test preparation. “Before” here refers to when they were in primary and secondary school, when English language was one of their compulsory subjects. In their previous national public examinations, the students had only been tested on their writing and reading skills in English. The MUET was their first language test that included all four language components.

The students were also asked to give their opinions on whether the MUET preparation class was helpful for them. Most students stated that they perceived the class as useful in preparing them to sit for the MUET:

Yes, it is quite useful in a way that the students are exposed to what the MUET is and how it is going to be conducted and how we are going to be scored in it.

(Syamimi)

For me the class is very beneficial and very helpful to prepare myself to sit for the MUET besides having the opportunity to experience the different interesting approaches taken by my teacher to teach us English.

(Marziana)

The students mentioned that the MUET preparation class helped them to familiarize themselves with the MUET in terms of its components and its format. The students stated that the teacher in their MUET preparation class helped them a lot, especially in their communication skills and in building up their courage and confidence to speak in English:

Very helpful because the teacher taught us a lot of new vocabulary, trained us how to be confident, how to look for ideas when we speak.

(Shahirah)

I am more confident now because the teacher always makes us sit in groups, just like in the MUET speaking test. It is comfortable now that I have become used to it and I feel more prepared.

(Haslinda)

Thus, regular exposure to how the actual test was going to be conducted helped Haslinda feel more confident and more prepared to sit for the test.

The second section covering the students' perceptions of their teachers in the questionnaire dealt with the specific activities conducted by their teachers during the MUET preparation lessons. For this section, the students were instructed to evaluate each activity on a 5-point Likert-type scale of frequency, with 1 for "never" and 5 for "always".

Table 4.5

Teachers' teaching activities in MUET preparation class as perceived by students

Item	N	Never (%)	Rarely (%)	Sometimes (%)	Often (%)	Always (%)
Organize group work or discussion (Q20)	137	3	19	14	46	18
Do mock exam-like activities (Q21)	137	5	15	22	39	19
Discuss textbook exercises (Q29)	137	12	12	16	34	26
Organize real-life language activities (e.g. mock interview, sketches, etc.) (Q30)	137	7	13	22	37	21

Table 4.5 shows that the highest percentage for each teaching activity is “often”, ranging from 34% to 46%, indicating that the students considered the teachers conducted all the listed activities most of the time. As depicted in Table 4.5, almost half of the students reported that their teachers conducted exam-like activities in their MUET preparation classrooms. During the interviews, the students were asked to describe their learning experience in the MUET preparation class at their respective pre-university institutions. There was an indication of washback in terms of how the teachers conducted the activities in the MUET preparation class as it seemed they followed the structure of the MUET speaking component, for which candidates are divided into groups of four and are asked to give an individual speech and participate in group discussion on assigned topics. Since the classes were intended to help the students prepare for the MUET, teaching to the test was expected. However, 5% (n = 7) reported their teachers never conducted such activities. In an interview, one student reported that frustratingly her teacher did not really emphasize their speaking skills in class. According to her:

We do not do any speaking activity at all. The teacher just asks us to speak in English in the class like when we want to ask questions, that is all. But doing activities like what will be tested in the exam, that does not happen.

(Umi)

She seemed rather frustrated with how her teacher conducted the MUET preparation class and expected to have more MUET-oriented speaking activities. Other than Umi's case, evidence of teaching to the test is prevalent in the findings of this study. The students reported that they did a lot of MUET practice exercises (see Appendix F) in their MUET textbook, which contains model MUET papers:

My teacher focuses more on the textbook that we use in the class. If anything, the teacher will just refer to that book, exercises, etc., everything from that book only.

(Umi)

We do a lot of exercises in the textbook. For writing, we practise writing essays. Once in the class we always write essays. For reading we always answer the questions in the textbook.

(Maisarah)

Similar findings were found in the quantitative data (see Table 4.5) for Item Q29 “*Discuss textbook exercises*”, with 34% (n = 46) of students reporting that their teachers “often” discussed textbook exercises in the class, followed by “always” at 26% (n = 36). The MUET preparation textbook was commonly used by the teachers to carry out their lessons, as reported by Haslinda:

We have one textbook that we bought, specialized for the MUET, to be used in the classroom. It is like an exercise book with a lot of exercises in it.

(Haslinda)

Table 4.5 also indicates that 12% (n = 16) of the students stated that their teachers never discussed textbook exercises in the classrooms. Some of the teachers rarely utilized the textbooks (12%, n = 16). Similarly, the interview data revealed that there were also teachers who did not rely excessively on the textbooks and used their own materials and techniques in teaching, as reported by Azleen:

We do use the MUET textbook but not so much. My teacher uses her own ways more to teach us compared to using the textbook. She is not bound by the textbook. For example, in the classroom, the teacher will ask us to present in class, to encourage us to speak more English in front of our friends. It feels like storytelling and lots of communication with the teacher.

(Azleen)

According to Azleen, her teacher employed more communicatively oriented activities to help enhance their communication skills. Thus far, it appears that most of the students had positive perceptions of their teachers, who focused on preparing them to score highly in the MUET by familiarizing them with the format of the test using the textbooks and activities similar to those that were going to be tested. The MUET was designed to require teachers to use communicatively oriented activities to prepare the students, especially in terms of the speaking component. Although the students did not explicitly indicate their eagerness to develop their language skills other than for the test, they were keen to improve their communicative skills to help them perform well in the MUET. Referring to the data presented in Table 4.5, the students reported that their teachers often

(36.5%, n = 50) and always (21.2%, n = 29) organized real-life language activities in the MUET preparation classroom. Although the evidence presented so far indicates clear evidence of teaching to the test, as seen by the students, it appeared that some of the teachers also tried to incorporate other elements that were not directly tested in the MUET as much as they could. In Shahirah's case, the example that she gave was that her teacher would normally ask one student to speak about any topic in front of the class:

For speaking, every day in the MUET preparation class, each and every student will come to the front to speak in English. In every class, it will be someone's turn to speak to the class. Although the format of the MUET speaking test is in a group, the teacher will still ask the students to go to the front and speak. It is to improve the way we speak and the way we look for ideas when we are delivering our short speech.

(Shahirah)

In the MUET speaking test, students are not required to speak in front of an audience. They only have to give a short speech in front of the other three candidates who are in the same group as them (see 2.10). Shahirah explained that the reason why her teacher made them do extra speaking practice in front of their classmates was not only to help them improve their speaking skills, but also to boost their confidence in their speaking skills. This finding indicates a certain degree of autonomy among the teachers in terms of the methodology they employed in their classroom. The motivation behind their decision is something that will be interesting to explore in the data from the teachers and their perspectives are discussed in detail in later sections.

Many students also reported that their teachers were very keen to organize group work or discussions, as depicted in Table 4.5 pertaining to Item Q20. More than half of the students agreed that their MUET teachers often (46%, n = 63) or always (18%, n = 25) did so during the lessons.

This was also supported by the qualitative data. From the interviews, it appears that a commonly used method by the teachers was collaborative learning, dividing the students into small groups for speaking activities:

Normally my teacher will put us in groups so that we will speak in English in that group.

(Haslinda)

My teacher conducts the speaking activities in groups and individually. She follows the format of the MUET but she does it in groups.

(Irdina)

According to Irdina above, the tasks involved in group work were closely related to the MUET, as the teacher used the format of the MUET to carry out speaking activities. Interestingly, aside from speaking, the students reported that the teacher also utilized collaborative learning to teach writing skills to the students:

We do our writing activity in groups, because in our class not all students are active, or are good in English language, so the teacher will mix all of us in groups so the higher proficiency students will teach the lower proficiency students.

(Haslinda)

My teacher always conducts our class in groups. For writing, for example, the teacher will divide us into groups, then she will assign each group tasks to be completed.

(Shahirah)

Although the format for the MUET writing component is not assessment in groups, their teachers had them work in groups to develop their writing skills because they wanted the students to be

able to learn from each other. Hence, it was shown that explicit teaching to the test was not the only method used to prepare the students for the MUET. Other means of instruction could be used if it helped the students to develop their language skills, thus focusing on mastery of the language as a whole, not solely preparing them to sit the exam.

4.2 Students' perceptions and language learning strategies

The second research question aimed to explore if the students' perceptions of test importance, test difficulty and their own self-efficacy would have any effect on their language learning strategies. The set of data used to answer this research question was from Group A participants ($n = 137$), as they were preparing for the MUET in Form 6. These students were asked to identify specific language learning strategies that they used when preparing for the MUET from the list of the language learning strategies provided in the student questionnaire and the students in the interviews were also asked to share their experiences of preparing for the MUET and the kinds of language learning strategies they employed.

Before delving into the relationship between the students' perceptions and their language learning strategies, the kinds of strategies they used most were studied first, followed by exploring if there were any significant differences in terms of gender and English language proficiency. Oxford (1990) divided strategies in the SILL into two categories – direct and indirect – and further into three types for each category: memory, cognitive and compensatory (direct) and metacognitive, affective and social (indirect). Direct strategies specifically involve the use of language, while indirect strategies do not directly involve using the language, but they support language learning (Ehrman & Oxford, 1990). These strategies can be specified as follows:

1. Memory strategies for remembering and retrieving new information.

2. Cognitive strategies for understanding and producing the language.
3. Compensatory strategies for using the language despite lack of knowledge.
4. Metacognitive strategies for coordinating the learning process.
5. Affective strategies for regulating emotions.
6. Social strategies for learning with others.

(Oxford, 1990, pp. 14–15).

A detailed explanation and description of each of the categories was presented in Chapter 3, Methodology (see 3.7.1). A 5-point Likert-type scale of frequency, with 1 representing “never” and 5 “always” was used. According to Oxford (1990, p. 291), a mean score above 3.5 on a SILL item is considered to reflect high usage of a given strategy; 2.5 to 3.4 indicates medium use; below 2.4 suggests low use of a strategy. In this study, the mean scores for all items range from 3.12 to 3.80, indicating that the students preparing for the test used the language learning strategies quite frequently. Table 4.6 shows the overall picture of the students’ reported language learning strategy usage in terms of these three categories.

Table 4.6

Students’ reported language learning strategy use according to frequency

Usage (mean)	N	Percentage (%)
High (> 3.5)	20	40
Medium (2.5–3.4)	30	60
Low (< 2.4)	0	0
Total	50	100

The majority of the students (60%) reported medium use of language learning strategies, whereas 40% reported high use when preparing for the MUET. None of the students reported low usage.

Next, item analysis was carried out to look at the overall pattern of the language learning strategies used by the students. The mean scores for all items were calculated and arranged in descending order. The top 10 items in the ranking are presented in Table 4.7.

Table 4.7

Students' language learning strategies in the top 10 ranked according to the mean scores

No	Item	Category	Mean	Median	Std. Dev.
Q48	I watch English language TV shows spoken in English or go to movies spoken in English.	Cognitive (D)	3.80	4.00	1.117
Q62	If I can't think of an English word, I use a word or phrase that means the same thing.	Compensatory (D)	3.74	4.00	.993
Q65	I pay attention when someone is speaking English.	Metacognitive (I)	3.74	4.00	1.171
Q44	I try to talk like native English speakers.	Cognitive (D)	3.68	4.00	1.144
Q80	I practise English with other students.	Social (I)	3.67	4.00	1.023
Q66	I try to find out how to be a better learner of English.	Metacognitive (I)	3.66	4.00	1.059
Q78	If I do not understand something in English, I ask the other person to slow down or say it again.	Social (I)	3.64	4.00	1.084
Q57	To understand unfamiliar English words, I make guesses.	Compensatory (D)	3.61	4.00	1.032
Q45	I practise the sounds of English.	Cognitive (D)	3.61	4.00	1.101
Q73	I encourage myself to speak English even when I am afraid of making a mistake.	Affective (I)	3.61	4.00	1.024

D = Direct; I = Indirect

Based on my own experience, both as a student and an educator, it was expected that students would use language learning strategies frequently, but the highest mean score reported was 3.80. Although this can be considered relatively high, it is still below 4.00, which was the level of expectation; as these students were preparing for a high-stakes language test, intense washback in the form of a high frequency of use of language learning strategies might be anticipated. However, I would argue that the notion that a high-stakes test will inevitably lead to some kind of enhanced use of strategies is a bit farfetched. Logically, though, if students know that they are going to be tested, they will do whatever they think might promote their English language mastery, although they might be aware or unaware that they are actually employing language learning strategies. These students made a conscious decision to increase their exposure to English in reaction to the pressure of preparing to sit a high-stake tests.

As can be seen in Table 4.7, the strategy that the students reported using the most was a cognitive strategy: *“I watch English language TV shows spoken in English or go to movies spoken in English”* (mean = 3.80). This strategy has an entertainment value, perhaps making it popular with the students. Another cognitive strategy that the students frequently used was *“I try to talk like native English speakers”* (mean = 3.68). The qualitative data also revealed that the students predominantly used cognitive strategies. The most popular (10 out of 11) cognitive strategy mentioned in the interview was reading English language materials:

I read English books all the time or rather I love to read English books, such as novels or story books.

(Haslinda)

I normally read a English newspaper called “The Stars.”

(Shahirah)

I read Manga comics (Japanese comics) online, the ones that have been translated into English, to prepare me to sit the MUET.

(Marziana)

The reading materials mentioned above indicate that the students were reading for pleasure, while still being conscious about the learning that took place. Husaini shared what he normally did when reading an English newspaper:

When I read an English newspaper, I'll have a dictionary beside me. If I find a new word, I'll look it up in my dictionary and I'll write the meaning in Malay and re-read it.

(Husaini)

This practice was also shared by four other interview participants. Another popular (10 out of 11) cognitive strategy employed was to work on English language exercises, or more specifically the MUET workbook and questions from past years. Practising model questions has commonly been reported in the literature. The students recounted using the same strategies to practise their reading and writing skills:

To practise my reading skills, I do a lot of exercises because the model questions are there in the textbook, I just have to choose which one I want to work on.

(Shahirah)

I read and revise the questions in the MUET textbook and finish the MUET tasks that are given by my teacher.

(Marziana)

I do the listening exercises using the recordings that my teacher gives us in a soft copy version for us to practise at home. So I just play that and answer the listening test in the textbook.

(Umi)

The qualitative data also showed that the students appeared to be keen on using strategies with an entertainment element, such as watching English films (5 out of 11) and listening to English songs (5 out of 11) to practise their English. According to Haslinda, it was her teacher's suggestion to start listening to English songs to familiarize their ears to English. However, some (3 out of 11) preferred to watch English news on the television instead for the following reason:

I normally watch movies to improve my listening. However, sometimes the actors use different English accents that I'm not familiar with. So, my teacher advised us to watch the news instead.

(Shahirah)

The students also reported frequently employing indirect language learning strategies, for example metacognitive “*I pay attention when someone is speaking English*” and social “*I practise English with other students*”.

In terms of metacognitive strategies, the students reported consciously searching for practice opportunities and planning for language tasks, even when the circumstances were not as supportive as those shared by Shahirah:

I always practise my speaking alone. If I do not have anything to say, or a partner to talk to, I still speak to myself. I will find the time to do this as I need to practise.

(Shahirah)

Not having a partner to speak to did not hamper Shahirah's efforts to practise her speaking skills. However, other students who were lucky enough to have company to practise with exploited their opportunity to speak in English to the fullest. This kind of strategy is grouped under the social category. The qualitative data revealed that 7 out of 11 respondents practised their speaking skills by using English with family, friends, teachers and native speakers:

Sometimes when I talk to my mom, my family, I use English just for fun. When I hang out with my friends, sometimes I talk to them in English and if there is any word that I use that they do not understand, I will explain the meaning of the word to them.

(Haslinda)

I practise my speaking skills by speaking in English at home with my family. Normally I will speak in English with my dad.

(Irdina)

This was very encouraging, as the students exhibited the awareness to start practising and incorporating English in their daily lives. Taking into account my own experience of preparing for the MUET 11 years ago, this was expected, especially for speaking skills. Husaini, on the other hand, challenged himself by talking to foreigners instead:

Yes, like I said before, whenever I find a tourist, I will try to speak in English to them, even though I speak in "broken" English. They are very friendly and I'm surprised that they are very nice to me. Actually, these tourists, they still understand what I am trying to say with my "broken" English, and they have never once laughed at me if I say something wrong, they only smile at me. If I say something wrong, they will help me get it right. That is how I practise my speaking skills.

(Husaini)

Despite acknowledging that his English was still weak, Husaini was not afraid to use it with native speakers, as according to him they were normally very encouraging and helpful. Similarly, Nazeerah and Marziana specifically mentioned that when it came to practising their speaking skills with friends, they were of the opinion that they needed a partner who had a better mastery of English than them because:

Once we have someone superior, someone who has a higher level of English compared to the one that I have acquired, it will be a motivation for me to improve myself.

(Nazeerah)

Next, the 10 items with the lowest mean scores are shown in Table 4.8. It can be seen that these items are still in the medium frequency range. Most of the strategies ranked in the bottom 10 were direct strategies, except for “*I write down my feelings in a language learning diary*”, which is an indirect strategy.

Table 4.8

Students' language learning strategies in the bottom 10 ranked according to the mean scores

No	Item	Category	Mean	Median	Std. Dev.
Q75	I notice if I am tense or nervous when I am studying or using English.	Affective (I)	3.30	3.00	1.107
Q74	I give myself a reward or treat when I do well in English.	Affective (I)	3.29	3.00	1.170
Q53	I try to find patterns in English.	Cognitive (D)	3.29	3.00	1.106
Q60	I read English without looking up every new word.	Compensatory (D)	3.28	3.00	1.050
Q59	I make up new words if I do not know the right ones in English.	Compensatory (D)	3.26	3.00	1.029
Q55	I try not to translate word for word.	Cognitive (D)	3.24	3.00	1.054
Q50	I write notes, messages, letters, or reports in English.	Cognitive (D)	3.20	3.00	1.090
Q76	I write down my feelings in a language learning diary.	Affective (I)	3.16	3.00	1.220
Q39	I use flashcards to remember new English words.	Memory (D)	3.14	3.00	1.119
Q67	I plan my schedule so I will have enough time to study English.	Metacognitive (D)	3.12	3.00	1.085

D = Direct; I = Indirect

As shown in Table 4.8, “*I plan my schedule so I will have enough time to study English*” has the lowest mean score of 3.12. This is a metacognitive strategy that deals with coordinating the learning process. In the interviews, when it came to metacognitive strategies, 5 out of 11 students mentioned that they purposely set aside at least one hour to study English:

Plan my own timetable.

(Umie)

Set up my schedule.

(Syamimi)

However, two students in the interviews mentioned that for them, the activities that they did in the class and the exercises given by their teachers were already enough to prepare them for the MUET. Hence, they did not allocate extra hours to study English outside class:

I do not really practise at home. It's just that when it's getting closer to the exam time, I will refer back to all the exercises that I did before. What we did in the classroom is already enough for me.

(Azleen)

I don't have the free time to do the preparation like my own style of preparation because I've been doing the work that she assigned. Because the whole semester would be the preparation for the MUET.

(Nazeerah)

Nazeerah mentioned that since she was busy studying for other subjects, the only time that she had for English language practice was allocated to doing her English homework, which she considered similar to preparing for the MUET. It seems understandable that students will not prioritize the MUET on top of other subjects, because despite being a high-stakes language test, the MUET is not the only requirement for these students to secure a place in one of the public universities in Malaysia. They still need to work on their major courses and subjects at pre-university level, which are equally important. The overall ranking demonstrated that the students used most language learning strategies frequently. However, the evidence presented thus far fails to show if the

language learning strategies used could be attributed to the washback effect of preparing for the MUET.

Next, differences between gender was explored in the use of language learning strategies when preparing for the MUET. The Mann–Whitney U test was carried out for each pair, as presented in Table 4.9. A non-parametric test was chosen as the independent variable was dichotomous in nature.

Table 4.9

Differences in learning strategies by gender

Strategies	Mean rank		p-value	Std. Dev.	Test value (Mann–Whitney U)
	Female	Male			
Memory	80.18	55.87	.000*	6.610	3159.0
Cognitive	79.66	56.48	.001*	10.229	3119.5
Compensatory	78.81	57.48	.002*	4.381	3057.0
Metacognitive	79.09	57.15	.001*	6.246	3077.5
Affective	79.14	57.09	.001*	4.494	3081.5
Social	79.03	57.21	.003*	4.202	3028.5

Direct strategies: Memory, Cognitive, Compensatory

Indirect strategies: Metacognitive, Affective, Social

Bonferroni's correction was used to adjust for multiple comparisons and to avoid Type 1 error at $p = 0.05$. Thus, the statistical results for differences in learning strategies across gender were considered significant only if the p-value was $\leq 0.05/6$, hence $p < .008$. As can be seen from Table 4.9, the differences between the male and the female students' use of language learning strategies in preparing for the MUET were all statistically significant at $p < .008$, with the female students having a higher mean rank score for all categories than the male students.

A Mann–Whitney U test was also carried out to see if there were any differences in the use of language learning strategies between high- and low-proficiency students. The data are presented in Table 4.10:

Table 4.10

Differences in learning strategies by proficiency level

Strategies	Mean rank		p-value	Std. Dev.	Test value (Mann–Whitney U)
	High proficiency	Low proficiency			
Memory	60.86	72.36	.123	6.610	2265.5
Cognitive	66.50	70.03	.636	10.229	2040.0
Compensatory	65.36	70.50	.489	4.381	2085.5
Metacognitive	78.11	65.24	.084	6.246	1575.5
Affective	61.27	72.19	.142	4.494	2249.0
Social	78.56	65.06	.066	4.202	1553.0

Table 4.10 shows that the low-proficiency students tended to use direct language learning strategies more than the high-proficiency students and the high-proficiency students tended to use more indirect language learning strategies compared to the low-proficiency students. However, the differences between these two proficiency groups were not statistically significant.

4.2.1 Relationship between perceived test importance and language learning strategies

As mentioned previously, the students' perception of test importance is one of the factors posited in Green's (2007a) washback model to influence washback intensity (see 2.3.4). Hence, Spearman's rank-order correlation was used to assess the relationship between the students' perceptions of test importance and the frequency of use of language learning strategies (see Table 4.11). All six groups of language learning strategies were found to be significantly correlated with

the students' perceived test importance, except for the direct compensatory strategy. The guidelines for determining the strength of the relationship proposed by Cohen (1988, pp. 79–81), i.e. .10 to .29 = small, .30 to .49 = medium and .50 to 1.0 = large, were used in this study. For example, the higher the correlation coefficient between perceived test importance and language learning strategies, the stronger the relationship.

Table 4.11

Correlations between perceived test importance and language learning strategies

Strategies	Correlation coefficient	Sig.	Result
Memory	.350*	.000	Medium positive correlation
Cognitive	.269*	.001	Weak positive correlation
Compensatory	.148	.085	Weak positive correlation
Metacognitive	.281*	.001	Weak positive correlation
Affective	.245*	.004	Weak positive correlation
Social	.259*	.002	Weak positive correlation

As shown in Table 4.11, medium to weak positive correlations were found between perceived test importance and reported use of direct language learning strategies: memory, cognitive and compensatory. Bonferroni's correction was applied to the p-value, with $p \leq 0.05/6$, hence $p < .008$. Memory strategies ($r = .350$, $p < .001$) presented the highest correlation, followed by cognitive strategies ($r = .269$, $p < .001$) and compensatory strategies ($r = .148$, $p > .005$). For indirect strategies, weak positive correlations were found between perceived test importance and metacognitive strategies ($r = .281$, $p < .001$), affective strategies ($r = .245$, $p < .001$) and social strategies ($r = .259$, $p < .001$). The findings are consistent with the washback hypothesis that there is a correlation between the importance of tests and the extent of washback (Alderson & Wall, 1993).

Next, the Kruskal–Wallis test was conducted to test for differences between the magnitude of the students’ perceived test importance, divided into three levels – “low” (n = 4), “moderate” (n = 13) and “high” (n = 120) – and the frequency of reported use of language learning strategies when preparing for the MUET. In other words, this test helped identify whether the students’ language learning strategies use differed based on their perceptions of test importance. The median values for language learning strategies used were statistically significantly different between groups: $\chi^2(2) = 12.180$, $p = .002$. Subsequently, pairwise comparisons were performed using Dunn’s (1964) procedure with Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed statistically significant ($p = .002$) differences in the frequency of use of language learning strategies between the groups perceiving moderate test importance (median = 2.84) and high importance (median = 3.61) , but not the group perceiving low test importance (median = 3.83). However, care must be taken in interpreting this finding as the division of the groupings was uneven due to the limited number of participants at each level of perceived importance.

4.2.2 Relationship between perceived test difficulty and language learning strategies

The second factor from Green’s (2007a) washback model is perceived test difficulty. The findings presented in Table 4.12 show weak negative correlations between students’ perceived test difficulty and the language learning strategies they employed, direct (memory, cognitive, compensatory) and indirect (metacognitive, affective, social).

Table 4.12

Correlations between perceived test difficulty and language learning strategies

Strategies	Correlation coefficient	Sig.	Result
Memory	-.080	.351	Very weak negative correlation
Cognitive	-.107	.215	Weak negative correlation
Compensatory	-.144	.092	Weak negative correlation
Metacognitive	-.107	.212	Weak negative correlation
Affective	-.098	.252	Very weak negative correlation
Social	-.151	.078	Weak negative correlation

A negative correlation indicates that the higher the perceived difficulty of the MUET, the lower the frequency of the reported use of language learning strategies. However, not all correlations were statistically significant. Furthermore, the strength of the correlation coefficients was very weak, ranging from $r = .080$ to $r = .15$.

A Kruskal–Wallis H test was conducted to determine if there were differences in the frequency of reported use of language learning strategies between groups that differed in terms of the level of perceived test difficulty: “easy” ($n = 19$), “moderate” ($n = 66$) and “difficult” ($n = 52$). The median scores for use of language learning strategies decreased from easy (median = 3.82), to moderate (median = 3.58), to difficult (median = 3.46), but the differences were not statistically significant ($\chi^2(2) = 2.028$, $p = .363$). This finding indicates that the students’ language learning strategies did not differ greatly across the three different levels of perceived test difficulty.

4.2.3 Relationship between self-efficacy and language learning strategies

To establish if there were any relationship between students’ perceptions of their ability (self-efficacy) in relation to their performance in the MUET with the language learning strategies used in preparing for the test, Spearman’s rank-order correlation was run. The correlation coefficients are presented in Table 4.13.

Table 4.13

Correlations between self-efficacy and language learning strategies

Strategies	Correlation coefficient	Sig.	Result
Memory	.319*	.000	Medium positive correlation
Cognitive	.302*	.000	Medium positive correlation
Compensatory	.180	.035	Weak positive correlation
Metacognitive	.260*	.002	Weak positive correlation
Affective	.249*	.003	Weak positive correlation
Social	.336*	.000	Medium positive correlation

Bonferroni's correction was applied to the p-value, with $p \leq 0.05/6$, hence $p < .008$. As can be seen from Table 4.13, there is a weak to medium positive correlation between the students' perceived self-efficacy and their language learning strategies and all correlations were found to be statistically significant, except for compensatory strategies.

A Kruskal–Wallis test was conducted to determine if there were differences in reported use of language learning strategies between groups differing in their level of perceived self-efficacy: “low” ($n = 4$), “moderate” ($n = 24$) and “high” ($n = 109$). The median scores for use of language learning strategies were statistically significantly different between the different levels of perceived self-efficacy groups ($\chi^2(2) = 9.336$, $p = .009$). Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with Bonferroni correction for multiple comparisons (statistical significance accepted at the $p < .0083$ level). This post hoc analysis revealed statistically significant differences in reported use of language learning strategies between the moderate (median = 3.10) and high (median = 3.63) groups ($p = .007$), but not the low group (median = 3.54), or any other group combination. Similar to the findings for perceived test importance, the students' language learning strategies differed significantly for those with high and moderate perceived self-efficacy in relation to their performance in the MUET, but not for those with low self-efficacy.

4.3 Washback length and intensity of the MUET

The third research question in this study concerned the washback length and intensity of the MUET and how these aspects appeared to influence washback on the learners. This was formulated in an attempt to address two important elements of washback, as discussed in 2.3.4. The data used to address the third research question were derived from the Group B students, who had already sat the MUET (see 3.4.2). The findings on washback length are discussed first, followed by washback intensity.

4.3.1 Washback length

Washback length refers to continuation of the influence of the test even after the students have sat the test. This study aimed to contribute to knowledge of washback length by exploring whether there were any significant differences between the use of language learning strategies reported by the Group B students when preparing for the MUET in the past and their current language learning strategies at their respective university. Oxford's (1990) SILL was again used to measure students' language learning strategies (see 3.7.1). Group B comprised 238 undergraduate students, who were asked to indicate if they had used each strategy when they prepared for the MUET in the past and if they were still using the same strategy in their English language learning at university. Tables 4.14 to 4.19 present the findings from McNemar's χ^2 tests for the students' use of language learning strategies "before" and "after" they sat the MUET. The percentage refers to the number of students using the language learning strategies. The differences in the *before* and *after* use of each strategy were determined by the significance ($p < .05$).

To explore this matter further, specifically with regard to the MUET, an open-ended question was given at the end of the student questionnaire for Group B to ask if they thought there

were any differences between the strategies they used when they were preparing for the MUET in the past (i.e. *before* taking the test) and those that they thought they were using in language courses at university (i.e. after they took the test). For ease of presentation, the findings concerning language learning strategies are reported in six separate tables according to type: memory, cognitive and compensatory (direct) and metacognitive, affective and social (indirect). Bonferroni's correction was applied to the p-value, with $p \leq 0.05/9$, hence $p < .005$. For memory strategies, it can be seen in Table 4.14 that the use of most strategies increased from "*before* the MUET" to "*after* the MUET".

Table 4.14

Students' language learning strategies before and after the MUET (memory)

No.	Item	Before (%)	After (%)	Sig.
31.	I think of relationships between what I already know and new things I learn in English.	92	90	.210
32.	I use new English words in a sentence so I can remember them.	78	83	.112
33.	I connect the sound of a new English word and an image or picture of the word to help remember the word.	64	68	.176
34.	I remember a new English word by making a mental picture of a situation in which the word might be used.	61	69	.010
35.	I use rhymes to remember new English words.	50	56	.045
36.	I use flashcards to remember new English words.	35	32	.361
37.	I physically act out new English words.	48	54	.015
38.	I review English lessons often.	65	56	.014
39.	I remember new English words or phrases by remembering their location on the page, on the board, or on a street sign.	62	63	1.00

However, none of the increases reported were statistically significant. Moreover, of all the strategies in the memory category, Item 38, "*I review English lessons often*", presented a substantial decrease in value from 65% to 56%. This might be attributed to the fact that the students

were no longer facing a high-stakes test and thus there was no longer any need to do extra revision. Although they would still have to take a test at the end of the English for Academic Purposes (EAP) class at the university, compulsory for all undergraduate students, this generally has small weighting in terms of the overall grade, ranging from 40% to 60%, unlike the MUET (100%).

In terms of cognitive learning strategies, Bonferroni's correction was applied to the p-value, with $p \leq 0.05/14$, hence $p < .0035$. As can be seen in Table 4.15, only three items pertaining to cognitive learning strategies increased statistically significantly from before the MUET to after the test: Item 47 "*I write notes, messages, letters, or reports in English*", Item 52 "*I try not to translate word for word*" and Item 53 "*I make summaries of information that I hear or read in English*".

Table 4.15

Students' language learning strategies before and after the MUET (cognitive)

No.	Item	Before (%)	After (%)	Sig.
40.	I say or write new English words several times.	65	67	.590
41.	I try to talk like native English speakers.	66	70	.312
42.	I practise the sounds of English.	74	81	.015
43.	I use the English words I know in different ways.	63	71	.007
44.	I start conversations in English.	54	61	.041
45.	I watch English language TV shows spoken in English or go to movies spoken in English.	86	90	.089
46.	I read for pleasure in English.	69	75	.025
47.	I write notes, messages, letters, or reports in English.	53	65	.001*
48.	I first skim an English passage (read over the passage quickly) then go back and read carefully.	71	74	.201
49.	I look for words in my own language that are similar to new words in English.	64	66	.511
50.	I try to find patterns in English.	52	56	.194
51.	I find the meaning of an English word by dividing it into parts that I understand.	72	78	.026
52.	I try not to translate word for word.	45	53	.002*
53.	I make summaries of information that I hear or read in English.	52	61	.001*

The students appeared to practise the sounds of English (see Item 42), start conversations in English (see Item 44) and read for pleasure in English (see Item 46) more than when they were preparing for the MUET. None of the items in this category decreased in terms of the percentage of usage. One possible explanation for the increase in terms of the use of cognitive strategies is the English language demands at university, motivating them to use these strategies more as such strategies allow the students to develop stronger schemas (knowledge structures) by practising in a naturalistic setting, in this case university, in which English is commonly used. In the open-ended question, when asked if they were still using the same language learning strategies that they had used when preparing for the MUET, some mentioned that they used different strategies to learn

English at university. This was because they no longer had the privilege of having teachers to guide them every step of the way when it came to English language learning, thus forcing them to be more independent:

At university I need to do the preparation by myself. I have to refer to the dictionary, Google, etc., to improve my English language mastery.

(Student 66)

Some students maintained the same strategies they had used when preparing for the MUET, but now that they were at university, they improvised and added more strategies:

I refer to the Internet, Google translate, prepare a book to record new English words because I understand and memorize better this way. I also watch English movies or TV series and listen to others who speak English.

(Student 16)

In terms of compensatory strategies, Bonferroni's correction was applied to the p-value, with $p \leq 0.05/6$, hence $p < .008$. As can be seen in Table 4.16, only Item 54, "*To understand unfamiliar English words, I make guesses*", yielded a significant difference. Similarly, in the qualitative findings, none of the strategies that the students mentioned in the interview were related to the compensatory type. All of the language learning strategies in this category showed an increase in usage from before the MUET to after the test. It is safe to assume that students might need to rely on these strategies more at university, as they are more exposed to English in higher education than at school.

Table 4.16

Students' language learning strategies before and after the MUET (compensatory)

No.	Item	Before (%)	After (%)	Sig.
54.	To understand unfamiliar English words, I make guesses.	72	78	.007*
55.	When I can't think of a word during a conversation in English, I use gestures.	75	77	.617
56.	I make up new words if I do not know the right ones in English.	57	63	.018
57.	I read English without looking up every new word.	49	49	1.00
58.	I try to guess what the other person will say next in English.	66	71	.080
59.	If I can't think of an English word, I use a word or phrase that means the same thing.	79	80	.719

After Bonferroni's correction was applied, the new p-value for the metacognitive group was $p < .005$. Similar to compensatory strategies, metacognitive strategies (see Table 4.17) also show a statistically significant increase for only one item, Item 61 "*I notice my English mistakes and use that information to help me do better*".

Table 4.17

Students' language learning strategies before and after the MUET (metacognitive)

No	Item	Before (%)	After (%)	Sig.
60.	I try to find as many ways as I can to use my English.	77	83	.014
61.	I notice my English mistakes and use that information to help me do better.	79	87	.001*
62.	I pay attention when someone is speaking English.	84	86	.405
63.	I try to find out how to be a better learner of English.	83	88	.045
64.	I plan my schedule so I will have enough time to study English.	43	46	.401
65.	I look for people I can talk to in English.	69	68	.766
66.	I look for opportunities to read as much as possible in English.	71	73	.635
67.	I have clear goals for improving my English skills.	60	67	.014
68.	I think about my progress in learning English.	74	81	.006

Purpura (1999) established that metacognitive strategies have “a significant, positive, direct effect on cognitive strategy use, providing clear evidence that metacognitive strategy use has an executive function over cognitive strategy use in task completion” (p. 61). Further analysis of the use of metacognitive strategies revealed that although the increase in the use of most items was not statistically significant, those with a noteworthy increase appeared to be related to the students’ own realization of and reflection on the need to improve their English for themselves by trying to find as many ways as they could to use their English (see Item 60) and be a better learner (see Item 63). There was also an increase for Item 67 “*I have clear goals for improving my English skills*” and Item 68 “*I think about my progress in learning English*”. Now that these students were at university, their English language learning goals were no longer prioritized based on their performance in language exams. It has to be noted here that a change in language learning context will strongly influence the choice of language learning strategies. This, to a certain extent, would

have minimized the impact of any washback deemed left from the MUET, leading to the question “Is it possible to measure washback length?” This will be discussed further in Chapter 5.

The analysis of the qualitative data showed that some of the students regarded language learning at university as more challenging:

[It is] very different because in the university, we use a higher level of English. There are many new words and terms used at this level. Extra preparation is very much needed.

(Student 35)

These students regarded the level of English used at university as higher compared to when they were at school. They stated that they needed to familiarize themselves with and learn new terms and vocabulary, making them feel that they should work harder at university. Another reason raised by the students who used different language learning strategies at university concerned practical purposes. English is the medium of instruction at most public universities in Malaysia. University students are not only required to use English within English language classrooms, but also for other courses, even if they are not English major students. Lectures and tutorials are commonly conducted in English and the students have to do presentations and undertake assignments in English. Most resources, such as books and lecture notes, are also in English. Thus, it appeared that the frequent exposure to English at university influenced the students, leading them to think that mastery of English was a necessity, rather than just being important to pass a test. The students seemed to have different learning goals at university as they emphasized the practicality aspect as opposed to getting a high score on a test. One skill that stood out was “speaking”:

So for now, I focus more on using English when I speak since I am at university.

(Student 34)

But now, I interact more with my other friends to improve my English.

(Student 38)

Concerning affective strategies, interestingly none of the items were statistically significantly different from “before the MUET” to “after the MUET”, as shown in Table 4.18, with $p < .0083$ after the application of Bonferroni’s correction.

Table 4.18

Students’ language learning strategies before and after the MUET (affective)

No	Item	Before (%)	After (%)	Sig.
69.	I try to relax whenever I feel afraid of using English.	75	80	.091
70.	I encourage myself to speak English even when I am afraid of making a mistake.	80	85	.100
71.	I give myself a reward or treat when I do well in English.	48	52	.109
72.	I notice if I am tense or nervous when I am studying or using English.	63	67	.216
73.	I write down my feelings in a language learning diary.	32	31	.839
74.	I talk to someone else about how I feel when I am learning English.	51	53	.441

Affective strategies deal exclusively with the students’ individual traits that directly affect learning strategies, such as identifying one’s mood and anxiety level, talking about feelings, rewarding oneself for good performance and using deep breathing or positive self-talk. The students mentioned that learning at university was more relaxing and enjoyable than it had been at school:

Yes, English language learning in the university is easier to understand and more fun.

(Student 44)

The English language usage now in the university is more relaxing and not burdensome.

(Student 47)

One interesting point of note is that the students mentioned their current English language learning not being “burdensome”. This might result from the relief of the pressure that the MUET imposed on them due to its high stakes and significant consequences. As discussed in Chapter 2, a certain amount of pressure can be beneficial depending on the individual, but it can also be detrimental when excessive.

As depicted in Table 4.19, the students did not report great changes when it came to the use of social strategies, despite learning in the university environment requiring greater use of English, both inside and outside of class.

Table 4.19

Students' language learning strategies before and after the MUET (social)

No.	Item	Before (%)	After (%)	Sig.
75.	If I do not understand something in English, I ask the other person to slow down or say it again.	84	85	.678
76.	I ask English speakers to correct me when I talk.	66	71	.015
77.	I practise English with other students.	72	70	.626
78.	I ask for help from English speakers.	69	68	.596
79.	I ask questions in English.	51	56	.134
80.	I try to learn about the culture of English speakers.	66	69	.440

Similarly, 27 respondents mentioned that they were still using the same strategies that they used when preparing for the MUET for their current studies at university:

Most of the English language learning strategies that I use now and when I was preparing for the MUET are not much different.

(Student 20)

The strategies that I used in the past and present are the same in learning English.

(Student 7)

They mentioned that as the four language skills tested in the MUET were the same as those they used at university, they did not see the need to change their learning strategies:

There is no difference in English language learning when preparing for the MUET as it is based on reading, listening, writing and speaking.

(Student 19)

The way I study now is the same as what I did for the MUET. For me, the four language aspects, reading, listening, speaking and the other one, are very suitable to improve my English language mastery.

(Student 40)

To conclude, it would appear that some of the respondents were still employing the same learning strategies as those when preparing for the MUET in the past. However, the intensity was somewhat lower than when they were preparing for the MUET due to the different stakes and consequences.

4.3.2 Washback intensity

The literature on washback suggests that its intensity depends on the stakes of the test (Cheng, 1998; Watanabe, 2004). It also relates to how much power the test has in influencing the teaching and learning that takes place in the process of preparing for a given test. In addition, as posited by

Green (2007a) in his washback model, perceptions of test importance and test difficulty will influence the intensity of washback for the participants. However, while the relationship between perceptions and intensity illustrated in Green's (2007a) washback model might seem straightforward, this is not the case due to the complexity of the phenomenon. Hence, the third research question aimed also to examine the relationship between the students' perceptions and washback intensity. Specifically, it aimed to examine the washback intensity of the MUET based on perceived test importance (see 4.1.1 and 4.2.1) and perceived test difficulty (see 4.1.2).

As washback is a complex phenomenon, it was also deemed necessary to consider other variables, namely the students' English proficiency level and their gender, as these may or may not contribute to the washback intensity of a test. Together with the main variables of the study – perceived test importance, perceived test difficulty and perceived self-efficacy – these were analysed using statistical tools to see if they contributed to the overall scores for use of language learning strategies in preparing for the MUET. The analysis of the quantitative data from the student questionnaire is presented first, followed by analysis of the qualitative data.

Hierarchical multiple regression was run to determine if the addition of perceived test importance, perceived test difficulty and self-efficacy improved the prediction of students' language learning strategies when preparing for the MUET over and above students' English proficiency and gender (see Table 4.20 for full details of each regression model).

Table 4.20

Regression analysis

Model	R	R ²	Adjusted R ²	Std. Error of the estimate	Change Statistics				
					R ² change	F change	df1	df2	Sig. F change
1	.107 ^a	.011	.004	.65273	.011	1.567	1	135	.213
2	.329 ^b	.108	.095	.62236	.096	14.496	1	134	.000*
3	.372 ^c	.138	.119	.61394	.030	4.704	1	133	.032*
4	.382 ^d	.146	.120	.61358	.007	1.153	1	132	.285
5	.397 ^e	.158	.126	.61159	.012	1.861	1	131	.175

a. Predictors: (Constant), English proficiency

b. Predictors: (Constant), English proficiency, gender

c. Predictors: (Constant), English proficiency, gender, importance

d. Predictors: (Constant), English proficiency, gender, importance, difficulty

e. Predictors: (Constant), English proficiency, gender, importance, difficulty, self-efficacy

f. Dependent variable: SILL

The full model of English proficiency, gender, perceived test importance, perceived test difficulty and perceived self-efficacy estimated in predicting students' language learning strategies was statistically significant ($R^2 = .158$, $F(5, 131) = 4.911$, $p < .0005$; adjusted $R^2 = .126$). The addition of gender (Model 2) and test importance (Model 3) in predicting the use of language learning strategies led to a statistically significant increase ($R^2 = .096$, $F(1, 134) = 14.496$, $p < .0005$) for model 2 and ($R^2 = .119$, $F(1, 133) = 4.704$, $p = .032$) for model 3. The addition of test difficulty (model 4) and self-efficacy (model 5) in predicting the use of language learning strategies led to an increase ($R^2 = .007$, $F(1, 132) = 1.153$, $p = .285$) for model 4 and ($R^2 = .012$, $F(1, 131) = 1.861$, $p = .175$) for model 5, but the increase was not statistically significant.

It would appear from the analysis that gender and perceived test importance were the strongest contributors to the students' frequency of use of language learning strategies when preparing for the MUET. However, care must be taken in interpreting these data as the division of

the groupings for gender (male and female) and proficiency level (high and low) were uneven due to the poor nature of the data collected (addressed further in the conclusions in Chapter 6).

SECTION II: MUET WASHBACK FROM THE TEACHERS' PERSPECTIVE

As pointed out by Alderson (2004), teacher-related factors play a crucial role in shaping the washback effect of a test, as their beliefs and understanding of the nature and rationale for the test will to a certain extent influence the way in which they prepare their students for the test. Although the main focus of this study was on the students, data from teachers were also collected using a questionnaire for triangulation purposes. It was decided to separate the student and teacher data for ease of organization and understanding due to the volume. However, these findings are discussed collectively in Chapter 5.

The aim of the teacher questionnaire was to obtain information on teachers' perceptions of the MUET and their attitudes towards aspects of learning when it came to preparing the students for the test. Both quantitative and qualitative data were obtained through the teacher questionnaire (see Appendix C). The teacher questionnaire was distributed online using Qualtrics, an online questionnaire system; this was feasible in terms of the time constraints on the study.

Teachers preparing students for the MUET were identified and contacted via email. These teachers included those teaching at the two schools from which the student participants from Group A were recruited and other schools with a similar educational setting. In all, 55 teachers were contacted and 36 responded to the online questionnaire, representing a 65.5% return rate. Demographic information concerning the teachers was provided in 3.4.2.

The questionnaire consisted of two main sections, closed and open-ended questions. The second part of the questionnaire comprised four open-ended questions, discussed in the section on

qualitative findings. Thus, qualitative data were obtained from the open-ended questions in the teacher questionnaire and classroom observation. In all, 23 out of the 36 teachers who participated in the teacher questionnaire responded to the open-ended questions, which were optional. The purpose of this was to obtain as many respondents as possible and there was a possibility that the teachers might refuse to respond to the questionnaire if it was deemed too demanding.

This section covers the four main categories of data obtained from the teacher questionnaire. First, the teachers' perceptions of the MUET are discussed, followed by their attitudes towards aspects of learning when preparing their students for the MUET. The data from these two sections are used to complement the student data in addressing the first research question concerning students' perceptions of the test and the factors that might influence these. Next, to understand the process of the washback, it is crucial to document what goes on in the classroom, both from the students' and the teachers' perspectives. Hence, the following sections cover the teachers' medium of instruction, difficulties experienced in teaching the MUET and the classroom observation data.

4.4 Teachers' perceptions of the MUET

The findings concerning teachers' perceptions of the MUET were derived from the open-ended questions in the teacher questionnaire. The findings are discussed in terms of (i) teachers' perceptions of the MUET as a university entrance exam and (ii) teachers' perceptions of the influence of the MUET on their students' futures. The teachers were first asked to give their opinions concerning the MUET as a university entrance exam. In terms of the validity and reliability of the MUET, most of the teachers agreed that the results could be used to gauge or assess students' English language proficiency as the test covers all four language skills. Some

teachers mentioned that having the information on their students' scores for each language skill enabled them to target specific skills that needed greater emphasis:

The MUET is able to compartmentalize my students' strengths more specifically into the different language skills, hence directing which areas require more emphasis.

(Teacher 14)

For instance, if most students were weak in writing skills, the teacher would focus on these based on the students' needs over and above the other language skills. According to the teachers, more often than not, the MUET band did reflect the students' actual English language ability:

I think MUET does that. Reasons? 1) MUET is under the control of the Majlis Peperiksaan Malaysia, a strong and experienced national exam council. 2) I've been a MUET speaking examiner for about 3 sessions and hence I dare say that the process of evaluating the students is pretty valid. 3) So far, my students' MUET scores do reflect their actual proficiency in real life.

(Teacher 2)

Teacher 2, who was a MUET speaking examiner, confirmed the validity of the MUET as monitored and administered by the Malaysian Examinations Council (MEC), established on 1 February 1980 under the Malaysian Law and Examinations Council Act 225. He also emphasized his belief in the reliability of the MUET band in terms of gauging students' mastery of English. Another teacher mentioned that since the MUET was developed specifically in the Malaysian context, the results are more valid compared to other international high-stakes test (e.g. IELTS, TOEFL, etc.) for the following reason:

Yes. Given that most, if not all, the materials in MUET are localized to the Malaysian setting, the MUET is the appropriate medium for assessing my students' English ability since they are not highly exposed to native speakers' ways of communicating.

(Teacher 15)

However, two teachers raised concerns that students' performance in the MUET also depends on their general knowledge, aside from their English language skills:

To some level, yes. But, sometimes it's just too exam oriented and the MUET not only tests students' English but also their general knowledge. If the students are lazy about reading, they will not do well in the exam.

(Teacher 4)

Sometimes it depends on the questions. Students will answer or speak better if the topic is within their content knowledge.

(Teacher 11)

Moreover, some teachers disagreed that the MUET measured the students' English language proficiency accurately. One teacher mentioned that the questions for the MUET reading component were too difficult for the students:

No for reading because it was bloody difficult. The articles can be considered authentic, but the MCQs [multiple choice questions] are so confusing. Not sure if the students are going to encounter crazy questions like the MUET reading in life.

(Teacher 5)

According to Teacher 5, it seemed unlikely that the students would encounter such difficult language materials as those presented in the reading component in real life. Another teacher described the reading questions in the MUET as “*total murder*”:

The MUET is a good evaluation of language proficiency and highly relevant but please tone down on the reading aspect. Not only does it carry the biggest weighting, but also the questions are total murder. If language practitioners find it difficult, how more so the students? Especially the low-intermediate students.

(Teacher 22)

According to Teacher 22, the reading items need to be revised as even he, a teacher, found them difficult. This argument was supported by Teacher 14, who mentioned that the test was more aimed at “*...testing whether the students are able to achieve understanding of perfect native-like texts*”.

Another teacher noted that the length of the reading test made it problematic:

To me, the reading test is a bit too long. Currently we are having 6 reading passages and 45 questions to answer. I would do it just with 4 passages but the number of questions should remain unchanged.

(Teacher 10)

It was interesting that the teachers, especially Teacher 5, mainly brought up the reading tasks in the MUET as being the most difficult. They were very keen on these being revised:

Reading also make students fail to get a good band. Hopefully, the multiple-choice question format can be amended accordingly.

(Teacher 5)

Two other teachers added:

As I have mentioned, I just hope the system is more [geared] to [assessing] students' ability to comprehend and perform understandable English in a daily setting.

(Teacher 9)

However, it is to be noted that the goal of the test is to ensure students are able to comprehend the language, not to achieve native-like mastery of the language, which for me is a bit unrealistic.

(Teacher 14)

In a similar vein, another teacher mentioned that the questions in the MUET, specifically the listening component, were staged and not applicable to real-life situations. This might be due to the type of the taped conversations used in the recordings, discussed further in Chapter 5.

When it came to the social status given to the MUET as one of the requirements for gaining entry to Malaysian public universities, the teachers seemed supportive of using it as a university entrance test because they felt that it assisted the students in surviving the English language demands of university. In tertiary education, students can no longer simply listen to the teacher in the classroom as they did at school. As undergraduates, they are required to make presentations, participate in discussion forums, conduct research, write reports, etc., which involve all four language skills. Teacher 3 mentioned that he knew a few students who quit their studies because they could not cope with the English language demands of university:

Of course, it is necessary. If one's English proficiency is very weak, how can one survive at university? Most of the lecturers and professors conduct classes in English. I do know a few students who quit university as their English level was very weak.

(Teacher 3)

This shows that the mastery of English is not only applicable for English language majors, but also students on other courses. Furthermore, resources and references are mostly published in English:

Yes, I think it's absolutely necessary because at university many references and journals are in English. A lack of proficiency in English means that the student will probably not be able to cope with his/her studies.

(Teacher 2)

Based on English being regarded as the second language in Malaysia and the global nature of the world today, the teachers noted the importance for students to have a high mastery of English:

I think university students must have a certain English standard to be able to function globally; hence yes, it is necessary.

(Teacher 13)

Using the MUET as one of the requirements for university entry, it was hoped that Malaysian graduates would be helped to excel in their respective fields and to function globally. Besides being used to measure students' English proficiency, the MUET results can also be used by both lecturers and the students themselves to identify specific skills that need more attention and improvement once they are at university.

One of the aims of this study was to investigate the washback intensity of the MUET. To explore this from the teachers' perspective, they were asked to comment on possible influences of the students' performance in the MUET on their future. The data revealed that most teachers agreed that the students' MUET results would influence their future if they intended to undertake further study at Malaysian public universities:

However, most students in my institution just take MUET for the sake of university entrance.

(Teacher 8)

According to some of the teachers, as the MUET results were only used for university entrance, there would be no effect if the students had plans other than pursuing degree studies:

Apart from that, if a student wants to be a successful Ramli Burger entrepreneur, I don't think MUET is necessary at all.

(Teacher 2)

Students who do not sit for MUET could still look for a job. Students who take the MUET are those who intend to further their study at university.

(Teacher 17)

However, when it comes to the students' future undertakings in terms of their career pathways, although they do not need their MUET results to apply for jobs, they still need them to get into their desired course. The kind of course they do in their degree will then influence their career pathway to a large extent:

The MUET influences the type of course and direction that they may embrace in terms of their academic development.

(Teacher 14)

Moreover, 4 out of 23 teachers did not view the MUET solely as a university entrance test. They viewed the MUET as a stepping stone for the students to enhance their English language mastery and motivating the students to use English was part of this:

Students have more motivation to speak English in the classroom.

(Teacher 3)

Even students who have lower proficiency and dislike the subject feel compelled to attend MUET classes in order to be able to answer the questions. Some learning takes place.

(Teacher 7)

Irrespective of the goals the students had in mind when preparing for the MUET, at the very least, they ended up practising the language. One teacher remarked that the process of taking the test in itself would benefit the students in the long term:

Despite constant reminders given by teachers about the importance of English, I think when they are in the process of taking the test, the real realization occurs in many students.

(Teacher 7)

Making the students realize the important of English in preparing for the test would thus benefit the students in the long term, as this would help boost their motivation to learn intrinsically rather than extrinsically.

4.5 Teachers' attitudes towards aspects of learning

This section on teachers' attitudes towards aspects of learning is divided into four sub-sections: (i) recommended learning strategies, (ii) proposed learning activities, (iii) strategies for motivating students to learn and (iv) teachers' perceptions concerning the use of mock tests.

4.5.1 Recommended learning strategies

The teachers were asked “*What are the learning strategies you would recommend to your students to prepare for the MUET?*” On a 5-point Likert-type scale, with 1 indicating “strongly disagree” and 5 indicating “strongly agree” (see Appendix C), the teachers were asked to rate their level of agreement or disagreement with the use of learning strategies listed. Again, for ease of data interpretation and presentation, the options for “strongly disagree” and “disagree” were grouped as “disagree” and the option for “agree” and “strongly agree” were grouped as “agree”. This applies to all charts in this dataset. Although the data were ordinal in nature and calculating the mean score is regarded as meaningless by some academics depending on their view of measurement theory, the means and standard deviations of ordinal data have the advantage of revealing statistical differences within smaller sample sizes and generate fruitful results. In this case, the mean score is also presented to help explain the findings in greater detail with care not to make any interval or ratio statements about the data. Nine items pertaining to learning strategies recommended by the teachers are listed in Table 4.21, according to the mean score in descending order.

Table 4.21

Teachers' recommended learning strategies (Q3)

Item ranking	N	Mean	Std. Dev.	Disagree (%)	Undecided (%)	Agree (%)
9. To communicate more in English	36	4.83	.447	0	3	97
7. To use English more in their daily life	36	4.81	.401	0	0	100
3. To learn to express their opinions in class	36	4.78	.422	0	0	100
8. To change from passive learning to active learning	36	4.69	.525	0	3	97
6. To be more active in classroom participation	36	4.64	.543	0	3	97
2. To expose themselves to various English media	36	4.61	.645	3	0	97
4. To put more emphasis on listening and speaking	36	4.42	.770	3	8	89
5. To learn to initiate questions	36	4.36	.762	3	8	89
1. To learn to take better notes	36	3.86	.931	14	8	78

From the high overall mean score of the strategies listed in Table 4.21, it can be seen that the teachers viewed all these strategies as important for the students in preparing for the MUET. With mean scores of 4.83 and 4.81 respectively, the teachers regarded “*To communicate more in English*” (Item 9) and “*To use English more in their daily life*” (Item 7) as the most important learning strategies that they would recommend to their students. The rest of the strategies yielded high to very high mean scores ranging from 4.36 (Item 5 – *To learn to initiate questions*) to 4.78 (Item 3 – *To learn to express their opinions in class*), suggesting that these strategies could be regarded as pragmatic ways for students to cope with the demands of the MUET as the test covers all four language skills. The learning strategies with the lowest mean score (3.86) was “*To learn to take better notes*” (Item 1). It appears that the teachers did not view taking better notes as a highly recommended learning strategy for students preparing for the MUET.

4.5.2 Proposed teaching activities

The teachers were also asked about their perceptions of the types of activities they thought should be involved in language learning in the context of the MUET preparation class. Specifically, they were asked to reflect on the kinds of teaching activities they employed during their MUET preparation classes. Table 4.22 presents seven activities according to the mean scores in descending order of perceived significance.

Table 4.22

Teachers' proposed teaching activities (Q4)

Item ranking	N	Mean	Std. Dev.	Disagree (%)	Undecided (%)	Agree (%)
3. Role play and group discussion	36	4.67	.478	0	0	100
4. Exposure to various English media	36	4.56	.695	3	3	94
1. Task-oriented activities	36	4.53	.560	0	3	97
5. Authentic materials	36	4.44	.652	3	0	97
2. Language games	36	4.39	.728	3	6	92
6. Training in basic language knowledge	36	4.33	.586	0	6	94
7. Extracurricular activities	36	3.58	.996	14	36	50

It was found that the mean scores for teaching activities that were communicative in nature topped the ranking, specifically with role play and group discussion (Item 3, mean = 4.67). Exposure to various English media (Item 4) came second with a mean score of 4.56, indicating the need for students to be exposed to various media so they could exercise all four language skills. Having been exposed to written materials over the 11 years of formal English education at school, the teachers thought that for the MUET there was a need to introduce a variety of English media, including audio and visual materials, to engage the students with all four language skills. The item with the lowest mean score (3.58) was Item 7 “*Extracurricular activities*”, which refers to activities pursued in addition to the normal course of study.

4.5.3 Strategies to motivate students to learn

The teachers were asked in what ways they would like to motivate their students to learn English.

There were eight items under this category, as presented in Table 4.23.

Table 4.23

Teachers' methods for motivating students to learn (Q5)

Item ranking	N	Mean	Std. Dev.	Disagree (%)	Undecided (%)	Agree (%)
6.To create a positive attitude toward language learning	36	4.69	.467	0	0	100
7.To provide students with effective language learning strategies	36	4.64	.487	0	0	100
5.To give students more encouragement to learn	36	4.56	.504	0	0	100
3.To organize real-life language activities	36	4.50	.609	0	6	94
4.To do more interesting language games	36	4.36	.723	0	14	86
2.To use more authentic materials	36	4.25	.732	6	0	94
8.To have better classroom discipline	36	3.97	.910	8	17	75
1.To do more mock exam papers	36	3.75	1.156	19	6	75

The first three items topping the ranking were those related to providing encouragement for the students to learn (Item 5, mean = 4.56) and creating a positive attitude to language learning (Item 6, mean = 4.69). Since the students who were preparing for the MUET were young adults, the teachers perceived it would be better to encourage them by making them realize the importance of English for their future undertakings rather than focusing on better classroom discipline (Item 8, mean = 3.97), which might not be an effective approach with young adults as opposed to adolescents or children. Surprisingly, doing more mock exam papers was the strategy least used

to motivate the students (Item 1, mean = 3.75). The use of mock exams is discussed in detail in the following section.

As mentioned previously, the qualitative findings revealed that teachers considered the pressure associated with sitting the MUET could motivate and encourage students to take their English language learning more seriously. It is common for students who are sitting a high-stakes test to experience test anxiety to a certain extent and this can be either beneficial or harmful. Indeed, the findings from the students' interviews confirmed that the pressure they felt prior to the exam was beneficial for them as they described this as a "positive challenge".

4.5.4 Perceptions of the use of mock tests

The teachers were also asked about their perceptions of the basic function of mock tests, the results of which are presented in Table 4.24.

Table 4.24

Teachers' perceptions of mock tests (Q6)

Item ranking	N	Mean	Std. Dev.	Disagree (%)	Undecided (%)	Agree (%)
5.To prepare students for public examination	36	4.50	.507	0	0	100
6.To identify area for re-teaching	36	4.44	.607	0	6	94
1.To give feedback to teachers	36	4.44	.504	0	0	100
2.To assess students' learning difficulties	36	4.42	.649	3	0	97
4.To direct students' learning	36	4.17	.941	8	3	89
3.To motivate students	36	3.58	1.105	17	31	53

Observing the patterns in the mean scores of the items presented in Table 4.24, it can be seen that the teachers perceived mock tests as a good tool for preparing students for a public examination

(Item 5, mean 4.50), in this case the MUET. They also thought that mock tests would be able to provide feedback (Item 1, mean = 4.44), identifying areas for re-teaching (Item 6, mean = 4.44) and assessing students' learning difficulties (Item 2, mean 4.42), so that they would know how to direct students' learning (Item 4, mean = 4.17). Item 3 "*To motivate students*" had the lowest mean score, 3.58, validating the findings from the previous section, namely that the teachers did not use mock tests to motivate their students.

4.6 Teachers' medium of instruction

As mentioned previously, to understand the process of washback, it is crucial to document what goes on in the classroom, not only from the students' perspective, but also from the teachers' perspective. Hence, the teacher questionnaire (see Appendix C) sought to explore the teachers' medium of instruction in terms of how they prepared their lessons, what they did in the classroom and what materials they used. The teachers were asked to evaluate each item on a 5-point Likert-type scale of frequency, anchored at 1 denoting "never" and 5 denoting "always". The following sections discuss each category in detail, beginning with the teachers' lesson preparation.

4.6.1 Lesson preparation

Seven items were employed to explore the aspects the teachers considered when preparing their lessons. The results are presented in Table 4.25.

Table 4.25

Teachers' lesson preparation (Q1)

Item ranking	N	Mean	Std. Dev.	Never (%)	Rarely (%)	Sometimes (%)	Often (%)	Always (%)
3.Tasks to be performed in teaching	36	4.69	.467	0	0	0	31	69
2.Content of teaching	36	4.67	.478	0	0	0	33	67
4.Skills to be taught	36	4.53	.560	0	0	3	42	56
6.How to motivate students to learn	36	4.47	.654	0	0	8	36	56
5.Any supplementary materials to be used	36	4.39	.728	0	3	6	42	50
1.Methods of teaching	36	4.22	.681	0	0	14	50	36
7.Homework given to students	36	3.50	1.028	0	19	31	31	19

According to the results in Table 4.25, the teachers appeared to pay more attention to the content of teaching (Item 2, mean = 4.67) as opposed to the methods of teaching (Item 1, mean = 4.22). Similar findings were reported in the previous washback literature, namely that teachers tend to change the content of teaching rather than their methodology when they teach for high-stakes exams. The findings also showed that the teachers were least concerned about giving homework to the students (Item 7, mean = 3.50). This might be due to the students' heavy study load, as they still need to prepare for their major subjects, which might require greater attention.

4.6.2 Teaching activities

To explore how often the teachers carried out the following activities in the MUET preparation class, aside from conducting classroom observation to collect data (discussed in 4.8), the teachers were also asked to rate nine activities on a 5-point Likert-type scale, anchored at 1 = never and 5 = always. The item ranking is presented in Table 4.26.

Table 4.26

Teaching activities (Q2)

Item ranking	N	Mean	Std. Dev.	Never (%)	Rarely (%)	Sometimes (%)	Often (%)	Always (%)
2.Demonstrate how to do particular language activities	36	4.36	.639	0	0	8	47	44
8.Organize group work or discussion	36	4.28	.815	0	6	6	44	44
1.Tell the students the aims of each lesson	36	4.22	.929	0	8	8	36	47
9.Organize integrated language tasks	36	4.14	.833	0	6	11	47	36
4.Explain specific language items such as words or sentences	36	3.97	.941	0	6	28	31	36
5.Explain textbook exercises	36	3.86	.899	0	8	22	44	25
3.Explain the meaning of the text	36	3.83	.878	0	6	31	39	25
7.Organize language games	36	3.72	1.059	0	14	31	25	31
6.Explain mock exams	36	3.67	1.042	3	11	25	39	22

From the findings, it can be seen that activity Item 2, “*Demonstrate how to do particular language activities*”, was implemented most by the teachers in the MUET preparation classrooms. This was closely followed by “*Organize group work or discussion*” (Item 8, mean = 4.28) and “*Tell the students the aims of each lesson*” (Item 1, mean = 4.22). The item that was associated with mock exams, Item 6, “*Explain mock exams*”, had the lowest mean score of 3.67. The overall mean score for all nine items indicated that the teachers frequently conducted various activities in the MUET preparation classroom besides teaching to the test, as also reported in the qualitative data and the classroom observation data.

One teacher mentioned that it seemed logical to teach to the test since the course itself was a MUET preparatory course:

Since I'm teaching a preparatory MUET class, of course I need to streamline the syllabus with my method of teaching and students' learning.

(Teacher 23)

Some of the teachers also mentioned that the MUET made their classes more exam-oriented, as they felt they needed to teach the students what was going to be tested in the MUET:

Yes. Because I tend to focus on the skills required in the MUET more.

(Teacher 4)

Yes, my teaching style is more exam oriented, focusing more on techniques, but integrating authentic materials.

(Teacher 16)

However, these teachers did not view teaching to the test as something negative. According to Teacher 22, the goal was to help the students do well in the test:

Yes, it becomes more exam-oriented as I use a lot of books and exercises to refer to. It is a good thing though, because the aim now, for them, is to get flying colours in their tests or exams.

(Teacher 22)

Most of the teachers who responded to the open-ended question (15 out of 23) agreed that the MUET influenced their teaching to a certain extent, especially the speaking component. As mentioned in the literature review (Chapter 2), past washback research indicates that teachers tend to change the content of their teaching, rather than their methodology. Most of the teachers in this study stated that they tended to emphasize speaking skills in their English classes to expose students to what would be involved in the MUET:

Yes, especially speaking skills. In my class, students are exposed to group discussion. This activity is relevant as it acts as exposure for students who have never taken MUET before.

(Teacher 10)

The MUET speaking component also encouraged the teachers to instil communication skills by conducting more interactive activities, thus “forcing” the students to speak more:

Yes, especially on the speaking component. In my class, I focus more on communicative English, so students are required to speak more and most tasks are geared towards speaking. Perhaps in school before this, students have not communicated much, so I will find activities that will force them to speak.

(Teacher 8)

Well, as a matter of fact, it does influence my method in teaching my English classes because it can lead my classes to engage in more interactive skills, most probably towards the oral communication skills.

(Teacher 6)

From the data, it appeared that the MUET speaking component encouraged, at least to a certain extent, more interactive and communicative activities in the classroom. However, the teachers did

not generally mention changing their methodology, but rather adding more language learning and teaching activities to cater to their students' needs in preparing for the MUET.

4.6.3 Teaching materials and resources

The teachers were also asked to indicate the use of the following teaching and learning aids in their teaching. The items are listed in Table 4.27 according to their mean scores in descending order.

Table 4.27

Teaching materials and resources (Q3)

Item ranking	N	Mean	Std. Dev.	Never (%)	Rarely (%)	Sometimes (%)	Often (%)	Always (%)
2. Supplementary materials	36	4.28	.741	0	3	8	47	42
7. Teaching syllabus	36	4.06	.893	0	6	19	39	36
8. Examination syllabus	36	3.94	1.013	3	3	28	31	36
1. Textbooks	36	3.75	1.079	3	6	39	19	33
6. Picture and/or cards	36	3.42	1.131	6	14	33	28	19
4. Newspapers	36	3.28	.974	3	14	50	19	14
3. Television/radio	36	2.97	1.207	14	19	33	22	11
5. Language laboratory	36	2.72	1.523	31	19	17	14	19

The teachers appeared to use more supplementary materials (Item 2, mean = 4.28) as opposed to textbooks (Item 1, mean = 3.75), television or radio (Item 3, mean = 2.97) and language laboratories (Item 5, mean = 2.72), which were the least used teaching and learning resources. This is understandable as not all pre-university institutions in Malaysia are equipped with language learning laboratories.

4.7 Teachers' difficulties in teaching the MUET

In terms of perceived difficulties in teaching the MUET, the quantitative findings (see Table 4.28) revealed that most of the teachers seemed to strongly agree that students' current English level (Q2, Item 1), with a mean score of 4.69, was the biggest obstacle or challenge when teaching the MUET.

Table 4.28

Teachers' difficulties in teaching the MUET (Q2)

Item ranking	N	Mean	Std. Dev.	Disagree (%)	Undecided (%)	Agree (%)
1.Students' current English level	36	4.69	.525	0	3	97
7.Inadequate time for students to practise English outside the language classroom	36	4.36	.899	8	3	89
6.Too heavy work load	36	3.86	1.046	14	19	67
2.Class size	36	3.78	1.333	25	8	67
5.Lack of teaching and learning aids and facilities	36	3.17	1.159	33	17	50
3.Inadequate textbooks and other teaching resources	36	2.75	1.180	56	17	28
4.Noisy learning environment	36	2.39	1.076	64	25	11

Similar findings were found in the qualitative data. Aside from aligning their teaching styles with the syllabus and instrumental goals, such as passing the MUET with flying colours, the teachers also took into consideration the students' level of proficiency when planning lessons:

The MUET influences how I should set my parameters in teaching them, so that it aligns with their current level of understanding of the language.

(Teacher 21)

This could be attributed to the teacher's effort to bridge the language gap between students of various proficiency levels in the classroom as there is no point of using high-level English in the class if the students are not able to follow the lesson. The quantitative data also show that the teachers perceived there was inadequate time for the students to practise their English outside the language classroom (Item 7, mean = 4.36) which could be attributed to the students' obligations in terms of other subjects or courses aside from preparing for the MUET. The teachers' heavy workload (Item 6) also contributed to the difficulties they faced when teaching the MUET. A noisy learning environment (Item 4, mean = 2.39) and inadequate teaching resources (Item 3, mean = 2.75) were not considered particularly problematic by these teachers.

However, surprisingly, class size seemed to be considered an aspect of difficulty by the teachers, with a mean score of 3.78. In this study, the average class size was 25 to 30 students. Although the teachers were not explicitly asked to describe how class size affected their teaching, it appeared that smaller class sizes were preferable, especially when teaching speaking skills. This deduction was based on the classroom observation data, as discussed later. From those data, it can be seen that the teachers were barely able to cater to all students when it came to speaking activities in the classroom and thus they resorted to having the students work in groups.

4.8 Classroom observation

Classroom observation was used in this study to explore the extent to which the MUET influenced (or not) how the three participating teachers carried out their MUET preparation classes. The data collected employing Part A of the COLT scheme (Spada & Frohlich, 1995) in five different MUET preparation classes are described and examined according to the three main categories of the

observation scheme: (i) participant organization, (ii) activities conducted in class and (iii) teaching materials.

As mentioned in Chapter 3 on the methodology of the study, the COLT scheme was used in this study as it provides an insight into what happens in the classroom, covering the kind of activities employed, the interactions and the content of teaching and learning. This study examined five MUET preparation classes: Class 1 conducted by Teacher 1, Class 2 conducted by Teacher 2 and Classes 3, 4 and 5 conducted by Teacher 3. It should be noted here that due to time constraints, only five classes were observed for this study and they all happened to be focusing specifically on speaking skills on the days of observation. Hence, the classroom observation data are rather limited, addressing the teaching and learning of speaking only.

4.8.1 Participant organization

Participant organization refers to the person doing the talking during the segments of the lesson as a percentage of class time (Spada & Frohlich, 1995). The participants involved in each class were the teacher and students.

Table 4.29

Participant organization

Name	Percentage of class time (%)*				Total
	Whole class		Group	Individual	
	T to S/C	S to S/C			
Class 1	34	22	22	22	100
Class 2	42	21	11	26	100
Class 3	33	17	25	25	100
Class 4	33	17	25	25	100
Class 5	39	23	15	23	100

T to S/C: Teacher to students or the whole class

S to S/C: Students to students or the whole class

*Based on a total of 100%

As can be seen from Table 4.29, the teachers in all the observed classes relied heavily on teacher-to-student or whole class (T to S/C) interaction (33–42%) rather than interaction between the students (S to S/C), indicating predominantly teacher-focused instruction in the MUET preparation classes. However, there was also a “healthy dose” of student-to-student interaction in the classroom, ranging from 17% to 23%, indicating that the teachers also applied a learner-centred approach to a certain extent in their MUET preparation classes. The students were required to work in groups as well as individually in class time.

In all five classes, the seating arrangement was in clusters so that the students were placed in groups. Figures 4.4 and 4.5 show screenshots of the seating arrangements from the classroom observations.



Figure 4.4 Seating arrangement from classroom observation



Figure 4.5 Seating arrangement from classroom observation

This type of seating arrangement shows that the participants were involved in collaborative work in MUET preparation classes. For example, in Class 1, the students were required to discuss their individual speaking tasks in their respective groups before sharing them with the whole class. The same was observed in the other four classes as well, with the teachers allocating 10 to 25 minutes for the students to discuss their assigned tasks or topics first before sharing them with the rest of the class. Group discussion (MUET Speaking Task B) is part of the speaking component, in addition to a short individual presentation (MUET Speaking Task A). During the group discussion, the teachers kept reminding the students to use English rather than their mother tongue.

The students from Classes 1 and 2 appeared to follow the rule set by their teachers, i.e. only to use English during group discussion. Indeed, in Class 2, when Teacher 2 left the class for a while, the students continued to use English in their group discussion. However, in Classes 3, 4 and 5, the students used their mother tongue more than English when discussing their speaking tasks in groups. Teacher 3 stated that the students in his classes were low-proficiency students, which might explain their low use of English compared to the students in Classes 1 and 2. Teacher 3 used a combination of English and Malay in his three classes, roughly half and half. For example,

in Class 4, every time he uttered a full sentence in English that he assumed the students did not understand, he would re-state the same sentence in the mother tongue, Malay. He did this throughout the entire class. In contrast, Teachers 1 and 2 mostly used English throughout their classes and would only use the mother tongue occasionally.

4.8.2 Activities conducted in the MUET preparation classrooms

The types of teaching and learning activities conducted by the teachers were also explored in order to see how teaching and learning was realised in the MUET preparation classrooms in Malaysia. This data was utilised to determine if the teaching and learning activities in the MUET preparation classrooms were highly influenced or shaped by the test itself in an attempt to validate if the washback of the MUET exists and if it did exist, how strong or how weak it was.

Classroom activity types were grouped into two categories, teacher activities and student activities. Findings from the classroom observation scheme on activity types were reported as a percentage of class time to indicate what types of activity were given priority in the MUET preparation classes as follow:

Table 4.30

Activities conducted in the MUET preparation classrooms

Activity type	Class	Percentage of class time (%)				
		Class 1	Class 2	Class 3	Class 4	Class 5
Teacher Activities:						
Pre-lesson activities		10	17	17	17	14
Lecturing, explaining, and guiding		10	42	32	32	29
Student Activities:						
Individual work: Speaking		40	25	17	17	29
Group work: Discussion		-	8	17	17	14
Class: Presentation		40	8	17	17	14
Total		100	100	100	100	100

As mentioned earlier, since all the five classes observed in this study happened to be focusing on speaking skills, data for the other language skills which are reading, writing, and listening were not available for analysis in this study. As depicted in Table 4.30, teacher activities which consisted of pre-lesson activities and lecturing, explaining and guiding the students occupied almost half of the entire class time except for Class 1 with only 20% of total class time. Pre-lesson activities that were observed in the MUET preparation classrooms includes greetings, recalling previous lessons, housekeeping, and setting out the objectives of the lesson on that day. There were a lot of teachers' guiding observed in all five classes especially when the students were doing their group discussion and individual oral presentation to the class. These findings show that the classes were dominated by the teachers to a certain extent.

Student activities are activities in which all students either individual, group or the whole class participated or carried out in the MUET preparation class. For student activities, most time was spent on students' individual task which was speaking (17% - 40%). Although the speaking task was carried out individually, the whole class was involved in the task as the students discussed their points in their respective group first before one representative from each group presented their tasks either to the whole class (Class 1, Class 2) or to the other groups (Class 3, Class 4, Class 5). The remaining time was allocated for students' group discussion (8% - 17%) and presentation to the whole class (14% - 40%).

4.8.3 Teaching materials

As far as teaching materials is concerned, solely from the classroom observation, the teacher in Class 1 and Class 2 used commercialized MUET preparation textbook which contained model questions for all components tested in the MUET. This was indicated by Teacher 2 who asked the

students to take out their textbooks and refer to a specific page before conducting their speaking tasks. As for Teacher 3, he wrote the questions for the speaking tasks on the whiteboard in the classroom and asked the students to copy the questions before proceeding with group discussion. As mentioned in the earlier section, since all the classes observed happened to be focusing on the speaking skills, hence, the findings on the teaching materials utilised in the class was limited to one that is related to speaking skills only for this study.

In summary, the type of teaching materials used by all three teachers for speaking tasks were test-oriented as it more or less followed the same format with the MUET speaking component. Although the resources were different, the format of the tasks given was the same. All topics covered in the speaking tasks were semi-pedagogical, utilizing real-life objects and texts but in a modified form for example a series of pictures from real newspapers accompanied with captions and exercises in text books (Frohlich, Spada, & Allen, 1985, p. 55).

4.8.4 MUET preparation classroom observation

This section reports the findings from further analysis of each teacher behaviour in the classroom and how they carried out their lessons.

Teacher 1 (Class 1)

One lesson from Teacher 1 was observed and analysed. Table 4.31 provides a summary of the data from the classroom observation scheme.

Table 4.31

Teacher 1's observation scheme data

Teacher 1	Recorded Observation	Percentage (%)	Total
Participant organization	Teacher to Student/Class	34	100
	Student to Student/Class	22	
	Group	22	
	Individual	22	
Activities conducted	<u>Teacher Activities:</u>		100
	i. Pre-lesson activities	10	
	ii. Lecturing, explaining & guiding	10	
	<u>Student Activities:</u>		
	i. Individual work: Speaking	40	
	ii. Group work: Discussion	-	
	iii. Class: Presentation	40	

It was discovered that of the 80 minutes of the lesson, Teacher 1 talked for 34% of the entire lesson time and the rest was dominated by the students, with student-to-student interaction at 22%, group discussion at 22% and individual interaction also at 22%. Table 4.31 also showed that with overall teacher activities taking up 20%, Teacher 1 did not dominate the lesson. She let her students carry out the speaking activities assigned to them – 40% speaking individually and 40% for class presentation.

Teacher 1 conducted her class entirely in English. In her class, the students were required to do individual presentations according to their groups. Teacher 1's approach to teaching speaking skills was to ask an individual student to come to the front and present points on selected topics for the speaking task. This approach seemed to be slightly different from the assessment in the MUET speaking component, in which students are supposed to present their tasks in their respective group, similar to the procedure in Teacher 2's and Teacher 3's classes (below). Teacher 1 was seen constantly asking the students questions to help them elaborate and clarify their points.

This helped the students who seemed to have run out of points or ideas to talk about. Tables 4.32 and 4.33 provide extracts from the classroom observation.

Table 4.32

Extract from classroom observation

Speaker	Excerpt	Note
Student 1A	:	Student 1A read out loud points from a piece of paper.
Teacher 1	: Can you link to the situation?	
Student 1A	: ...	Student 1A was silent.
Teacher 1	: So that's why the students...	Teacher 1 helped Student 1A to start her sentence.
Student 1A	: So that's why the students...	
Teacher 1	: Are under...? Refer to the situation.	Teacher 1 pointed to the student's note.
Student 1A	: That's why the students are under pressure because of money problems	

Table 4.33

Extract from classroom observation

Speaker	Excerpt	Note
Student 1B	:	Student 1B hesitated to start his presentation.
Teacher 1	: Okay how can a poor family make the students feel under pressure or stress?	
Student 1B	: Because umm... umm	
Teacher 1	:	Teacher 1 signalled Student 1B to face the audience.
Student 1B	: Because umm.. The family umm.. Because umm..	Student 1B laughed and shook his head.
Teacher 1	: Because the parents? The parents do not make... enough money...	Teacher 1 tried to help Student 1B deliver his points.
Student 1B	: The parents do not make enough money... and...	
Teacher 1	: Therefore... To pay for what?	
Student 1B	: To pay for the university or college fees.	
Teacher 1	: You have to relate to the situations. Relate to the situation.	

Teacher 1 not only focused on the students' language, but also kept reminding the students to face the audience in order to help them build their confidence when speaking in English, as can be seen in the screenshots in Figures 4.6, 4.7 and 4.8.



Figure 4.6 Screenshot from classroom observation



Figure 4.7 Screenshot from classroom observation



Figure 4.8 Screenshot from classroom observation

From the classroom observation data, Teacher 1 did not seem to be very particular about her students making grammatical mistakes when they gave their oral presentations. She would let the students finish their presentations first and if necessary, or whenever she had the time, she would let the students know. Towards the end of the lesson, Teacher 1 wrapped up by reminding the students how to handle the MUET speaking Task A and Task B. She also mentioned what the examiner would do on the test day, elaborating specifically on the procedures that the students would have to follow:

Just a tip for your Task A, individual presentation/speech for 2 minutes, you will be given the questions on the table. You read them and then the examiners will let you find points for 2 minutes that is your preparation before you start your presentation. So you will be given a piece of paper, a pencil and an eraser. So what you can do, you divide the paper into four sections and then you write down A, B, C and D. That means, for example, if you are Candidate A, you write down your notes or your three points here in column A. After your 2 minutes preparation is up, you are going to start your presentation. Each candidate will have 2 minutes for presenting. Let's say you are Candidate D, when your friend, Candidate A, presents his or her view, you have to list what your friend is saying so that you can use the points in Speaking Task B, the Group Discussion. When you present your points, you have to give examples.

(Teacher 1)

Teacher 2 (Class 2)

Table 4.34 provides a summary of the observation scheme for Teacher 2 in Class 2.

Table 4.34

Teacher 2's observation scheme data

Teacher 2	Recorded Observation	Percentage	Total
Participant organization	Teacher to Student/Class	42	100
	Student to Student/Class	21	
	Group	22	
	Individual	22	
Activities conducted	<u>Teacher Activities:</u>		100
	i. Pre-lesson activities	17	
	ii. Lecturing, explaining & guiding	42	
	<u>Student Activities:</u>		
	i. Individual work: Speaking	25	
	ii. Group work: Discussion	8	
	iii. Class: Presentation	8	

From the classroom observation data, during the 80 minutes of the lesson, Teacher 2 was observed to talk for 42% of the time. Since the lesson was scheduled for speaking skills, the whole lesson was devoted to speaking activities. The teaching material that Teacher 2 used was confined to semi-pedagogical materials from the MUET workbook. It can be seen from Table 4.34 that half of the class time was dominated by the teacher (59%).

From the classroom observation, it was clear that Teacher 2 incorporated some elements from the MUET speaking task into her speaking lessons, but she did not exactly follow the test format. Teacher 2 asked her students to give their individual presentations in front of the whole class instead of in their respective group, as they would be tested in the MUET. Doing this allowed the students to build their confidence and self-esteem and also helped other students to learn from

each other. One of the students who presented in Class 2 even interacted with the rest of the class when she was presenting by asking them questions to make sure that they were paying attention to her presentation. This made the class interactive and also gave the chance for other students to participate.

Similar to Teacher 1, Teacher 2 tried to keep her class interactive by probing the students with questions to encourage them to speak more in the class as shown in the extracts in Tables 4.35 and 4.36.

Table 4.35

Extracts from classroom observation

Speaker	Excerpt	Action
Student 2A	: For the first one is...	
Teacher 2	: Was.	Correction attempt by Teacher 2.
Student 2A	: ...was sports and recreation such as <i>congkak</i> , <i>gasing</i> and <i>silat</i> . When the... <i>bangsa</i> ?	Student 2A was not sure about the English word for “ <i>bangsa</i> ”, which means race in English.
Teacher 2	: Races.	
Student 2A	: When the races play with their peers, they can share their traditional sports with others. Other races will give support when they participate in their traditional games.	
Teacher 2	: Support in terms of? Participating together?	
Student 2A	: Yes, participating together. Some of the sports that can be shared are <i>silat</i> and Tai chi. That’s all.	
Teacher 2	: That’s all? How many points did you deliver just now? Any other points? Anything else?	

She frequently corrected her students instantaneously whenever they committed glaring grammatical or pronunciation mistakes.

Table 4.36

Extract from classroom observation

Speaker	Excerpt	Note
Student 2B	: Good afternoon to teachers and friends.	
Teacher 2	: How many teachers are there?	Correction attempt by Teacher 2
Student 2B	: Teacher and friends. I am going to present this topic... how to handle stress.	
Teacher 2	: On how to handle...	Correction attempt by Teacher 2
Student 2B	: On how to handle stress. That we know...	
Teacher 2	: As we know.	Correction attempt by Teacher 2
Student 2B	: As we know, our examination is just around the corner, right?	

As meticulous as it might appear to be, Teacher 2's approach to correcting her students did not seem to distress them. On the contrary, they appeared to be appreciative whenever their teacher corrected their mistakes. At the end of the lesson, Teacher 2 reminded the students to practise speaking at home.

Teacher 3 (Class 3, Class 4, Class 5)

Teacher 3 taught three MUET preparation classes in his school. Table 4.37 provides a summary of the classroom observation scheme for all three classes.

Table 4.37

Teacher 3's observation scheme data

Teacher 3	Recorded Observation	Class 3	Class 4	Class 5	Total	
Participant organization	Teacher to Student/Class	33	33	39	100	
	Student to Student/Class	17	17	23		
	Group	25	25	15		
	Individual	25	25	23		
Activities conducted	<u>Teacher Activities:</u>					
	i. Pre-lesson activities	17	17	14	100	
	ii. Lecturing, explaining & guiding	32	32	29		
	<u>Student Activities:</u>					
	i. Individual work: Speaking	17	17	29		
	ii. Group work: Discussion	17	17	14		
	iii. Class: Presentation	17	17	14		

For Teacher 3's classes, the distribution of the percentage of time spent on activities was similar across classes. Teacher 3 was consistent in how he managed his English language lessons. In Table 4.37, Teacher 3 spent an average of 47% on teacher activities, including briefing the students before he began his lessons. These classes appeared to be dominated by Teacher 3, as depicted in the extract provided in Table 4.38. This could be attributed to the fact that these students were lower proficiency and hence they were more reserved and passive compared to the classes with higher proficiency students.

Table 4.38

Extract from classroom observation

Speaker	Excerpt	Note
Teacher 3	: Good morning and welcome to the MUET lesson. Today we are going to do some speaking activities. I have one simple question to ask you about our topic for today. Which one do you prefer, studying in local universities or studying abroad? If you prefer studying abroad raise your hand.	Teacher 3 gave a short briefing about what they were going to do in the class.
Class	:	The whole class remained silent. One student raised his hand.
Teacher 3	: Okay Bahtiar. What about others? Girls? Boys? Would you like to study in foreign universities?	
Bahtiar	: Local university.	
Teacher 3	: So, you prefer local university. I think the majority of you prefer local university. Why do you prefer local university? Alif? Why do you prefer local university?	
Alif	:	Alif kept silent.
Bahtiar	: Because...	Another student, Bahtiar, decided to help.
Teacher 3	: Okay, Bahtiar?	
Bahtiar	: Why I prefer local university? Because the cost is cheaper than foreign university.	
Teacher 3	: Okay very good answer. Because it's cheaper compared to the foreign university. What about you Faisal? Quite a handsome boy Faisal. What do you think? Why do you prefer local university?	Teacher 3 tried to lighten up the mood by making joke.
Faisal	: Because the cost is lower.	Faisal gave the exact same answer as the previous student and Teacher 3 did not say anything.
Teacher 3	: What about you Hani? Aida?	
Class	:	Both students just smiled and kept silent.
Teacher 3	: Okay, today your task is quite simple. Just copy the situation from the white board, "It is better to	Teacher 3 proceeded with giving the students

study in the local than in the foreign university. instructions on their
What are some of the advantages?" You have to speaking tasks.
discuss...

Unlike Teacher 1 and Teacher 2, who asked the students to provide points for discussion, Teacher 3 gave out the points first and elaborated these points with his students as an example of how they were supposed to present their points. Only then were the students assigned topics to be discussed in their respective groups.

It was also observed that 51% to 57% of the total class time was devoted to students' activities, with a similar division of time, except for Class 5 in which 29% was devoted to speaking, as opposed to Class 3 and Class 4 at 17%. At the beginning of his classes, Teacher 3 asked one representative from each group to present their points to other groups concurrently and moved from one group to another. Although he used model speaking tasks similar to those in the MUET speaking test, he did not simulate the MUET speaking test in his class. This might be due to time constraints as each lesson only lasts for 80 minutes and he would have needed more time to act as the examiner and monitor each group during the speaking exercises. Unlike Teachers 1 and 2, Teacher 3 just observed as the students carried out their presentations, without stopping them or asking them to clarify their points further until the whole class was done with their parts. The students from Classes 3, 4 and 5, the lower proficiency students, seemed to be reading from their notes rather than presenting their points orally, as can be seen in the screenshot in Figure 4.9.



Figure 4.9 Screenshot from classroom observation

This approach might be necessary as the low-proficiency students could be overwhelmed if they were required to do something that they perceived as beyond their English language capabilities. The students from Classes 1 and 2 referred to their notes when they conducted their oral presentations, but only occasionally.

The students in Teacher 3's classes seemed less interactive and more passive compared to Classes 1 and 2 (see Table 4.38). The teacher kept asking probing questions to encourage them to speak, but they only responded with very short answers. Even when they were engaged in group discussion, they used their mother tongue and if they happened to ask for Teacher 3's assistance, it was to help them translate their points from their mother tongue to English. Teacher 3 appeared to be receptive to the whole situation, not forcing the students to use English due to their lower proficiency. Indeed, Teacher 3 was reluctant to be observed at first because he mentioned that he was teaching lower proficiency students and recommended another teacher who was teaching higher proficiency students. He later agreed when he was informed that the aim of this study was not to evaluate the students' performance in the classrooms, but to explore the washback effect of the MUET. For this study, having a mixture of both high- and low-proficiency students would

provide a variety of phenomena to be observed and would help to see washback from a wider perspective.

Teacher 3 used English and Malay roughly half and half in his classes. He would translate what he said into Malay for the students to understand. Similar to Teachers 1 and 2, Teacher 3 also allocated some time for the students to discuss among themselves first on the points that they were going to present. However, before the students started their group discussion, he would discuss the topic first, providing several sample answers. He even allowed the students to use their smartphones to find extra information on their tasks, which would not happen in the MUET speaking test. When it came to giving tips on how to do well in the speaking test, in one of his classes (Class 5), Teacher 3 listed the characteristics of a Band 4 student, who he described as “confident and clear”. Similar to Teacher 2, Teacher 3 explained to the students what was going to happen during the MUET speaking test and what sorts of tasks they would need to carry out.

After you have finished with your Task A, let's say you're Candidate A, you straight away prepare for Task B, don't waste your time. Finish your presentation, straight away prepare for Task B. Boleh? (Can you?). Any question? Ada soalan ke? (Any question?).

(Teacher 3)

Teacher 3 also mentioned that if the student happened to get Option 4 during the speaking test, meaning the student was candidate number 4, he or she was considered unlucky because according to him, it was believed that Option 4 is always the most difficult.

If you look at these four options here, which one do you think is the most difficult? Number four, of course, is obviously more difficult than the other options. In the speaking test, it

depends on your luck. If you get Option A, Option B, Option C, you are very lucky. But, if you manage to get the last one, you have to try your best, try to think outside of the box.

(Teacher 3)

This might be discouraging for students on the day of the test if they happened to get Option 4 and might make them more anxious than they already were. However, this might be one of the odd cases of bad coaching, as a high-stakes test like the MUET would not be designed in such a manner. This point will be discussed further in Chapter 5. Teacher 3 also kept reminding the students to treat the speaking activities that they did in the classroom as if they were sitting for the real test. One interesting point to note is that the students in Class 2, Class 3, Class 4 and Class 5 all started their presentations by addressing the examiner; for example, for each presentation, they would start with “Good morning to the examiner and fellow friends”. It seemed that they were following a certain template when practising their speaking tasks for the MUET.

4.9 Summary

The findings presented thus far have indicated the existence of a relationship between the students’ perceptions of the MUET in terms of perceived test importance, perceived test difficulty and perceived self-efficacy. It was reported that the students’ perceptions play a big role in mediating the washback effect of the MUET, as their perceptions shaped their goal, which subsequently influenced their course of action and – in this study – their language learning strategies. The findings on washback length also give rise to the need for further analysis to be discussed further in the following chapter.

CHAPTER 5: DISCUSSION

5.1 Introduction

The purpose of this chapter is to interpret and describe the significance of the findings of this study, addressing all three research questions. These aimed to investigate the washback effect of the Malaysian University English Test (MUET), a university entry test in Malaysia, by exploring Malaysian students' perceptions of the MUET and their own self-efficacy and exploring the relationship with the language learning strategies employed when preparing for the test. In addition, this study aimed to explore the washback length of the MUET by also including students who had already sat this test. The following sections of this chapter attempt to integrate the findings from the various sources of data to address the three research questions. The chapter sequentially deals with students' perceptions of the MUET, followed by the relationship between the students' perceptions of the test and their language learning strategies when preparing for the MUET and finally, the washback length and intensity of the MUET. Before further discussion, it has to be noted here that all the quantitative and qualitative data collected in this study, except for classroom observation, were self-reported by the students and the teachers. Hence, the research findings should be interpreted and accepted with caution.

5.2 RQ1: What are students' perceptions of the MUET and what are the factors, if any, that seem to influence such perceptions?

For perceived test importance, unsurprisingly, the test under investigation, the MUET, was regarded as a very important test by most of the students in this study. Having a direct significant effect on university applications implies that the MUET is a high-stakes test with important

consequences for the students. Inducing positive washback on the teaching and learning of English at the pre-university level in Malaysia was one of the main intentions of the MoE in introducing the MUET. This action was deemed necessary to prepare the students for tertiary education, as English is the main medium of instruction in most Malaysian public universities.

As discussed in the literature review (see 2.3.1), the washback phenomenon was addressed by Alderson and Wall as long ago as 1993 with the proposition of 15 washback hypotheses. Among these, hypothesis 12 “*Tests that have important consequences will have washback*” and hypothesis 13 “*Tests that do not have important consequences will have no washback*”, suggest that a test with important consequences, i.e. a high-stakes test, could influence the students’ perceptions of the test, which will then affect how they react to it. However, it is necessary to clarify what is meant by the “*important consequences*” of a test in relation to washback due to the variability of perceptions among stakeholders. A test might be considered important to one student but not another as they may have different views of the same test based on their personal goals and what they aim to obtain from the test, as established in the previous washback literature (see Green, 2007a; Hamp-Lyons, 1987; Hughes, 1993, 2002; Mehrens, 1998). Moreover, this study was interested in looking into other possible factors affecting students’ perceptions of test importance, namely gender and English language proficiency.

Perceived test importance is closely associated with the social and political functions of a test given by the authorities (Shohamy, 2014). This social power that is given to a test can influence how the students respond to it. The more important a test is to students, the more effort they will put into preparing for the test. In this study, while the majority of the students regarded the MUET as a very important test, a small number were either undecided or disagreed. Although the percentages were very low, it was interesting to explore in greater depth the possible factors that

influenced their views. Closer examination revealed that even when a test is designed to have important consequences, it does not guarantee similar perceptions or reactions among all the test takers. This could be attributed to the fact that some of the students considered they had other alternatives if they performed poorly in the MUET. These included retaking the MUET until they got their desired band; the test is held three times a year (March, July and September) and there is a fee of MYR 125 (equivalent to GBP 23). Alternatively, they could consider pathways other than pursuing tertiary education, or settle for whichever courses and institutions were available to them according to their MUET band. The availability of these options seemed somewhat to “dilute” the power that the MUET had on some of the students in this study, particularly in view of their different goals. Indeed, the aforementioned alternatives clearly reduced the level of the “important consequences” of the MUET for those students who were undecided about the importance of the MUET. Furthermore, those who did not plan to enrol in tertiary education showed a tendency to be indifferent about performing well in the MUET as the result is only used for university entry. As it is common practice for pre-university institutions in Malaysia, such as the Form 6 public schools in this study, to include the MUET in their syllabus, all pre-university students are required to sit the test, regardless of their plans for their future undertakings. This partially explains why some of the students in the study did not perceive the MUET as an important test for them.

With regard to using the MUET results for enrolment into desired courses, some of the students were willing to change the course they were planning to apply for if they did not meet the MUET requirement. Two of the students in the interviews explicitly expressed their belief that the MUET band signified their actual English language ability, which would determine what course of study would suit them. They added that this would actually be beneficial to them in coping with their studies at tertiary level. To them, their MUET result could be used as a basis for their

decisions in terms of which course of study they were going to undertake. These students did not seem to have a strong commitment to specific courses of study and were willing to forgo their desired field of study just because they could not fulfil the English language requirement. It appeared that the possibility of not being able to enrol in their desired field of study was not a strong enough motivator to induce some of these students to work hard on improving their English. This could be attributed to Malaysian learners' negative attitudes towards English language learning: as reported in Ganapathy and Ying's (2016) qualitative study on the attitudes and motivation of secondary school students in Penang, Malaysia, towards learning English as a second language, it is regarded as just another subject that they need to pass.

Ganapathy and Ying (2016) found that although the majority of the students in their study had positive attitudes towards learning the language, a few had developed negative attitudes. Focus group interviews with 20 students revealed that the reasons for these negative attitudes were lack of proper teaching and resources and inadequate efforts to instil awareness of the significance of the language (Ganapathy & Ying, 2016) beyond its instrumental value simply in terms of passing tests. It seemed that some of these students were neither driven nor motivated to pursue their goals and ambitions, something that was also evident in a few of the students in this present study. The students' motivational level varied based on their future plans. Ganapathy and Ying (2016) also found that the students who found the language interesting and believed that English was essential to further their studies appeared motivated to learn, while those who were aiming to study courses not requiring a high level of proficiency in English struggled to learn as they were already demotivated (Ganapathy & Ying, 2016). These students were more extrinsically than intrinsically motivated and were only learning English for examination purposes, as has also been reported in

other studies in the Malaysian context (Bidin et al., 2009; Thang, 2004). As long as they were able to enter a university, no matter what course they would be doing, they were unconcerned.

This has the potential to cause further harm to the students, as they may struggle with motivation when studying disciplines they were not originally interested in. This has been a problem among Malaysian students for decades. Thang (2001), in her doctoral study exploring Malaysian undergraduates' conceptions of their learning processes and their perceptions of their English language courses in a tertiary learning context, stated that some of the Malaysian students in her study appeared to be uncertain about the reasons why they had decided to pursue studies at tertiary level. She attributed this to the fact that many of them did not gain entry to the courses they applied for when they were admitted to the university. This undiscerning attitude is typical of Malaysian students and the complexity of the university admission system in Malaysia is no help, leaving these students with little choice other than to undertake whichever courses are offered to them, attend private universities with more lenient admission rules, or avoid tertiary education altogether and start working instead.

In this study, some of the teachers were also found to have similar attitudes to some of these students, perceiving the MUET as unnecessary for students not planning to pursue tertiary education. Positive washback is intended from the MUET, as clearly stated by the MEC: the MUET syllabus seeks *“to consolidate the English language ability of pre-university students to enable them to perform effectively in their academic pursuit at tertiary level, in line with the aspirations of the National Education Philosophy”* (Malaysian Examination Council, 2001, p. 11). It appears that the MUET is taken at face value, seen only as an enabler for university entrance, not as a tool to help induce positive washback on English language learning among the students. This finding is echoed in Zhan's (2009) qualitative case study, in which he attempted to explore

systematically how a small group of non-English major students experienced washback from the 2006 revised College English Test Band 4 (CET-4) in China. He found that the students performed “superficially” because they did not understand the real intentions of the test designers, let alone internalize them. For them, the test was there just to assess them. Such reactions from test takers both in Zhan’s (2009) and this study are expected and quite understandable, especially in the context of exam-oriented education systems such as those in China and Malaysia, where tests are seen simply as hurdles to be overcome in order to progress to the next level. Hence, it is deemed unrealistic to expect washback awareness from the teachers, let alone the students.

What the stakeholders in the aforementioned studies and this research failed to realize is that the importance of a test goes beyond university entrance. It is actually used to stream students based on the English language demands of specific courses. As mentioned in the Introduction (see 1.3), students applying for critical courses that require a high level of English mastery, such as medicine, law, engineering and English language courses, will need to obtain at least Band 4 in the MUET. The MUET is actually there to encourage them to work harder on their English language learning to prepare them to meet the language demands at tertiary level. This is not an empty claim made solely to justify the objective of executing the MUET in the first place. The significance of having adequate English language mastery at tertiary level was communicated by one of the teachers in this study, Teacher 3, in the open-ended question on the questionnaire: he mentioned that there were cases in which his students had to drop out of university, not because they were not smart, but because their English was very weak and they could not cope with the teaching and learning in the university as most lessons are conducted in English.

This study also found a gender effect that appeared to play a significant role in determining how test importance is perceived, with more female students regarding the test as important than

male students. This is in contrast to Fan, Ji, and Song's (2014) finding in their study, which explored the roles of gender and English language ability in shaping students' reported washback on the Fudan English Test (FET), a university-based English test in China. They found that gender was an insignificant factor in shaping students' perceptions of the test. However, it has to be noted here that the FET is not a high-stakes language test. Hence, care must be taken when comparing the FET and the MUET. Given the scarcity of studies examining the role of gender in washback, both for high-stakes and low-stakes testing, it was considered relevant to examine it in this study and contribute more information on this little investigated matter. More discussion on the gender factor is provided in the later parts of this chapter.

Aside from gender, the students' English proficiency level seemed to be a significant factor in determining students' perceived test importance. It was interesting to discover that lower proficiency students significantly perceived the MUET as important to them, more so than the higher proficiency students. Although there is no evidence from the data to explain this finding further, it is logical that the low-proficiency students would be more concerned with how well they were going to perform in the test. Low proficiency could induce a certain amount of test anxiety, as reported by Chen and Hsieh (2011) (see 2.6.2). This study extends the findings reported in Shih's (2007) intensive study, which found that low-proficiency students showed a higher tendency to worry about the stakes of the test than higher proficiency students. Indeed, it has been reported in numerous washback studies that level of proficiency is one of the significant variables that will influence the washback of a test (see Chu, 2009; Ferman, 2004; Pan, 2014; Shih, 2007; Shohamy et al., 1996; Watanabe, 2001, 2006), as also presented in this study specifically for the Malaysian context. Therefore, this study concludes that the students' perceptions of the MUET play a significant role in shaping the washback from the test.

With regard to students' perceptions of the difficulty of the MUET, similar to perceived test importance, Green's (2007a) washback model posited that the more difficult a test is perceived to be, the more intense the washback will be, regardless of the actual level of difficulty of the test. The findings from the student questionnaire indicated that the students were of the opinion that the MUET was moderately difficult, inclining towards difficult. Closer examination based on the interview data revealed that the majority of the students were particularly concerned about the new language components, listening and speaking, tested for the first time. These two skills are not tested throughout the 11 years of learning English formally at school. Speaking was specifically highlighted as the main cause of the MUET being perceived as difficult. However, it is interesting to note that some of the students described the MUET as "challenging" as opposed to "difficult", indicating that the difficulty level is viewed as being within an acceptable range.

This study also found that gender and students' level of proficiency did not significantly influence students' perceptions of test difficulty, unlike perceived test importance. This finding was unexpected as this study hypothesized that the students' proficiency might be one of the factors influencing perceived difficulty, such that students of higher proficiency would regard the MUET as easier compared to low-proficiency students. However, it transpired that higher proficiency students rated the MUET as very difficult, unlike the low-proficiency students, although the difference was not statistically significant. The question that then arises is: "*What makes the test be perceived as difficult by these students?*" Adding new components to a test has been reported to produce washback both on teaching and learning (Allen, 2016a; Watanabe, 2004). Watanabe (2004) discussed washback specificity briefly in his conceptualization of washback dimensions. According to him, washback can either be general or specific and one way of achieving specific washback is by introducing a new element to a test with the intention of focusing

on that particular skill in teaching and learning. In the case of the MUET, during the interviews the students explained that they focused more on their speaking and listening skills, as most of their previous English language examinations had only tested them on their reading and writing skills. Other than the novelty effect of introducing new test components, the weighting of each component plays an important role in determining students' behaviours when preparing for the test. The higher the weighting that is assigned to a test component, the more attention the students will pay to mastering it (Xie & Andrews, 2012).

However, this study found that the students' language learning strategies for test preparation were not based on the language component with the highest weighting (reading 45%), but on the language skills of which they had least experience, in this case speaking (15%) and listening (15%). The fact that test preparation was not based on the test weighting would be surprising were it not for the introduction of the new elements, speaking and listening, which could have made a considerable difference. The fact that reading has the highest weighting, almost half of the overall score for the MUET, does not seem to encourage students to focus more on this aspect as intended by the test designers. As explained in Chapter 2, the MEC stated that the reading component has the highest weighting of the overall MUET band because it is deemed to be the skill used most at university, followed by writing (25%) and speaking (15%) and listening (15%). The finding that assigning different weighting to the components of the test did not seem to encourage the students to focus more on the targeted skills partially corroborates Allen's (2016a) study, which showed that the students tended to focus more on productive skills (speaking and writing) than receptive skills (listening and reading); this he attributed to students' perceptions of these skills as difficult. Students have a tendency to perceive unfamiliar parts of a test as difficult and hence pay extra attention to the new test components. Thus, in the case of the MUET, it can

be concluded that the students paid more attention to speaking and listening skills because they perceived them as difficult compared to writing and reading, which they had been doing for the past 11 years. Surprisingly, the teachers seemed to think otherwise. Further analysis of the qualitative data revealed that some of the teachers stressed over and over again that the reading component in the MUET was very difficult. One of the teachers in the teacher questionnaire, Teacher 5, in his own words, described the MUET reading component as “*bloody difficult*”, “*total murder*” and “*crazy questions*”. He also went to the extent of describing the MUET reading component as the factor that “*make students fail to get a good band*”. Nevertheless, the findings revealed that most of the teachers focused on speaking skills when it came to teaching the MUET preparation classes. Again, it has to be reiterated that the novelty effect and the lack of familiarity with the new test components were stronger motivators encouraging teaching and learning of these targeted skills than the weighting or perceived difficulty of a test component.

With regard to ranking the importance of each skill to optimize performance in the MUET, it was hypothesized that the students would rank speaking and reading highly, the former due to its novelty effect and the latter due to it having the highest weighting for the overall band. From the findings, speaking was ranked the most important skill, followed by reading, as speculated, which indicates that these students were aware of the weighting of each component and the skills that needed more attention if they wanted to score highly in the MUET. Nevertheless, reading was actually mentioned less by the students in the interviews than speaking and even listening. The students in the interviews mostly talked about practising their speaking and listening very specifically and significantly, but not their reading, despite it having the highest weighting. One plausible explanation concerns the nature of the tasks in the test, namely in relation to productive or receptive skills. For example, to answer the questions in the reading component of the MUET,

the students only need to circle the answer that they think is correct. Even if they do not know the answer, they can still circle any of the answers provided in the question booklet. For speaking and writing, the task demand is higher as the students need to produce something. Thus, this concerns the demands of performance tests as opposed to more traditional test formats. In the MUET speaking test, it is understandable if the students feel more pressure compared to other components because they are directly assessed by the examiner in front of three other students in their group. The last thing they would want to do is remain silent during the speaking test. Hence, it is safe to conclude that besides the novelty effect of the test tasks, the nature of the task also exerts a great influence on students' perceptions and behaviour with regard to a test, not so much the weighting. Nonetheless, from the questionnaire and interview data, it was apparent that many students cared about their mastery of the English language skills rather than just focusing on getting a high band in the MUET.

It was made apparent from the literature review, one element that is often neglected in washback research is student's perceived self-efficacy, i.e. how the students perceive their ability in relation to performing in a test. The findings concerning the students' perceived self-efficacy might be able to shed light on the perceived difficulty of a test. In this study, most of the students reported high self-efficacy in relation to their ability to perform well in the MUET preparation class and the MUET itself. The high self-efficacy among students relates to the findings for perceived test difficulty, as they consider the MUET lies within their range of capability to perform well. It was interesting to note that gender was a significant factor in students' self-efficacy, but not their level of proficiency. The female students appeared to have higher self-efficacy compared to the male students. This is commonly reported in self-efficacy studies, which have shown that

female students tend to have higher self-efficacy in fields related to language and the arts than male students (Pajares, 2002).

5.3 RQ2: How does washback of the MUET operate and to what extent do students' perceptions seem to have a washback effect on their language learning strategies?

The second research question specifically aimed to explore how does washback of the MUET operate and if the students' perceptions of test importance, test difficulty and their own self-efficacy would have any effect on their language learning strategies. The MUET was introduced to help prepare students for the language demands of university, specifically by encouraging them to enhance the four major language skills. With this goal in mind, the students who take pre-university studies (e.g. diploma, Form 6, A-level, matriculation, etc.) are required to sit the test to gain entry to university for their bachelor degrees.

Before delving deeper into this matter, it is best to explore and discuss the English language learning that took place both inside and outside of the classroom in order to understand how the washback of the MUET operate in the context of this study. When exploring the overall direction of washback of the MUET, whether it brings about positive or negative washback to the learners, the kinds of language learning activities taking place in the preparation process were taken into consideration. Although the focus of this study was on the learners, data from the teachers were also collected as learning takes place both in and outside the classroom. Including data on the teachers' perspectives aimed to yield a more holistic view of the kinds of language learning activities conducted in the MUET preparation classrooms.

5.3.1 English language learning inside the classroom

It is undeniable that in the classroom, teachers generally have the greatest authority in conducting the lessons. The teachers' data and the classroom observations were used to triangulate and verify the findings from the students. This study found that the MUET produced both positive and negative washback on the students' learning. Most of the positive washback reported in this study did not differ greatly from that previously reported in the literature (see Ferman, 2004; Ren, 2011; Xie & Andrews, 2012). Quite a number of the students, based on the questionnaire and interview data, were found to be motivated to learn English and used various language learning strategies frequently in the process of preparing for the test. As mentioned in 4.2.1, direct strategies involve the use of language, while indirect strategies do not directly involve using the language, but support language learning (Ehrman & Oxford, 1990). The findings from the questionnaire showed that both direct and indirect strategies were used equally by these students. The interview data revealed coherent findings, with more emphasis given to test preparation strategies, for example practising model MUET questions to hone their reading and writing skills and listening to English songs to improve their listening skills, to name but a few (see 4.2.1). Some of the highly motivated students also seemed to be independent in their learning, consciously seeking out practice opportunities to hone their mastery of English, even when the circumstances were not particularly supportive. As mentioned in 5.2, one of the factors inhibiting students' motivation to learn was a lack of resources and learning support (Allen, 2016b; Ganapathy & Ying, 2016). As pointed out by Hughes (1993, p. 3), one of the five conditions that need to be met to promote positive washback is that "The necessary resources for successful test preparation must be available". In relation to this study, it was revealed that there was a lack of support for the students to practise their speaking. However, some of the students showed initiative by involving their family members and friends. They

appeared to be developing interest and awareness when it came to practising and incorporating English language in their daily lives.

The findings from the student data suggested that the majority were driven to work hard for the MUET by their teachers. However, the students reported that the teachers did not limit them to working only on the specific language skills in the test. The teachers also assisted them in building up their courage and confidence to speak in English. This finding was supported by the teacher data, with the majority of teachers in the teacher questionnaire unanimously agreeing that the strategies most frequently used to motivate students to learn were creating a positive attitude towards language learning in general and giving students encouragement to learn on top of preparing them for the MUET. Generating awareness and instilling a positive attitude towards the whole process of learning and acquiring English during lessons depends greatly on the teachers.

As much as test-oriented activities were conducted in the classroom, a variety of other teaching methods were also employed, balancing test preparation with developing mastery of the language skills. This shows that preparing for a high-stakes test does not have to be all drilling and practice, as indicated by the teachers involved in this study. The findings from the student data showed that the language learning activities conducted in the classroom were not restricted to those outlined in the textbook; rather, the teachers used their own methods to encourage the students to communicate in the classroom. They conducted interactive activities to “force” the students to speak and be active in classroom participation, linked to the introduction of listening and speaking elements in the MUET. Engaging in such activities, many of which were not aspects of the MUET, nonetheless helped the students prepare for the test without neglecting the need to master the language holistically, not only for the sake of sitting the test.

It must be noted that the lessons examined in this study were MUET preparation classes and hence elements of teaching to the test were expected. One issue highlighted in this study was the negative connotation attached to teaching to the test due to a lack of understanding of what teaching to the test actually is. In this study, it was found that the language learning activities carried out in the MUET preparation classrooms could be categorized as teaching to the test, with the teachers requiring the students to practise all four language skills tested; indeed, the teachers confirmed that they applied an exam-oriented teaching style by focusing on the skills required in the MUET. Teaching to the test is not a foreign concept in the washback literature and is expected. However, what this study learned from the teachers was that despite the expectation that teaching to the test is central, the MUET seemed to push the teachers to incorporate other elements of teaching and learning while still preparing the students for the test, for example by having the students work collaboratively on their writing skills. Working in groups encouraged the students to communicate and discuss what they were doing, which indirectly also prepared them for the speaking tasks in the MUET. Injecting other elements did not seem particularly costly in terms of time spent in the classroom or on test preparation.

It is clearly important to familiarize the students with the format of the test, but it is also important to ensure they learn and acquire the language not just for the sake of sitting the exam and this is possible if the teachers themselves have appropriate goals in mind. Interestingly, the teachers' questionnaire data revealed that doing mock exam papers came last in the ranking of strategies used to motivate students to learn, although it was expected they would carry out activities closely related to the test as they were teaching test preparation classes. For these teachers, mock exams were a good means of preparing the students for the test, but were not deemed efficient in motivating them, as reported in the teacher questionnaire data (see Table 4.23,

4.5.3). For them, the main purpose of doing mock exams was to prepare students for the MUET and to identify areas for re-teaching, but not to motivate the students.

I would argue that perhaps doing mock exams and employing the test grading and marking system would just hamper the students' motivation and confidence level, particularly if they happened to get a low mark. Although this study did not explicitly elicit responses from the teachers concerning mock examination results, the findings from the literature might help to explain the findings of the present study further. As reported by Chamberlain, Daly, and Spalding (2011) in their study on students' experiences of test anxiety when taking A-level examinations in the south of England, one of the triggers of pre-exam anxiety was the impact of unexpectedly poor mock test results. They found that as much as mock tests might be beneficial in terms of familiarizing the students with the test format and demand, obtaining unexpectedly poor results will not only counteract some of the potential benefits of mock tests, but also make the students anxious about their performances in the actual test. Some of the students in their study reported feeling stress upon receiving poor mock test results and this worsened their test anxiety. In another study on the effect of mock tests with Iranian EFL learners conducted by Khodabakhshzadeh, Zardkanloo, and Alipoor (2017), it was found that mock tests had a positive effect in terms of preparing IELTS candidates, specifically in terms of improving their scores. They also concluded that practising test-taking strategies seemed more effective than teaching course content as far as a high-stakes tests like the IELTS is concerned. As can be seen, the findings of these two studies on mock tests are contradictory and it is therefore not possible to conclude whether mock tests have potential benefits and/or disadvantages per se. However, based on reports in the literature and in this study, I consider that to reap the greatest benefit from mock tests, it is probably best to conduct them at the beginning of the class: the sooner the better. Doing so will help the students

gauge their current performance and how much effort they need to put in to achieve a higher mark in the actual test. Even though it can be argued that doing so might affect students' motivation negatively if their scores are low, they will still have plenty of time to work on their skills; for all one knows, this could actually motivate them to work even harder.

In the current study, the teachers admitted that the students felt compelled or obligated to attend the MUET preparation class. In the questionnaire, Teacher 7 mentioned that students, regardless of their proficiency (high or low) or their preferences (like or dislike) for English language as a subject, "...felt compelled to attend MUET classes in order to be able to answer the questions. Some learning takes place". He added that as long as the students "...are in the process of taking the test, real realization occurs in many students". Similar findings were reported in Ren's (2011) study on the washback effect of CET-4 conducted in five universities in Tianjin, China. In his study, which involved 210 students, he found that introducing a test with important consequences such as CET-4 was able to drive most students, whether they loved English language as a subject or not, to work hard on learning for it (Ren, 2011). It is common sense that students should not be "forced" or "tricked" into going to classes, but more often than not, the students in this study were not fully aware of the importance of acquiring English language mastery specifically in the Malaysian context. By having them sit a high-stakes test, it is hoped there will be a degree of pressure on them to work harder. There might be negative energy in the beginning, due to the sense of threat that their future could be jeopardized if they do not do well in the test. However, this is deemed a necessary trigger and source of stress that will drive the students to start learning; once they are attuned to it, with adequate support and learning resources, hopefully they will actually enjoy the process and benefit from it, as was evident in this study. Two of the students in the interviews explicitly stated how the pressure of having to sit the test encouraged them to put

more effort into learning. In one of the student's own words, Azreen mentioned *“Preparing for the MUET is stressful, but it is a positive stress. It makes me want to learn the English language more”*.

As expected in any test preparation classroom, the teachers' data revealed that among their main aims was ensuring that their students would pass the MUET with flying colours. Hence, the strategies used in their classes were mostly geared towards preparing the students to perform well in the test, also known as test-wiseness, rather than fostering deep learning in acquiring the language skills. It can be considered natural to see this kind of teaching as it is perceived as helpful, especially by the students with instrumental goals. These students appeared frustrated and felt neglected if the teachers did not explicitly “train” them to sit the test. This has also been reported in the literature, especially among low-proficiency students, with students regarding preparation for the test as a form of education. In a recent study, Yang and Badger (2015) stated that IELTS preparation courses which developed test-wiseness gave students a sense of security, as most students want to learn how to tackle the test and gain high scores. However, there was one aspect of concern in this study, when Teacher 3 talked about the “the tricks of the trade” and suggested that being the fourth and last candidate to speak in the test was “unlucky” as Option 4 was always the most difficult (see 4.8.4). This might be discouraging for students assigned Option 4 on the day of the test and might make them more anxious than they already were. In this case, the teacher had disseminated poor information about the test based on a misconception: it is highly unlikely that the MUET, carefully designed as a high-stakes test, would have different levels of difficulty for different candidates as it would jeopardize the test's reliability and validity. Although this only arose in the case of one teacher and the teacher meant no harm, it could cause negative washback

to the students, as they might start worrying about being assigned Option and it places unnecessary extra pressure on the students, adding to the burden they already feel in sitting the test.

5.3.2 English language learning outside the classroom

A majority of the students appeared motivated to learn English, consciously searching for opportunities to practise, even when the circumstances were not supportive, for example lacking partners with whom to practise speaking. As shared by Shahirah, *“I always practise my speaking alone. If I do not have anything to say, or a partner to talk to, I still speak to myself. I will find the time to do this as I need to practise”*. Awareness of the need to start practising and incorporating English in their daily lives also was apparent, as they started taking the initiative to use the language; they knew that by doing so, their mastery of the language would increase. This also reflects preparation for the MUET as their speaking skills were going to be tested for the first time in a high-stakes test. Aside from the variation in the methods employed by their teachers in the classroom, the students also applied variation in their language learning strategies. It was refreshing to see how the students tried their best to make full use of the skills and knowledge they already had in practising their English when preparing for the test. Despite describing the MUET as challenging and noting the new burden that comes with it, the students appeared to have positive attitudes towards preparing for the MUET. As reported in the qualitative data from the student interviews and open-ended question in the questionnaire, some of the students who had previously hated learning English started watching English films and picking up English books. They read English novels and storybooks and even comics during their free time. They read for pleasure, while also being conscious about the learning that took place. The more serious students, like Husaini, even read newspapers with a dictionary next to him. He stated that *“...whenever I find a*

new word, I will look it up in my dictionary and I will write the meaning in Malay and re-read it”.

This finding confirms that the high-stakes test was a powerful tool in motivating the students to practise English.

As expected, some test-related practices took place even outside the classroom when the students were preparing for the MUET. The findings showed that they utilized model MUET papers to the fullest, on top of what the instructors did with them in the MUET preparation classes. Previous literature on washback (see Andrews, 1994; Cheng, 1998; Pan, 2014; Zhan & Andrews, 2014) has widely discussed superficial washback, with learning taking place on the surface level as opposed to deep meaningful learning, the main goal being to pass the test rather than to improve mastery of the language. However, this study found that this was not the case, as the students seemed to be employing roughly equal amounts of direct and indirect strategies when preparing for the MUET. There may have been instances of the use of learning strategies specifically geared towards achieving high marks in the test, but this was balanced by other approaches. No cases of rote learning were reported in the data, which would not work due to the design of the test. This study suggests that the test washback from the MUET was such that to achieve a high overall band, the students had to work hard learning and practising the language. This was because of the way the test was set up and designed. Hence, I suggest that well-designed tests could actually limit the opportunities for training to develop test-wiseness. However, care must be taken in the interpretation of this finding as there was no way of knowing whether the students who used test-wiseness strategies could also have achieved a high overall band in the MUET.

When it comes to determining if the students’ learning when preparing for the MUET could be deemed superficial or “shallow”, as opposed to deep (see 2.7), it appeared to be a mixture of both. However, there were cases inclining towards superficial washback, with some learners trying

to change what they learned rather than the way they learned (Cheng, 1998) and making minimal effort. This was largely due to limited time and other academic obligations, as the MUET was not the only requirement for securing a place at the tertiary level. Hence, they would resort to “shortcuts” as opposed to deep learning. Moreover, as already noted, some students made a strategic decision not to put much effort into preparing for the MUET because they knew it was not so important for them and thus there appeared to be instances of negative washback. This is actually an issue with how the test results are used, also known as consequential validity (Messick, 1989, 1996). Nevertheless, it was encouraging to see that more than half of the students allocated at least one to two hours per week to study English, despite their other academic obligations. As far as test validity is concerned, a test is considered valid not only when the test measures what it claims to measure, but also if the test scores do indeed mean what the test designers intend them to mean (McNamara, 2006). This study confirms that in terms of validity, it is not only how tests are designed that is vital but also how the results are used if positive washback is intended.

5.3.3 Students’ perceptions of the MUET and their language learning strategies

One of the overarching aims of this study was to find supporting evidence for earlier findings from washback studies concerning students’ perceptions of a test and the washback that follows. It was hoped that this would help untangle the complexity of washback in order to gain a better understanding of the phenomenon by examining a specific context and circumstances. Previous washback studies have suggested that students’ proficiency might be related to how they prepare for the test; however, this study found otherwise. There was some indication from the questionnaire data that the low-proficiency students used direct strategies more frequently, while

the high-proficiency students used indirect strategies more frequently, but the difference was not significant.

Gender, on the other hand, showed a significant correlation with the frequency of the language learning strategies used by the students. Specifically, female students appeared to use both direct and indirect strategies more frequently, which could also be attributed to the finding discussed earlier in relation to RQ1, namely that female students appeared to regard the test as more important than the male students. Akbari (2012) noted that gender could pose a risk as an illegitimate factor affecting test takers' performance if a test systematically results in better scores for males than for females and vice versa. However, it was beyond the scope of this study to look at the students' performance according to gender in order to verify if there is a risk of gender bias in the MUET. This will be discussed further in the recommendations for future research in Chapter 6.

As far as students' perceptions of the test are concerned, this study found that only perceived test importance and self-efficacy correlated significantly with the language learning strategies employed by the students, not perceived test difficulty. Although the correlation between perceived test difficulty and language learning strategies was not significant, it was interesting to see that the relationship was negatively correlated, indicating that the more difficult the test is perceived to be, the less frequently students use language learning strategies. One possible explanation for this is the fact that the test has to be within a range of difficulty that the students perceive as doable or achievable. A test that is perceived as too difficult will hamper the students' efforts and could be rather counterproductive, as confirmed by previous findings that low-proficiency students are not motivated to learn because they feel that no matter how hard they try, they will not be able to excel in the test (Chu, 2009; Watanabe, 2001). Equally, the high-

proficiency students in Shih's (2007) study who thought the test was too easy were not motivated to work hard and were not driven to prepare for the test. As reported in the literature (see 2.6.1) and discussed in 5.2, if students think that the test is beyond their capacity, they will give up from the very beginning because of low self-efficacy and the belief that they will not perform well on the test, no matter how much effort they put into preparation, a point that was evident from the quantitative data in this study. It was found that the students' perceived self-efficacy was statistically significantly correlated with their language learning strategies. Although the students' perception of the difficulty of the MUET did not correlate significantly with their use of language learning strategies, it was statistically significantly correlated negatively with the students' self-efficacy. The more difficult the test was perceived to be, the lower the students' self-efficacy. Hence, the findings of this study support the notion that the perception of test difficulty beyond perceived capacity leads to diminished self-efficacy, which then possibly affects students' use of language learning strategies. The perceived difficulty of the test on its own was found to be a weak factor driving the washback impact of the test as opposed to the students' self-efficacy. In other words, it does not matter how difficult they think the test is, if they believe they can manage it, they will increase their efforts to study, in this case by employing more English language learning strategies.

5.4 RQ3: What is the intensity of the washback effect of the MUET and what is its length?

How do these appear to influence washback on the learners?

The third research question of this study was formulated in an attempt to investigate two other dimensions of washback, length and intensity. As far as this study is concerned, washback length, namely how long the washback effect lasts after the test, is one of the dimensions that is under-

explored and the information available thus far on this matter is rather limited (see 2.4.3). This might be due to the difficulty of establishing washback length. Washback length can be categorized as short term or long term (Watanabe, 1997, 2004). As defined by Watanabe (2004):

...if the influence of an entrance examination is present only while the test takers are preparing for the test, and the influence disappears after entering the institution, this is short-term washback. However, if the influence of entrance exams on students continues after they enter the institution, this is long-term washback. (p. 20)

However, taking this definition at face value, the concept of length is problematic and somewhat loose. The factors contributing to washback length have not been empirically investigated and there are likely to be intervening variables influencing the duration of the effect. For instance, students may or may not have to continue to take English classes after the test, or may have to use English at the university level. This may also influence their later use of English language learning strategies for example. Moreover, the difficulty of establishing the length of washback, which would ideally entail longitudinal study with students before and after the test, means that care must be taken in discussing findings on washback length.

In this study, as far as washback length was concerned, the findings appeared to be inconclusive. There were significant changes in students' learning strategies, but these cannot be attributed solely to the washback of the MUET without taking into consideration other possible intervening variables, as mentioned earlier. Perhaps the only way of establishing the washback length of the MUET would be to collect data immediately before the students start tertiary education in their respective university. This study used data from students who had already enrolled, which potentially affected the reliability of the findings on washback length. However,

the qualitative data can be considered useful as the students were specifically asked to compare what they did when they were preparing for the MUET and the English language strategies they were employing at university. The qualitative findings reported that there were differences to a certain extent when it came to the use of English language learning strategies before and after the MUET. In this study, either strategy use increased in quality and quantity or decreased in quality and quantity. For those who made better use of strategies when preparing for the MUET but ceased doing so once they entered university, the reason given was that the pressure of the exam was no longer there. Although they would have to sit the end of the semester exam for their English language course, the pressure was different than in the MUET; unlike for the MUET, they did not see the need to obtain a high score in the exam and hence they no longer felt the need to continue doing what they had in the past.

In contrast, another group of students in this study reported maintaining and even increasing the quality and quantity of their language learning strategies once they entered university. To a certain extent, they attributed their current language learning strategies to the MUET, but they also mentioned the language demands in the university, not only in English language courses but also other major courses taught in English. Hence, they felt the need to continue improving their English language mastery even without the MUET. Therefore, even though they were still using the same language learning strategies as when they were preparing for the MUET, it is not safe to conclude that this reflects long-term washback, as it was rather other demands at university level driving their behaviour.

According to the model proposed by Green (2007a), the washback intensity of a test is influenced by its perceived importance and perceived difficulty. He further listed three criteria for determining when participants will experience the most intense washback, i.e. when they:

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

(Green, 2007a, pp. 24–25)

As discussed in relation to RQ1, washback intensity may be “diluted” by certain factors, such as in this case the availability of other options: the students being allowed to repeat the MUET until they get their desired band, or choosing courses or universities commensurate with their MUET band. Gates (1995) cautions that the availability of other alternatives may alter the students’ perceptions of the stakes associated with a test, which will have implications for washback in at least two respects. First, other test options will allow students to choose which test best fits with what they are learning in their English language classes. Second, the opportunity to sit a similar or a different test again if they do not score well the first time will more or less affect their perceptions of how important the test is for them, as found in this study. The intensity of washback is also likely to be seasonal, increasing with the approach of the test date (Bailey, 1999; Watanabe, 1997). As the test draws closer, “desperate” learners will rely on their old test preparation methods and use the language learning strategies that they know will work best for them, even though there is only a short-term effect, also evidenced in this study. However, this study discovered that other than perceived test importance, perceived self-efficacy is a greater mediator of washback intensity than perceived test difficulty.

CHAPTER 6: CONCLUSION

6.1 Introduction

One of the aims of this study was to contribute more knowledge to the field of washback by tapping into the less explored areas of this phenomenon, specifically how the students' perceptions of a test affect the washback of the test. It was established by Alderson (2004) that factors such as teachers' beliefs about teaching and learning, their degree of professionalism, their experience of teaching and dealing with tests and the adequacy of training, to name but a few, play a vital role in the washback effect of a test. Using Green's (2007a) model of washback as the main framework to guide this study, it was posited that the washback intensity of a test could be influenced by students' perceptions of the test importance and test difficulty. With these two independent variables in mind, this study argued that perceived self-efficacy might be a good addition to the equation. Based on Zhan's (2009) washback study of learners' perceptions and how their possible selves helped to steer their actions when preparing for the test, the concept of self is considered a factor on which washback studies should start focusing more. Hence, the three possible factors driving washback based on students' perceptions investigated in this study were perceived test importance and perceived test difficulty, adapted from Green's (2007a) washback model, and perceived self-efficacy, based on Zhan's (2009) doctoral dissertation on washback and possible selves.

6.2 Summary of the main findings

It seems apparent that perceptions of a test play a considerable role in determining students' use of language learning strategies when preparing for the test. In the context of this study, the

washback of the Malaysia University English Test (MUET) appeared to be more or less influenced by how students viewed the test as their perceptions affected their self-efficacy. Using a high-stakes test as a lever of change, as intended by the Ministry of Education in Malaysia, appears to be useful to a certain extent, the outcome is perhaps not what the stakeholders involved might have hoped, as minimal change was reported. However, acquiring and mastering a language takes time and hence, expecting one test to improve learners' language proficiency dramatically is deemed too ambitious. Moreover, as has repeatedly been demonstrated in research on washback, it is far from straightforward; indeed, the more washback is explored, the more complex it is found to be. Perhaps then, it is fair to conclude from what has been found thus far, both in this study and in the washback literature in general, that yes, a high-stakes test can indeed help motivate students to focus more on certain skills and to work harder, but no, it cannot drastically improve students' language proficiency; if that is so, improved proficiency might not be an appropriate measure of whether a test causes positive or negative washback. Even if the data show such an improvement, with what degree of confidence could one conclude that the students' achievement is solely the result of introducing a high-stakes test? It is extremely difficult, if not impossible, to control other variables when it comes to researching learners in their natural setting. Hence, while it might have been useful for washback researchers to look at students' outcomes in the past in an attempt to determine the existence of any washback, be it positive or negative, now that the phenomenon has been explored more extensively, it has become increasingly apparent that it is more realistic to examine washback in terms of process, not outcome. In particular, students' efforts can show if the thought of sitting a high-stakes test encourages or discourages them in terms of learning. In this regard, their perceptions can help researchers to explain their behaviours.

From this study, as reported in the discussion in Chapter 5, it appeared that the students' perceptions of the test did to a certain extent influence the washback effect of the MUET. In this study, washback was examined in terms of the process involved, specifically in terms of the influence on the students' use of language learning strategies. This study argued that washback intensity could be considered high when language learning strategies were used frequently by the students when preparing for the test. In this study, it appeared that perceived test importance and perceived self-efficacy significantly correlated with and affected the students' use of language learning strategies when preparing for the test. However, perceived test difficulty did not seem to be a good determinant of washback intensity compared with test importance and self-efficacy. While perceived test difficulty did correlate well with perceived self-efficacy, it did not affect the students' use of language learning strategies directly, unlike perceived test importance and perceived self-efficacy. Hence, it might be necessary to incorporate self-efficacy in Green's (2007a) model as it appeared to be a better predictor of washback than test difficulty. However, perceived test difficulty was not altogether without influence as it was related to how the students shaped their self-efficacy.

Another overarching aim of this study was to fill an existing gap in terms of the aspect of the length of washback. Washback length refers to the continuation of the influence of the test even after the students have sat the test. Other than the definition of the concept, literature on washback length is lacking and to the best of my knowledge, there has been no attempt to investigate and measure it. This might be due to the difficulty of investigating this washback mechanism or it might not even exist. Hence, this study aimed to measure the length of washback from the MUET by involving a whole different group of students, who had already taken the test, to see if they were still affected by it once they had started university. The students were asked to

reflect on the English language learning strategies that they used previously in preparing for the test and compare them to those they were using at the time of the study. The results are inconclusive, primarily due to the many other possible intervening variables that need to be taken into consideration. Thus, it is not possible to make a strong claim that whatever the students were doing in terms of the use of language learning strategies at university was not because of other variables but solely because of the MUET. Separating out these variables seems impossible, as there are so many other factors that need to be addressed, such as the context, the students' course of learning, the students' attitudes, etc. However, an attempt was made to interpret the findings concerning the washback length of the MUET and in this regard, the qualitative data from the students were of great help. Although there were both increases and decreases in terms of the language learning strategies used before and after the MUET, they were not significant. This could be attributed to the fact that learning English is an ongoing process of lifetime learning, especially for ESL speakers. While the existence of a test might to a certain extent influence the intensity of the language strategies used, students are likely to continue to employ them (to a greater or lesser degree) as long as they are beneficial. Hence, it does not seem possible to equate an increase or decrease in the intensity of use with a short-term or long-term washback effect. Furthermore, there was a possibility that the students who passed the MUET using a certain range of strategies might retain or maintain similar strategies on the basis that these had helped them pass the high-stakes test in the past and perceived these strategies to be working for them.

6.3 Limitations of the study

Although the study shed light on the students' perceptions and how these relate to the washback of the MUET, a number of caveats need to be noted. Like any empirical study, there were some limitations in this study that need to be acknowledged.

First, in terms of the generalizability of the findings, the quantitative study of students comprised a sample of only 375: 137 pre-university students in Group A and 238 undergraduate students in Group B. With such a small sample size, caution must be applied in attempting to generalize the findings to the wider population of pre-university and undergraduate students in Malaysia, let alone beyond. It can also be seen that the ratios of students in terms of gender and proficiency level were uneven. This was partly due to the limited number of participants, especially in the individual interviews, with only one male student and ten female students. For the teacher questionnaire, 36 teachers participated (31 females and only 5 males). Initially, gender was not taken into consideration as one of the main variables under investigation until the data analysis revealed the need to so. This issue could have been avoided if it had been established that gender and proficiency level would need to be addressed before data collection commenced. At the time, gender and language proficiency were solely collected as demographic information on the participants, as opposed to treating them as valid variables, which seemed ambitious given the scope of the study. However, incorporating them in the sampling criteria would have yielded more reliable results.

As already established, to date, there have been few reliable instruments available to measure washback accurately, be it in terms of intensity or length. Hence, based on Green's (2007a) washback study, Oxford's (1990) Strategy Inventory for Language Learning (SILL) was used in an attempt to measure the intensity of washback from the MUET, looking at the frequency

of the language learning strategies used. Other constructs in the student questionnaire were either partially adopted or adapted from previous relevant studies.

For the components measuring washback length, both empirical studies and instruments pertaining to washback length were very scarce, which posed a major challenge. Hence, the design of the instrument to measure washback length had to be done from scratch based on my assumptions and limited prior knowledge regarding washback length. It proved possible to collect data on washback length, but strong conclusions could not be drawn as the findings were rather inconclusive. Nevertheless, the open-ended question posed at the end of the questionnaire helped contribute to an understanding of the washback length of the MUET to a certain extent.

The student questionnaire was only piloted once due to time constraints: the questionnaire needed to be ready as soon as possible as it was almost the end of the semester by the time the data were collected. Waiting for the pre-university students to return from their semester break was not considered a favourable option as they would already have sat the MUET. This limitation prevented advanced statistical analysis of the pilot data and might have affected the reliability and validity of the findings had the study only relied on one set of data. The triangulation of data helped immensely to validate the findings.

It should also be pointed out that the study used a small convenience sample for the student interviews. The interview sample comprised 11 pre-university students who indicated their willingness to be interviewed in the questionnaire and 10 out of the 11 students were female. Such a small, imbalanced sample might have led to a certain degree of bias in the findings. Furthermore, the interviews were conducted via telephone due to time constraints. It should be noted here that Malaysian ESL students tend to be reserved when it comes to voicing their opinions. Due to this, it was quite challenging to engage with the interview participants over the telephone and they

tended to give very short responses. A better approach would perhaps have been to conduct the interviews face to face, particularly as I was the “primary instrument of data collection and analysis” (Merriam, 1998, p. 42). Moreover, the inability to probe participants’ responses further was costly in terms of the quality of the data collected.

Next, the data gathered for this study were mostly based on self-reports (student questionnaire, student interviews and teacher questionnaire). This raises the issue of how reliable the data are in terms of the match between self-reported information and actual behaviours. More in-depth study with various stakeholders involved in the test would help to alleviate this problem, but would require a team of researchers and would be beyond the scope of research at the PhD level. Instead, triangulation through the collection of classroom observation data was used as a way of minimizing this weakness, but discrepancies and mismatches between data from different sources gave rise to the question of how reliable the self-report data were and which data should be trusted. In this study, due to limited time and accessibility it was only possible to conduct classroom observation in five classes from two different schools. Initially, two teachers had agreed to be observed in one of the schools involved in this study. However, over the course of the study, one of the teachers was admitted to hospital, causing her to withdraw from the research. Furthermore, the fact it was approaching the end of the semester by the time of data collection hindered the collection of observation data somewhat and it was not possible to extend the research duration. As far as the teachers’ data collected in this study was concerned, it could be argued that the role of teacher data was not central to this study and some parts of the data (i.e. classroom observation data) was not fully utilized. Teachers’ data was initially planned to be used as a complement to the students’ data as the focus of this study was on the washback on the learners. However, due to mismanaging and poor planning during data collection, additional data from the

teachers suggested that it could actually stand on its own, not as a complementary element to the students' data. However, poor planning resulted in a time consuming process causing the data not being exploited to the full.

6.4 Implications of the study

Despite the limitations, the findings of this study have a number of important implications for future practice. The study has given several insights to both theory and practice of the washback impact of the MUET on learners. From the practical perspective, the study has created awareness of the importance of the MUET as evident in this study. It was revealed that the test is doing what it was designed to do, which were to (1) bridge the gap in English language needs between secondary and tertiary education, and to (2) consolidate and enhance the English language proficiency of students preparing to enter Malaysian public universities. The findings revealed that the MUET may have a positive effect on students' approaches to language learning (in terms of language learning strategies adopted), and to classroom practice (encouraging practice in oral and aural as rather than just written skills). The findings suggest that the MUET provides the students with an avenue to actually practice their four language skills, especially the two skills that were not formally tested before during primary and secondary school. Not only do they get to practice new language skills when preparing for the MUET, they also get to improve their other language skills as well, which would support their study at the tertiary level later on.

Although this study did not specifically look at the design of the MUET, there are some implications that can be taken from the findings of this study in relation to its design and content. The teachers' data showed that the MUET reading component was considered too difficult for the students. The teachers' concerns were mainly related to the practicality of the MUET reading

component and its applicability in real life, as students would not be encountering such difficult texts as those in the text outside their learning context. In addition, a teacher mentioned that the reading test was too long and suggested that this component be reduced to four reading passages as opposed to six. The findings of this study also revealed that the recordings used in the MUET listening test were viewed as staged and not applicable to real-life situations. Also, native English speakers are normally used to record the conversations for the MUET listening test and some of the students in this study found it difficult to comprehend their accents. Perhaps the recordings for the listening test should be done by local Malaysians. Essentially, the take-home message from the findings of this study is that the practicality and applicability of the language skills to real life should come first.

From the theoretical perspective, this study adds to the available literature on washback on the learners specifically in terms of its relationship with students' perceptions of the test. It demonstrated the importance of taking into consideration individual differences when trying to understand how washback operates at the process level. The study revealed that the students' perceptions of a test could influence its washback. It was found that the students perceived newly tested language skills, in this case speaking, as difficult. This was due to the novelty effect and to a certain extent it affected their self-efficacy, which can either reduce or increase the washback intensity of the test. To maximize the intended positive washback of a test, it may well be more effective to deal with the students' perceptions as opposed to altering the test itself. This study also found that assigning different weighting to specific test components of the MUET to encourage the stakeholders, both students and teachers, to pay more attention to certain language skills was unsuccessful. However, this finding might be unique to the context of the present study, as it has been reported in previous washback research that weighting components does to a certain extent

influence students' test preparation strategies, with a greater focus reportedly given on components with the higher weighting. In the case of the MUET, this did not seem to be the case, as the test component that was given the most focus appeared to be speaking, which only contributes 15% to the overall MUET band. Reading, on the other hand, which has the highest weighting of 45%, was less emphasized.

The study also revealed that despite being a high-stakes test with important consequences, the MUET did not guarantee similar perceptions or reactions among test takers. A small number of the students in this study did not seem to be greatly affected by the MUET, as they considered that if they did not pass the MUET with flying colours, there were other options available to them (see Chapter 5). This rather "diluted" the power of the MUET, as it was viewed only as one of the requirements for students to enrol in public university in Malaysia. However, recently, Malaysia has had a change of government after the opposition party won the 2018 election and this leaves to many policies being revised, including the English language policy. On 6 June 2018, it was announced that the federal government would introduce an English competency test for senior civil servants to improve the quality of public services in the country. The newly elected Prime Minister of Malaysia, Mahathir Mohamed, announced that Cabinet members had expressed their concern about the importance of English, especially among senior civil servants and said "*Senior civil servants must be proficient in the English language because they often have to negotiate with outsiders on matters of government policies*" (6 June 2018, Putrajaya). This move could well provide a boost to the motivation for Malaysian students to start taking English language learning more seriously, not only for the sake of taking a test. It was revealed in this study that some of the students and even teachers were of the opinion that the MUET was not important if the students were not planning to pursue tertiary education in Malaysia as it would not affect their future

undertakings. Hopefully, this new policy will help change these students' mindsets about English language learning for the sake of their future, encouraging them to make full use of the MUET to help improve their English language mastery. As revealed in this study, once perceptions change, the course of action will follow, hopefully for the better.

As far as washback length is concerned, to my knowledge, such a study which attempted to empirically measure the washback length of a test has not been conducted. The findings therefore, would provide a conceptual basis for further research and exploration on washback length. Based on what was uncovered in this study, it appeared that the result was inconclusive due to the difficulty to identify washback length. This may be attributed to the limitations of this study in terms of its data collection instrument for Group B's students and fairly limited qualitative data. Furthermore, the confounding variables suggest that identifying this effect appeared to be impractical other than providing a list of language learning strategies that might be effective for students.

6.5 Suggestions for future research

This study highlights potential avenues for research on perceptions and washback. In this study, the number of participants involved was limited and they only came from three different secondary schools for Group A students and one public university for Group B students. The variability of the participants was rather limited and the findings of the study could not be generalized. It would be beneficial to carry out studies of a similar nature in more schools and universities, covering a wider geographical area and both urban and rural. More classroom observations should also be conducted to compensate for the use of self-report data as potential issues with the reliability of such data have always been a concern in washback studies.

It has been established that there is very little information, let alone empirical study, on washback length. In the present study, the student questionnaire was not fruitful in providing useful data on washback length. While using a questionnaire seemed practical at the time of data collection, asking the students to recall what they did almost a year ago posed a risk for the reliability of the data, especially as a list of language learning strategies was provided in the questionnaire. However, the qualitative data seemed to be a better prospect in uncovering the length of washback. Instead of providing the students with a list of language learning strategies for them to choose from, it would perhaps be more appropriate to ask the students to write down what strategies they used before and after the MUET. Hence, a qualitative approach within a longitudinal study design would seem to be a better option to collect data on washback length. A longitudinal study design would also help in terms of allowing researchers to follow the same group of participants, as opposed to asking students to recall what they did before they sat the test and what they are currently doing, which was the case in the present study. Doing so potentially lessened the reliability of the data as the students might not have been able to give an accurate depiction of what they had done in the past. Furthermore, qualitative data would help to mitigate the effect of possible intervening variables, since the students would be asked to reflect on their own test preparation or language learning strategies in relation to a specific test under investigation. Such extensive studies would enrich the findings of the study and would hopefully contribute more knowledge on this scarcely explored area of washback. In summary, targeting washback, as being emphasized multiple times throughout this study, is quite complex because the testing effects are indirect, unpredictable and context-dependence. There are still some underexplored issues worth re-exploring in the washback cycle especially in a context that is

'drenched' with exam-oriented practice like in Malaysia. Washback on learning is undoubtedly a promising research area for us in the future.

Appendices

Appendix A – Student Questionnaire (Group A)

UNIVERSITY *of* York

Department of Education

Heslington, York, YO10 5DD

Tel: (01904) 323460

Web: <http://www.york.ac.uk/education>

Dear Student,

My name is Najihah Mahmud, and I am conducting a study for my doctorate at the University of York in the United Kingdom. I would like to ask you for your opinions of learning English and preparing for the Malaysian University English Test (MUET). The purpose of my study is to critically examine the washback effect of MUET on the teaching and learning of ESL in Malaysia in relation to learners' perceptions and language learning strategies. To help me, please fill in this questionnaire based on your own experience. It should take no more than 20 minutes to complete. Thank you so much.

Students' Perceptions of the Malaysian University English Test (MUET)

Background Information - Please tick one appropriate answer or provide written answer for each item.

- Gender Male Female
- English Malaysian Certificate of Education (MCE/SPM) result
 A+ A A- B C+ C D E F

Part I : Perception of the MUET – Please circle the appropriate answer based on your experience preparing for the MUET.

- The MUET is an important test to me.
- It is very important for my future undertakings that I do well in the MUET.

	<i>Strongly disagree</i>	<i>Disagree</i>	<i>Undecided</i>	<i>Agree</i>	<i>Strongly agree</i>
1	1	2	3	4	5
2	1	2	3	4	5

3. If I do poorly in the MUET, my chance to get into top universities will be affected.
4. If I do poorly in the MUET, my chance to enrol in my desired course will be affected.
5. I believe I will receive an excellent grade in the MUET.
6. Taking into consideration of its difficulty, I think I can perform very well in the MUET.
7. Taking into consideration of my ability, I think I can perform very well in the MUET.
8. I'm confident I can do an excellent job on the assignment and tasks in the MUET class.
9. My MUET teacher makes me practise my writing skills more than before.
10. My MUET teacher makes me practise my listening skills more than before.
11. My MUET teacher makes me practise my speaking skills more than before.
12. My MUET teacher makes me practise my reading skills more than before.

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

13. On a scale from 1 to 5, with 1 = Very Easy, and 5 = Very difficult, how difficult do you think the MUET is for you? Circle the number to indicate your answer below:



14. Which of the following language areas you think are the most important to score highly in MUET (Tick all suitable answers)

	Very Important	Moderately Important	Not Important
Listening			
Speaking			
Reading			
Writing			

37. I look for words in my own language that are similar to new words in English.	1	2	3	4	5
38. I try to find patterns in English.	1	2	3	4	5
39. I find the meaning of an English word by dividing it into parts that I understand.	1	2	3	4	5
40. I try not to translate word-for-word.	1	2	3	4	5
41. I make summaries of information that I hear or read in English.	1	2	3	4	5
42. To understand unfamiliar English words, I make guesses.	1	2	3	4	5
43. When I can't think of a word during a conversation in English, I use gestures.	1	2	3	4	5
44. I make up new words if I do not know the right ones in English.	1	2	3	4	5
45. I read English without looking up every new word.	1	2	3	4	5
46. I try to guess what the other person will say next in English.	1	2	3	4	5
47. If I can't think of an English word, I use a word or phrase that means the same thing.	1	2	3	4	5
48. I try to find as many ways as I can to use my English.	1	2	3	4	5
49. I notice my English mistakes and use that information to help me do better.	1	2	3	4	5
50. I pay attention when someone is speaking English.	1	2	3	4	5
51. I try to find out how to be a better learner of English.	1	2	3	4	5
52. I plan my schedule so I will have enough time to study English.	1	2	3	4	5
53. I look for people I can talk to in English.	1	2	3	4	5
54. I look for opportunities to read as much as possible in English.	1	2	3	4	5
55. I have clear goals for improving my English skills.	1	2	3	4	5
56. I think about my progress in learning English.	1	2	3	4	5
57. I try to relax whenever I feel afraid of using English.	1	2	3	4	5
58. I encourage myself to speak English even when I am afraid of making a mistake.	1	2	3	4	5
59. I give myself a reward or treat when I do well in English.	1	2	3	4	5
60. I notice if I am tense or nervous when I am studying or using English.	1	2	3	4	5
61. I write down my feelings in a language learning diary.	1	2	3	4	5
62. I talk to someone else about how I feel when I am learning English.	1	2	3	4	5
63. If I do not understand something in English, I ask the other person to slow down or say it again.	1	2	3	4	5
64. I ask English speakers to correct me when I talk.	1	2	3	4	5
65. I practice English with other students.	1	2	3	4	5
66. I ask for help from English speakers.	1	2	3	4	5
67. I ask questions in English.	1	2	3	4	5
68. I try to learn about the culture of English speakers.	1	2	3	4	5

If you would like to take part in a telephone interview on your experience learning English (which will be conducted in Malay or English depending on your preference), please fill in your details below and I will get in touch with you.

Nick name : _____

Phone Number : _____

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE!

Appendix B – Student Questionnaire (Group B)

UNIVERSITY *of* York

Department of Education

Heslington, York, YO10 5DD

Tel: (01904) 323460

Web: <http://www.york.ac.uk/education>

Dear Student,

My name is Najihah Mahmud, and I am conducting a study for my doctorate at the University of York in the United Kingdom. I would like to ask you for your opinions of learning English and preparing for the Malaysian University English Test (MUET). The purpose of my study is to critically examine the washback effect of MUET on the teaching and learning of ESL in Malaysia in relation to learners' perceptions and language learning strategies. To help me, please fill in this questionnaire based on your own experience. It should take no more than 20 minutes to complete. Thank you so much.

Students' Perceptions of the Malaysian University English Test (MUET)

Background Information - *Please tick one appropriate answer or provide written answer for each item.*

1. Gender Male Female
2. English Malaysian Certificate of Education (MCE/SPM) result
A+ A A- B C+ C D E F

Language Learning Strategies – Please circle TWO appropriate answers for each item for the language learning strategies that you used when you were preparing for the MUET and your current language learning strategies.	MUET		NOW	
	True	Not true	True	Not True
1. I think of relationships between what I already know and new things I learn in English.	1	2	1	2
2. I use new English words in a sentence so I can remember them.	1	2	1	2
3. I connect the sound of a new English word and an image or picture of the word to help remember the word.	1	2	1	2
4. I remember a new English word by making a mental picture of a situation in which the word might be used.	1	2	1	2
5. I use rhymes to remember new English words.	1	2	1	2
6. I use flashcards to remember new English words.	1	2	1	2
7. I physically act out new English words.	1	2	1	2
8. I review English lessons often.	1	2	1	2
9. I remember new English words or phrases by remembering their location on the page, on the board, or on a street sign.	1	2	1	2
10. I say or write new English words several times.	1	2	1	2
11. I try to talk like native English speakers.	1	2	1	2
12. I practice the sounds of English.	1	2	1	2
13. I use the English words I know in different ways.	1	2	1	2
14. I start conversations in English.	1	2	1	2
15. I watch English language TV shows spoken in English or go to movies spoken in English.	1	2	1	2
16. I read for pleasure in English.	1	2	1	2
17. I write notes, messages, letters, or reports in English.	1	2	1	2
18. I first skim an English passage (read over the passage quickly) then go back and read carefully.	1	2	1	2
19. I look for words in my own language that are similar to new words in English.	1	2	1	2
20. I try to find patterns in English.	1	2	1	2
21. I find the meaning of an English word by dividing it into parts that I understand.	1	2	1	2
22. I try not to translate word-for-word.	1	2	1	2
23. I make summaries of information that I hear or read in English.	1	2	1	2
24. To understand unfamiliar English words, I make guesses.	1	2	1	2
25. When I can't think of a word during a conversation in English, I use gestures.	1	2	1	2
26. I make up new words if I do not know the right ones in English.	1	2	1	2
27. I read English without looking up every new word.	1	2	1	2
28. I try to guess what the other person will say next in English.	1	2	1	2

29.	If I can't think of an English word, I use a word or phrase that means the same thing.	1	2	1	2
30.	I try to find as many ways as I can to use my English.	1	2	1	2
31.	I notice my English mistakes and use that information to help me do better.	1	2	1	2
32.	I pay attention when someone is speaking English.	1	2	1	2
33.	I try to find out how to be a better learner of English.	1	2	1	2
34.	I plan my schedule so I will have enough time to study English.	1	2	1	2
35.	I look for people I can talk to in English.	1	2	1	2
36.	I look for opportunities to read as much as possible in English.	1	2	1	2
37.	I have clear goals for improving my English skills.	1	2	1	2
38.	I think about my progress in learning English.	1	2	1	2
39.	I try to relax whenever I feel afraid of using English.	1	2	1	2
40.	I encourage myself to speak English even when I am afraid of making a mistake.	1	2	1	2
41.	I give myself a reward or treat when I do well in English.	1	2	1	2
42.	I notice if I am tense or nervous when I am studying or using English.	1	2	1	2
43.	I write down my feelings in a language learning diary.	1	2	1	2
44.	I talk to someone else about how I feel when I am learning English.	1	2	1	2
45.	If I do not understand something in English, I ask the other person to slow down or say it again.	1	2	1	2
46.	I ask English speakers to correct me when I talk.	1	2	1	2
47.	I practice English with other students.	1	2	1	2
48.	I ask for help from English speakers.	1	2	1	2
49.	I ask questions in English.	1	2	1	2
50.	I try to learn about the culture of English speakers.	1	2	1	2

51. Is there any difference between your current English language learning strategies with the strategies that you used when you were preparing for the MUET in the past? Comment below.

THANK YOU FOR COMPLETING THIS QUESTIONNAIRE!

Appendix C – Teacher Questionnaire

Department of Education
 Heslington, York, YO10 5DD
 Tel: (01904) 323460
 Web: <http://www.york.ac.uk/education>

UNIVERSITY *of York*

Dear Teachers,

My name is Najihah Mahmud, and I am conducting a study for my doctorate at the University of York in the United Kingdom. I would like to ask you for your opinions of teaching English and preparing your students for the Malaysian University English Test (MUET). The purpose of my study is to critically examine the washback effect of MUET on the teaching and learning of ESL in Malaysia in relation to learners' motivation, self-efficacy and language learning strategies. To help me, please fill in this questionnaire based on your own experience. It should take no more than 20 minutes to complete. Thank you so much.

Part 1 : Please tick appropriate answer.

1. Your Gender:

Lelaki Perempuan

2. Number of years you have been teaching:

1-3 4-6 7-9 10 and above

3. Number of periods you teach English per week:

16-21 22-27 28-33 above 33

Part 2 : Please grade the following on a 5-point scale format where

1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

Please put 1, 2, 3, 4 or 5 in the box provided.

1. What do you find the most difficult aspects of teaching the MUET if any?

1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree

- i. Students' current English level
- ii. Class size
- iii. Inadequate textbooks and other available teaching resources
- iv. Noisy learning environment
- v. The lack of teaching and learning aids and facilities

- vi. Too heavy work load
 - vii. Inadequate time for students' practice of English outside the language classroom
2. What are the learning strategies you would recommend to your students to prepare for MUET?
1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree
- i. To learn to take better notes
 - ii. To expose themselves to various English media
 - iii. To learn to express their opinions in class
 - iv. To put more emphasis on listening and speaking
 - v. To learn to initiate questions
 - vi. To be more active in classroom participation
 - vii. To use English more in their daily life
 - viii. To change from passive learning to active learning
 - ix. To communicate more in English
3. What types of activities do you think should be involved with language learning?
1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree
- i. Task-oriented activities
 - ii. Language games
 - iii. Role play and group discussion
 - iv. Exposure to various English media
 - v. Authentic materials
 - vi. Training in basic language knowledge
 - vii. Extracurricular activities
4. In what ways do you think you would like to to motivate your students in learning English?
1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree
- i. To do more mock exam papers
 - ii. To use more authentic materials
 - iii. To organise real life language activities
 - iv. To do more interesting language games
 - v. To give students more encouragement to learn
 - vi. To create a positive attitude toward language learning
 - vii. To provide students with effective language learning strategies

- viii. To have better classroom discipline
5. What do you think are the basic functions of mock tests in school?
1=Strongly disagree, 2=Disagree, 3=Undecided, 4=Agree, 5=Strongly Agree
- i. To give feedback to teachers
 - ii. To assess students' learning difficulties
 - iii. To motivate students
 - iv. To direct students' learning
 - v. To prepare students for public examination
 - vi. To identify area of re-teaching

Part 3 : Please grade the following on a 5-point scale where

1=never, 2=seldom, 3=sometimes, 4=often, 5=always, and put 1, 2, 3, 4, or 5 in the box provided.

1. How often do you consider the following aspects when you prepare your lessons?
1=never, 2=seldom, 3=sometimes, 4=often, 5=always
 - i. The methods of teaching
 - ii. The contents of teaching
 - iii. The tasks to be performed in teaching
 - iv. The skills to be taught
 - v. Any supplementary materials to be used
 - vi. How to motivate students to learn
 - vii. Homework to give to students

2. How often do you do the following activities in class?
1=never, 2=seldom, 3=sometimes, 4=often, 5=always
 - i. Tell the students the aims of each lesson
 - ii. Demonstrate how to do particular language activities
 - iii. Explain the meaning of the text
 - iv. Explain specific language items such as words or sentences
 - v. Explain textbook exercises
 - vi. Explain mock exams
 - vii. Organise language games
 - viii. Organise group work or discussion
 - ix. Organise integrated language tasks

3. How often do you use the following teaching and learning aids in your teaching?

1=never, 2=seldom, 3=sometimes, 4=often, 5=always

- i. Textbooks
- ii. Supplementary materials
- iii. Television/Radio
- iv. Newspapers
- v. Language laboratory
- vi. Picture and/or cards
- vii. Teaching syllabus
- viii. Examination syllabus

Part 4 : Please answer the following questions

1. Do the MUET influence the way in which you teach your English class? (If yes, how? If no, why?)

2. Do you think the MUET assessed your students' English ability appropriately? (If yes, how? If no, why?)

3. Do you think that the MUET is necessary for university entrance? (If yes/no, why?)

4. Do you think that the MUET influence the future of Malaysian students? (If yes, how? If no, why?)



“THANK YOU FOR YOUR TIME”

Appendix D – Student Interview Questions

Student Telephone Interview Questions

Perception on the MUET

1. What do you think of the MUET compared to other English language test that you've taken before?
 - a. Do you feel pressured about taking the MUET? (Probe question)
 - b. Do you think you are motivated to study English harder because you have to sit for the MUET? (Probe question)
2. How confident are you when it comes to the MUET?
3. Do you think that the MUET is necessary for the university entrance exams?
4. Do you think that the MUET influence the future of high school students (e.g., profession, family life, and personal development)?
5. What are the consequences that you will face if you get a low score in the MUET/English language test?
6. Which of the four language skills do you consider to be the most important to survive or to do well in the university?

Preparation for the MUET

1. How do you prepare for the MUET?
 - a. Did you prepare for the MUET outside of English classes in school? If yes, please describe what kinds of preparation you did. (Probe question)
 - b. Do you need to prepare differently for the MUET? (Probe question)
2. How would you describe your MUET preparation class?
 - a. Do you think your MUET preparation class is useful to prepare you for the MUET? (Probe question)

CONFIDENTIAL*

2

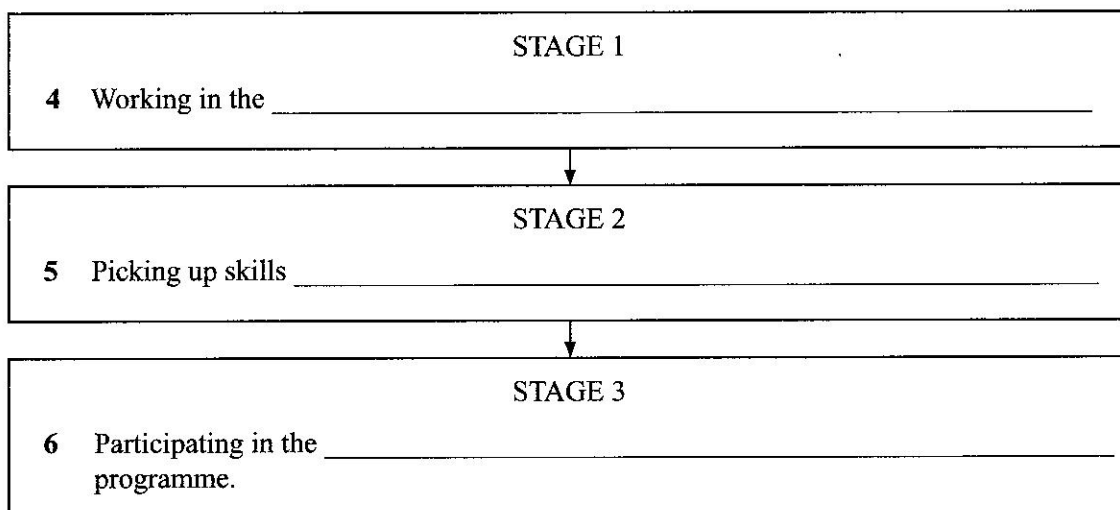
Part I**[8 marks]****For questions 1 to 3, write your answer in NOT MORE THAN THREE WORDS.**

1 The Financial Literacy Campaign is conducted through _____

The two mottos used in the Financial Literacy Campaign are:

2 _____

3 _____

For questions 4 to 6, write your answer in NOT MORE THAN FIVE WORDS.**Training Programme****For questions 7 and 8, circle the correct answer.**

7 The most important skill the trainee gained from the workshop was the ability to

- A plan the budget
- B build public relations
- C work as a team

8 The speaker became a trainee by applying through

- A an e-mail to the bank
- B the bank's website
- C a university

CONFIDENTIAL*

3

Part II

[6 marks]

For questions 9 to 12, circle the correct answer.

- 9 About 30% of Malaysian children are likely to get allergies because
- A it is difficult to control a child's food intake
 - B parents are not aware of preventive measures
 - C they are exposed to changes in the environment
- 10 Allergies can be treated
- A and can be cured
 - B but can reappear
 - C if detected early
- 11 Mothers are encouraged to breast feed in order to
- A exclude cow's milk in the diet
 - B reduce the likelihood of eczema
 - C delay the introduction of solid food
- 12 The speaker's main intention is to inform the audience about
- A the symptoms of allergies
 - B the treatment of allergies
 - C how to prevent allergies

For questions 13 and 14, choose one letter A to F from the box below to complete each sentence.

- | | |
|---|-----------------|
| A | milk formula |
| B | solid food |
| C | dust mites |
| D | house pets |
| E | cigarette smoke |
| F | family genetics |

- 13 According to the speaker, respiratory allergies are likely to be caused by _____
- 14 Parents can control causes of allergies except _____

CONFIDENTIAL*

4

Part III

[6 marks]

For questions 15 to 20, write your answer in NOT MORE THAN FIVE WORDS.

15 The cooking competition is being held to _____

16 The winner will be given shopping vouchers and _____

17 The settlement in the Bujang Valley is older than _____

18 The jetty and iron furnaces show that people in the Bujang Valley _____

19 School officials solved behavioural problems on the school bus by putting in a _____

20 An unexpected result of the experiment was that the students _____

CONFIDENTIAL*

3

Candidate A**Instructions to candidates:**

This test consists of two tasks, Task A and Task B.

Task A is carried out first followed by Task B.

*You are given **one** minute to read the instructions and the tasks given.*

Task A: Individual Presentation (2 minutes)

- *You have **two** minutes to prepare your response.*
- *You have **two** minutes to present your views.*
- *Listen to the others while they are making their presentations and take down notes for the group interaction in Task B.*

Situation

Malaysia needs to produce more world-class sportsmen and sportswomen. Who can contribute towards achieving this?

Task A: You think that **parents** play an important role in producing more world-class sportsmen and sportswomen. Elaborate.

Tear off along the perforated line.

CONFIDENTIAL*

5

Candidate B**Instructions to candidates:**

This test consists of two tasks, Task A and Task B.

Task A is carried out first followed by Task B.

*You are given **one** minute to read the instructions and the tasks given.*

Task A: Individual Presentation (2 minutes)

- *You have **two** minutes to prepare your response.*
- *You have **two** minutes to present your views.*
- *Listen to the others while they are making their presentations and take down notes for the group interaction in Task B.*

Situation

Malaysia needs to produce more world-class sportsmen and sportswomen. Who can contribute towards achieving this?

Task A: You think that **schools** play an important role in producing more world-class sportsmen and sportswomen. Elaborate.

Tear off along the perforated line.

CONFIDENTIAL*

7

Candidate C**Instructions to candidates:**

This test consists of two tasks, Task A and Task B.

Task A is carried out first followed by Task B.

*You are given **one** minute to read the instructions and the tasks given.*

Task A: Individual Presentation (2 minutes)

- *You have **two** minutes to prepare your response.*
- *You have **two** minutes to present your views.*
- *Listen to the others while they are making their presentations and take down notes for the group interaction in Task B.*

Situation

Malaysia needs to produce more world-class sportsmen and sportswomen. Who can contribute towards achieving this?

Task A: You think that **the public** plays an important role in producing more world-class sportsmen and sportswomen. Elaborate.

Tear off along the perforated line.

CONFIDENTIAL*

9

Candidate D**Instructions to candidates:**

This test consists of two tasks, Task A and Task B.

Task A is carried out first followed by Task B.

*You are given **one** minute to read the instructions and the tasks given.*

Task A: Individual Presentation (2 minutes)

- *You have **two** minutes to prepare your response.*
- *You have **two** minutes to present your views.*
- *Listen to the others while they are making their presentations and take down notes for the group interaction in Task B.*

Situation

Malaysia needs to produce more world-class sportsmen and sportswomen. Who can contribute towards achieving this?

Task A: You think that **the Ministry of Youth and Sports** plays an important role in producing more world-class sportsmen and sportswomen. Elaborate.

CONFIDENTIAL*

10

Candidate A, B, C and D**Task B: Group Interaction (10 minutes)**

- *You have **two** minutes to prepare for the discussion.*
- *You may maintain or change your views presented in Task A.*
- *In your discussion, you may support or oppose the other candidates' views.*
- *At the end of the discussion, try to come to a group decision.*
- *You are given **ten** minutes for the discussion.*

Situation

Malaysia needs to produce more world-class sportsmen and sportswomen. Who can contribute towards achieving this?

Task B: Discuss which of the following play(s) **the most important role** in producing more world-class sportsmen and sportswomen.

- Parents
- Schools
- The public
- The Ministry of Youth and Sports

CONFIDENTIAL*

2

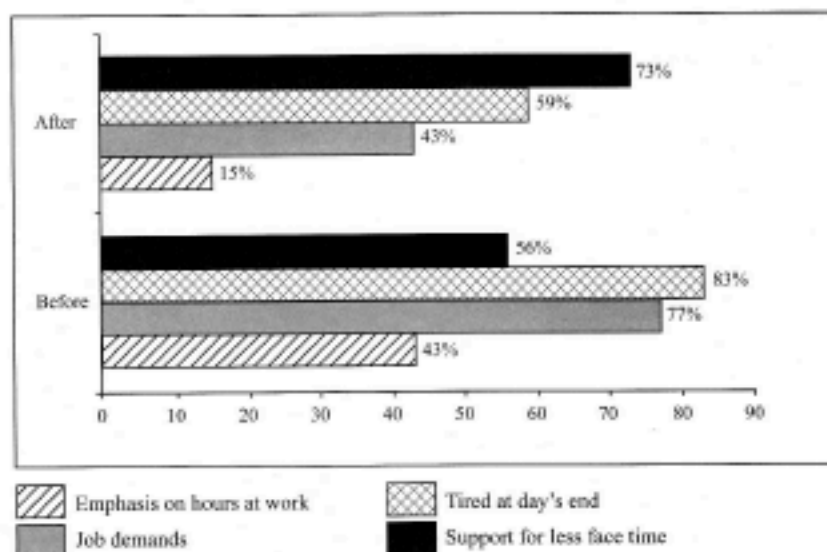
Question 1 to 7 are based on the following passage.

1 The hotel business is relentless. The management has to provide twenty-four hours service, 365 days a year, and every single day is just as important as any other day. Not surprisingly, M&M Hotel, which prides itself on providing excellent customer service, had for many years a deeply-ingrained culture of 'face time' – the more hours you put in, the better. That philosophy of 'see and be seen' was effective for serving customers, but it had a price: the management were finding it increasingly tough to recruit talented people and some existing managers were leaving, often because they wanted to spend more time with their families. 5

2 In the following year, M&M Hotel implemented a test programme to help managers strike a better balance between their professional and personal lives while maintaining the quality of its customer service and the bottom line of its financial results. They found a lot of quick fixes by eliminating redundant meetings and other inefficient procedures. For instance, they learnt that managers could file certain business reports less frequently and that many of the regular scheduled meetings were unnecessary. They also re-examined certain hotel procedures they were following, traditionally. For instance, the scheduled overlap time of front desk manager's with the person on the next shift was reduced from one hour to only fifteen minutes. Additionally, managers were given better Information Technology (IT) support so that they could communicate with customers through email and get connected to relevant sections within minutes to get immediate assistance. 10 15 20

3 At the end of the test programme, managers reported working an average of five hours less each week. Perhaps, more important, was the change in attitudes (Figure 1).

Figure 1: Attitude Adjustment



800/3/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

3

- 4 Before the test programme, 77% of managers felt that their jobs were so demanding that they could not take adequate care of their personal and family responsibilities. At the end of the programme, that percentage had plummeted to 43%. In addition, the percentage of managers who felt that the emphasis was on hours worked, plunged from 43% to 15%. One of the most important things shown was that people could be just as productive when they worked fewer hours. This is so because they are extra-motivated to get things done and they do not waste time in doing what they need to do. 25 30

(Adapted from *Harvard Business Review*, November 2001)

- 1 In the 'face time' work culture, the longer a manager spends time at work, the better it is.
 - A True
 - B False
 - C Not stated
- 2 The main objective of the test programme was to change the employees' attitude towards their job.
 - A True
 - B False
 - C Not stated
- 3 The test programme that was implemented reviewed the work procedures.
 - A True
 - B False
 - C Not stated
- 4 In Figure 1, the test programme showed an increase in the percentage of managers who felt tired at the end of the day.
 - A True
 - B False
 - C Not stated
- 5 From Figure 1, it can be inferred that the managers were happy with the changes made.
 - A True
 - B False
 - C Not stated
- 6 Work productivity declined with fewer hours at work.
 - A True
 - B False
 - C Not stated

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*

CONFIDENTIAL*

4

- 7 It can be concluded from the passage that working in a hotel industry is rewarding.
- A True
 - B False
 - C Not stated

CONFIDENTIAL*

5

Questions 8 to 14 are based on the following passage.

Caffeine Myths

- 1 Through the years, the public has been buffeted by much misguided information about caffeine and its most common source, coffee. In March, the Centre for Science in the Public Interest published a comprehensive appraisal of scientific reports in its *Nutrition Action Health* newsletter. Its findings and those of other research reports follow. 5
- 2 **Hydration.** It was long thought that caffeinated beverages were diuretics, but studies reviewed last year found that people who consumed drinks containing up to 550 milligrams of caffeine produced no more urine than when drinking fluids free of caffeine. Above 575 milligrams, the drug was a diuretic.
- 3 So even a Starbucks Grande, with 330 milligrams of caffeine, will not send you to a bathroom any sooner than if you drank 16 ounces of pure water. Drinks containing usual doses of caffeine are hydrating and, like water, contribute to the body's daily water needs. 10
- 4 **Cancer.** Panic swept this coffee-dependent nation in 1981 when a Harvard study tied the drink to a higher risk of pancreatic cancer. Coffee consumption temporarily plummeted, and the researchers later concluded that perhaps smoking, not coffee, was the culprit. 15
- 5 In an international review of 66 studies last year, scientists found coffee drinking had little if any effect on the risk of developing kidney cancer. In fact, another review suggested that compared with people who do not drink coffee, those who do have half the risk of developing liver cancer. 20
- 6 And a study of 59 000 women in Sweden found no connection between coffee, tea or caffeine consumption and breast cancer.
- 7 **Weight loss.** Here's a bummer. Although caffeine speeds up metabolism, with 100 milligrams burning an extra 75 to 100 calories a day, no long-term benefit to weight control has been demonstrated. In fact, in a study of more than 58 000 health professionals followed for 12 years, both men and women who increased their caffeine consumption gained more weight than those who did not. 25
- 8 Probably the most important effects of caffeine are its ability to enhance mood, mental and physical performance. At consumption levels up to 200 milligrams, consumers report an improved sense of well-being, happiness, energy, alertness and sociability. Roland Griffiths of the Johns Hopkins School of Medicine reported that higher amounts sometimes cause anxiety and stomach upset. 30
- 9 Millions of sleep-deprived Americans depend on caffeine to help them make it through their day and drive safely. The drug improves alertness and reaction time. In the sleep-deprived, it improves memory and the ability to perform complex tasks. 35
- 10 For the active, caffeine enhances endurance in aerobic activities and performance in anaerobic ones, perhaps because it blunts the perception of pain and aids the ability to burn fat for fuel. 40

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*

CONFIDENTIAL*

6

- 11 Another review found that compared with non-coffee drinkers, people who drank four to six cups of coffee a day, with or without caffeine, had a 28 percent lower risk of Type 2 diabetes. This benefit probably comes from coffee's antioxidants and chlorogenic acid.

(Adapted from *New York Times*, August 5, 2008)

- 8 The research findings presented in this article are taken from a study carried out by the Centre for Science in the Public Interest.

A True
B False
C Not stated

- 9 A person who drinks 700 milligrams of coffee will pass more urine than someone who drinks the same amount of a caffeine-free drink.

A True
B False
C Not stated

- 10 Scientists say that coffee hydrates our body better than water.

A True
B False
C Not stated

- 11 Findings of the study on coffee and cancer concluded that

A there is no clear link between coffee drinking and cancer
B there is a clear link between pancreatic cancer and caffeine
C more research is needed to draw the connection between caffeine and breast cancer

- 12 The writer uses the phrase, *Here's a bummer* (line 24) to

A expose side effects
B express dissatisfaction
C debunk a misconception

- 13 In paragraph 8, it can be inferred that

A a higher level of caffeine will lead to greater alertness
B one should not consume more than 200 milligrams of caffeine
C consuming caffeine will affect one's mood more than physical performance

- 14 After reading the text, the reader could be persuaded to

A drink more coffee
B stop drinking coffee
C reduce coffee consumption

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

7

Questions 15 to 21 are based on the following passage.

- 1 When Ariel Lugo takes visitors to the rainforests of Puerto Rico, he likes to play a little trick. First, the ecologist shows off the beautiful surroundings: the diversity of plant life on the forest floor; the densely-packed trees merging into a canopy, high overhead. Only when his audience is suitably impressed does he reveal that they are actually in the midst of what many conservationists would dismiss as weeds – a collection of non-native species growing uncontrolled, on land once used for agriculture. 5
- 2 His guests are almost always taken aback, and who wouldn't be? For years we have been told that invasive alien species are driving native ones to extinction and eroding the integrity of ancient ecosystems. The post-invasion world is supposed to be bleak, biologically-impooverished wasteland, not something you could mistake for untouched wilderness. 10
- 3 Lugo is one of a small but growing number of researchers who think much of what we have been told about non-native species is wrong. Alien species, they argue, are rarely as monstrous a threat as they have been painted. In fact, in a world that has been dramatically altered by human activity, many could be important allies in rebuilding healthy ecosystems. Given the chance, alien species may just save us from the worst consequences of our own destructive actions. 15
- 4 Many conservationists cringe at such talk. They view non-native species as ecological tumours, spreading uncontrollably at the expense of natives. To them the high rate of accidental introductions – hundreds of alien species are now well established in ecosystems from the Mediterranean Sea to Hawaii – is one of the biggest threats facing life on Earth. Mass extinction of native species is one fear. Another is the loss of what many regard as the keys to environmental health: the networks of relationships that exist between native species after thousands or even millions of years of co-evolution. 25
- 5 Such concerns have fuelled an all-out war. Vast sums are being spent on campaigns to eradicate or control the spread of highly-invasive exotics. Conservation groups enlist teams of volunteers to uproot garlic mustard from local parks. Government agencies fill waterways with poisonous chemicals to halt the advance of Asian carp. Most governments have no choice but to join the fight; under the terms of the Convention of Biological Diversity. 30
- 6 Advocates for non-native species do not deny that they can sometimes create major problems, particularly in cases where disease-causing microbes are introduced into a new host population. But they argue that often the threat is overblown. For one thing, many species are not nearly as problematic as they are made out to be. 35
- 7 The notorious cane toad, introduced into Australia in the 1930s to control pests of the sugar cane crops, is considered a major threat to the continent's unique fauna. Its highly-toxic skin has long been seen as a death sentence for native predators, while its rapid spread is thought to have occurred at the expense of other amphibians. Yet, the first serious impact study on the cane toad recently concluded that they may in fact be innocent of all charges. 40

(Adapted from *New Scientist*, January 20, 2011)

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over

CONFIDENTIAL*

CONFIDENTIAL*

8

- 15 What is the *little trick* (line 2) played by Ariel Lugo when he takes visitors to the rainforests of Puerto Rico?
- A He shows the visitors uncontrolled weeds instead of the rainforests.
 - B He makes them believe that what they are seeing is native to the land.
 - C He takes them to see the impoverished wasteland and not the wilderness.
- 16 In paragraph 2, the guests are described as *almost always taken aback* (line 8). This is because
- A weeds have overtaken former agricultural land
 - B non-native species have flourished alongside native species
 - C the beautiful surroundings are actually inhabited by non-native species
- 17 The main idea of paragraph 3 is that
- A non-native species could be allies in the fight to save the ecosystems
 - B it is a popular belief that alien species are a threat to the environment
 - C the effects of human activity can be more disastrous than that of alien species
- 18 Conservationists are of the opinion that
- A non-native species will destroy the natives
 - B relationships between native species will be altered with time
 - C alien species are now more established than natives in the ecosystems
- 19 The writer mentions the use of volunteers to uproot garlic mustard from local parks to
- A lend support to the work of the Convention of Biological Diversity
 - B illustrate the attempts at controlling the spread of non-native species
 - C show the extent of money spent on campaigns to eradicate alien invaders
- 20 In paragraph 6, advocates for non-native species claim that
- A the dangers of introducing non-native species have been proven
 - B problems created by alien species are not as serious as made out to be
 - C introducing disease-causing microbes into the host population can be fatal
- 21 The following are true of the cane toad **except**
- A its skin is poisonous
 - B it kills other amphibians
 - C it was introduced into Australia to control pests of the sugar cane

CONFIDENTIAL*

9

Questions 22 to 29 are based on the following passage.

- 1 Today, when we think of the world's teeming billions of humans, we tend to think of overpopulation, poverty, disease, instability and environmental destruction. Humans are the cause of most of the planet's problems. What if that were to change? What if the average humans were able to contribute more than consume? To add more than subtract? Think of the world as if each person drives a balance sheet. On the negative side are the resources they consume without replacing. On the positive side are the contributions they make in the form of resources they produce, the artifacts of value they build, and the ideas and technologies that might create a better future for their families, their communities, and the planet as a whole. Our future hangs on whether the sum of those balance sheets can turn positive. 5 10
- 2 What might make that possible? One key reason for hope is that so far we have barely scraped the surface of human potential. Throughout history, the vast majority of humans have not been the people they could have been. Take this simple thought experiment. Pick your favourite scientist, mathematician, or cultural hero. Now imagine that instead of being born when and where they were, they had instead been born with the same abilities in a poverty-stricken village. Would they have made the same contribution they did make? Probably not. They would not have received the education and encouragement it took to achieve what they did. 15
- 3 If only we could find a way of unlocking that potential. Two keys might be enough: knowledge and inspiration. If you learn how to transform your life for the better and you are inspired to act on that knowledge, there is a good chance that your life will indeed improve. 20
- 4 There are many scary things about today's world, but what is thrilling is that the means of spreading both knowledge and inspiration have never been greater. Five years ago, a teacher or professor who is able to change the lives of his or her students, could realistically hope to reach maybe a hundred of them a year. Today, that same teacher can communicate through video to millions of eager students. The cost of distributing a recorded lecture anywhere in the world via the Internet has effectively fallen to zero. This has happened with breathtaking speed and its implications are not yet widely understood. But it is surely capable of transforming global education. 25 30
- 5 For one thing, the realization that today's best teachers can become global celebrities will boost the calibre of those who teach. For the first time, it is possible to imagine ambitious, brilliant eighteen-year-olds putting 'teacher' at the top of their career choice list. Indeed, the very definition of 'great teacher' will expand, as numerous people outside the profession who can communicate important ideas find new incentive to make that talent available to the world. Additionally, teachers can amplify their own abilities by inviting into their classrooms, on video, the world's greatest scientists, visionaries and tutors. 35

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*

CONFIDENTIAL*

10

- 6 Now, think of this from the pupils' perspective. In the past, your success depended on whether you were lucky enough to have a great mentor or teacher in your neighbourhood. The vast majority have not been that fortunate. But a young girl born in Africa today will probably have access, in ten years' time, to a cell phone with a high-resolution screen, a Web-connection, and more power than the computer you own today. We can imagine her obtaining face-to-face insight from her choice of the world's great teachers. She will get the chance to be what she can be.

(Adapted from *This Will Change Everything*,
John Brockman (Ed.) HarperCollins, 2010)

- 22 The writer introduces his article with a gloomy picture of the world to
- A contrast it with his vision of a better world
 - B explain that the world's problems are man-made
 - C highlight the reality that it would be difficult to change the world
- 23 The question *To add more than subtract?* (lines 4 and 5) _____ the previous question.
- A restates
 - B illustrates
 - C elaborates
- 24 What might make that possible? (line 11)
- What is the answer to the question above?
- A More resources should be produced
 - B Human potential should be exploited more fully
 - C Contributions should balance with consumption
- 25 The writer brings in the *simple thought experiment* (lines 13 and 14) to
- A show that people have different abilities
 - B compare the contributions of past heroes with today's heroes
 - C support the idea that the poor have limited opportunities to excel
- 26 Paragraph 4 lists the following reasons why the Internet can transform education globally **except**
- A it can quickly reach huge student populations
 - B it is a cheap means to distribute recorded lectures
 - C it adopts teaching techniques that students are familiar with

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

11

- 27 The main outcome of making great teachers accessible to the masses through the Web is
- A that they will become celebrities
 - B people will learn to communicate better
 - C the teaching profession would attract more talent
- 28 In paragraph 6, the writer makes the following points **except**
- A children can decide whom they want to learn from
 - B every child will have equal opportunity to education
 - C even children from remote areas can learn from great teachers
- 29 Which of the following best expresses the central idea of the passage?
- A A Web-empowered revolution is taking place in education
 - B Web-based education has replaced the traditional classroom
 - C The Web has the power to inspire and disseminate knowledge

CONFIDENTIAL*

12

Questions 30 to 37 are based on the following passage.

- 1 Stories and psychological description are effective ways of building emotional appeal. Emotional appeal works best when people want to be persuaded.
- 2 Even when you need to provide statistics or numbers to convince the careful reader that your anecdote is a representative example, telling a story first makes your message more persuasive. Experiments with both high school teachers and quantitatively-trained Master of Business Administration (MBA) students show that people are more likely to believe a point and more likely to be committed to it when points were made by examples, stories and statistics; the combination was more effective than statistics alone. In another experiment, attitude changes lasted longer when the audience had read stories than when they had only read numbers. Recent research suggests that stories are more persuasive because people remember them. 5 10
- 3 As with other appeals, the emotional appeal should focus on the reader. To describe its service of gathering up and renting good-quality used cardboard boxes, Boomerang Boxes could focus on its innovative thinking, but its Website appeals to readers by telling them they can "Save time, save money and save trees!" The company tells its story with descriptive language: "No longer do you have to drive around aimlessly searching for good quality boxes behind supermarkets and liquor stores. No longer do you have to contribute to the destruction of strong healthy trees, just so more cardboard boxes can be made, used (often only once) and thrown away." That story is likely to resonate with many apartment-renting students who have scrounged boxes for a low-cost moving day. 15 20
- 4 Sense impressions – what the reader sees, hears, smells, tastes, feels – evoke a strong emotional response. Psychological description means creating a scenario rich with sense impressions so readers can picture themselves using your product or service and enjoying its benefits. You can also use psychological description to describe the problem your product will solve. Psychological description works best early in the message to catch readers' attention. 25
- 5 In psychological description, you're putting your reader in a picture. If the reader doesn't feel that the picture fits him or her, the technique backfires. To prevent this, psychological description often uses subjunctive verbs ("if you like..." "if you were...") or the words *maybe* and *perhaps*. 30
- 6 The best phrasing depends on your relationship to the reader. When you ask for action from people who report directly to you, orders ("Get me the Ervin file.") and questions ("Do we have the third-quarter numbers yet?") will work. When you need action from co-workers, superiors, or people outside the organisation, you need to be more forceful but also more polite. 35
- 7 How you ask for action affects whether you build or destroy positive relationships with other employees, customers and suppliers. Professor and consultant, Dan Dieterich, notes that the calls to action in many messages are: 40
- Buried somewhere deep in the middle of the correspondence.
 - Disguised as either statements or questions.
 - Insulting because they use "parental language".

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

13

- 8 Such messages, Dieterich points out, "lower productivity within the organisation and reduce or eliminate the goodwill customers have toward the organisation... Those two things...can put the organisation out of business." 45
- 9 Avoiding messages that sound parental or preachy is often a matter of tone. Saying "Please" is a nice touch, especially to people on your level or outside the organisation. Tone will also be better when you give reasons for your request or reasons to act promptly. 50
- 10 When you write to people you know well, humour can work. Just make sure that the message is not insulting to anyone who does not find the humour funny.

(Adapted from *Business and Administrative Communication*,
7th ed. New York: McGraw Hill, 2006)

30 Even when you need to provide statistics or numbers to convince the careful reader that your anecdote is a representative example, telling a story first makes your message more persuasive (lines 3-5). This phrase means

- A facts and figures are more convincing than stories
 - B persuasive message has more anecdotal elements than facts and figures
 - C a combination of statistics, numbers and examples will improve communication
 - D telling a story before presenting facts and figures will make the message more effective
- 31 The story told by Boomerang Boxes is likely to appeal to many apartment-hunting students. This is because the story
- A is told in descriptive language
 - B is communicated through a website
 - C reflects their concerns and experiences
 - D presents an innovative way of doing things
- 32 In paragraph 4, the following are recommendations to evoke a strong emotional response **except**
- A making a connection with the senses
 - B highlighting the benefits of the product
 - C introducing the psychological description early
 - D focusing on the seriousness of related problems
- 33 The phrase, *the technique backfires* (line 30) means that the technique brings
- A adverse effects
 - B expected results
 - C surprising feedback
 - D unpredictable consequences

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*

CONFIDENTIAL*

14

- 34 The message, *Get me the Ervin file* (line 34) is considered appropriate if given by
- A clients
 - B a friend
 - C a superior
 - D colleagues
- 35 Which of the messages below is likely to be labelled as parental or preachy?
- A If necessary, please call us at this number.
 - B Can't you leave the gym shoes in the basket?
 - C Everyone is expected to comply with these regulations.
 - D Even on casual days, visitors expect us to dress professionally.
- 36 Which of the following is **not** a condition for humour to work?
- A If it is really funny
 - B If it is not insulting
 - C If it is not in writing
 - D If it is used with people you know well
- 37 The main topic under discussion in the passage is
- A writing to persuade
 - B avoiding emotions in writing
 - C the reasons for requests made
 - D the structure of product advertisement

CONFIDENTIAL*

15

Questions 38 to 45 are based on the following passage.

- 1 Last year the skyrocketing cost of food was a wake-up call for the planet. Between 2005 and 2008, the price of wheat and corn tripled, and the price of rice climbed five-fold, spurring food riots in nearly two dozen countries and pushing 75 million more people into poverty. But unlike previous shocks driven by short-term shortages, this time, the high prices were a symptom of a larger problem. Simply put: For most of the past decade, the world has been consuming more food than it has been producing. After years of drawing stockpiles, in 2007, the world saw global stocks fall to 61 days of global consumption, the second lowest on record. 5
- 2 This was not the first time the world had stood at the brink of a food crisis. At 83, Gurcharan Singh Kalkat has lived long enough to remember one of the worst famines of the 20th century. In 1943, as many as four million people died in the Bengal Famine. For the following two decades, India had to import millions of tons of grain to feed its people. Then came the green revolution. In the 1960s, as India was struggling to feed its people during yet another crippling drought, an American plant breeder named Norman Borlaug was working with Indian researchers to bring his high-yielding wheat varieties to Punjab. Borlaug was born in Iowa and saw his mission as spreading the high-yielding farm methods that had turned the American Midwest into the world's breadbasket to impoverished places throughout the world. His new dwarf wheat varieties with short stems supporting full, fat seeds were a breakthrough. They could produce grain like no other wheat ever seen – as long as there was plenty of water and synthetic fertilizer and little competition from weeds or insects. To that end, the Indian government subsidized canals, fertilizer, and the drilling of tube wells for irrigation. The new wheat varieties quickly spread throughout Asia, changing the traditional farming practices of millions of farmers, and were soon followed by new strains of 'miracle' rice. The new crops matured faster and enabled farmers to grow two crops a year instead of one. 10 15 20 25
- 3 Today, though, the miracle of the green revolution is over in Punjab: Yield growth has flattened since the mid-1990s. Over-irrigation has led to steep drops in the water table while thousands of hectares of productive land have been lost to water-logged soils. Forty years of intensive irrigation, fertilization, and pesticides have not been kind to the fields of Punjab. Nor, in some cases, to the people themselves. In the farming village of Bhuttiwala, home to some 6000 people, village elder, Jagsir Singh adds up the toll: "We've had 49 deaths due to cancer in the last four years," he says. "Most of them were young people. The water is not good. It's poisonous, contaminated water. Yet, people drink it. The green revolution has brought us only downfall. The government has sacrificed the people of Punjab for grain." 30 35
- 4 Others, of course, see it differently. Rattan Lal, a soil scientist believes it was the abuse – not the use – of green revolution technologies that caused most of the problems. That includes the overuse of fertilizers, pesticides, and irrigation and the removal of all crop residues from the fields. "I realize the problems of water quality and water withdrawal," says Lal. "But it saved hundreds of millions of people. We paid a price in water, but the choice was to let people die." In terms of production, the benefits of the green revolution are hard to deny. India has not experienced famine since Borlaug brought his seed to town, while world grain production has more than doubled. 40 45

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*

CONFIDENTIAL*

16

- 5 Many crop scientists believe the solution to our food crisis lies in a second green revolution, based largely on our newfound knowledge of the gene. Plant breeders now know the sequence of nearly all of the 50 000 or so genes in corn and soybean plants and are using that knowledge. Robert Farley, chief technology officer for the agricultural giant Monsanto, is convinced that genetic modification, which allows breeders to bolster crops with beneficial traits from other species, will lead to new varieties with higher yields, reduced fertilizer needs and drought tolerance. He believes biotechnology will make it possible to double yields of corn, cotton, and soybeans by 2030. 50
- 6 But is a reprise of the green revolution the answer to the world's food crisis? Last year, a six-year study concluded that the production increases brought about by science and technology in the past 30 years have failed to improve food access for many of the world's poor. The study called for a paradigm shift in agriculture toward more sustainable and ecologically-friendly practices that would benefit the world's 900 million small farmers, not just agribusiness. And so a shift has already begun to small, under-funded projects scattered across Africa and Asia. Some call it agroecology, others sustainable agriculture, but the underlying idea is revolutionary: that we must stop focusing on maximizing grain yields at any cost and consider the environmental and social impacts on food production. Vandana Shiva, an agroecologist, argues that small-scale, biologically-diverse farms can produce more food with fewer petroleum-based inputs. Her research has shown that using compost instead of natural-gas-derived fertilizer increases organic matter in the soil. "If you are talking about solving the food crisis, these are the methods you need," adds Shiva. 60 65 70
- 7 Regardless of which model prevails – agriculture as a diverse ecological art, as a high-tech industry, or some combination of the two – the challenge of putting enough food in nine billion mouths by 2050 is daunting.

(Adapted from *National Geographic*, June 2009)

- 38 The main point of paragraph 1 is
- A the reasons for a global food crisis
 - B shortage of food leads to high prices
 - C the effects of long-term food shortages
 - D consumption of food exceeds production
- 39 The following are reasons for the skyrocketing cost of food **except**
- A low yield of crops
 - B depleted food stockpiles
 - C steeply-rising population
 - D increasing dependence on imported grains

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

17

- 40 Which of the following did Borlaug introduce to India?
- I A high-yielding wheat strain
 - II Growing of two crops together
 - III The building of tube wells for irrigation
 - IV The American industrial farming method
- A I and III
B I and IV
C II and III
D II and IV
- 41 In paragraph 3, the writer's intention is to
- A explain the failure of the green revolution
 - B compare the pros and cons of the green revolution
 - C account for the effects of the green revolution on water
 - D draw attention to the poor soil condition as a result of the green revolution
- 42 In paragraph 4, the writer began with *Others, of course, see it differently* (line 38). What is the different view?
- A The green revolution is too costly to sustain.
 - B The green revolution has caused much damage.
 - C The abuse of green revolution technologies was the culprit.
 - D The problem of shortage of food was resolved by the green revolution.
- 43 The most distinguishing feature in the *second green revolution* (lines 47 and 48) is
- A doubling the yield of grains
 - B reduction in the use of fertilizers
 - C crops are no longer affected by drought
 - D genetically-modified crop varieties are used
- 44 The study called for a *paradigm shift* (line 59) in agricultural practices. The following are attributes of the shift **except**
- A less funding is needed
 - B more focus on increasing grain yield
 - C use of compost as the main fertilizer
 - D more emphasis on small-scale farming

800/3/N

*This question paper is CONFIDENTIAL until the test is over.

[Turn over
CONFIDENTIAL*]

CONFIDENTIAL*

18

- 45 In the last paragraph, the writer is of the opinion that
- A agriculture is essentially more a science than an art
 - B the agricultural models all have their merits and demerits
 - C applying a combination of agricultural technologies will solve the food crisis
 - D no matter which model is adopted, it will be difficult to feed the rising population

CONFIDENTIAL*

2

Question 1

You are advised to spend about **40 minutes** on this task.

A survey was carried out on the quality of facilities available in three colleges. Using only the information provided, write about the results of the survey. In your writing, you are to link the information in Table 1 and Table 2. You should write **150 to 200 words**. [40 marks]

Table 1: Rating of College Facilities

Facility	Rating of facility		
	College A	College B	College C
Wireless connection	1	2	1
Bookshop	2	2	1
Printing	2	2	1
Sports centre	2	1	2
Bank	1	2	1
Co-curricular centre	2	3	1
Food court	1	3	2
Library	2	2	1
Accommodation	2	2	1
Transport	2	3	1

Note: 1 – Good; 2 – Average; 3 – Poor

Table 2: Fees, Scholarships and Number of Registered Students

	College A	College B	College C
Fees per semester	RM6000	RM4000	RM10 000
Number of scholarships offered per year	500	100	10
Total number of registered students	7000	1000	5500

800/4/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

CONFIDENTIAL*

3

Question 2

You are advised to spend about 50 minutes on this task.

"Women make better leaders than men." Do you agree? Discuss. You should write **at least 350 words**. [60 marks]

800/4/N

*This question paper is CONFIDENTIAL until the test is over.

CONFIDENTIAL*

References

- Akbari, R. (2012). Validity in language testing. In *The Cambridge guide to second language assessment* (pp. 30–36).
- Akpınar, K. D., & Cakildere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and DS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81–94.
- Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. ix–xii). Routledge.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). Language test construction and evaluation. *Cambridge Language Teaching Library*. <https://doi.org/10.5897/ERR12.035>
- Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*. <https://doi.org/10.1177/026553229601300304>
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*. <https://doi.org/10.1093/applin/14.2.115>
- Ali, F. D. (2014). Washback Effects of Malaysian University English Test (MUET) on Teachers' Perceptions on Their Classroom Teaching and Teaching Materials. *Teaching and Learning Language: Current Trends and Practices*, 30(1).
- Allen, D. (2016a). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(1), 7. <https://doi.org/10.1186/s40468-016-0030-z>
- Allen, D. (2016b). Japanese cram schools and entrance exam washback, 3(1), 54–67.
- Allen, P., Frohlich, M., & Spada, N. (1983). *The Communicative Orientation of Language Teaching: An Observation Scheme*.
- Andrews, S. (1994). Washback or washout? The relationship between examination reform and curriculum innovation. In *Bringing about change in language education: Proceedings of the International Language in Education Conference*.
- Andrews, S. (2004). Washback and curriculum innovation. In *Washback in Language Testing: Research Contexts and Methods* (pp. 37–59). <https://doi.org/10.4324/9781410609731>
- Ary, D., Jacobs, L., Sorensen, C., & Walker, D. (2013). *Introduction to research in education*. Cengage Learning.
- Atkins, L., & Wallace, S. (2012). *Qualitative research in education*. Sage Publications.

- Bachman, L. F., & Palmer, a. S. A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests. Oxford Applied Linguistics.* <https://doi.org/10.2307/328718>
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257–279. <https://doi.org/10.1177/026553229601300303>
- Bailey, K. M. (1999). Monograph Series Washback in Language Testing.
- Bandura, A. (1986). The Explanatory and Predictive Scope of Self-Efficacy Theory. *Journal of Social and Clinical Psychology*. <https://doi.org/10.1521/jscp.1986.4.3.359>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.
- Barnes, M. M. (2010). The Washback of the TOEFL iBT on English Language Programs in Vietnam, (October), 246. <https://doi.org/10.14221/ajte.2016v41n7.10>
- Bidin, S., Jusoff, K., Aziz, N. A., Salleh, M. M., & Tajudin, T. (2009). Motivation and Attitude in Learning English among UiTM Students in the Northern Region of Malaysia. *English Language Teaching*, 2(2), 16–20. <https://doi.org/10.5539/elt.v2n2p16>
- Biggs, J. B. (1995). Assumptions underlying new approaches to educational assessment: Implications for Hong Kong. *Curriculum Forum*, 1, 5–22.
- Booth, D. K. (2012). *Exploring the Washback of the TOEIC in South Korea A sociocultural perspective on student test activity*. The University of Auckland.
- Borg, W. R., & Gall, M. D. (1979). *Educational research, an introduction* (3rd ed.). New York: Longman.
- Borgers, N., Hox, J., & Sikkels, D. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral mid-point. *Quality and Quantity*, 38(1), 17–33. <https://doi.org/10.1023/B:QUQU.0000013236.29205.a6>
- Brown, H. D. (2000). *Principles of language learning and teaching* (4th Ed.). New York: Longman.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge University Press.
- Buck, G. (1988). Testing listening comprehension in Japanese university entrance examinations. *Journal of the Japan Association of Language Teachers JALT*, 10(1), 15–42.
- Burrows, C. (1998). *Searching for washback: an investigation of the impact on teachers of the implementation into the Adult Migrant English Program of the assessment of the Certificates*

- in Spoken and Written English*. Macquarie University, Sydney.
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28(4), 523–539. [https://doi.org/10.1016/S0346-251X\(00\)00036-1](https://doi.org/10.1016/S0346-251X(00)00036-1)
- Chen, I. W.-L., & Hsieh, J. J.-C. (2011). English language in Taiwan: An examination of its use in society and education in schools. In *English language education across Greater China* (pp. 70–94).
- Cheng, L. (1997). How does washback influence teaching? Implications for hong kong. *Language and Education*, 11(1), 38–54. <https://doi.org/10.1080/09500789708666717>
- 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. [https://doi.org/10.1016/S0191-491X\(98\)00018-2](https://doi.org/10.1016/S0191-491X(98)00018-2)
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. *Studies in language testing*. Cambridge University Press.
- Cheng, L. (2008). Washback, impact and consequences. *Encyclopedia of Language and Education, 2nd Edition, Volume 7: Language Testing and Assessment*, 7, 349–364. https://doi.org/10.1007/978-0-387-30424-3_186
- Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 3–17). Routledge. <https://doi.org/10.4324/9781410609731>
- Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. *Washback in Language Testing: Research Contexts and Methods*. <https://doi.org/10.4324/9781410609731>
- Cho, D. (2004). Use of Standardized Tests as University Graduation Requirement. *English Teaching*, 59(1), 251–266.
- Chu, H. Y. (2009). *Stakes, needs and washback: An investigation of the English benchmark policy for graduation and EFL education at two technological universities in Taiwan*. National Taiwan Normal University.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*. *Education*. https://doi.org/10.1111/j.1467-8527.2007.00388_4.x

- Cohen, L., Manion, L., & Morrison, K. (2011). *Research Method in Education. Educational Research*. <https://doi.org/10.1017/CBO9781107415324.004>
- Council, M. E. (2001). *Malaysian University English Test*. Kuala Lumpur.
- Creswell, J., & Clark, V. (2011). *Designing and conducting mixed-methods research. The Sage handbook of qualitative research*.
- Creswell, J. W. (2008). *Educational Research*. Los Angeles, London: Sage.
- Creswell, J. W. (2014). *Research Design: Qualitative, quantitative, and mixed methods approaches - 4th ed. SAGE Publication, Inc.*
<https://doi.org/10.1017/CBO9781107415324.004>
- Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*. <https://doi.org/10.3102/00346543058004438>
- Cummins, R. A., & Gullone, E. (2000). Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. In *Second International Conference on Quality of Life in Cities*.
- Czaja, R. (1998). Questionnaire pretesting comes of age. *Marketing Bulletin-Department of Marketing Massey University*, 9, 52–66.
- Daly, A. L., Chamberlain, S., & Spalding, V. (2011). Test anxiety, heart rate and performance in A-level French speaking mock exams: An exploratory study. *Educational Research*, 53(3), 321–330. <https://doi.org/10.1080/00131881.2011.598660>
- Damankesh, M., & Babaii, E. (2015). The washback effect of Iranian high school final examinations on students' test-taking and test-preparation strategies. *Studies in Educational Evaluation*, 45, 62–69. <https://doi.org/10.1016/j.stueduc.2015.03.009>
- Deepak Singh, R. S., Noriah Ismail, & Intan Safinas, M. A. A. (2012). A Study on the Relationship between Learners' Informal Exposure to English and their Achievement in the Malaysian University English Test (MUET). *English Language Journal*, 5, 114–122. Retrieved from <http://ajelp.upsi.edu.my/index.php/archive/85-vol-5-2012>
- Denscombe, M. (2014). *The Good Research Guide: for small-scale social research projects*. McGraw-Hill Education.
- Denzin, N. K., & Lincoln, Y. S. (2003). Part II: Paradigms and perspectives in transition. In N. K. Denzin & Y. S. Lincoln (Eds.), *The landscape of qualitative research* (2nd ed., pp. 245–252). London: Sage.

- De Vaus, D. (2001). *Research design in social research*. Sage.
- Dörnyei, Z. (2009). The L2 Motivational Self System. *Motivation, Language Identity and the L2 Self*, (1959), 9–42. <https://doi.org/9786611973452>
- Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*. <https://doi.org/10.1080/00401706.1964.10490181>
- Ehrman, M., & Oxford, R. (1990). Adult language learning styles and strategies in an intensive training setting. *The Modern Language Journal*, 74(3), 311–327.
- Embse, N. von der, & Hasson, R. (2012). Test Anxiety and High-Stakes Test Performance Between School Settings: Implications for Educators. *Preventing School Failure: Alternative Education for Children and Youth*, 56(3), 180–187. <https://doi.org/10.1080/1045988X.2011.633285>
- Erickson, F. (1991). Advantages and disadvantages of qualitative research design on foreign language research. *Foreign Language Acquisition Research and the Classroom*, 338–353.
- Fan, J., Ji, P., & Song, X. (2014). Washback of university-based English language tests on students' learning: A case study. *The Asian Journal of Applied Linguistics*, 1(2), 178–191.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 191–210). Routledge. <https://doi.org/10.4324/9781410609731>
- Frederiksen, J. R., & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher*, 19(9), 27–32. <https://doi.org/10.3102/0013189X018009027>
- Frohlich, M., Spada, N., & Allen, P. (1985). Differences in the communicative orientation of L2 classrooms. *TESOL Quarterly*, 19(1), 27–57.
- Ganapathy, M., & Ying, G. C. (2016). ATTITUDE AND MOTIVATION OF STUDENTS TOWARDS LEARNING ENGLISH AS SECOND LANGUAGE IN A SECONDARY SCHOOL IN PENANG Malini Ganapathy and Gooi Chiew Ying, 5(2), 39–56. <https://doi.org/10.24200/mjll.vol5iss2pp39-56>
- Gates, S. (1995). Exploiting washback from standardized tests. *Language Testing in Japan*, 101–106.
- Gebril, A., & Brown, G. T. L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles,*

- Policy and Practice*, 21(1), 16–33. <https://doi.org/10.1080/0969594X.2013.831030>
- Gebril, A., & Eid, M. (2017). Test Preparation Beliefs and Practices in a High-Stakes Context: A Teacher's Perspective. *Language Assessment Quarterly*, 14(4), 360–379. <https://doi.org/10.1080/15434303.2017.1353607>
- Gibbs, A. (1997). Focus groups. *Social Research Update*, 19(8), 1–8.
- Gipps, C. (2002). *Beyond testing: Towards a theory of educational assessment*. Routledge.
- Glatthorn, A. A., & Joyner, R. L. (2005). *Writing the winning thesis or dissertation: A step-by-step guide*. Corwin Press.
- Goleman, D. (1996). Emotional intelligence: Why it can matter more than IQ. *Nature*. <https://doi.org/10.1016/j.paid.2003.12.003>
- Goşa, C. M. . (2004). Investigating washback: a case study using student diaries, (July).
- Gray, D. E. (2004). *Doing research in the real world*. Sage. London. Thousands Oaks. New Delhi. <https://doi.org/10.1016/j.neulet.2004.06.080>
- Green, A. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*. <https://doi.org/10.1016/j.asw.2006.07.002>
- Green, A. (2007a). *IELTS Washback in Context: Preparation for Academic Writing in Higher Education*. *Studies in Language Testing*.
- Green, A. (2007b). 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. <https://doi.org/10.1080/09695940701272880>
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. *Exploring Language Assessment and Testing: Language in Action*. <https://doi.org/10.4324/9781315889627>
- Guba, E. G. (1990). *The paradigm dialog*. Sage Publications.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing Paradigms in Qualitative Research. In *Handbook of qualitative research* (pp. 163–194). <https://doi.org/http://www.uncg.edu/hdf/facultystaff/Tudge/Guba%20&%20Lincoln%201994.pdf>
- Hamid, B. A. (2001). Malu. Shyness and shyness behaviour in the English language classroom: Who, what, when, how, and why. *The English Teacher*, 21.

- Hamp-Lyons, L. (1987). *Testing second language writing in academic settings*. University of Edinburgh.
- Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing*. <https://doi.org/10.1177/026553229701400306>
- Hanapiah, M. F. (2004). English Language and the language of development: A Malaysia Perspective. *Jurnal Kemanusiaan*, 2(1), 106–120.
- Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In *Washback in Language Testing: Research Contexts and Methods* (pp. 97–111). <https://doi.org/10.4324/9781410609731>
- Hoque, E. M. (2011). *Washback of the public examination on teaching and learning English as a foreign language (EFL) at the higher secondary level in Bangladesh*. Jahangirnagar University. Retrieved from http://encore.fama.us.es/iii/encore/plus/C__SCreswell_y_Miller%2C_2000%3A125__Orightresult__U__X0;jsessionid=C96A76095727367593D006B91190D172?lang=spl&link=http%3A%2F%2F0-search.ebscohost.com.fama.us.es%2Flogin.aspx%3Fdirect%3Dtrue%26site%3Ded-live%26sco
- Hsu, H.-F. (2010). The impact of implementing English proficiency tests as a graduation requirement at Taiwanese universities of technology. *Educational Studies*, (November). Retrieved from <http://etheses.whiterose.ac.uk/576/>
- Hughes, A. (1993). *Backwash and TOEFL 2000*. University of Reading.
- Hughes, A. (2002). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Hung, S.-T. A. (2012). A washback study on e-portfolio assessment in an English as a Foreign Language teacher preparation program. *Computer Assisted Language Learning*, 25(1), 21–36. <https://doi.org/10.1080/09588221.2010.551756>
- Idris, R., & Alias, S. A. (2019, February 23). Hapuskan MUET. *Berita Harian*. Retrieved from <https://www.pressreader.com>
- Jäger, D. J., Maag Merki, K., Oerke, B., & Holmeier, M. (2012). Statewide low-stakes tests and a teaching to the test effect? An analysis of teacher survey data from two German states. *Assessment in Education: Principles, Policy and Practice*, 19(4), 451–467. <https://doi.org/10.1080/0969594X.2012.677803>

- Jalaluddin, N. H., Awal, N. M., & Bakar, K. A. (2008). The Mastery of English Language among Lower Secondary School Students in Malaysia: A Linguistic Analysis. *European Journal of Social Sciences*, 7(2), 106–119.
- Jianrattanapong, A. (2011). Positive Washback from Thai University Entrance Examinations. *Language Testing in Asia*, 1(1), 50. <https://doi.org/10.1186/2229-0443-1-1-50>
- Kassim, H., & Ali, F. (2010). English communicative events and skills needed at the workplace: Feedback from the industry. *English for Specific Purposes*, 29(3), 168–182. <https://doi.org/10.1016/j.esp.2009.10.002>
- Kaur, N., & Nordin, R. (2006). A case for reconstruction of the pedagogy of the Malaysian University English test (MUET) through thematic units instruction. *Journal of Institutional Research South East Asia*, 4(1), 5–16.
- Keeves, J., & Sowden, S. (1992). Analyzing qualitative data. *Methodology and Measurement in International Educational Surveys*.
- Kent, R. (2015). *Analysing quantitative data: Variable-based and case-based approaches to non-experimental datasets*. Sage Publications Ltd.
- Khodabakhshzadeh, H., Zardkanloo, R., & Alipoor, I. (2017). The effect of mock tests on Iranian EFL learners ' test scores. *International Journal of Education & Literacy Studies*, 5(3), 47–51.
- Kober, N. (2002). Teaching to the test: The good, the bad, and who's responsible. *Test Talk for Leaders*, 1.
- Lahmers, A. G., & Zulauf, C. (2000). The secret to academic success: Hours and hours of study. *Journal of College Student Development*, 41(5), 545–554.
- Langdridge, D. (2004). *Introduction to Research Methods and Data Analysis in Psychology*. Pearson Education.
- Lee, K. S. (2004). Exploring the connection between the testing of reading and literacy: The case of the MUET. *GEMA Online Journal of Language Studies*, 4(1), 41–50.
- Li, H., Zhong, Q., & Suen, H. K. (2012). Students ' perceptions of the impact of the college English test. *Language Testing in Asia*, 2(3), 77–94. <https://doi.org/10.1186/2229-0443-2-3-77>
- Lin, S. (2010). *The requirement of English language proficiency for graduation in Taiwanese universities: Its impact on non-English majors and their English curriculum*. University of Bristol.

- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Liu, Y. (2014). Motivation and Attitude: Two Important Non-Intelligence Factors to Arouse Students' Potentialities in Learning English. *Creative Education*, 5(14), 1249-1253.
- Lublin, J. (2003). Deep, surface and strategic approaches to learning. *Centre for Teaching and Learning*, 806–825.
- Luong-Phan, N.-H., & Effeney, G. (2015). TOEFL iBT and language learning motivation: An investigation into teaching styles and influential factors for Vietnamese adolescents. *International Journal of Research Studies in Language Learning*, 4(3), 3–18. Retrieved from http://search.proquest.com/docview/1718058049?accountid=14548%5Cnhttp://metadata.lib.hku.hk/hku?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ:llba&atitle=TOEFL+iBT+and+language+learning+motivation:+An+investigat
- Madaus, G. F. (1988). The Distortion of Teaching and Testing: High-Stakes Testing and Instruction. *Peabody Journal of Education*, 65(3), 29–46. <https://doi.org/10.1007/s10869-007-9037-x>
- Marton, F., & Saljo, R. (1976). On Qualitative Differences in Learning: I - Outcome and Process. *British Journal of Educational Psychology*. <https://doi.org/10.1111/j.2044-8279.1976.tb02980.x>
- May, K. A. (1991). Interview techniques in qualitative research: Concern and challenges. In J. M. Morse (Ed.), *Qualitative nursing research: A contemporary dialogue* (pp. 188–201). Newbury Park, CA: Sage.
- McEwen, N. (1995). Educational accountability in Alberta. *Canadian Journal of Education*, 20(1), 27–44.
- McNamara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- McNamara, T. (2006). Validity in Language Testing: The Challenge of Sam Messick's Legacy. *Language Assessment Quarterly*. https://doi.org/10.1207/s15434311laq0301_3
- Mehrens, W. A. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives*, 6(13), 1–30.
- Menon, S., & Patel, N. (2012). Importance of English language in corporate world. *Journal for Research in Management and Pharmacy*, 1(1), 42–55.
- Merriam, S. B. (1998). *Qualitative research and case study application in education: Revised and*

- expanded from case study research in education*. San Francisco: Jossey-Bass Publishers.
- Mertens, D. M. (2010). *Research and evaluation in education and psychology: integrating diversity with quantitative, qualitative, and mixed methods* (3rd ed.). Thousand Oaks: Sage Publications.
- Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Sage Publications.
- Messick, S. (1989). Meaning and Values in Test Validation: The Science and Ethics of Assessment. *Educational Researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256. <https://doi.org/10.1177/026553229601300302>
- Mohd Nopiah, Z., Ismail, N. A., Khatimin, N., Abdullah, S., & Mustafa, M. M. (2011). MUET score and loadings hour: An analysis on the relationship towards academic performance. *Procedia - Social and Behavioral Sciences*, 18, 103–109. <https://doi.org/10.1016/j.sbspro.2011.05.015>
- Morgan, D. L. (1988). *Focus groups as qualitative research*. (Sage, Ed.). Newbury Park, CA.
- Morrow, K. (1986). The evaluation of tests of communicative performance. *Innovations in Language Testing*, 1–13.
- Muñoz, A. P., & Álvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*. <https://doi.org/10.1177/0265532209347148>
- Nambiar, M., & Ransirini, S. (2012). “Teaching MUET, not English”: A study of the washback effect of the Malaysian University English Test(MUET). In M. D. Z (Ed.), *English in Multicultural Malaysia. Pedagogy and Applied Research* (pp. 49–62). Malaysia: University of Malaya Press.
- Nazari, M., & Nikoopour, J. (2011). The Washback Effect of High School Examinations on Iranian EFL Learners’ Language Learning Beliefs. *Journal of Language and Translation*, 2(1), 29–49.
- Newby, P. (2014). *Research methods for education* (2nd ed.). Routledge.
- Nur Sakinah, T., & Nooreen, N. (2013). Relationship between Malaysian University English Test (MUET) achievements and trait emotional intelligence (Trait EI) among matriculation students. *Graduate Research in Education (GREDUC 2013)*, (Greduc).

- Omar, A. H. (1992). *The Linguistic Scenery in Malaysia*. Malaysia: Dewan Bahasa dan Pustaka, Ministry of Education.
- Onaiba, A. M. E. M. (2013). *Investigating the washback effect of a revised EFL public examination on teachers' instructional practices, materials and curriculum*. University of Leicester.
- Othman, J., & Nordin, A. B. (2013). MUET as a predictor of academic achievement in ESL teacher education. *GEMA Online Journal of Language Studies*, 13(1), 99–111.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle Publishers.
- Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System*, 23(1), 1–23. <https://doi.org/10.1159/000104095>
- Pajares, F. (2002). Gender and Perceived Self-Efficacy in Self-Regulated Learning. *Theory Into Practice*, 41(2), 116–125.
- Pan, Y.-C. (2014). Learner Washback Variability in Standardized Exit Tests. *TESL-EJ: Teaching English as a Second or Foreign Language*, 18(2), 1–30. Retrieved from http://anu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwnV1LSwMxEB6UXgTx_baQg3hbmya7jXsSkRYPvb14LZvHwiKs2t2K-uudSba19qDgLRAlA0lmJl_mmw9AiiserfgEmcYYh3jiuDVa9YWwzhhbOYPTHEJmnK9o wgdRPSEHY7bmT9J7bPhsCzXt9JZNBzPGC3Ly8RiQjRd-trabGOnQw1U5Jy0Fl6eIFJmVgSHJFVCE
- Pan, Y.-C., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching*, 9(1), 108–122.
- Pan, Y., & Newfields, T. (2011). Teacher and student washback on test preparation evidenced from Taiwan's English certification exit requirements. *International Journal of Pedagogies and Learning*, 6(3), 260–272. <https://doi.org/10.5172/ijpl.2011.6.3.260>
- Parahoo, K. (1997). *Nursing research: Principles, process and issues*. Palgrave Macmillan.
- Pearson, I. (1988). Tests as levers for change. *ESP in the Classroom: Practice and Evaluation*, 128, 98–107.
- Petrides, K. V., Furnham, A., & Mavroveli, S. (2007). Trait emotional intelligence: Moving forward in the field of EI. In G. Matthews, M. Zeidner, & R. Roberts (Eds.), *Emotional*

- intelligence: Knowns and unknowns*. Oxford: Oxford University Press.
- Philip, B., & Koo, Y. L. (2006). Social-Motivational Metacognitive Strategy Instruction (SMMSI): Breaking the Silence in Language Reading Classes. *Brometey Philip and Koo Yew Lie*. 3L: *Language, Linguistics, Literature, 11*, 171–190.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). Motivated Strategies for Learning Questionnaire (MSLQ). *Mediterranean Journal of Social Sciences*. <https://doi.org/10.5901/mjss.2015.v6n1p156>
- Popham, W. J. (1987). The Merits of Measurement-Driven Instruction. *The Phi Delta Kappan*, 68(9), 679–682.
- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modelling approach*. Cambridge University Press.
- Putwain, D. (2008). Do examinations stakes moderate the test anxiety-examination performance relationship? *Educational Psychology*, 28(2), 109–118. <https://doi.org/10.1080/01443410701452264>
- Qi, L. (2005). *Stakeholders' conflicting aims undermine the washback function of a high-stakes test*. *Language Testing* (Vol. 22). <https://doi.org/10.1191/0265532205lt300oa>
- 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. <https://doi.org/10.1080/09695940701272856>
- Qian, D. D. (2014). School-based English language assessment as a high-stakes examination component in Hong Kong: Insights of frontline assessors. *Assessment in Education: Principles, Policy and Practice*, 21(3), 251–270. <https://doi.org/10.1080/0969594X.2014.915207>
- Rashid, A. F. D. (2014). Washback Effects of Malaysian University English Test (MUET) on Teachers' Perceptions on Their Classroom Teaching and Teaching Materials. *Teaching and Learning Language: Current Trends and Practices*, 30.
- Read, J., & Hayes, B. (2003). The impact of the IELTS test on preparation for academic study in New Zealand. In *IELTS Research Reports* (Vol. 5, pp. 153–206). Canberra: IELTS Australia.
- Ren, Y. (2011). A study of the washback effects of the College English Test (band 4) on teaching and learning English at tertiary level in China. *International Journal of Pedagogies and Learning*, 6(3), 243–259. <https://doi.org/10.5172/ijpl.2011.6.3.243>

- Resnick, L. B., & Resnick, D. P. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. In *Changing Assessments* (pp. 37–75). Springer. https://doi.org/10.1007/978-94-011-2968-8_3
- Rethinasamy, S., & Chuah, K. M. (2011). The Malaysian University English Test (MUET) and Its Use for Placement Purposes : A Predictive Validity Study The Malaysian University English Test (MUET) and its Use for Placement Purposes : A Predictive Validity Study, 8(September 2017), 234–245. <https://doi.org/10.2139/ssrn.2146007>
- Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the Number of Categories in Agree-Disagree Scales. *Sociological Methods and Research*, 43(1), 73–97. <https://doi.org/10.1177/0049124113509605>
- Ribiero, N. F., & Yarnal, C. M. (2010). The Perceived Difficulty Assessment Questionnaire (PDAQ): Methodology and Applications for Leisure Educators and Practitioners. *Schole: A Journal of Leisure Studies and Recreation Education*, 25(1), 111–115. <https://doi.org/10.1080/1937156X.2010.11949658>
- Richard Ryan, & Brown, K. (2005). Legislating Competence: High-Stakes Testing Policies and Their Relations wiht Psychological Theories and Research. In *Handbook of Competence and Motivation*.
- Roderick, M., & Engel, M. (2001). The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-Stakes Testing. *Educational Evaluation and Policy Analysis*. <https://doi.org/10.3102/01623737023003197>
- Runswick-Cole, K. (2011). Interviewing. In P. Banister (Ed.), *Qualitative Methods in Psychology: A Research Guide* (pp. 88–99). McGraw-Hill Open University Press.
- Salovey, P., & Mayer, J. D. (1990). Emotional Intelligence. *Imagination, Cognition and Personality*. <https://doi.org/10.2190/DUGG-P24E-52WK-6CDG>
- Sato, T., & Ikeda, N. (2015). Test-taker perception of what test items measure: a potential impact of face validity on student learning. *Language Testing in Asia*, 5(1), 10. <https://doi.org/10.1186/s40468-015-0019-z>
- Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
- Scaramucci, M. V. R., & Kobayashi, E. (2013). Washback effect of Cambridge english: Key for schools in são Paulo, brazil: A case study. *Research Notes*, (54), 13–20.

- Schuman, H., & Presser, S. (1996). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. Sage Publications.
- Schunk, D. H., & Pajares, F. (2009). Self-Efficacy Theory. *Handbook of Motivation at School*.
<https://doi.org/doi:10.4324/9780203879498>
- Shih, C.-M. (2007). A New Washback Model of Students' Learning. *Canadian Modern Language Review*. <https://doi.org/10.3138/cmlr.64.1.135>
- Shih, C. (2007). A New Washback Model of Students ' Learning. *The Canadian Modern Language Review*, 64(1), 135–161.
- Shih, P.-C. (2013). *The English Benchmark Policy for Graduation: An Investigation of Perception, Motivation, and Approaches to Learning at a University of Technology in Central Taiwan*. Durham University.
- Shohamy, E. (1992). Beyond Proficiency Testing : A Diagnostic Feedback Testing Model for Assessing Foreign Language Learning. *The Modern Language Journal*, 76(4), 513–521.
<https://doi.org/10.1111/j.1540-4781.1992.tb05402.x>
- Shohamy, E. (2014). *The power of tests: A critical perspective of the uses of language tests*. Routledge.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*. <https://doi.org/10.1177/026553229601300305>
- Sirat, M., Bakar, R., Lim, H. E., & Katib, M. N. (2004). Pencapaian akademik & kebolegunaan tenaga siswazah institusi pengajian tinggi. *USM IPPTN Monograf*, 3.
- Sirat, M., Buang, A. A., Isa, A. M. M., Pandian, A., Abdullah, M. A., Ibrahim, M. D., ... Lee, M. N. N. (2004). Masalah pengangguran di kalangan siswazah. *USM IPPTN Monograf*, 2.
- Smith, J. K. (1983). Quantitative Versus Qualitative Research: An Attempt to Clarify the Issue. *Educational Researcher*. <https://doi.org/10.3102/0013189X012003006>
- Spada, N., & Frohlich, M. (1995). COLT observation scheme: Communicative orientation of language teaching observation scheme coding conventions & applications. *NCELTR Australia*.
- Spector, P. E. (1993). Research Designs. In M. L. Lewis-Beck (Ed.), *Experimental Design and Methods. International Handbook of Quantitative Applications in the Social Sciences* (pp. 1–74). London: Sage.
- Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research.

- Gastroenterology*, 126, 124–128.
- Sperber, A. D., Devellis, R. F., & Boehlecke, B. (1994). Cross-cultural translation methodology and validation. *Journal of Cross-Cultural Psychology*, 25(4), 501–524.
- Spolsky, B. (1994). The examination-classroom backwash cycle: Some historical cases. *Bringing about Change in Language Education*, 55–66.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*.
<https://doi.org/10.1191/1362168805lr152oa>
- Statistics, M. E. (2017). *Quick Facts: Malaysia Educational Statistics*. Putrajaya.
- Stecher, B. M. (2002). Consequences of large-scale, high stakes testing on school and classroom practice. *Making Sense of Test-Based Accountability in Education*, 79–100.
- Stoneman, B. W. H. (2006). *The impact of an exit English test on Hong Kong undergraduates: A study investigating the effects of test status on students' test preparation behaviours*. ProQuest Dissertations and Theses. The Hong Kong Polytechnic University. Retrieved from https://search.proquest.com/docview/304909295?accountid=10673%0Ahttp://openurl.ac.uk/redirect/athens:edu/?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+%26+theses&sid=ProQ:ProQuest+Dissertations+%26+Theses+Global&at
- Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys*. Chicago: Aldine.
- Sullivan, G. M., & Artino, A. R. (2013). Analyzing and Interpreting Data From Likert-Type Scales. *Journal of Graduate Medical Education*. <https://doi.org/10.4300/JGME-5-4-18>
- Swain, M. (1985). Large-scale communicative language testing: A case study. *New Directions in Language Testing*, 35–46.
- Takagi, A. (2010). *A critical analysis of English language entrance examinations at Japanese universities*. University of Exeter.
- Talib, O. (2013). *Asas penulisan tesis, penyelidikan & statistik*. Universiti Putra Malaysia Press.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Thang, S. M. (2001). *Malaysian learners' conceptions of their learning processes and their perceptions of their English as a second language (ESL) courses in a tertiary distance learning context*. University of Nottingham.
- Thang, S. M. (2004). Learning English in multicultural malaysia: Are learners motivated? *Journal*

- of Language and Learning*, 2(2), 142–153.
- Thomas, R. M. (2005). *High-stakes testing: Coping with collateral damage*. Routledge.
- Tsagari, D. (2007). Review of Washback in Language Testing: What Has Been Done? What More Needs Doing? Retrieved 19 March, 2016 From <http://Files.Eric.Ed.Gov/Fulltext/ED497709.Pdf>.
- Tsagari, D. (2009). Revisiting the concept of test washback: Investigating FCE in Greek language schools. *Cambridge ESOL, Research Notes*, 35, 5–9.
- Tsagari, K. (2006). *Investigating the washback effect of a high-stakes EFL exam in the Greek context: Participants' perceptions, material design and classroom applications*. Lancaster University.
- Tsai, Y., & Tsou, C. (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. <https://doi.org/10.1080/09695940903319711>
- Wall, D. (1997). Impact and Washback in Language Testing. *Encyclopedia of Language and Education*, 7, 291–302.
- Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*. [https://doi.org/10.1016/S0346-251X\(00\)00035-X](https://doi.org/10.1016/S0346-251X(00)00035-X)
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge University Press.
- Wall, D. (2012). Washback in Language Assessment. In *The Encyclopedia of Applied Linguistics*. <https://doi.org/10.1002/9781405198431.wbeal1274>
- Wall, D., & Alderson, J. C. (1993). Examining washback: the Sri Lankan Impact Study. *Language Testing*, 10(1), 41–69. <https://doi.org/10.1177/026553229301000103>
- Watanabe, Y. (1997). *The washback effects of the Japanese university entrance examinations of English: classroom-based research*. University of Lancaster.
- Watanabe, Y. (2001). Does the university entrance examination motivate learners? A case study of learner interviews. *Trans-equator Exchanges: A Collection of Academic Papers in Honour of Professor David Ingram*, 100–110.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 19–36).

Routledge.

- Watanabe, Y. (2006). The importance of motivation theories for understanding washback to the learner. *Bulletin of the Center for Educational Research and Practice, Faculty of Education and Human Studies, Akita University*, 28, 1–6.
- Wong, B. E., & Chan, S. H. (2009). Test innovation and implementation: A critical examination of the MUET impact. In F. S. Abdullah, M. H. Abdullah, & B. H. Tan (Eds.), *Critical Perspectives on Language and Discourse in the New World Order* (pp. 252–274). Cambridge Scholars Publishing.
- Xie, Q., & Andrews, S. (2012). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*.
<https://doi.org/10.1177/0265532212442634>
- Yang, T., & Sc, B. (2013). Rethinking washback : The interplay of beliefs and contextual factors to mediate pedagogy, (June).
- Yang, Y., & Badger, R. (2015). How IELTS preparation courses support students: IELTS and academic socialisation. *Journal of Further and Higher Education*, 39(4), 438–465.
<https://doi.org/10.1080/0309877X.2014.953463>
- Yu, C. (2018). Reliability of self-report data. Retrieved June 18, 2018, from <http://www.creative-wisdom.com/teaching/WBI/memory.shtml>
- Yunus, M. M., & Salehi, H. (2012). The Washback Effect of the Entrance Exam of the Universities (EEU) on the Iranian Pre-university Students' English Learning. *International Journal of Learning*, 18(7).
- Zhan, Y. (2009). Washback and possible selves: Chinese non-English-major undergraduates' English learning experiences. https://doi.org/10.5353/th_b4394377
- Zhan, Y., & Andrews, S. (2014). Washback effects from a high-stakes examination on out-of-class English learning: Insights from possible self theories. *Assessment in Education: Principles, Policy and Practice*, 21(1), 71–89. <https://doi.org/10.1080/0969594X.2012.757546>