Student Motivation on a Diagnostic and Tracking English Language Test in Hong Kong

TSANG Hoi Ka, Carrie
Institute of Education
University of London
Doctor in Education

Abstract

Performance in an assessment is not the reflection of just one's knowledge and skills; motivation also plays a part. When the stakes of the assessment are low, it is logical to assume that students will have lower motivation to perform well in it. The Diagnostic English Language Tracking Assessment (DELTA) diagnoses and tracks students' English language progress during their years of study at three universities in Hong Kong. Although the DELTA is a low stakes assessment, students get a report with their DELTA measure and detailed feedback on their performance. This study provides insights into test motivation as well as how useful students find a diagnostic report is to their language learning by ways of questionnaire survey and group interview, so as to explore students' perceptions of test stakes and test value. The survey includes the Student Opinion Scale by Sundre and Moore (2002), which measures students' motivation during the test; and a feedback usefulness scale specifically designed for this study to measure students' perceptions of the usefulness of the diagnostic report. The results show that both scales are valid instruments to be used in this context and students are not motivated whilst sitting the test although they find the DELTA report quite useful. Data from the students' interviews provide further information as to students' motivation before and after the DELTA. In general they are not motivated before the test and their motivation to work on their English after the test largely depends on their perceived usefulness of the DELTA report. Lastly, as L2 motivation is a dynamic entity which will not remain constant over time, the study also demonstrates how Dörnyei and Ottó's (1998) process model of L2 motivation can be adapted in explaining students' test preparation and test taking process in low stakes diagnostic tests.

I hereby declare that, except where explicit attribution is made, the work presented in this thesis is entirely my own.

Word count (exclusive of the statement, table of contents and figures and tables, references and appendices): 44,732 words.

TSANG Hoi Ka, Carrie
May 2013

Acknowledgements

I would like to take this opportunity to thank some very important people here who have in different ways made it possible for me to come to this stage.

First among all, thank you to Dr Amos Paran for his amazingly helpful, sensible and wise supervision; and also to his generous and kind encouragement all along.

Thank you to all my colleagues in the Language Testing Unit of the Hong Kong Polytechnic University and the DELTA project team; and in particular, Dr Alan Urmston and Dr Michelle Raquel, who provided me valuable suggestions and support whenever I needed.

Thank you also to Prof Paul Newton who, as the second reader of my thesis, gave me a lot of useful comments; Prof Trevor Bond who inspired me in my data analysis; and Prof Charles Alderson, who in one occasion reassured that I was on the right track.

Last but not least, thank you to Alan Chiu, my better half, for his unconditional support and love, which help me survive the ups and downs throughout all these years of study.

Contents

List of tables 8

List of figures 9

Statement 11

Chapter 1 Introduction 16

1.1 English as the prestigious language 16

1.2 The examination-led education system 17

1.3 Hong Kong Chinese Learners of English 18

1.4 Autonomous learning in tertiary education 19

1.5 The rise of the Diagnostic English Language Tracking Assessment 20

1.6 Research rationale—Motivation in low stakes diagnostic assessment 22

1.7 Chapter summary 24

Chapter 2 Literature review 26

2.1 Some dominant psychological motivation theories 26

2.1.1 Achievement Goal Theory 26

2.1.2 Self-Determination Theory 27

2.1.3 Expectancy-Value Theory 28

2.2 Second Language Motivation 29

2.2.1 The Social Education Model 30

2.2.2 Self-Determination Theory 33

2.2.3 Expectancy-Value Theories 34

2.2.4 The dynamics of second language motivation 36

2.3 Test motivation 42

2.3.1 Measuring test-taking motivation 44

2.3.2 Test motivation and test validity in low-stakes tests 46

2.4 Diagnostic assessment 52

2.4.1What is diagnostic assessment 53

2.4.2 DELTA as a formative test 55

2.4.3 RASCH measurement and the DELTA 59

2.4.4 Usefulness of feedback 63

2.5 Chapter summary 66

Chapter 3 Methodology 68
3.1 Research questions 68
3.2 Methods of data collection and analysis 69
3.3 Questionnaire survey 71
3.3.1 The construction of the questionnaire 71
3.3.2 Piloting of questionnaire 73
3.3.3 Participants of the questionnaire survey 77
3.3.4 Methods of analysis 77
3.4 Interview 79
3.4.1 The structure of the interview 79
3.4.2 Interviewees 81
3.4.3 Methods of analysis 82
3.5 Ethical issues 84
3.6 Chapter summary 86
Chapter 4 Findings 87
4.1 Questionnaire surveys 87
4.1.1 The Student Opinion Scale (SOS) as an instrument of measuring motivation 88
4.1.1.1 Internal structure evidence of the SOS 88
4.1.1.2 Validating the SOS using the Rasch model 92
4.1.1.3 Findings of SOS 106
4.1.2 The Report Usefulness Scale as an instrument of measuring perceived usefulness of diagnostic report 111
4.1.2.1 Internal structure evidence of the Report Usefulness Scale 111
4.1.2.2 Validating the Report Usefulness Scale using the Rasch model 113
4.1.2.3 Findings of the Report Usefulness Scale 118
4.1.2.4 Correlations between the SOS and the Report Usefulness Scale 122
4.2 Findings of interviews 123
4.2.1 Reasons for misfitting items 125
4.2.2 Constructing and validating the test motivation process model 127
4.2.2.1 Understanding the preactional phase of the process model 131
4.2.2.2 Understanding the postactional phase of the process model 137

4.3 Chapter summary 141
Chapter 5 Discussion 142
5.1 Summary of results 142
5.1.1 Research question1: Are students motivated when they are taking the DELTA? 142
5.1.2 Research question 2: Do they find the DELTA report useful? 144
5.1.3 Research question 3 and 4: Are students motivated before and after taking the DELTA? How can Dörnyei and Ottó's (1998) process model of L2 motivation be applied in explaining the process of test motivation in a low stakes diagnostic test? 146
5.2 The role of DELTA in L2 learning 149
5.3 Limitations of the study 151
5.4 Implications of the study 153
5.5 The way forward 154
5.6 Potentials for researchers and administrators 157
5.7 Chapter summary 160
References 161
Appendices 169
I Sample DELTA report 169
II Questionnaire survey 176
III List of interview questions 183
IV Sample interview transcription and translation 186

List of tables

Table 2.1 Comparison of L2 motivation theories 35
Table 3.1 Student Opinion Scale items 71
Table 3.2 Test Blueprint for SOS 72
Table 3.3 DELTA Report Usefulness—questions on action 76
Table 3.4 DELTA Report Usefulness—questions on information 76
Table 3.5 List of themes and categories 83
Table 4.1 Rotated Component Matrix of SOS 90
Table 4.2 SOS-Effort: Item-Total Statistics 91
Table 4.3 Fit Statistics of SOS 96
Table 4.4 Rotated Component Matrix of SOS (Q7 deleted) 98
Table 4.5 Summary of Category Structure of SOS 102
Table 4.6 Table of Measures of Items on Importance 105
Table 4.7 Table of Measures of Items on Effort 105
Table 4.8 Responses to the SOS-Importance questions (n=141) 106
Table 4.9 Responses to the SOS-Effort questions (n=141) 106
Table 4.10 Summary statistics of SOS 107
Table 4.11 Characteristics of students based on SOS Winsteps analysis 110
Table 4.12 Component Matrix of Report Usefulness Scale (8Qs) 113
Table 4.13 Fit Statistics of Report Usefulness Scale 115
Table 4.14 Component Matrix of Report Usefulness Scale (7Qs) 116
Table 4.15 Summary of Category Structure of Report Usefulness Scale 116
Table 4.16 Table of Measures of Report Usefulness Scale 118
Table 4.17 Responses to the Report Usefulness Scale-Action (n=134) 118
Table 4.18 Responses to the Report Usefulness Scale-Information (n=134) 119
Table 4.19 Summary statistics of Report Usefulness Scale 119
Table 4.20 Students perceived usefulness of the DELTA report 122
Table 4.21 Correlations between Importance, Effort and Report Usefulness 123
Table 4.22 Correlations between DELTA measure, SOS and Report Usefulness (16 students) 125

List of figures

Figure 2.1 Dörnyei and Ottó's (1998:48) process model of L2 motivation 38
Figure 2.2 Test motivation process model (Compulsory test: GSLPA) (Tsang, 2011) 41
Figure 2.3 Test motivation process model (Voluntary test: IELTS) (Tsang, 2011) 42
Figure 2.4 Sample (simulated) DELTA track 56
Figure 2.5 Sample DELTA Component Skills Profile 57
Figure 2.6 Sample DELTA Component Diagnostic Report on Reading 57
Figure 2.7 Sample DELTA bubble chart 62
Figure 4.1 Scree Plot of SOS Factor Analysis 89
Figure 4.2 Person-Item Map of SOS 94
Figure 4.3 Bubble Chart of SOS 95
Figure 4.4 Scree Plot of SOS with 9 items 97
Figure 4.5 Person-Item Map of SOS (Q7 deleted) 99
Figure 4.6 Person-Item Map of SOS with Thresholds 101
Figure 4.7 Probability Curves of SOS 104
Figure 4.8 Item-Person Map in 3 motivation levels 109
Figure 4.9 Scree Plot of Report usefulness Scale (8Qs) 112
Figure 4.10 Person-Item Map of Report usefulness Scale 114
Figure 4.11 Bubble Chart of Report Usefulness Scale 115
Figure 4.12 Probability Curve of Report Usefulness Scale 117
Figure 4.13 Item-Person Map of 134 students on Report Usefulness Scale 120
Figure 4.14 Distribution of SOS measures of the 16 students 124
Figure 4.15 Dörnyei and Ottó's (1998) actional phase 128
Figure 4.16 Tsang's (2011) actional phase of compulsory exit test 128
Figure 4.17 Actional phase of diagnostic and tracking test 128
Figure 4.18 Test motivation process model 130
Figure 4.19 Location of students with comparatively higher and lower SOS measure 132
Figure 4.20 Process mdoel of diagnostic test—Preactional stage 136
Figure 4.21 Location of students who find report more or less useful 138
Figure 5.1 Test motivation process model (reproduction of Figure 4.17) 147

Figure 5.2 Interactions of different process models with the L2 learning process (Tsang, 2011) 150
Figure 5.3 DELTA and the L2 learning process 151

Statement

Why Doctor of Education (EdD) at the Institute of Education, University of London?

I commenced my EdD in October 2007. Before that, I was an English Instructor at a Hong Kong tertiary institution, during which I had always been looking for opportunities for further studies. When it came to the decision of whether I should go for a PhD or EdD, after careful consideration and comparison of the nature of the two, I decided to go for the latter because I liked the way that the programme was structured and the progress tracking mechanism at different stages; and I particularly enjoyed having a group of 'classmates' and the opportunities to meet, exchange ideas and learn from each other during the research weeks. More importantly, the flexibility that the EdD programme offered by the Institute was just perfect for people who desired to maintain their full time work in their home country while taking their doctorate abroad. I completed my MA at the School of Oriental and African Studies more than 10 years ago and greatly appreciated the superb study culture and environment as well as the immensely high academic standard there. Therefore I had no hesitation at all when an opportunity to study again in London arose.

The four EdD taught modules

i. Foundations of Professionalism

The first module provided us a valuable opportunity to reflect on our own career. It introduced models of professionalism and different expectations of key stakeholders in education. My assignment was about 'professional identity' which was once one of my biggest concerns when I worked as an instructor of self-financed higher diploma programmes (higher diploma programmes together with associate degrees are referred to as sub-degrees in Hong Kong). In the process of working on the assignment, I was able to position myself in a bigger picture of the education environment and was inspired to reconsider what mattered to me the most in the teaching profession—what was vital as an English instructor should be how I could contribute to the sector that I was working at. This reflection helped formulate my assignment topic in the next module.

ii. Methods of Enquiry I (MoEI)

In the MoEI, I learnt how to write a comprehensive and feasible research outline. The course was of great value to students like me who had not received formal research training before. My research question for the assignment in this module was greatly influenced by my first assignment and was concerned about whether an English language course offered in the sub-degree programme I was teaching was able to help students improve their English proficiency. Although the assignment only required us to submit a research design and no real data was collected, the whole process of brainstorming, writing and reviewing of the assignment helped me develop a solid understanding of the structure of a research proposal and the concerns in designing a research project. As an initial training on how to plan a research project and as a preparation practice for the MoEII module to be followed, the MoEI was a very useful and valuable exercise. The practice gained from the writing up of the research methods and ethical issues sections were particularly useful.

iii. Specialist Course in International Education

In this module, we were required to complete two tasks: together with the short assignment on the Education for All game, I submitted another essay on the 'Internationalization of Hong Kong universities in the post colonial era'. With 'Internationalization' as the broader perspective, I shifted my focus from the current English standard at sub-degree level to whether tertiary students in Hong Kong could maintain their English proficiency. The implementation of an English exit test was considered to be an effective initiative in this regard. While working on this assignment, my initial interests in examining the effects of language tests started to grow which led to my research topic in my MoEII.

iv. Methods of Enquiry II (MoEII)

The last taught module of the programme was a continuation of the MoEI. In the MoEII, we were required to plan and carry out a small-scale research design. Having considered the experience gained in the MoEI and the Specialist Course, I revised my research question in the MoEII to a more practicable one, i.e. 'Should an English exit test be introduced into the sub-degree curriculum—Higher Diploma students' perspectives in Hong Kong'. My research objective was to find out from the students' perspectives whether they welcomed an inclusion of

an exit test in their syllabus, and if the answer was affirmative, then what kind of exit test would be the most preferred and why. My research was just a small-scale one that covered the final year students of one of the programmes that I taught. I used questionnaire survey and my own self-reflection as a sub-degree English instructor as the research method. The findings of this study in short were that students' opinion varied but a majority of them supported having an English exit test; they believed that having an exit test could give them the motivation to work harder in order to improve their English and that having good results in the exit test could help them articulate into degree programmes.

Upon completion of the MoEII, I developed a strong interest in investigating the effect of an English exit test on students' learning of English and learnt practical methodological skills such as the techniques in compiling a questionnaire and handling ethical issues as well as data analyzing skill (e.g. grouping similar information and using SPSS to create different charts). MoEII also helped me properly shape my research topic of the IFS, and more importantly, offered me a chance to carry out a mini pilot study which formed a stepping stone for my IFS research.

The Institution Focused Study

From the MoEI, in which we were required to create a research design, to the MoEII in which we had to carry out a full-blown research project, I learnt that there were many issues to consider before a research study could be carried out. The modules gave me practical training such as how to write a good and feasible research topic, how to set research questions and how to collect data. In particular, the MoEII helped me crystallize as to what I would like to investigate in the coming research stage. I started to realize that language exit testing was playing an increasingly important role in tertiary education especially in the Hong Kong context where English had a vital role in maintaining the prosperity of the society in the post colonial time. In 2002, the Hong Kong government introduced the Common English Proficiency Assessment Scheme, under which the Government sponsored university students to take IELTS as an English exit assessment on a voluntary basis. During the time when I was writing my IFS proposal, I moved from teaching sub-degree programmes in one institution to teaching degree programmes in the

Hong Kong Polytechnic University (HKPU). There had always been suspicion that the English language standard among Hong Kong students was decreasing over time especially after the transfer of sovereignty to China. There had also been doubts whether an exit test could help with the decreasing English standard among Hong Kong students; and much controversies as to what kind of English exit test should be introduced in Hong Kong tertiary education - should it be an internationally recognized test such as TOEFL and IELTS, or should it be a territory-wide test particularly designed to be used in the Hong Kong context. HKPU gave me the best environment in carrying out my research because at that time, it was the only university in Hong Kong which had developed its own exit test (known as GSLPA) and it was compulsory for students to take the test before their graduation (the test was considered as low stakes as the test results did not affect students' GPA); while students were also encouraged to take the government funded IELTS (IELTS was considered as higher stakes as a lot of employers and institutions gave much weight to IELTS results) in their final year of study. From my MoEII study, I had developed a better understanding about the importance of motivation in one's L2 learning process. I therefore decided to combine the two elements, test and motivation, in my IFS research. My IFS topic was 'A study of two English exit tests on students' motivation to learn English at a Hong Kong university'. Apart from the impact of the two exit tests, by modifying Dörnyei and Ottó's (1998) process model in L2 learning, I proposed a process model in test preparation and test taking. My main argument was that while the stakes of a test affect the intensity of students' motivation, test preparation and taking were in fact a dynamic process in which motivation fluctuated in the course of the process regardless of the stakes of the tests. This argument was supported by my questionnaire and interview findings.

The Thesis

After my IFS, I started to realize that it was formative assessment that could better help with L2 teaching and learning. At that time, apart from teaching, I was also the Project Manager of an inter-institutional research project in Hong Kong in which three institutions jointly developed an online diagnostic and tracking test (DELTA). Hence, naturally, I had re-structured my thesis and looked into the motivation of students in the DELTA and their perceived usefulness of a diagnostic report; and what role formative assessment such as the DELTA was playing in students' dynamic L2 learning process.

Reflection on the programme and my professional development

The EdD experience had a huge impact on my career. Not only did it guide me in discovering my ultimate research interest but it also gave me adequate academic training and provided me an opportunity to evaluate the different aspects in the tertiary education environment in Hong Kong. After going through the four modules, the IFS and the thesis, I now understand the phenomena of the internationalization of Hong Kong's tertiary education and the emergence of the sub-degree layer in the education system more than I did before. The successful internationalization of tertiary education was mainly due to Hong Kong's historical development, geographical location and bilingual environment; while the emergence of the sub-degree layer was caused by the change in government's policy. I also recognize the value and effectiveness of the introduction of English exit test at the tertiary level; and most importantly, I found my research interest—motivation in second language testing and the role of diagnostic tests in students' L2 learning.

Throughout the years of the EdD training, I had developed a deep interest in the research of testing motivation and this allowed me to reflect and reconsider my own career path. My current job as an instructor and a member in the Language Testing Unit of the English Language Centre provides me with the best environment for conducting my EdD research. I am able to have easy access to all kinds of resources and data relating to different English tests and can be in touch with the students taking the test; I have the opportunity to meet leading experts and scholars in the field; I have the chance to present in regional and international conferences in relation to testing issues; and more importantly, I learnt a lot about practical issues in running and analyzing tests such as the need to conduct rater training and double marking. My EdD is indeed a perfect fit for me as the program and my job are mutually beneficial to each other.

I am more determined in developing my expertise further in this special area after the EdD programme. As mentioned in the final section of my thesis, I have already laid down a list of research plans based on my thesis study and have the support from my unit head. The EdD to me is not an end in itself but a means to guide me in my future academic life.

Chapter 1 Hong Kong Chinese learners of English in a university setting

1.1 English as the prestigious language

English has always been the prestigious language in Hong Kong since it was a British colony as proficiency in the language represents the gain of horizontal and vertical access in the society. It is the language to be used in professions from lawyers to accountants and from doctors to government officials; it is also the language to be used by the managerial level within an organization. An English-dominant education system has been in practice in the territory in the last century and has produced an elite bilingual social group whose cultural identities are constructed through their successful investments in an English-medium education, a mastery of the English language and their familiarity with and membership in English-based modern professional institutions (e.g. the various English-based professional associations of accountants, lawyers, doctors and engineers, and English-mediated professional accreditation mechanisms) (Choi, 2003; Lin, 2005; Pennington & Yue, 2007; Tung, Lam, & Tsang, 1997). In the late years of colonialism, the government expanded the higher education sector—from a two-university system (admitting about 2% of age-appropriate students) to eight publicly funded universities, most of which maintained English as the teaching medium (admitting about 18% of age-appropriate students) (Lee & Gopinathan, 2003).

In July 1997, Hong Kong's sovereignty was returned to China by the British. After that the Hong Kong government announced the 'biliterate and trilingual policy' for all schools (students have to be proficient in both written English and Chinese; and fluent in spoken English, Cantonese and Putonghua) (Lai, 2001). Since then, there is a rising demand on the use of written Chinese and Putonghua, but English still remains as a symbol of success. The impression that English is the esteemed language and the key to a prosperous future is deeply rooted in all Hong Kong students' minds.

1.2 The examination-led education system

The Hong Kong educational system is characterized as an examination-led system where what goes on in the classroom is largely dictated by what happens in the public examination halls (Forlin, 2007; Fullilove, 1992). This phenomenon happens as early as in primary schools when students have to participate in the Secondary School Places Allocation exercise and is intensified with the Hong Kong Certificate of Education Examinations (HKCEE) at the end of Secondary 5 to select students to proceed to further studies and the Hong Kong Advanced Level Examinations (HKALE) at the end of Secondary 7 for tertiary selection (Cheng, 1997). (Starting from 2012, a new 3+3+4 academic structure is in place in which all students are expected to complete three years of junior secondary education, followed by 3 years of senior secondary education. A single baccalaureate-style examination, the Hong Kong Diploma of Secondary Education (HKDSE) has replaced the two Examinations (HKCEE and HKALE). A proportion will then proceed to four-year undergraduate degree programs in universities (Hill & Wan, 2006).)

Assessments and examinations are perceived by the public in general as the most equitable and impartial method of determining achievement and so, while they place enormous pressure on students, they are seen as stepping stones to success (Kennedy, Fok & Chan, 2006; Lai, 2009; Tang & Biggs, 1996). The examination-driven system has resulted in exam-oriented classrooms and students who only learn for the sake of good marks (Kennedy, Fok & Chan, 2006). This exam heavy culture is also spreading to the tertiary sector.

As English is regarded as one of the prerequisites for Hong Kong to continue as an international financial centre as well as one of the strengths of Hong Kong people over the thousands of millions of mainlanders, it is believed that maintaining a high level of proficiency in the language is paramount. Due to the overwhelming concerns that there is a drop in the English proficiency among students in Hong Kong in the last few decades, the HKSAR government has implemented a list of remedial policies. Among them, the one that directly related to the tertiary education sector is the introduction of English exit test in the university curriculum (Coniam & Falvey, 2002; Lumley & Qian, 2003; Qian, 2007) (A more detailed description of the rise and fall of exit tests in the tertiary sector in Hong Kong will be provided in section 1.5).

1.3 Hong Kong Chinese learners of English

The historical development of the importance of English and the examination-led education system as discussed in the previous sections, coupled with the traditional Chinese culture, contribute to the unique character of Hong Kong Chinese learners of English:

i) Learning English for pragmatic reasons

Due to the examination-driven education setting and the importance of English being one of the most decisive criteria for the success of students both at university or at work (Cheng, 1997; Fullilove, 1992), Chinese students in Hong Kong were motivated to learn English mostly for pragmatic reasons – to pass examinations, to get into a good tertiary education and eventually to get a good, high-paying job (Watkins, 2009). Students are mostly extrinsically motivated to learn English. Lai (2009) compared the findings of four studies conducted in 1998, 1995, 1992 and 1990 on Hong Kong secondary school students and concludes that students have a strong instrumental and career-related motive in learning English. Apart from this practical reason, Hong Kong students' learning attitudes are also shaped by traditional Chinese culture. Lee (1996) studied the impact of Confucian values on Hong Kong Chinese learners and concludes that the value of pursuit of self-perfection through learning is considered the highest achievement in life and it is also the gateway for family honour, social contribution, and upward social mobility.

ii) Passive learners

The picture that often emerges from the research literature on Chinese learners is a caricature of rote-learning, memorization and passivity which is said to be also determined by their 'Confucian heritage' (Kennedy, 2002). Murphy (1987:43) also suggests that 'Hong Kong students display an almost unquestioning acceptance of the knowledge of the teacher . . . may be a transfer of the Confucian ethic of filial piety, coupled with an emphasis on strictness of discipline and proper behaviour'. According to a study by Tsui (1996), over 70% of a group of 38 teachers identified getting more student oral response as one of their major problems. These teachers described their students as "passive", "quiet", "shy", "unwilling to speak English" and

so forth. Ferris and Tagg's study (1996) reviewed that 12 professors specifically mentioned Asian students as having cultural differences which inhibited their oral participation in class and their willingness and ability to ask questions.

However, these preconceptions may not be telling the whole picture. Liu and Littlewood (1997) and Littlewood (2000) reveal that students actually prefer an active speech role in class and their apparent reticence is only due to their lack of experience and confidence in speaking in English. Cheng (2000) shares the same view and argues that the allegation of passivity of Chinese learners is over-generalized and even if some Chinese learners indeed appeared to be quieter, the causes are situation specific such as due to the differences between teaching methodologies and lack of required language proficiency rather than a pre-set culture. Clark and Gieve (2006) warn that the generalization of Chinese as passive learners is a kind of racial stereotype and has reduced individuals to inadequately understood group characteristics.

Despite the above divergence of opinion regarding Chinese students' characteristics, students' learning attitudes and expectations towards their learning approach are likely to be different when they come to tertiary level. Lee (1999) argues that students entering universities begin to realize that tertiary education is no longer didactic and requires them to think critically. Lai (2000) reviewed a few studies that there appears to be stronger link between motivation, effort, and achievement in that high achievement triggers intrinsic motivation to learn English. Breaking free of secondary school life seems to give them license to question traditional approaches to English teaching and express strong preference for collaborative learning (Gieve & Clark, 2005; Littlewood, 2001). Kennedy (2002) conducted survey and case studies and his results suggest that Hong Kong adult learners are receptive to new modes of learning and adopt learning styles quite different from those they deployed in school. Students also have adopted more flexible language learning strategies if they are in unfamiliar learning environments (Gao, 2006).

To conclude, Hong Kong learners of English at tertiary level are pragmatic and they can appear to be passive due to their lack of confidence; but when they reach tertiary level, there can be possible changes in their learning attitude.

1.4 Autonomous learning in tertiary education

The examination-oriented and highly competitive education system in Hong Kong has reinforced a learning model where the teacher is in full control of the learning process, giving explicit directions for almost every learning activity. The total learning environment is seen as one in which independence and individuality is neither required, valued or nurtured (Biggs, 1987). Researchers have suggested that typical Hong Kong Chinese students see knowledge as something to be transmitted by the teacher rather than discovered by themselves (Watkins & Biggs, 1996). According to Chan (2001) Hong Kong students are used to the educational culture conditions very early and it is extremely difficult for any change in learning habits to take place when they enter university.

Although the established spoon-feeding practice in secondary school has shaped the passive learning attitude of students; given the higher teacher-student ratio, less contact hours in English language classes and the difference in teaching focus, it is the wish of the universities to call for students to be more autonomous (Chan, Spratt, & Humphreys, 2002). Littlewood (1996) defines an autonomous person as one who has an independent capacity to make and carry out the choices which govern his or her actions while capacity depends on both ability and willingness. According to Morrison (2011), independent learning is sometimes used synonymously with terms such as ‘autonomous learning’, ‘self-directed learning’, ‘independent study’ and ‘self-regulated learning’. In this study, the terms ‘autonomous learning’ and ‘independent learning’ refer to the same entity.

The aim of independent language learning is to develop the ability of learners to engage with, derive benefit from and contribute to learning environments not directly mediated by a teacher (White, 2011). It is the wish of the English Language Centres in universities in Hong Kong that with the ability and knowledge of independent language learning, students can further develop into autonomous learners who have the skill, will and self-regulation to survive and thrive in different academic or training environment (Weinstein et al, 2011:42). It is one of the English Language Centres’ goals to alter students’ perceptions to English learning and making them autonomous learners that are intrinsically motivated, and eventually successful language learners

1.5 The rise of the Diagnostic English Language Tracking Assessment

As discussed briefly in section 1.2, there is an overwhelming concern from the public that English proficiency among Hong Kong students is dropping (Boyle, 1997; Hirvela & Law, 1991; Lu, 2002; Qian, 2008). In light of this, the Hong Kong government has implemented a list of remedial policies in the past decade. The introduction of an English exit test in the university curriculum is one of them. In 2002, the University Grants Committee (hereinafter referred to as 'UGC'), a governmental body that oversees funding and policies of tertiary institutions in Hong Kong, implemented an initiative known as the Common English Proficiency Assessment Scheme (hereinafter referred to as CEPAS). Under this Scheme, students are encouraged to take the IELTS in their last year of university studies by being able to get a full refund of the examination fees for the test. If they apply for the refund, the date on which they have taken the IELTS will be shown on their academic transcripts.

Though exit test is regarded as one of the useful tools in enhancing students' English language proficiency, there are doubts as to whether it is sufficient in fulfilling this purpose on its own as well as whether IELTS should be used as "the Test". In May 2011, the UGC announced that after careful considerations and review, they plan to move beyond the use of IELTS to a more effective scheme for enhancing the language proficiency of students:

> "CEPAS has already achieved its original purposes of enhancing students' awareness of the importance of English language proficiency and providing a wealth of information on students' strengths and weaknesses in English. It is time to move beyond encouraging students to take language tests to an arrangement that can provide direct funding support for institutions' joint projects that aim to enhance students' language proficiency." (UGC, 2012)

Prof Lyle Bachman was engaged as consultant by the UGC to review institutions' language enhancement activities in 2008/09. In his review, he commented that:

> "over-emphasis on test preparation [IELTS] might undermine efforts to help students genuinely improve their language proficiency… The Tertiary English Language Test (now renamed as The Diagnostic English Language Tracking Assessment) is well-researched, solidly grounded and hence should receive serious attention. As the test is specifically designed for use as a diagnostic tool in local institutions, students and institutions will likely find it useful and practical." (Bachman, 2010)

The Diagnostic English Language Tracking Assessment (DELTA), mentioned in Prof Bachman's report, was previously known as the Tertiary English Language Test (TELT). It was developed by the Hong Kong Institute of Education (HKIEd) in 2007 with an aim to place students entering the Institute into one of three bands for the purpose of English language proficiency enhancement. In mid-2009, the HKIEd decided to further develop the TELT for the purposes of diagnosing and tracking and in that year Lingnan University of Hong Kong and the Hong Kong Polytechnic University joined to form a collaboration in the further development of the TELT into a web-based assessment. To reflect the change in the nature of the TELT, it was renamed DELTA. Since mid-2010, the fourth institution, the City University of Hong Kong, has also joined the development project.

In light of the above, the UGC therefore decided to fund the Diagnostic English Language Tracking Assessment team (hereinafter referred to as 'DELTA') in further developing the test as well as to encourage more institutions to try out the test by providing funding for the installation of the DELTA system in those institutions starting 2012.

1.6 Research rationale— Motivation in low stakes diagnostic assessment

Performance in an assessment is the result of the interplay among prior knowledge, speed of information processing and test motivation (Baumert & Demmrich, 2001). It is logical to assume that students will be less motivated to perform well on low-stakes assessments. As Wise and DeMars (2005:1) acknowledge, 'without consequences for performance, many students will not give their best effort to such "low-stakes" tests; as a result, their assessment test scores may not

serve as valid indicators of what they know and can do'. It is questionable, especially for low-stakes assessments, whether they are assessing students' real proficiency or how much of their ability the students would like or be able to show. Not acknowledging students' motivation in the assessment situation and the impact of motivation on performance may pose a threat to the validity of the interpretation and use of assessment results (Eklöf, 2010).

The DELTA is a type of formative assessment which diagnoses and tracks students' progress during their years of study at the university. The Association of Language Testers in Europe (1999) defines a diagnostic test as a test which is used for the purpose of discovering a learner's specific strengths or weaknesses and the results of the test may be used in making decisions on future training, learning or teaching. Although the DELTA is a low stakes assessment as the results of it will not count towards students' grade point average, students get a report with their DELTA measure and detailed feedback on their performance. With the report, they are able to tell their strengths and weaknesses in the areas of English, compare their own performance with the other students in their cohort each year and keep track of their own development in English. Since the DELTA is a low stakes assessment, even if students' motivation is low, it may not seem to be very problematic. However, low motivation in fact creates problems in different aspects. The most important implication is that the test may not be able to serve its diagnostic purpose: if students did not pay enough effort in taking the test, the report would not be able to reflect accurately their performances in the different areas of English. Since students did not take the test seriously, naturally, they would not read the report seriously or use it as a reference when they do further studies. Besides, it is also useful for university administrators, teachers or language researchers to know if students are motivated in the test so that when they interpret students' DELTA score, such as when they compare performances across institutions or different groups of students, they can also take students' motivation level into consideration. The main focus of my research therefore, is to investigate the motivation that a student has to perform to the best of their ability whilst sitting the test.

In my Institution Focused Study (IFS), I reported that test preparation and test taking are in fact a dynamic process during which motivation will fluctuate regardless of the assessment stakes. By looking into students' motivation in two tertiary exit assessments (summative assessments), one

low stakes and the other relatively higher stakes, I found out that Dörnyei and Ottó's (1998) process model in L2 motivation can be used to explain one's test motivation (see section 2.2.4 for a brief summary of my IFS or for details, see Tsang, 2011). The value of DELTA is to provide a diagnostic report which the students can refer to when they want to do self study in improving their English and to provide a tracking function of their English proficiency throughout their years of university study, thus, the major purpose of the test is to motivate and help students in their preactional and postactional phase within the process model (see again section 2.2.4). Therefore apart from students' motivation whilst sitting the test, I also wanted to find out, first, if students find the diagnostic report useful, and second, whether the DELTA provides them with the motivation to prepare for sitting the test as well as the motivation to work on their English/ to have a better score in the next DELTA. Finally, based on the results from my study, I aimed at coming up with a modified process model which could explain the test motivation in diagnostic tests and I believe that my study can shed some lights as to the role of diagnostic test in students' English learning.

In short, this study is mainly about the three types of motivation involved in the test-taking process:

i. Motivation in the actional phase: the motivation that a student has to perform to the best of their ability whist sitting the test;
ii. Motivation in the preactional phase: the motivation that a student has to prepare for sitting the test;
iii. Motivation in the postactional phase: the motivation that a student has to improve their English proficiency after the test and to attain a better score in the next test.

1.7 Chapter summary

This chapter set out the background of the education system in Hong Kong-- being English-based and examination-led. It explained the history and background which constitutes the importance of English in Hong Kong; and how the government has tried to promote Hong Kong

students' English standard by introducing exit tests at the tertiary level. It then went on to give details about the development of the DELTA as the government has realized the significance of formative assessment. The last section of the chapter explained the research rationale and the aims of the study, i.e. to find out if students are motivated to perform to the best of their ability whilst sitting the test; if they are motivated to prepare for sitting the test; and if they are motivated to improve their English after the test.

In the next chapter, I will provide a review of the literature from motivation theories to second language motivation and diagnostic tests; as well as a detailed description of the DELTA itself.

Chapter 2 Motivation and low stakes diagnostic tests

As explained in my research rationale in section 1.6, motivation plays a vital role in how much of a student's ability the test results can reflect. The aim of this thesis is to find out the three different types of students' motivation in a low stakes diagnostic test and evaluate the role of diagnostic test in students' English learning. Motivation in a language test is an interplay of many different factors. In order to unveil what actually motivates a student in the test taking process, in this chapter, I start with the presentation of (i) the array of the most dominant psychological motivation theories and (ii) why people are motivated to learn L2, before I get into the literature on test motivation. After that I talk about how previous studies have been trying to measure test motivation, particularly in low stakes tests and diagnostic tests. Finally, an account of diagnostic tests, a detailed description of the DELTA and research on the usefulness of diagnostic feedback are presented.

2.1 Some dominant psychological motivation theories

Three of the most dominant motivation theories in psychology are presented here and they are namely: the Achievement Goal Theory, Self-Determination Theory and the Expectancy-value Theories.

2.1.1 Achievement Goal Theory

Achievement Goals refer to the purposes or reasons an individual is pursuing a task, which are most often operationalized in terms of academic learning tasks, although they can be applied to other achievement contexts such as athletic or business settings . The term 'goal orientation' is often used to represent the idea that achievement goals are not just simple target goals or more general goals, but to represent a general orientation to the task that includes a number of related beliefs about purposes, competence, success, ability, effort, errors and standards (Pintrich, 2000).

Ames (1992) distinguishes between two achievement goal constructs— 'mastery' and 'performance'. Mastery-oriented students focus on learning, understanding, developing skills and mastering information whereas performance goal oriented students focus on managing the impression that others have regarding one's ability. The latter attempt to create an impression of higher capability and avoid giving the others an impression that they are incapable (Dweck, 1986).

2.1.2 Self-Determination Theory

According to Self-Determination Theory, there are two general types of motivation; one based on intrinsic interest in the activity and the other based on rewards extrinsic to the activity itself. The theory of Intrinsic and Extrinsic Motivation is undoubtedly one of the most eminent motivation concepts under the Self-determination Theory. The two types of motivation are not categorically different but rather lie along a continuum (Noels, Pelletier, Clement & Vallerand, 2003). Intrinsic and Extrinsic motivation bear resemblance to the achievement goal constructs proposed by Ames (1992) as explained in the previous section; where 'mastery-oriented' students can said to be more intrinsically motivated in doing a task and those who are 'performance-oriented' are more concerned about the extrinsic expectation or requirement imposed on them.

Ryan and Deci (2000) commented that probably no other better phenomenon could reflect the positive potential of human nature than intrinsic motivation, which is defined as the inherent tendency to seek novelty and challenges, to extend and exercise one's capacities, to explore, and to learn. Ryan and Deci point out that although intrinsic motivation is an important type of motivation, an adult's action is no longer intrinsically motivated when there are increasing social pressure or responsibility for them to do things that they are not interested in. By way of comparison, extrinsic motivation refers to the performance of an activity for the purpose of attaining some separate outcome. Deci and Ryan (1985) introduced the Organismic Integration Theory (OIT) to detail the different forms of extrinsic motivation. They divided extrinsically motivated behaviour into four types in the degree which they are autonomous: first, externally

regulated behaviour means one performs an act to satisfy an external demand; second, introjected regulated behaviour involves taking in regulation and acting to avoid guilt or anxiety; third, regulation through identification, when the action is identified as personally important; lastly, integrated regulation when identified regulations are evaluated and in line with one's other values and needs. The authors also contrasted all types of intrinsic and extrinsic motivation with amotivation. Amotivation appears when people see no relation between their actions and the consequences of those actions, and thus have no intrinsic nor extrinsic motivation to perform the activity. They would be expected to give up the activity as soon as possible.

2.1.3 Expectancy-Value Theories

The Expectancy-Value Theories of motivation are one of the most influential and well researched frameworks in the area. According to Schunk, Pintrich and Meece (2008), expectancies are people's beliefs and judgments about their capabilities to perform a task successfully; while values refer to the students' beliefs about the reasons for engaging in a task.

Atkinson can be regarded as one of the leading scholars who shape the Expectancy-Value Theories. His concept of Achievement Motivation is the model which had dominated the field for a few decades and is still considered influential nowadays. Atkinson (Atkinson, 1964; Atkinson & Birch, 1978) suggests that human behavior is determined by four components. In addition to the expectancy of success and incentive values, he adds two more components in his model— the need for achievement and the fear of failure. Atkinson and Birch described the need for achievement as a capacity for reacting with pride in accomplishment and the fear of failure as a capacity for reacting with shame and embarrassment when the outcome of performance is failure (1978:96).

A few researchers came up with various theories with an aim to explain the 'expectancy' component in the Expectancy-Value framework. Weiner (1986, 1992) contributed to the Attribution Theory which asserts that people's motivation to succeed will be affected by their own past successes and failures with attributions as the perceived causes of outcomes. Self-

Efficacy is deemed important by Bandura (1977, 1986, 1993, 1997) in the determination in performing an action. Bandura defines Self-Efficacy as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances". Bandura (1977) hypothesizes that students with a high sense of efficacy for accomplishing an educational task will participate more readily, work harder and persist longer when they encounter difficulties than those who doubt their capabilities.

The second component of the Expectancy-Value Theories is the 'value' factor. Atkinson (1957) defines value as the relative attractiveness of success or failure of a task while Eccles et al (1983) list four components of subjective values: attainment value, intrinsic value, utility value and cost. Attainment value is the importance of doing well on a given task; intrinsic value is the enjoyment one gains from doing the task; utility value refers to how a task fits into an individual's future plans; and cost refers to what the individual has to give up to do a task as well as the anticipated effort one will need to put into task completion (Wigfield & Eccles, 2002). There is a major disagreement between Atkinson's assertion and the findings of Eccles, Wigfield and their colleagues as to what constitutes a valued task for individuals. Atkinson claims that individuals tend to value the tasks which are difficult for them to do; whereas Eccles and Wigfield (1995) and Wigfield et al. (1997) suggest that people actually value the tasks on which they think they can succeed.

The Expectancy-Value Theories, together with the Self-Determination Theory in the previous section, are two most prominent theories in psychology which explain motivation in general. In the next section, I will move on to talk about second language motivation.

2.2 Second Language Motivation

Second Language Motivation (hereinafter referred to as 'L2 motivation') is a very complex subject. Second language can be a subject in school in which one's performance level can be reflected in tests and exams while unlike other school subjects, second language is also a social and cultural issue. Hence, abundant research with diverse theories, approaches and focuses have

been carried out in this area. The earliest and most highly regarded theory is certainly Robert Gardner's Social Education Model.

2.2.1 The Social Educational Model

The Social Educational Model, as it is named, puts focus on the social interaction with members of the L2 group. According to the Model (Gardner & Lambert, 1972), a person's second language learning attitude is greatly influenced by how that person views that second language and the people and the community who speak that language. In their early studies into the second language learning situation in Canada, Gardner and Lambert (1959) identified two types of orientation to learn a second language— 'integrative' orientation and 'instrumental' orientation. Instrumental orientation refers to the pragmatic reasons in learning a second language, such as to get a job or to perform better in an examination. Integrative orientation is the desire to communicate in the second language or even to get into the community of people who speaks that language, and it is further characterized into 'integrativeness', 'attitude towards a learning situation' and 'motivation'. It is the interaction of these three components that determine the outcome of second language learning.

Gardner and his co-workers constructed an instrument, the Attitude/Motivation Test Battery (AMTB) (Tremblay and Gardner 1995:505; Gardner, Smythe and Clément, 1979), to measure the above three components which they believe can determine second language learning outcome. The AMTB has over 130 items and was written in the Canadian French setting. Apart from the three components, the AMTB also included the measurement of language anxiety and some other attributes. Gardner (1985) believes that integrative orientation is the most important type of motivation in second language learning and it is also the reason why he was being criticized as 'creating a false split' (Oxford, 1996) between integrative motivation and instrumental motivation.

In order to respond to the criticisms, Tremblay and Gardner (1995) expand the Social Educational Model by investigating the relationships between the motivational variables in the

Model with measures of motivation derived from the general psychological literature. The authors introduced the distinction between motivational behaviour and motivational antecedents. Motivational behaviour is defined as the characteristics of an individual that can be perceived by an observer while the motivational behaviour that cannot be perceived by an observer but is self-reportable by the actor is referred to as motivational antecedents (Tremblay and Gardner 1995:506). Tremblay and Gardner argue by reviewing the literature that the Social Educational Model has indeed included motivational elements from many other existing psychological literatures that describe motivation and proposed an expanded motivational model based on their review. According to the authors, motivational behaviour in the Model incorporates the components of 'effort', 'persistence' and 'attention' (which they named as motivational behaviour); while the characteristics of theories and concepts in intrinsic and extrinsic motivation, expectancy-value, goal-orientation theory etc., are motivational antecedents as these theories explain the internal characteristics of the individual that cannot be perceived by an observer. Apart from the literature review, Tremblay and Gardner (1995) conducted a study with 75 francophone secondary school students in Canada by asking them to complete the AMTB plus a few other motivational and attitudinal measures taken from other research or newly written by the authors for this study, and then analyzed together with the students' final grades in their French course. Using structural equation modeling analysis, the authors confirm that their proposed motivational model is sound, i.e. other elements of motivation can be incorporated into the Social Educational Model.

Gardner (2007) tries to distinguish between two types of motivation, namely the language learning motivation and classroom motivation, specifically the language classroom; as well as between two contexts, namely the cultural context and the educational context. He proposed a model to demonstrate the effects of cultural and education contexts on motivation in second language learning and named the variable for the cultural context and educational context as 'integrativeness' and 'attitudes toward the learning situation' respectively. He points out the difference between a motive or orientation and motivation and emphasizes that for him, the type of motivation is not that important, i.e. the distinctions between integrative and instrumental motivation and/or intrinsic and extrinsic motivation etc. are not helpful in explaining how motivation affects second language learning; it is the intensity of the motivation that matters.

Although Tremblay and Gardner (1995) extend the coverage of the Social Educational Model and argue that the Model in fact has incorporated the motivational components of other psychology theories, the 'intergrative motive' in second language learning is still the key in this Model. Gardner reiterates in his paper that an individual is integratively motivated when (2006:19):

a. the individual is motivated to learn the other language
b. the individual is learning the language because of a genuine interest in communicating with members of the other language (either because of positive feelings toward that community or members of that community, or because of a general interest in other groups)
c. the individual has a favourable attitude toward the language learning situation.

While Gardner's Model may explain the situation of second language motivation in some countries perfectly well, I am inclined to think that his Model is not the best in explaining the Hong Kong context which, as set out in Chapter 1, is examination oriented. Hong Kong students are passive learners and learn English mostly for pragmatic reasons. I share the view of Clément and Kruidenier (1983), who investigate the effects of ethnicity and target language etc. on the development of orientations to the study of language, that contrary to what has been proposed by earlier researchers such as Gardner and Lambert, integrative orientation is not fundamental to the L2 process but has relevance only in specific sociocultural contexts. In Clément and Kruidenier's study, 871 grade 11 students were asked to rate the 37 reasons for learning the target language. Correlations and factor analysis were run and the authors found out that it is another four types of orientations which may be seen to sustain motivation, and they are i) travel, ii) friendship, iii) knowledge and, iv) instrumental orientations. Their findings however, has not been followed up with a conceptual rationale describing a psychological mechanism to account for the importance of the above four orientations for L2 motivation (Noels, Pelletier, Clément & Vallerand, 2003) .

Moreover, though I agree with Gardner that the intensity of motivation should be the ultimate factor shaping one's second language learning; research on the types of motivation would contribute to our understanding of students' behavior and hence help teachers and university administrators in curriculum planning as well and thus should also be very helpful. In the next

sections, I will review two other important second language motivation theories and their suitability as the framework in this study.

2.2.2 Self-Determination Theory

Deci and Ryan's self-determination theory and intrinsic and extrinsic motivation has been so influential that quite a number of researchers have tried to apply it to L2 motivation. According to Ryan and Deci (2000), educational activities in schools are mostly not intrinsically interesting and should therefore fall into one of the four types of extrinsic motivation, i.e. externally regulated behavior, introjected regulated behavior, regulation through identification and integrated regulation. They raised the importance of the fostering of internalization and integration of values and behavioural regulations, where internalization is the process of taking in a value or regulation, and integration is the process by which individuals more fully transform the regulation into their own so that it will emanate from their sense of self.

Noels, Pelletier, Clément and Vallerand (2003) carried out a study on language learning orientations using the self-determination theory. The project involved the development of an instrument to assess the different subtypes of intrinsic and extrinsic motivation and it also explored the link between these subtypes and the orientations proposed by Clément and Kruidenier (1983), i.e., travel, friendship, knowledge and instrumental orientations. The instrument contains three scales: the first one is to measure the above four subtypes of orientations by Clément and Kruidenier (1983); the second scale is to measure the constructs of Self-determination Theory, i.e. intrinsic, extrinsic and amotivation; and the third scale is to measure the antecedents of motivation such as self-perception of L2 proficiency, perception of autonomy and anxiety. 159 students registered in the English psychology classes at a French-English bilingual university participated in the study. The results show that instrumental orientation is most highly correlated with external regulation while travel, knowledge and friendship are positively and highly correlated with the more self-determined form of motivation. The authors concluded that learner motivation can be validly assessed using Deci and Ryan's

intrinsic and extrinsic subtypes and the pattern of correlations reflects a continuum of self-determination.

2.2.3 Expectancy-Value Theories

Although no real expectancy-value model has been proposed in L2 motivation research, several components associated with the expectancy-value framework have been incorporated into various L2 research paradigms (Dörnyei, 2001:55).

Gardner's social motivation theory incorporates the idea of intrinsic value and utility value of the expectancy-value model. For example, the 'integrative motive' (which in essence is a kind of intrinsic motivation under the expectancy-value model) under the social motivation theory is the 'motivation to learn a second language because of positive feelings toward the community that speaks that language' (Gardner, 1985:82-83). Utility value under the expectancy-value model is also included in Tremblay and Gardner's model as 'Instrumental orientation'.

Clément and his colleagues studied extensively the interrelationship of motivational factors, self-confidence and L2 acquisition process (Clément, 1980; Noels & Clément, 1996). Their interpretation of self-confidence share similarities to the self-efficacy in the expectancy-value theories; but apart from being a pure cognitive component, they also see self-confidence as a socially defined construct. According to Noels and Clément (1996:4), self-confidence is defined as the self-perceptions of communicative competence and concomitant low levels of anxiety in using the second language. With increased second language competence, the individual will come to identify with the second language community. Clément and his colleagues (1994) also believe that when there is little direct contact with native speakers of L2, indirect contact with the L2 culture can also constitutes as a motivational factor. Their idea of self-confidence is in line with Gardner's integrative motive in the social motivation theory.

The L2 motivation theories/concepts discussed above have different philosophies and constructs, but they do share similarities with each other.

Table 2.1 Comparison of L2 motivation theories

	Internal	External	Self-evaluation	Others
Social Educational Model	Integrative (positive sentiment towards the L2 culture)	Integrative (Social need); instrumental		
Self-determination Theory	Intrinsic	Extrinsic		Amotivation
Expectancy-value theories	Intrinsic; attainment value (personal satisfaction)	Utility value; attainment value (fulfilling social expectation)	Expectancy (Self efficacy)	Cost
Clément and Kruidenier	Travel; friendship; knowledge	Instrumental		
Clément, Noels and colleagues	Self-confidence (positive sentiment towards the L2 culture)		Self-confidence	

Table 2.1 shows possible connections among them. From the table, we can see that 4 of the 5 major theories distinguish between internal and external factors while the Expectancy-value theories are comparatively more comprehensive because the concept of self-evaluation (self efficacy) and cost are being taken into account. As defined in section 2.1.3, cost refers to what an individual has to give up to do a task as well as the anticipated effort one will need to put into task completion (Wigfield & Eccles, 2002). Even within the internal factor, the Expectancy-value theories are more inclusive because apart from the intrinsic element which is similar to the other theories, i.e. learning L2 out of personal interests either towards the culture or towards the people who speaks that language, attainment value which represents personal satisfaction is also included. My study applied the Expectancy-value theories as the theoretical framework. I believe that this construct is particularly useful in explaining a students' test-taking motivation in the Hong Kong context: the pragmatic mentality of Hong Kong students can be reflected in the utility and cost components; the fact that English is the prestigious language in the territory can be shown in the attainment value (external), i.e. fulfilling social expectation.

2.2.4 The dynamics of second language motivation

The theories discussed in the previous sections all seem to be based on the assumption that motivation is a stable entity. However, a number of researchers in recent years have pointed out that second language learning is a sustained learning process and during such a prolonged process, it is quite impossible for motivation to remain constant. Dynamic L2 motivation is a fairly new idea in L2 research. Yet, there have been traces of the concept embedded in some of the education psychology theories much earlier. Weiner points out before explaining his research study that "it is evident, however, that the underlying premise of our conceptual approach is that thought does influence action. Further, a temporal sequence of events is presumed, in which action follows thought" (1974:4). Another significant example is the Action Control Theory suggested by Heckhausen and Kuhl (1985) which turns out to be very important for the later developments in the theory of dynamic L2 motivation. Heckhausen and Kuhl (1985) advocated the division of the 'predecisional' phrase and the 'postdecisional' phase. The predecisional phase is when initial wishes and desires are formed; while in the postdecisional phase, one has to exercise action control during the actual action.

Ushioda (1996, 1998) highlighted the importance of incorporating the "dynamic" perspective when researching L2 motivation. In her study in 2001, she tried to explore the motivational thinking of language learners. In her interviews with 20 students who were taking French as a second language as part of their degree programmes at the Trinity College Dublin, she focused her questions on four aspects in which two were related to the dynamics of motivation. She looked at students' motivational evolution over time and their motivational perspectives on L2 development over time. After a detailed content analysis of the interviews, she came up with eight motivational features: i) academic interest, ii) language related enjoyment, iii) desired levels of L2 competence, iv) personal goals, v) positive learning history, vi) personal satisfaction, vii) feelings about French-speaking countries or people and viii) external pressure/incentives and five motivational profiles of learners as well as developed a schematic representation of how learner's conception of motivation can be defined within a theoretical framework of varying temporal perspectives. Ushioda concluded that (2001:122):-

> "Motivation is thus viewed not simply as cause or product of particular learning experiences, but as process—in effect, the ongoing process of how the learner thinks about and interprets events in relevant L2 learning and L2 related experience and how such cognitions and beliefs then shape subsequent involvement in learning".

With reference to the work by Heckhausen and Kuhl and Ushioda, Dörnyei and Ottó (1998) developed a process model of L2 motivation (Figure 2.1). This model organizes the motivational influences of L2 learning along a sequence of discrete actional events within the chain of initiating and enacting motivated behavior.

Dörnyei (2001) suggests that time is relevant to motivation in two areas. Firstly, motivation is a complex mental process which forms gradually from initial planning, then to implementation and finally to evaluation. Different phases in the process may be caused by different motives. Secondly, when the action in question is a sustained long-term activity such as second language learning, it is not possible for motivation to remain constant throughout the whole process.

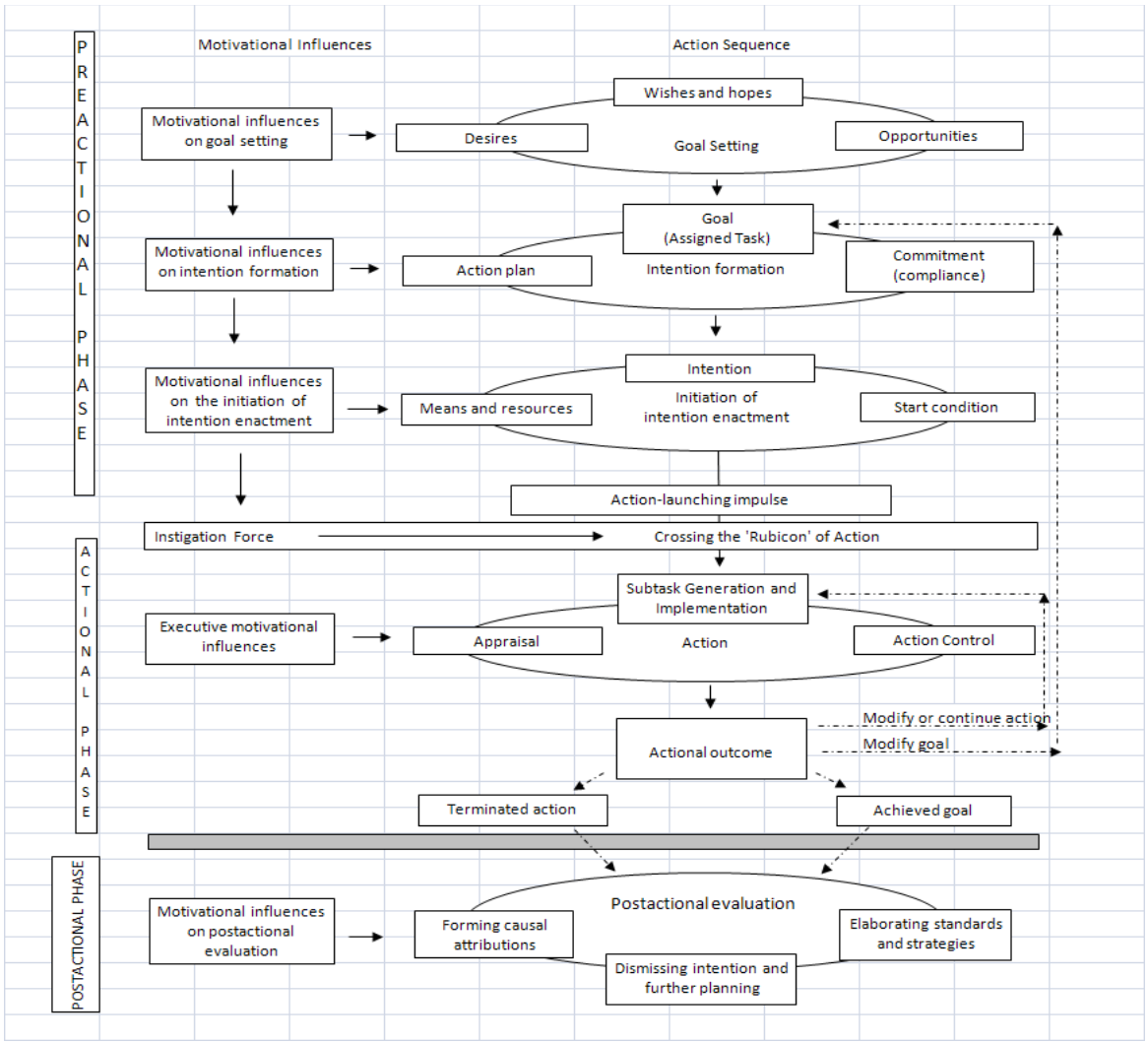


Figure 2. 1 Dörnyei and Ottó's (1998:48) process model of L2 motivation

Dörnyei and Ottó's (1998) process model of L2 motivation (Figure 2) contains basically two dimensions, namely the *action sequence* and the *motivational influences*; and three tiers, namely the *preactional phase*, *actional phase* and *postactional phase*. Each tier is affected by its respective motivational influences. Within the *preactional phase*, there are three subprocesses, namely goal setting, intention formation and initiation of intention enactment. Desires, wishes and hopes and the availability of opportunities together form the precondition of goal setting. However, the goal itself does not necessarily generate any action; the antecedent of goal is intention, which involves commitment and the forming of an action plan. When an intention has been formed, once again it does not immediately lead to an action. Rather, the availability of the necessary means and resources as well as the existence of the 'start' condition are essential before an action will finally take place. The motivational influences in each phase of this process

model are a mixture of the essence of different motivational theories. Apart from constructing a process-oriented perspective of motivation, a second target of Dörnyei and Ottó is to "synthesise a number of different lines of research in a unified framework, thereby construing a non-reductionist, comprehensive model (Dörnyei, 2001:85)".

Dörnyei and Ottó (1998) used Heckhausen and Kuhl's (1985) metaphor of "*crossing the Rubicon*" to represent the transition from decision making to real implementation of action. It indicates that an individual has committed themselves to a course of action. In this *actional phase*, an individual has to go through the appraisal process when one evaluates the action concerned, generates and implements subtasks which may be accompanied with the main action as well as exert action control to prevent oneself from distractions and/or other temptations which may cause oneself to slow down or even terminate the action. Finally, the continued action will produce an actional outcome, which may or may not be the end of the phase. Dörnyei and Ottó believed that if the motivational foundation of the initial wish or desire is strong enough, the individual may continue his/her action with revised strategies or generate new subtasks of the action; or one may even go back to the *preactional phase*, revise his/her goal and form new intention.

When the action comes to an end, no matter whether the goal has been achieved or not, the individual will enter the *postactional phase*. Postactional evaluation will take place when one forms causal attributions about the extent to which the intended goal has been reached. An individual will also evaluate his/her standards and the effectiveness of his/her strategies to the action concerned. When new wishes and desires appear and a new intention is formed, the individual will go through a new cycle. During the whole process, motivational influences which energize the specific phase or subprocesses in the actional sequence operate simultaneously and impact on how the subprocesses would be carried out.

As explained in my Research Rationale in section 1.6, I believe that test preparation and test taking is a dynamic process in which motivation will fluctuate in the course of the process regardless of test stakes. In my IFS, I tried to find out if Dörnyei and Ottó's (1998) L2 process model can be applied in the explanation of one's L2 test motivation. I conducted a study with the

final year students in my Institution who took the Graduating Students Language Proficiency Assessment (GSLPA), a compulsory language exit test which they have to take in their final year; and IELTS, which was a government funded de facto exit test common for all final year students in Hong Kong. I received 581 questionnaire surveys and conducted focus group interviews with 8 students. My findings were that, exit tests can motivate them to work harder in their English learning process and in line with all other research; test-stakes affect the level of motivation. Dörnyei and Ottó's (1998) L2 process model can be used to explain students' test motivation but slight adaptations have to be made and the process models for compulsory and voluntary exit test are slightly different as well.

Figures 2.2 and 2.3 below are my suggested process models for compulsory and voluntary tests respectively. The differences between these two models and that of Dörnyei and Ottó's are that, the 'Action' phase is now renamed as 'Test Preparation' and there is an additional stage of 'Test Taking'. Two dotted lines after the *postactional phase* (in both compulsory and voluntary test models) are also added and the reason for doing this is that the experience of the test preparation and taking process will create motivational influences to a student's other L2 learning actions, either positive or negative, thus affecting the action sequence of the student's other L2 actions. The two process models I suggested are very similar except that in the compulsory test model, the process starts with the 'Intention formation stage' instead of the 'Goal setting stage' since students are compelled to take the test, they do not need to form "wishes and hopes" to sit for the test in the first place, nor do they require the 'desire' and 'opportunities'.

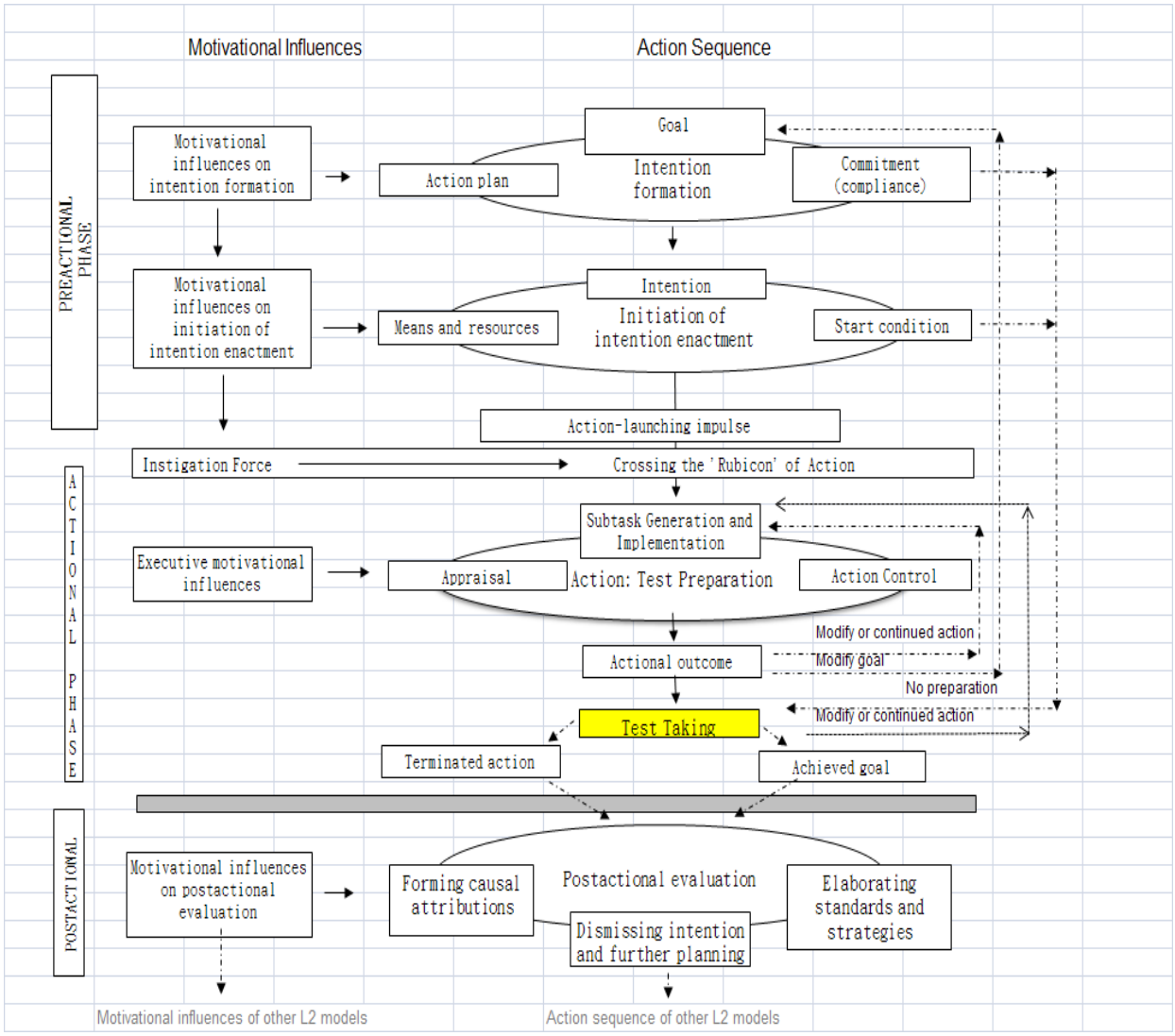


Figure 2.2 Test motivation process model (Compulsory test: GSLPA) (Tsang, 2011)

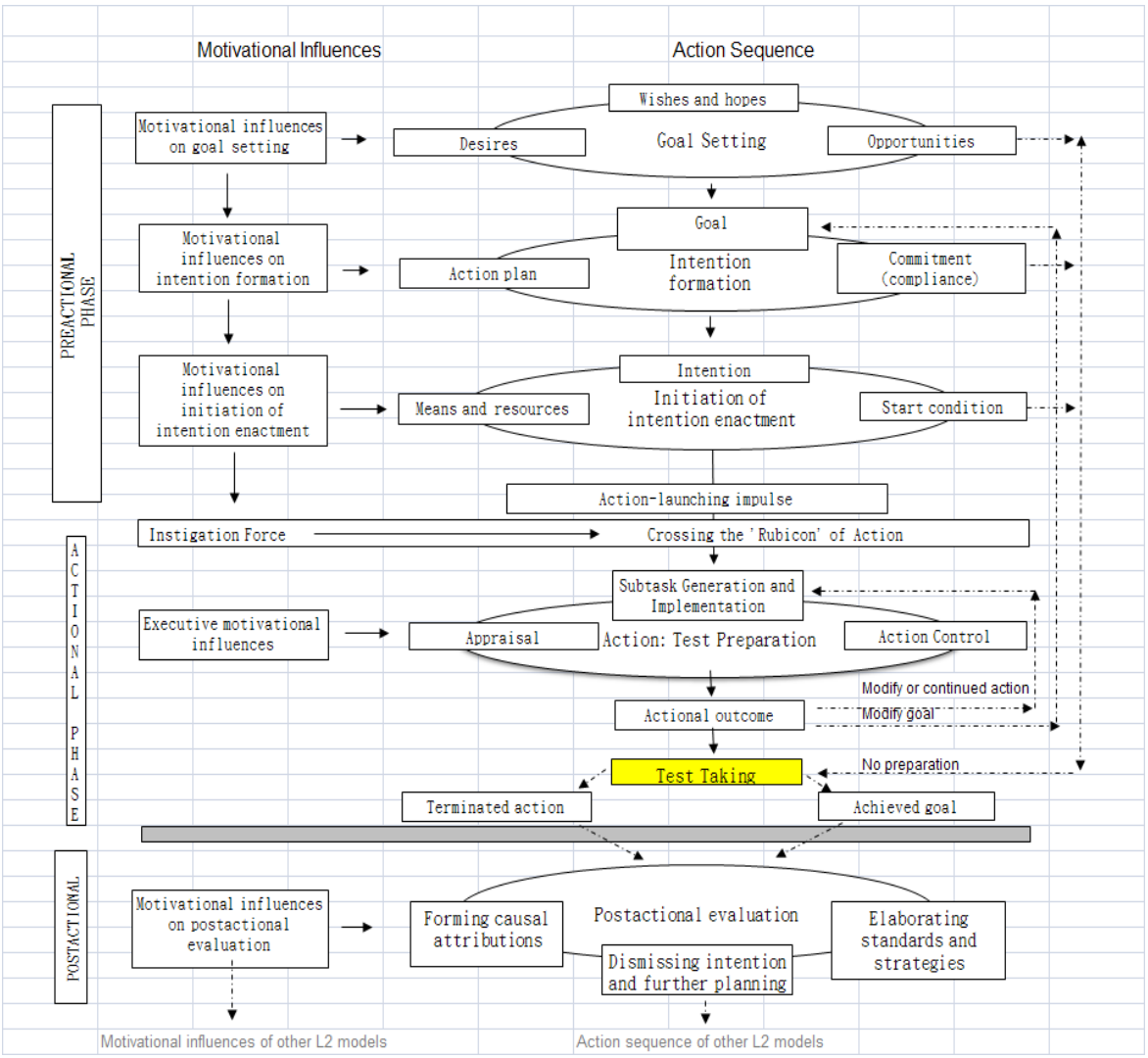


Figure 2.3 Test motivation process model (Voluntary test: IELTS) (Tsang, 2011)

In this current study, I use the same process model to explain the DELTA and to examine how the process model can also work on a formative diagnostic and tracking test.

2.3 Test Motivation

In order to demonstrate any knowledge that has been acquired, e.g. in an assessment environment, a certain amount of motivation and effort is needed (H. Eklöf, 2010). Test-taking motivation according to Baumert and Demmrich (2001) can be defined as 'the willingness to

engage in working on test items and to invest effort and persistence in this undertaking'. It has been a perpetually hot topic in the educational setting if tests/exams do any good in positively motivating students and in turn raise standards. Harlen and Crick (2003) presented a systematic review of test motivation research in education and concluded that on the one hand, testing does raise standards, while on the other hand, testing, in particular high stakes testing, can have a negative impact on motivation for learning that may adversely affect the preparation for lifelong learning. While it is debatable as to whether tests/exams do more harm than good or vice versa, there is no doubt that tests/exams do affect one's motivation.

There are two basic assumptions in test taking motivation studies. First, if two students of identical abilities are taking the same test, the student with higher level of motivation will perform better. Second, if the stakes of a test are low, students will generally value the test less and tend not to try their best in the test. Hence, when a student takes a test, their performance is a reflection of their ability plus how much effort they have put in, or in other words, how motivated they are. This is particularly true for low-stakes tests as the results of the tests have no direct consequence to the test takers. In order to ensure the validity of a test, that the test performance can truly reflect one's ability; motivation should be accounted for when considering one's test results.

In this section, I provide a literature review on how test-taking motivation can be measured in other studies; then an overview of studies in test motivation and test validity particularly in low-stakes tests. Whether a test is valid or not, from a popular quote from Lado (1961:321), depends on whether the test measures what it claims to measure. Messick (1989) provides the concept of validity an influential interpretation. He defines validity as "an overall evaluative judgment of the degree to which evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores" (1989:3). According to Messick (1998:37), content validity, criterion-related validity or consequential validity etc, are in actual fact one single entity, namely construct validity. All the other validity types contribute to the overall judgment of construct validity because what needs to be valid are the inferences made about score meaning, i.e. score interpretation and its actual implication for test use. When students are not motivated in doing a test and their test results are used in making decisions such as

comparisons of proficiencies among countries or schools, or in determining the usefulness of a certain programme etc, the issue of validity of test score interpretation and implication of test use comes into play.

2.3.1 Measuring test-taking motivation

In order to better explain the different studies on test motivation and test validity in low-stakes test (section 2.3.2), I will first look at how different scholars have attempted to use different instruments to measure test motivation. In terms of the potential impact on test validity and research value of test motivation, the development of test motivation measurement tools has a relatively short history. Researchers have adopted different ways in measuring test-taking motivation so as to increase the validity of the interpretation and the use of the results from tests. The three abovementioned major theoretical frameworks in motivation (the Achievement Goal Theory, Self-Determination Theory and Expectancy-Value Theories) are what most researchers based on when they attempt to explain motivation. Different instruments aiming at measuring the degree of motivation in tests have also been developed and tested based on these theoretical frameworks, however, these tools are mostly context specific and therefore so far none of them can said to be better than the others or being seen as 'the' instrument.

Barkoukis et al (2008) examined the validity of the **Academic Motivation Scale** developed by Vallerand et al. (1992; 1989) based on Deci and Ryan's (1985) **Self-Determination Theory**. They conducted two studies to examine the psychometric properties of the AMS with Greek high school students and the scale showed satisfactory levels of internal consistency and temporal stability which support the use of the Greek version of the AMS for the assessment of intrinsic motivation, extrinsic motivation and amotivation.

Young (2007) applies the **Multiple Goals Theory Measure (MGTM)** constructed based on Pintrich's (2000) **Achievement Goal Theory** to 257 undergraduates at a U.S. university and concludes that the MGTM is useful in predicting college grades.

Compared to the Self-Determination Theory and the Achievement Goal Theory, the **Expectancy-Value Theory** (Atkinson 1957; Wigfield and Eccles 2000) is the one which has often been used in understanding and explaining student test-taking motivation. Based on this theory, Sundre and Moore (2002) have developed the **Student Opinion Scale (SOS)** which is a self-report tool to measure the test-taking motivation of examinees. The SOS is a modification of the motivation scale that has 8 likert-type items developed by Wolf and Smith (1995) for their own research study. Sundre and Moore basically add two items and change some of the wordings to make it 10 items in a five-point likert scale. Five of the items constitute the Importance scale and the others form the Effort scale.

With reference again to the Expectancy-Value Theory, Sanchez, Truxillo and Bauer (2000) developed a 10-item multidimensional measure of test-taking motivation called the **Valence, Instrumentality, Expectancy Motivation scale (VIEMS)**. They conducted two studies to test the scale, the first being a post-test VIEMS survey with the applicants for an entry-level police officer written selection exam and the second study, a pretest and a posttest of VIEMS to another group of applicants sitting for a separate administration of the same test. Mixed results were revealed from the studies. In the first study, VIEMS was related to the actual test performance but in the second study, neither pretest nor posttest motivation reports were related to actual test performance. The authors suspect that it may be due to the disconnect experience between test performance and job obtainment. The applicants in study 2 were divided into three groups: group 1 had previously passed the test and was not hired, group 2 had previously taken the test but failed it and therefore was not hired, and group 3 had not taken the test before. Results show that those who had not taken the exam before reported higher instrumentality than those who did not pass the test. The authors also suggest that the discrepancies in research results deserve further attention in a wider range of selection settings for example by using selection methods other than written tests such as interviews.

Eklöf(2006), again, using the expectancy-value theory as the theoretical framework, developed a self-reporting instrument known as the **Test- Taking Motivation Questionnaire (TTMQ)** which measures students' test-taking motivation. A sample of 350 Swedish eighth-grade students who took part in the 2003 Trends in International Mathematics and Science Study (TIMSS) was

used. The TTMQ contains 24 items which were first generated pursuant to the expectancy-value theory and after a review of previous research on measurement of students' test-taking motivation. The study shows that task value perceptions are distinct from task performance expectancies and test-taking motivation is distinct from the general attitude toward a subject. Exploratory factor analysis was done to assess the internal structure and discriminant validity of the instrument and results suggest that most of the items in the TTMQ asking for value perceptions and perceived feelings of test-taking motivation could be used as a measure of student test-taking motivation.

2.3.2 Test motivation and test validity in low-stakes tests

The literatures in this section are grouped into three major categorizations. Firstly, I start with the studies on the effects of test stakes, how it contributes to test anxiety, students' test taking strategies as well as motivation and performance. Then I go on to discuss motivation in low-stakes test. Lastly, remedies suggested by some researchers regarding the possible validity issues brought by low-stakes test are also presented.

Wolf and Smith (1995) investigated the relationships of test consequence, motivation, anxiety and performance. The underlying argument of their research is that in a high-stakes test, students work harder but at the same time they may have more test anxiety which adversely affects their performance. 158 undergraduate students majoring in psychology participated in their research. Two experimental conditions were created. In the first condition, the students were told that the exam results would be counted as part of their course grade while in the second condition, their results would not be counted. Students were asked to fill in the Achievement Anxiety Test (AAT) by Alpert & Haber (1960) five days before the exam and a motivation questionnaire which consisted of eight likert-type items developed for this study immediately after the exam (these eight items were then adopted by Sundre and Moore (2002) and form the SOS). Results from the studies demonstrate that high motivation/high anxiety produced roughly the same levels of performance as low motivation/low anxiety. Surprisingly, many students actually performed

better in low-stakes condition (1/3 of them) meaning that anxiety does adversely influence performance at least in terms of the course exams in this study.

Baumert and Demmrich (2001) raised the concern that the variability of test compliance (the socio-cultural context in which a test or testing program is embedded, the test taker's familiarity with the assessment, and the positive or negative consequences associated with a test) and test motivation, pose a threat to the validity of the tests, especially when the test results are used to compare cross-country performances. According to the authors,

> "There is widespread concern that assessments which have no direct consequences for students, teachers or schools underestimate student ability, and that the extent of this underestimation increases as the students become ever more familiar with such tests." (Baumert and Demmrich, 2001:442)

They conducted studies with 467 9th grade students who participated in the OECD's Programme for International Student Assessment (PISA) and explored whether test motivation and performance can be improved when the stakes of the test are raised. They set up three separate experimental conditions to investigate the impact of informational feedback, grading and performance-contingent financial rewards on test performance. In the three experimental conditions and the control group, all students received the same instructions, except for the parts of the instruction that dealt with the purpose of the study and the consequences for students. Test and self-reported questionnaire (Online Motivation Questionnaire by Boekaerts (1987)) results were analyzed. Apart from the questionnaire, students had to imagine a situation that would be extremely important to them in which they would pay maximum effort. This maximum effort has a value of 10 on a 1 to 10 scale. Then they needed to rate how much effort they had paid in the test comparing with the maximum effort situation. The key finding of this study is that in low-stakes testing situations, financial rewards, grading of students and evaluation of teachers can lead to increase in test motivation and performance; while in the case of prominent national or international assessment, the incentive of the societal value of participating in it is in itself sufficient enough for the generation of test motivation. The motivation in the control group was proved to be the same as those participated in the experimental conditions: the personal value

and performance orientation allocated to the test by the prospect of individual feedback or grade and the perception of test utility as performance-contingent rewards resulted in the same level of motivation as the societal value of participating in an international assessment.

Sundre and Kitsantas (2004) explored the psychology of examinees by looking at the differences in their exam-taking strategies and motivation under high-stakes and low-stakes test conditions. 62 undergraduate students of a psychology course in the States were asked to take two parallel classroom tests consisting of multiple-choice questions and an essay. One of the tests counted towards the students' class grade (consequential) and the other did not (non-consequential). After the test, the students were asked to attend a structured interview with 15 questions on self-regulation strategies and complete the Student Opinion Scale (SOS). The research resulted in a number of findings. Firstly, test-taking motivation cannot predict high achievers' performance in high-stakes multiple choice testing conditions as this kind of testing context has high expectancy, value and motivation components and so motivation is equally high for all students; while in low-stakes multiple-choice test, self-regulation and motivation can predict test performance. Secondly, in high-stakes testing condition, self-regulated strategies can predict students' performance in essay writing; while motivation can significantly predict essay performance in low-stakes test.

A number of studies had been conducted in the last decade on motivation in low-stakes test. Eklöf (2007)used the Test-Taking Motivation Questionnaire (TTMQ) with a sample (n=343) of the Swedish grade 8 students who took the TIMSS 2003 (mathematics) (N=4,256) and regressed their self reported level of test-taking motivation with their test scores. The TIMSS is generally regarded as a low-stakes test for the test takers as the results of the test have no impact on students' grades and students will not receive individual results of their own. In this study, only the items in the TTMQ which measure test-taking motivation (known as the TTM) plus an open-ended question were used. The results from the TTM and the interviews indicated that students were motivated to do their best in the TIMSS 2003 although they understood that it was a low-stakes test for them. Using a regression analysis, the author found that test-taking motivation was not significantly related to mathematics achievement in this sample. Eklöf points out that her study results are in agreement with the study presented by Baumert and Demmrich (2003) who

also found that students were motivated on tests even if they are low-stakes to them. Eklöf explanation for this is that students may not necessarily think of tests such as the TIMSS and the PISA as low-stakes.

Cole, Bergin and Whittaker (2008) conclude after a detailed literature review that motivation is a significant predictor for exams with low stakes, whereas for high stakes exams, the average score on motivation is quite high with low standard deviation. The authors then, using the expectancy-value theories, measured the component of 'importance' within the 'value' in the expectancy-value theories and then used it to compile a total motivation scale. They did not take the expectancy of success into account in the study because they believed that test-takers in low stakes tests seldom discovered if they were successful in the tests or not as they either did not receive their scores at all or received them some months later. Besides, the authors argue that when individuals had high expectancy for success, i.e. they felt competent that they could complete a task, but their task value was low, they might still choose not to participate. The authors proposed that based on expectancy-value theory, the impact of the three components which make up the 'task value', i.e. 'interest', 'usefulness' and 'importance' would significantly predict test-taking effort and effort would significantly predict test performance. The participants of the study were 1005 undergraduate students who completed a standardized general education exam (CBASE) with four subtests (English, Mathematics, Science and Social Studies) which carried no consequences to them. First, 'Effort' was measured by asking students to rate their perceived effort on each subtest. Then, the 'Interest' and 'Usefulness' in 'task value' came from the items in the Motivated Strategies learning Questionnaire (MSLQ) created by Pintrich, Smith, Garcia and McKeachie (1991); while the 'importance' scale in the 'task value' was adapted from the Student Opinion Survey (SOS) modified by Sundre and Moore (2002). The studies conclude that 'interest' in taking the test was significant and negatively related to 'effort' for English and Mathematics but unrelated to Science and Social studies and that 'importance' was acting as a restraining variable on 'interest' for English. In general, students' perceived 'usefulness' and 'importance' for taking a low stakes test are important predictors of test-taking effort; 'effort' is a strong predictor of performance. The results of this study fit well into the expectancy-value theories.

Eklöf (2010) conducted another study with a Swedish sample (n=163) participating in the TIMSS advanced 2008 field study. A total of 9 items, adapted from the items used in the 2003 study as well as the Student Opinion Scale by Sundre and her colleagues (Sundre & Finney, 2002; Sundre & Moore, 2002) was administered to this new group and the results from the 2 studies (the 2003 and the 2008 group) differed. The results in the 2003 study were not biased by low motivation among students while for the 2008 group, 58% did not agree that they felt motivated to do their best on the test. Eklöf suggested two possible reasons for this difference: i) if the TIMSS main study (2003 data) is a low-stakes test, then the TIMSS field study (2008 data) is basically a no-stakes test to the students, as the purpose of the field study is only to try out and evaluate items but not to evaluate student proficiency; ii) students who participated in the 2003 study were in grade 8 while students in the 2008 students were in grade 12 and according to Eklöf, research has shown that with increasing age, years of schooling and experience, students are more aware of how their effort should be invested.

Due to the possible validity issues in low-stakes tests, some researchers provide suggestions and remedies as to how to raise the stakes of the test or how to increase students' motivation even in tests which carry no consequences to them. Wise and DeMars questioned the validity of 'low stakes' assessments - whether the results of these assessments can really reflect the real picture of students' abilities and performances. The authors presented an extensive review of previous research which all shows that low student motivation will result in a substantial decrease in test performance. Therefore, "to the extent that some students do not give their best effort, assessment results will underestimate what a group of students know and can do, and the validity of the entire assessment program will be threatened" (2005:3). In order to deal with such a threat, the authors put forward seven recommendations to avoid any possible challenge to the validity of the assessments results: (a) raise the stakes of the test by imposing consequences for performance, (b) provide incentives by for example offering financial reward, (c) make tests not too mentally taxing for example using multiple choice items instead of essay writing, (d) make tests more intrinsically motivating, (e) data analyses to reduce the effects of lower motivation, (f) statistical adjustments and (g) motivation filtering.

Barry and Finney (2009) question the common practice that data collected in low-stakes tests are always used to examine the psychometric functioning of the test items. If behaviours like rapid-guessing are accounted for as well, items that have low item difficulties will appear more difficult and more discriminating as there will be higher chances that students will get these easy items wrong (more about rapid-guessing to be followed). The authors look at the impact of context on examinees' test-taking behaviour when tests are of no-stakes to them. Five samples of data were collected with different testing conditions from students completing online version of the test at home without time limit (n=3562) to students doing the test in a classroom in the presence of an invigilator (n=237), using the College Self-Efficacy Inventory (Solberg, O'Brien, Villareal, Kennel and Davis, 1993) at a mid- Atlantic university. The findings of the study are that a testing condition with a higher degree of control, in terms of time and invigilation, tends to increase students' motivation even when the test is of low-stakes in itself. Thus, the authors suggested that one possible way to alleviate the problem of validity in low-stakes tests is to increase the level of control in the testing condition.

In order to account for the possible rapid-guessing behaviour in a test, Wise and Kong (2005) developed an index called Response Time Effort (RTE) for measuring an examinee's overall test taking effort. The RTE operates on the assumption that in each encounter, an examinee makes a choice to engage in either solution or rapid-guessing behaviour which can be determined by the time the examinee takes to respond to an item. For example in each item, there is a specific threshold that represents the response time boundary between the rapid-guessing behaviour and solution behaviour. The sample of this study was 306 freshmen who took a computer-based test as the university general education assessment. Wise and Kong set the threshold for each item of the assessment based on item length and whether or not an item used a figure, illustration or other supplementary reading materials. Two different examinee behaviours were found with occurrences of rapid-guessing behaviour yielding item responses whose accuracy did not exceed chance levels. The authors suggest applying the RTE as motivation filtering for tests of similar kind in order to enhance validity.

Wise and DeMars (2006) examined the psychometric characteristics of the effort-moderated model which incorporates the RTE developed by Wise and Kong (2005), relative to those of a

standard IRT model (for more of IRT, please see section 2.4.3). The sample in this study was 524 mid-year sophomores who took a computer-based low stakes assessment in a university. The assessment was a 60-item version of the Information Literacy Test (ILT) which is a locally developed test to assess students' information literacy knowledge and skills. Response time was collected for each examinee-item encounter. The results of this study show that in the presence of rapid-guessing behaviour, the effort-moderated model is less biased and can provide more valid proficiency estimates than the standard IRT model. Using the standard model, the items appeared more difficult and more discriminating than they actually were. The authors recommend that low-stakes tests should take into account item response time, either merely monitor the effort examinees give toward their tests (i.e., using RTE scores), remove the data from examinees who exhibited low effort (i.e., motivation filtering), or to incorporate response time into proficiency estimation using an effort-moderated IRT model as in their study.

As the research surveyed above shows, studies on test-taking motivation to date have been mostly on developing and validating motivation measurement scales and confirming that motivation suffers in most low-stakes tests. My current study is to look into students' motivation pattern in a new type of low-stakes test. i.e. online diagnostic test, which is a fairly new product in the testing field.

2.4 Diagnostic assessment

Diagnostic testing is a much neglected area within the general field of language testing (Alderson, 2007). This is evident from the limited number of diagnostic tests available in the market. There is a need to conduct more research to the construction and application of diagnostic tests, and the usefulness and improvement on diagnostic reports especially in view of the huge potential benefits it can bring about to students in comparatively lower administration costs when it is made online.

2.4.1 What is diagnostic assessment

There has been a divergence of opinion as to what is a diagnostic test. According to Bachman (1990), virtually any language test has some potential for providing diagnostic information. Alderson and Clapham (1995:12) share similar thoughts and define diagnostic tests as tests which seek to identify areas in which a student needs further help. The authors believe that diagnostic tests can be fairly general, and show, for example whether a student needs particular help with one of the four main language skills; or they can be more specific, seeking perhaps to identify weaknesses in a student's use of grammar. According to Moussavi (2002), if diagnostic testing is defined as providing feedback to teachers and students regarding their strengths and weaknesses, then almost any test would be diagnostic. Bachman and Palmer (1996) provide a more specific definition of what is considered as diagnosis: Diagnosis involves identifying specific areas of strength or weakness in language ability so as to assign students to specific courses or learning activities. Alderson (2005) , which is still to date the only book dedicated to discussing second language diagnostic tests, provides a set of hypothetical features of diagnosis (p.11). My research target, the DELTA, satisfies points 1 to 8 (see below). Since the DELTA is an English proficiency test, the content of it would not be covered particularly in instruction; however, the content of the assessment is general and academic based, so students are expected to have knowledge upon.

1. Diagnostic tests are designed to identify strengths and weaknesses in a learner's knowledge and use of language.
2. Diagnostic tests are more likely to focus on weaknesses than on strengths.
3. Diagnostic tests should lead to remediation in further instruction.
4. Diagnostic tests should enable a detailed analysis and report of responses to items or tasks.
5. Diagnostic tests thus give detailed feedback which can be acted upon.
6. Diagnostic tests provide immediate results, or results as little delayed as possible after test-taking.
7. Diagnostic tests are typically low-stakes or no-stakes.
8. Because diagnostic tests are not high-stakes they can be expected to involve little anxiety or other affective barriers to optimum performance.

9. Diagnostic tests are based on content which has been covered in instruction, or which will be covered shortly.

One of the most established and well recognized diagnostic tests is the DIALANG (http://www.lancs.ac.uk/researchenterprise/dialang/about) (Alderson, 2000, 2007; Huhta et al., 2002; Zhang & Thompson, 2004). DIALANG stands for diagnostic and language test and is a suite of computer-based diagnostic tests in 14 European languages. The test is based upon the Common European Framework of Reference for Languages (CEFR) (http://www.coe.int/t/dg4/education/elp/elp-reg/cefr_EN.asp) and has components in reading, listening, writing, grammar and vocabulary. Anyone can take the test online anytime and will get a report in no time right after the test. The report, without any score, provides test takers with a CEFR level (A1 to C2) and diagnoses test takers' ability within each sub-skill. Test takers can also check their original responses to see why they have got something correct or not. The test also allows test takers to assess their own abilities in terms of the CEFR and explanations are provided as to why there might be a mismatch between self assessment and test results, and advice is also given on how they can improve if they want to progress to the next CEFR level.

Some other well established diagnostic tests apart from the DIALANG are the Diagnostic English Language Assessment (DELA) (http://services.unimelb.edu.au/academicskills/services/dela) in Australia and the Diagnostic English Language Needs Assessment (DELNA) (http://www.delna.auckland.ac.nz/uoa/) in New Zealand. The University of Melbourne developed the DELA with an aim to help students identify academic language skills that they may need further development in order to do well in their studies. The test lasts 2 hours and has reading, writing and listening components. All first-year undergraduates coming in without reaching a certain English level are requested to do the test, for e.g. those with IELTS score 7 or below. Students will receive a diagnostic report shortly after the test and if they have problems understanding their report or want academic advice, they can then go to the student centre or contact the DELA representative of each respective faculty. DELNA is the diagnostic test in the University of Auckland. It is in fact divided into two phases: the DELNA screening plus the DELA (which is called the DELNA diagnosis in University of Auckland). All first-year students are required to do the 30-minute computer based DELNA

screening regardless of their language qualifications. If their results are unsatisfactory, they will be asked to do the 2-hour pen and paper DELA and will be given a report of their language profile. Students will need to discuss their results with language advisers who would give them suggestions on how they can improve their English in their years of university study.

2.4.2 DELTA as a Formative Test

Formative assessment (FA) refers to assessment that is specifically intended to generate feedback on performance to improve and accelerate learning while summative assessment (SA) is a judgment which encapsulates all the evidence up to a given point (Sadler, 1998). IELTS is an international benchmark exam which aims at assessing the language ability of students who need to study at undergraduate or postgraduate level in the medium of English. Being a proficiency exam, the Hong Kong government through CEPAS has been using it as the de facto summative assessment, an exit test, common to all final year undergraduates. The effectiveness of whether IELTS is a suitable test for this purpose and whether a summative test is useful in enhancing students' language proficiency has aroused questions and debates (Qian, 2007).

Black and Wiliam (1998) carried out a substantial review of FA and concluded that it is effective in promoting student learning across a wide range of education setting (disciplinary areas, types of outcomes, levels). FA requires feedback which indicates the existence of a 'gap' between the actual level of the work being assessed and the required standard. It also requires an indication of how the work can be improved to reach the required standard (Taras, 2005). Although diagnostic and formative assessment have different roots, the meanings of the two terms begin to approach each other (Huhta, 2008). FA is critically important for student learning. Without formative feedback on what they do, students will have relatively little by which to chart their development (Yorke, 2003). Another central argument is that, in higher education, formative assessment and feedback should be used to empower students as self-regulated learners. Different to the IELTS, the DELTA is a formative test that provides diagnostic reports to students as well as tracking them throughout their tertiary studies.

The DELTA consists of individual multiple-choice tests of listening, vocabulary, reading and grammar (writing and speaking components are under development) and lasts 1 hour and 30 minutes. It is expected that students will take the DELTA once every year and will be provided with an individual report of their performance every time they take it. After the analysis of students' performances adopting methods of Rasch measurement and the Winsteps software package (Linacre & Wright, 2000), results are disseminated to them in the form of a diagnostic score report. Students can view their report online or download it as a PDF file. The report shows the track of their overall proficiency level each year, which is known as the DELTA measure (Figure 2.4); a component skills profile (Figure 2.5), which shows their performance in each skill relative to their overall proficiency; and component diagnostic reports (Figure 2.6), which show breakdowns of the questions in each of the four skills that the students did in descending order of difficulty. A sample of a full DELTA Report is in Appendix 1.

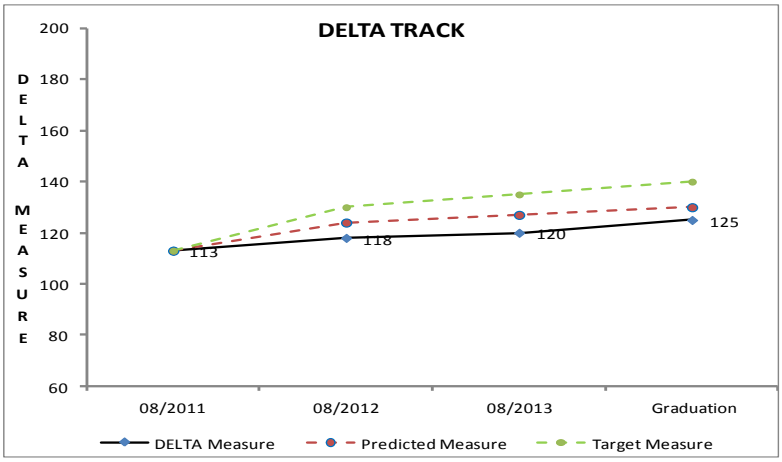


Figure 2.4 Sample (simulated) DELTA Track

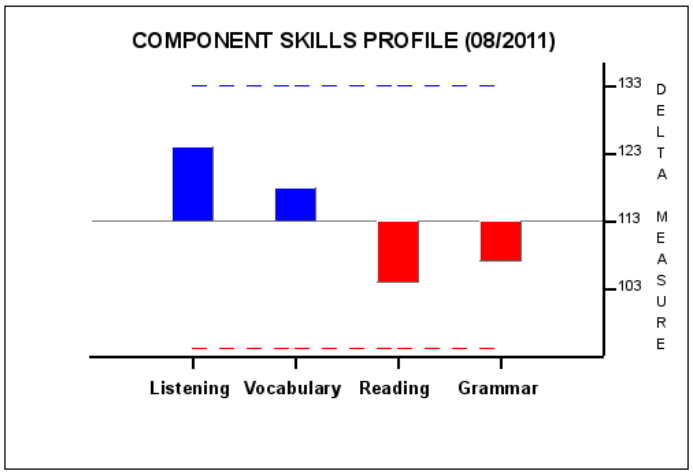


Figure 2.5 Sample DELTA Component Skills Profile

		Subskills tested	Text type	Theme
More	✗	Identifying specific information	News article	Animal ethics
	✗	Understanding main ideas and supporting ideas	Academic article	History and culture
	✗	Understanding information and making an inference	News article	Environmental issue
	✗	Interpreting an attitude or intention of the writer	Fiction	History and culture
	✗	Understanding main ideas and supporting ideas	Academic article	History and culture
	✗	Interpreting a word or phrase as used by the writer	Academic article	History and culture
	✓	Identifying specific information	News article	Environmental issue
	✗	Identifying specific information	News article	Animal ethics
	✗	Understanding main ideas and supporting ideas	News article	Animal ethics
	✗	Understanding main ideas and supporting ideas	Academic article	History and culture
	✗	Interpreting a word or phrase as used by the writer	News article	Animal ethics
				DELTA 113
Difficulty	✗	Identifying specific information	Fiction	History and culture
	✓	Understanding main ideas and supporting ideas	News article	Animal ethics
	✓	Understanding information and making an inference	Academic article	History and culture
	✓	Identifying specific information	News article	Environmental issue
	✓	Understanding information and making an inference	Academic article	History and culture
	✗	Identifying specific information	Academic article	History and culture
	✓	Identifying specific information	News article	Environmental issue
	✓	Interpreting a word or phrase as used by the writer	Academic article	History and culture
	✗	Identifying specific information	Fiction	History and culture
	✓	Interpret a word or phrase as used by the writer	News article	Environmental issue
Less	✗	Inferring the writer's reasoning	Fiction	History and culture
	✓	Understanding main ideas and supporting ideas	Fiction	History and culture
	✓	identifying specific information	News article	Environmental issue
	✓	Understanding information and making an inference	Fiction	History and culture
	✓	Interpreting a word or phrase as used by the writer	News article	Environmental issue
	✓	Identifying specific information	News article	Environmental issue

Figure 2.6 Sample DELTA Component Diagnostic Report on Reading

The DELTA track (Figure 2.4) plots a student's DELTA measure at the time they took the DELTA as well as their predicted DELTA measure the next time they take it. The DELTA

measure is a scaled score corresponding to the overall English language proficiency. The predicted measure is calculated based on available information in the system, i.e. the student's previous performance and other students' (with similar characteristics including stage of degree, discipline studied, etc) performance. The student will also need to input their target DELTA measure before they enrol for the test each time and the default setting is their current DELTA measure +1.

The Component Skills Profile (Figure 2.5) shows the contributions that the score on each component makes to the DELTA Measure. This gives an indication of a student's relative strength and weakness in the four components.

In the Component Diagnostic Report (Figure 2.6), items (questions) above their level of proficiency (in this example, the student has a DELTA measure of 113 and the items above the line in Figure 2.6 shows the items above their ability level) are those that they would not be expected to get correct; items below their level are those that they should be getting correct and those they didn't are highlighted as areas in need of attention. After receiving the reports, students are then advised to make use of the report to help with their English enhancement and teachers are briefed on how to help students understand the report and advise them on suitable action to take. The whole system has been available online since mid-2011. Whenever a student takes the Assessment (after the first time), based on their performance in their previous attempt, the system will generate for them an assessment that is targeted to their proficiency level, thus enabling them to have a useful testing experience.

In 2012, the UGC provided funding for the further development of DELTA as well as the trialing of the test in other institutions in Hong Kong. It is expected that the DELTA will be gaining recognition and importance as more tertiary institutions in Hong Kong get to trial the test.

2.4.3 RASCH measurement and the DELTA

Classical Test Theory (CTT) and Item Response Theory (IRT) are two very different measurement frameworks. The CTT provides information on item difficulty in terms of the proportion of candidates getting an item correct. In CTT, item discrimination is based on the assumption that a difficult item should be answered correctly by more capable candidates; and person ability in terms of the total scores on the test. These values: item difficulty, item discrimination and person ability, are dependent on the difficulty of the test items as well as the ability of the group of candidates that is tested (McNamara, 1996). If the percentage of candidates passing or achieving given grades is kept constant, then individual students' performance will become dependent on the ability level of the whole group of test candidates. This implies that in Classical Test Theory, we need to take into account the relative ability of the cohort in which the students are tested in order to know the true picture of their ability.

DELTA test scores are analysed by a software package called Winsteps. Winsteps is a Windows-based software which enables one to do Rasch analysis, a method for obtaining objective, fundamental, linear measures (qualified by standard errors and quality-control fit statistics) (Linacre & Wright, 2000) under the Item Response Theory (IRT). Unlike the Classical Test Theory, which is sample dependent, the IRT models produce item statistics independent of examinee samples and person statistics independent of the particular set of items administered (Fan, 1998).

The IRT attempts to "model statistically patterns in data from performances by candidates on dichotomously scored test items in order to draw conclusions about the underlying difficulty of items and the underlying ability of candidates" (McNamara, 1996:258). DELTA analysis follows the IRT one-parameter (1PL) or Rasch model (Rasch, 1960) and is concerned with defining a single dimension on which to measure two variables: candidate ability and item difficulty.

The characteristics and difficulties of test items are very important in estimating candidates' ability. The degree of difficulty of a test is important in comprehending the meaning of one's achievement. A middling score on a difficult test is more impressive than a high score on an easy

one (McNamara, 1996). In The Rasch 1PL model, candidates' ability is related to item difficulty: Rasch measurement enables estimates of candidates' underlying ability to be made by analyzing the candidate's performance on a set of items, after allowance has been made for the difficulty of the items and how they were matched to the candidate's ability level. Similarly, the underlying difficulty of items can be estimated from the responses of a set of candidates, by taking into account the ability of the candidates and the degree to which there was a match between the ability of the whole group of candidates and the difficulty of the items. (McNamara, 1996)

In essence, the Rasch model proposes that the more able the candidate, the higher his or her chance should be of getting the answer correct. If an item is easy and the candidate's ability is high, there will be a high likelihood of a correct response. On the other hand, if the item is difficult, and the candidate's ability is low, it is unlikely that he or she will make a correct response. Naturally, the Rasch model is a theoretical model of probability, it is always possible for the actual observed response to be not as predicted and therefore there is always an estimated error associated with their scores. 'Fit' statistics provided by the Rasch model give information on whether the test in question forms a good basis for estimating the observed scores for each person in all items. If an item is 'overfitting', it means all the students having an ability level higher than the item difficulty get the item correct. It is 'overfit' because the outcome is too good to be true. If an item is 'misfit', it means that a proportion of the students having an ability level higher than the item difficulty are not getting the item correct. By checking the fit statistics, item diagnosis can be performed. The DELTA team reviews 'misfitting/overfitting' items to determine whether they need to be revised or deleted from the item bank.

Item difficulty and person ability are jointly estimated in Rasch analysis. The estimates of person ability are referred to as 'person measures', and are placed on an objective measurement scale constructed in the analysis. Both item and person measures are plotted on the same scale. The unit of measurement in which the measures are expressed is the logit. By convention, the average difficulty of items in a test is set at zero logits (McNamara, 1996). The Rasch model allows us to see whether a test is too easy or too difficult for a group of test takers.

The two-parameter IRT model (2PL) and the three-parameter IRT model (3PL) include the additional item parameters of item discrimination and item discrimination plus guessing factor respectively. The justification of using the 1PL model is that the 2PL or the 3PL models are more 'expensive' (as it involves more data and therefore more time consuming) to run while they show no benefits over the 1PL (Wright, 1995). Wright (1995) compared the results of the analysis of the 1992 National Adult Literacy Study using the 1PL and the 3PL models. Correlations of the results from the two models were no less than .92, and principal component analysis showed that the first factor absorbed 88% of the total variance and the second factor less than 3%. The 3PL infers the presence of guessing behaviour in items while Rasch reviews person outfit to detect lucky guessing. Wright (1995) believes that the third parameter, guessing behavior, penalizes all respondents, especially the low performers, who indeed know the answer. A preferred strategy is to remove the misfitting person or item from the data analysis.

McNamara provides an effective summary of the Rasch approach (McNamara, 1996):

1. It estimates ability by considering data from an individual in the context of data from the whole data matrix, that is, the responses of all candidates to all items;
2. It relates person ability and item difficulty by estimating the likelihood of responses of particular persons to particular items;
3. The difficulty of an item and the ability of a person to answer that item correctly are mapped onto the same probability scale such that a person with an ability level of X has a 50% probability of answering correctly an item of difficulty X.

Calibration of items includes investigation of the fit statistics (if an item is 'misfit' or 'overfit') of each item. Winsteps bubble chart (Figure 2.7) illustrates the degree of fit against item measure.

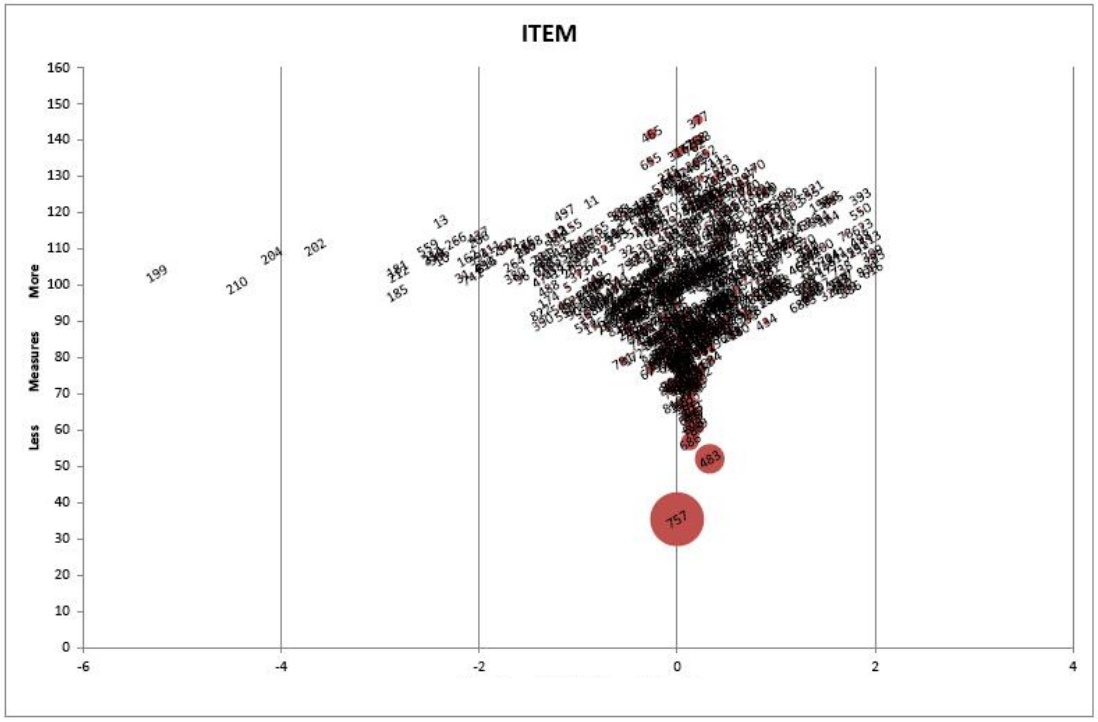


Figure 2.7 Sample DELTA bubble chart

Each circle in the diagram represents a test item. The x-axis shows the Infit Zstd (Zstd stands for standard deviation) and the y-axis shows the DELTA measure of the items. Items with an Infit Zstd of more than 2 are considered to be 'misfitting' while items with an Infit Zstd of less than -2 are considered to be 'overfitting'. Misfitting items are eliminated (as in Figure 2.7) while items which 'overfit' normally would be kept as they would not do much harm. The analysis is re-run until the items satisfactorily fit the Rasch model and result reports can be generated.

Item diagnosis is also performed as part of the item calibration process. Each component of the test is divided into testlets and items within testlets that do not fit the Rasch model are highlighted for further inspection. These items are reviewed to investigate their quality (i.e. revision of items or deletion from the item bank). Items that do not fit the Rasch model are excluded from analysis to prevent inaccurate and imprecise measurement of student abilities.

2.4.4 Usefulness of feedback

The quality and usefulness of feedback to students is an under-researched area regardless of its importance in the learning process. It is common for teachers to complain that students are only concerned about their grades without paying much attention to the feedback and comments that were given. Some researchers believe that students are "instrumental consumers of education, driven solely by the extrinsic motivation of the mark and as such desire feedback which simply provides them with 'correct answers'" (Higgins, Hartley & Skelton, 2002:53). Carless (2006) studied the differences in perception in the feedback process and conducted a large-scale questionnaire survey with students and teachers across the eight universities in Hong Kong. His findings show that the views of teachers differ from those of the students: while teachers believe that the feedback they give is useful, some students (if not all), did not bother to collect marked assignments. Conversely, students think that the lack of useful feedback is the key problem in an assessment process.

Despite the important role students play in the feedback process, much of the feedback research has put teachers at centre stage, focusing on the strategies teachers use in giving feedback and their stances and perspectives etc (Ferris, 1997; Hyland & Hyland, 2001). It is therefore important to investigate in more detail; how students actually perceive the feedback they receive so as to find out why it seems that they usually would not pay much attention to them; and how students respond to feedback which might help in explaining why some believe that feedback could not improve learning.

Hartley and Chesworth (2000) administered a questionnaire to 102 second-year psychology students at Keele University, UK in order to understand students' perception of feedback. They found that very often students have difficulty in understanding and interpreting the feedback and that feedback is very often given too late to be helpful. Weaver (2006) reported on his research which focuses on students' attitudes, beliefs and perceptions in relation to the written feedback received by tutors. In his study, he asked 44 students in the faculties of Business and Art & Design at Nottingham Trent University UK to indicate how confident they were in understanding the feedback given. The second section of his survey was about students' perceptions on the use

and helpfulness of feedback and the third section is about the perceived effectiveness of feedback. The results of the survey revealed that although a large majority of students believed that the feedback they received was too brief or general to be helpful, they admitted that feedback encouraged and motivated them to improve. Poulos and Mahony (2008) conducted a research on students' perceived effectiveness of feedback. They analyzed their data in three dimensions: perceptions of feedback, impact of feedback and credibility of feedback. Their study reviews that students do not hold a homogenous view of what effective feedback is and how it could be used. From the above studies we can see that there is not an overall pattern of how students view the feedback that they received.

Apart from finding out students' perceived usefulness of feedback, some researchers go a step further and look into what makes effective feedback. Brown and Glover (2006) looked at formative written feedback for science students in two UK universities based on Gibbs and Simpson's (2004) criteria of effective feedback:

- Is frequent, timely, sufficient and detailed enough;
- Can be linked to the purpose of the assessment task and criteria;
- Is understandable, given the students' level of sophistication; and
- Focuses on learning rather than marks by relating explicitly to future work and tasks.

They interviewed 6 module leaders and 13 students and came up with the conclusion that the 2 parties held different perceptions of feedback effectiveness and the above 4 criteria of effective feedback were not met as frequently as originally believed.

Lizzio and Wilson (2008) investigated students' perceptions of written assignment feedback. 57 students reflected on the feedback they had received on a range of assessment tasks and described aspects of helpful and unhelpful assessor comments; while 277 students returned a questionnaire to evaluate assessment feedback. They concluded that students' perceptions of assessment feedback can be meaningfully understood in terms of three dimensions: developmental, encouraging and fair feedback. The authors had also developed a Feedback Effectiveness Scale which is a seven point scale on 3 items: 1. how effective has the written feedback provided been in facilitating learning; 2. how effective has the written feedback been in facilitating competence as a learner and 3. how effective has the written feedback been in

facilitating confidence as a learner. This three-item scale demonstrated high internal consistency to be employed as a uni-dimensional measure of perceived effectiveness of assessment feedback.

From the literature review above, we can see that most of the current studies in this area are mainly on the effectiveness of teachers' written feedback on students' assignment and the assignment/assessment may not be necessarily diagnostic. Students were perceived to be passive learners. In the recent decades, more emphasis has been put on the promotion of formative assessment and self-regulated learning which is believed to be more beneficial to learners. Nicol and Macfarlane-Dick (2006) point out that in higher education, formative assessment and feedback should be used to empower students as self-regulated learners. The authors developed a model which is different from the usual understanding of feedback in that students are assumed to occupy a central and active role in all feedback processes and they are actively involved in monitoring their own performance and construct their own understanding of feedback messages derived from external sources. Based on the research literature on formative assessment and the authors' proposed self-regulation model, they come up with 7 principles of good feedback practice that they believe if implemented, would contribute to the development of self-regulation in learning (Nicol & Macfarlane-Dick, 2006:205):

Good feedback:
1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape teaching.

We can see that the 7 principles can basically be grouped under two categories: providing information and promoting further action. These 7 principles provide a good guideline when teachers or tests administrators try to provide feedback to students. However, there is not to date

a scale which we can use with students if we want to measure their perceived feedback usefulness especially on diagnostic feedback.

2.5 Chapter Summary

This chapter outlines the literature on the theories and studies relating to the understanding of motivation in low stakes diagnostic tests. I started the chapter by first looking at the most dominant and frequently used psychology theories in explaining motivation, i.e. the Achievement Goal Theory, Self-Determination Theory and the Expectancy-Value Theories. I then examined the application of such theories in L2 motivation and introduced the process model of L2 motivation proposed by Dörnyei and Ottó in 1998. Literature on test motivation, the issue of test validity in low-stakes test and the measuring of test motivation are also presented in order to create a complete background for my current study. Lastly, the history of diagnostic assessment and usefulness of feedback are also being discussed. After reviewing the relevant literature in the area, I am convinced that my study which is mainly about the measuring of test motivation in low-stakes online diagnostic tests (pre test, during test and post test motivation) and the perceived usefulness of diagnostic feedback is an area which is well under researched and that my work can fill the gap in the current knowledge in 5 ways: 1) there is no diagnostic test such as the DELTA which is developed in the Asian region, and I believe that it is a valuable research topic because of the huge development potential due to the comparatively low administration costs and high educational value in terms of the diagnostic feedback and tracking function it provides; 2) there are studies on the measuring of motivation in low stakes tests but very few that I am aware of is about motivation in a diagnostic test as well as the usefulness of a diagnostic report; 3) current studies on the measuring of motivation in low stakes tests is to measure students' motivation whilst sitting the test, the pre test and post test motivation is an area being neglected; 4) current research on perceived feedback usefulness is mostly about teachers' written feedback on students' writing, again, nothing that I am aware of is on students' perceived usefulness of diagnostic report.; and 5) although people are aware of the dynamic nature of motivation, none of the existing studies has tried to explain a students' test taking

motivation using a process model. In the next chapter, I will present the methodology of my study.

Chapter 3 Methodology

3.1 Research questions

The previous chapter set out the background and current development of the literature on the related theories and studies in the understanding of motivation in low stakes diagnostic test. My observation is that the topic concerned is clearly under researched. While there are abundant research on each individual area such as L2 motivation, measurement of motivation, usefulness of feedback etc; none of the studies seem to have linked them together. This study aims at finding out the test motivation in a low stakes diagnostic English test, as well as the pre and post test motivation generated by the test, and how useful students find a diagnostic report is to their language learning so as to explore students' perceptions of test stakes and test value. In addition, the L2 process model by Dörnyei and Ottó (1998; Dörnyei, 2000, 2002, 2003, 2006) is adapted to explain students' test taking motivation in this context. The research questions that guide my study are:

Are students motivated in the [dynamic] DELTA taking process?

i. Are students motivated to perform to the best of their ability whilst sitting the DELTA?

ii. Do they find the DELTA report useful?

iii. Are students motivated before and after taking the DELTA?

iv. How can Dörnyei and Ottó's (1998) process model of L2 motivation be applied in explaining the process of test motivation in a low stakes diagnostic test?

In my Institution Focused Study, I adapted Dörnyei's process model and came up with two slightly different process models of test motivation of exit test (summative assessment), one for compulsory and the other for voluntary test (Tsang, 2011). I substantiated my arguments through a detailed study (by means of questionnaire survey and focus group interviews) of an institution-specific exit test and IELTS, as well as the student's approaches in handling the two tests. In this study, I attempt to apply the same process model in explaining the test motivation of a diagnostic test (formative assessment) and see if any adaptation is needed. The role of diagnostic tests in students' L2 learning will also be explored.

3.2 Methods of data collection and analysis

Wise and DeMars (2006) list a number of methods of measuring examinees' effort and discuss the pros and cons of each method. The most commonly used method is to administer post-test self-report scale questionnaires but the drawback of this method is that the scales can be biased as the validity of responses is reliant on the truthfulness of the examinees. Person-fit statistics is another way of revealing effort that is not difficult to manage. Most person-fit statistics compare a person's observed and expected item scores across test items while the expected item scores are determined on the basis of an Item Response Theory (IRT) model (Meijer & Sijtsma, 2001). The underlying assumption of IRT is that test takers with higher values of attribute measured by the test should have higher probabilities of positive responses than test takers with lower values (Drasgow & Hulin, 1990). Investigation of person fit may help to obtain information about the answering behaviour of a person. However, the drawback is that it is difficult to conclude whether a misfit in a person's performance is purely caused by a lack of motivation. For example, a person may produce an answering pattern that is unlikely to happen under the IRT model because he has preknowledge of the correct answers to some of the most difficult items in the test (Glas & Meijer, 2003). They might have come across the difficult questions in another occasion and therefore knew the answers simply because of the 'accidental' preknowledge. As discussed in the literature review (Section 2.3.1), Wise and DeMars (2006) suggest a third method, Response Time Effort (RTE), introduced by Wise and Kong (2005). The RTE is based on the assumption that when students take a computer-based test they will either engage in

solution behavior or rapid-guessing behavior. When students are not motivated, they simply scroll down the screen and rapidly and randomly click the responses by guessing. Therefore by monitoring their response time, we can deduce whether a student is paying attention/devoting effort when answering the questions. Ranging from 0 to 1, the higher the RTE the more the students engaged in solution behavior.

The RTE can be a useful method of detecting examinees' behaviour in a computer-based test. However, it cannot be applied to the DELTA. There are four sections in the DELTA, one on each subskill (listening, reading, grammar and vocabulary). Listening comes first and the system locks the pages to other sections until the listening recording is finished, i.e. all students have to complete the listening section at the same time. While for reading, unlike some other on-line tests in which questions pop up one by one as the student proceeds, the DELTA reading text together with all the questions on that particular text will be shown at the same time parallel to each other on the same screen. That is to say it is not possible to measure how much time a student spends on a particular item. It is the same case for grammar. The whole DELTA test is conducted in a language laboratory. Students are not allowed to leave the test venue until the test session is finished so as to create the least disruption to the other candidates. In this case, students who do the test by wild guessing may in fact go back to the test questions again and revise their answers when they realize that they have too much time left. There is simply no practical way to calculate how much time a student has spent on each item.

Having considered the practicability as well as the strengths and drawbacks of the above methods, a sequential mixed research method (Creswell, 2009) was used in this study which begins with quantitative post-test self report questionnaire surveys in which results are generalized, followed by qualitative interviews to supplement the quantitative data and to provide further details. Motivation is a theoretical construct: a latent variable that it is not possible to observe directly but one that has to be operationalized and inferred from verbalizations or behavior (H. Eklöf, 2010). Therefore, it is only by detailed analysis of students' self reports via questionnaire and interview that we are able to explain the cognitive processes of the examinees.

3.3 Questionnaire survey

3.3.1 The construction of the questionnaire

There are 2 sections in the questionnaire survey. Section 1 is taken from the Student Opinion Scale (SOS) by Sundre and Moore (2002) while section 2 examines the perceived usefulness of the DELTA report.

The first section, which is the SOS, helps to investigate if students are motivated in taking the DELTA. A sound measure of motivation is prerequisite in order to accomplish this task. It is worthwhile to test how well a developed scale like the SOS works in the diagnostic assessment context in Hong Kong. The TTMQ developed by Eklöf (2006) was also considered. However, since the TTMQ was developed based on the Swedish TIMSS and a number of the questions are context specific and therefore not applicable to this study, the SOS naturally seems to be a more suitable choice.

Sundre has provided a copy of the instrument, a manual on how to administer the SOS as well as how to analyze its scores for free download at www.jmu.edu/assessment/resources/Overview.htm. Tables 3.1 and 3.2 are taken from Sundre's *The Student Opinion Scale Test Manual* (2007):

Table 3.1 Student Opinion Scale items

1. Doing well on this test was important to me.
2. I engaged in good effort throughout this test.
3. I am not curious about how I did on this test relative to others. *
4. I am not concerned about the score I receive on this test. *
5. This was an important test to me.
6. I gave my best effort on this test.
7. While taking this test, I could have worked harder on it. *
8. I would like to know how well I did on this test.
9. I did not give this test my full attention while completing it. *
10. While taking this test, I was able to persist to completion of the task.
* Denotes items that are reversed prior to scoring.

Table 3.2 Test Blueprint for SOS

Subscale	# Items	Items
Importance Definition: How important doing well on the test is to the student (the consequence of the test for the student).	5	1, 3, 4, 5, 8
Effort Definition: The perceived degree of work or mental taxation put forth in completing the test.	5	2, 6, 7, 9, 10

The SOS is comprised of two subscales: Importance and Effort (see Table 3.2). According to the Manual, scores should be added up separately for each subscale and therefore the range of possible scores for each subscale is 5 to 25. Higher scores on both subscales indicate greater self-reported motivation.

Instead of following the Manual and adding up the scores of each subscale, the Rasch model is used in this study in analyzing students' motivation level. The Rasch model emphasizes item-level information, i.e. it models the probabilistic distribution of an examinee's success at the item level (Fan, 1998). Basically the more ability possessed by an examinee and the easier the item, the higher the chances that an examinee would perform successfully (Rasch, 1960).

The second section of the questionnaire aims at finding out if students find the DELTA report useful. There were originally 10 questions in the second section of the pilot questionnaire which were reduced to three main ones in the final questionnaire (with one of the questions, question number 13, incorporating eight embedded questions). Details and reasons for the changes from pilot to actual version will be explained in the coming section. Again, the Rasch model is used in analyzing students perceived usefulness of diagnostic report.

A copy of the questionnaire can be found in Appendix II. The questionnaire is in English as my original argument was that English is the medium of instruction in all Hong Kong universities and students are very familiar with and proficient in reading in English from university

announcements, lecture notes, reading materials for classes to all kinds of tests and exams. It is a norm for all written documents and correspondence to be in English. What is more, it is a follow up questionnaire after the DELTA and the DELTA is an English proficiency test, so using an English questionnaire would seem to be logical.

3.3.2 Piloting of questionnaire

At the time when I was collecting data for this study, DELTA was in the piloting stage. It is in full operation starting academic year 2012/2013. Around 300 students who took the Preliminary English for University Studies course in academic year 2011/2012 were required to sit for the DELTA pilot test at the beginning of the semester. The first version of the questionnaire was sent to these students a week after they had taken the test. One of our partner institutions is also piloting the DELTA test before its full implementation in 2012/2013. Around 500 students took the test in the first semester of 2011/2012 and were asked to fill in the same questionnaire. These two groups of students were used as the pilot group for the questionnaire.

In the pilot study of the questionnaire, 186 responses (52 from my institution and 134 from our partner institution) were collected. After the collection of responses, 4 teachers from the English Language Centres (2 from my institution and the other 2 from our partner institution) were invited to a meeting and reviewed the survey questions in detail to discuss the appropriateness of the questions. After the meeting, the questions were revised accordingly. The 10 questions taken from the Student Opinion Scale remained unchanged. For the second section of the questionnaire, some of the questions in the pilot version were modified or deleted for better conciseness and clarity. The two questions that were deleted are:

* The Component Skills Profile shows my relative strengths and weaknesses in the four components: Listening, Vocabulary, Reading and Grammar. This information on my English Proficiency is useful.

Strongly agree Agree Neutral Disagree Strongly disagree

**The breakdown of the subskills (The Component Diagnostic Reports) provide useful information on my English proficiency.

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Listening					
Reading					
Grammar					
Vocabulary					

The reason for the deletion of the above 2 questions is that the original question (*) had generated results that were too easy to predict. The Component Skills Profile is simply a bar graph informing students of their relative strengths/weaknesses in the four skills (see Component Skills Profile sample in Figure 3.1).

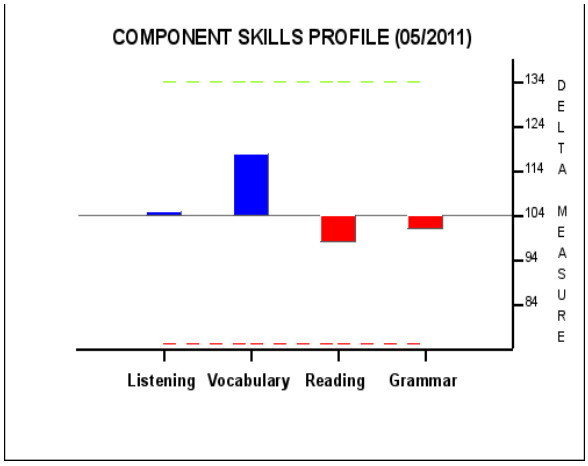


Figure 3.1 Sample Component Skills Profile

In the pilot, while the majority (70.6%) agreed with the statement, 14.7% of them indicated "neutral". In both questions, the term 'useful' is also too general and cannot help understanding in what way the report is useful to students' learning, therefore the statement is now deleted.

There were originally 6 questions on the specific usefulness of the DELTA report and they were combined into one embedded question (and 2 additional statements were added) after the pilot

using the same stem so that it is now more user-friendly for the students to complete. The list of the revised questions, on a 5-point Likert scale same as the SOS in section one, is listed below:

11.	The DELTA report as a whole is easy to understand.
12.	In the DELTA report under each component (Listening, Reading, Vocabulary and Grammar) shows a breakdown of the subskills (e.g. Listening: identifying specific information). I understand the meanings of the subskills listed in each of the components.
13.	The DELTA report can tell me if I am making any progress. tell me my English proficiency. tell me my strengths and weaknesses. allow me to compare my results with other students. allow me to refer to it if I want to do self-study on English. guide me in how to prioritize my English learning. encourage me to seek help from an English teacher. motivates me to improve my English.

With reference to the principles of good feedback by Nicol and Macfarlene-Dick (2006:205) in section 2.4.4, I developed my own definition and measurement tool of students' perceived usefulness of diagnostic feedback and applied it in my study. The 8 embedded questions of question 13 of the questionnaire contribute to the perceived usefulness of the DELTA diagnostic report as a whole and are more specific as to what kind of information students find useful (4 questions) and if the DELTA report encourages students to carry out further action (4 questions). The 8 questions form the Report Usefulness Scale.

Table 3.3 DELTA Report Usefulness- questions on action

Report usefulness	Item	Statement
Action	1	I will refer to my Report if I want to do self-study on English.
	2	Having a DELTA report will encourage me to seek help from an English teacher.
	6	The DELTA report allows me to compare my results with other students.
	8	The DELTA report motivates me to improve my English.

Table 3.4 DELTA Report Usefulness- questions on information

Report usefulness	Item	Statement
Information	3	The DELTA report can tell me if I am making any progress.
	4	The DELTA report can tell me my English proficiency
	5	The DELTA report can tell me my strengths and weaknesses.
	7	The DELTA report can guide me in how to prioritize my English learning.

To sum up, after the revision of the questionnaire, each of the questions is now directed to answering the specific research questions:

The first section i.e., the 10 questions from the SOS, is about students' test motivation whilst sitting the test (research question 1).

The second section is about feedback on the DELTA report (research question 3).

The questionnaire results also help in answering the last research question: the data collected via the SOS and the Report Usefulness Scale would help in the construction and validation of the process model of test motivation.

3.3.3 Participants of the questionnaire survey

After the pilot, the revised questionnaire was sent via e-mail to the students who took the DELTA in the second semester of 2011-2012 (January/February 2012). 40 students from my institution took the DELTA in this round. They were students of the English for University Studies course under the English Language Centre. The DELTA test was used as one of the subject activities. 400 students from our partner institution also took the DELTA during this period. There were also 70 students from another institution in Hong Kong which was trialing DELTA in that semester to see if the test was suitable for their use. If the feedback from their stakeholders were positive, they might consider joining the DELTA collaboration project. In total, the finalized questionnaire was sent to these 510 students.

3.3.4 Methods of analysis

According to the *Standards for educational and psychological testing* (1999), validation of an instrument involves evidence from four sources: content, response processes, internal structure and relations to other variables. 'Content' evidence refers to the extent to which the items included in the instrument adequately represent the content domain of the concept of interest and is largely a matter of judgment usually by scale developers (Squires, Estabrooks, Newburn-Cook, & Gierl, 2011). This current study would not be looking at content evidence as it is assumed that both the SOS and the Report Usefulness Scale which was developed based on the literature and discussions with a team of experienced teachers should satisfy this criterion. 'Response processes' evidence is most often derived from observations or interviews to test if the respondents' behaviour is congruent with their responses to the instrument. My interview data would be able to confirm if what students said is consistent with their SOS and report usefulness scores. 'Internal structure' evidence refers to the relationships between the items in an instrument. Factor analysis and Cronbach's Alpha were used as a quality check in this regards. Lastly, 'relations to other variables' refer to the relationships between the instrument and other external variables which measure the same, related or different concepts. Although I did not correlate the two scales with other scales which measure the same concepts, a student's perceived usefulness

of the diagnostic report should contribute to the 'value' component under the Expectancy-value theories; correlations of the SOS and the report usefulness scale will be examined to see if this provides further evidence in validating the scales.

Together with the Rasch analysis, the SOS and Report Usefulness Scale as valid instruments of measuring test motivation and students' perceived usefulness of diagnostic report is demonstrated. Further explanation of why and how these methods were applied and the justification of using them are in section 3.3.4 and in Chapter 4.1.

In order to validate the 'internal structure' of the scales, firstly, factor analysis (principal-components analysis) was run. The purpose of factor analysis is to "explore the underlying variance structure of a set of correlation coefficients. Thus, factor analysis is useful for exploring and verifying patterns in a set of correlation coefficients. . . " (Brown, 2001, p. 184). Descriptive methods of factor extraction assume that the subjects and variables to be analyzed represent the populations of interest only (Tinsley and Tinsley, 1987, p.418) . Principal-components analysis, being one of the descriptive methods of factor extraction, factor analyzed 100% of the variance (which is technically a transformation of the data into a set of uncorrelated variables) and is the most commonly used type. Then, the internal consistencies of the two subscales under the SOS were examined by checking their Cronbach's alpha. Cronbach's alpha is the most common measure of scale reliability which is based on the idea that individual items should produce results consistent with the overall questionnaire (Field, 2009).

After that, I analyzed the SOS and the questions on report usefulness by using the Rasch model under the one-parameter IRT to determine if the SOS and the report usefulness scale is a well functioning scale for this targeted group of students. The software that is used in this study is called Winsteps (version 3.72). The justification for using the Rasch model instead of following the traditional way of analysis or what the SOS manual instructs, i.e. by simply adding up the total score in each subscale, is that if scores are simply added up together as a total score, this indicates that "the relative value of each response category across all items is treated as being the same, and the unit increases across the rating scale are given equal value" (Bond & Fox, 2007: 102) . The strength of using Rasch with likert-scale items is that Rasch establishes the relative

"difficulty" of each item of a questionnaire. Each item/question, will be given a "difficulty" estimate. Rasch analysis has 2 major differences when compared with simple addition of scores: first, each question can vary in the level of motivation that it can actually detect based on the way the group of respondents actually responded to that item; second, the scale may not be an interval one, i.e. the distance between each response category (Strongly agree, Agree, Neutral, Disagree and Strongly disagree) may not be the same in between each other. For example, the increase in motivation implied by the move from Disagree to Neutral may be less than the increase required by the move from Agree to Strongly agree. Whereas the traditional way of adding up total scores (SA=5, A=4, N=3, D=2, SD=1) assumes that SA has a value five times greater than that of SD. Therefore, a person's overall SOS score calculated by simple addition of the scores in each subscale and that by using Rasch where the relative value of each response category of each question are taken into account, can be quite different. A sample of how the two methods of calculation compared can be found in Sections 4.1.1.2.

The running of the Rasch analysis provides another way of validating the SOS as a useful motivation measurement in addition to the studies carried out by Sundre and her colleagues (Sundre & Finney, 2002; Thelk et al, 2009) . The same comprehensive analysis has also been used to validate my questions on report usefulness (see section 4.1.2).

3.4 Interview

3.4.1 The structure of the interview

The interview was divided into 3 sections: the first section was to clarify the ambiguities encountered during the data analysis of the questionnaire. The second section was about the DELTA and the expectancy-value model of motivation (to provide further evidence in order to support the SOS and the feedback usefulness data). In the third section, the process model of test motivation was used to investigate/explain if students were motivated before and after taking the DELTA. Finally, with the help of the questionnaire survey and interview, I examined how the process model of test motivation can be adapted and applied to DELTA.

DELTA is a diagnostic and tracking assessment which aims at tracking students throughout the 4-year university curriculum and each round of test preparation and test-taking is a dynamic process on its own where motivation might fluctuate during the course. Questionnaire survey can only inform students' motivation level at certain point(s) in time while interview can elicit thoughts, feelings and reasoning behind. The SOS is a measure of students' motivation while they are taking the test, i.e. motivation in the actional phase. I wanted to find out from the interviews with students what is happening in their preactional and postactional phases of a test which provides a detailed diagnostic report as well as a track on their performance throughout the years. As discussed in the literature review, diagnosis involves identifying specific areas of strength or weakness in language ability so as to assign students to specific courses or learning activities (Bachman and Palmer, 1996). This is what makes a diagnostic and tracking test different from other test types such as exit tests and proficiency tests in which the test itself is usually an end in the process. Therefore, the significance of a diagnostic and tracking test should lie in the preactional (whether it motivates and helps students in setting goals and forming intention) and postactional phases (whether the report motivates or helps students in elaborating standards and strategies to further improve their English).

As explained in Section 3.3.3, there were 40 students who took the DELTA in the second semester in my institution and volunteers to participate in the interview were invited from this group. The 40 students had also completed the questionnaire survey. One of the volunteers was invited to come for an interview a week before the other volunteers and the session was used as a pilot for the other coming interviews. The pilot interview was successful and proved that the questions could generate rich data. As for the rest of the interviews, I by design allowed the students to have the freedom to choose if they want to come alone or if they want to form into their own groups of 2-3. The idea of allowing students to come alone or in small groups was that some students may be shy to talk in front of a group and prefer private interviews; while some may feel more confident and secure if they come along with their classmates. Therefore I believe allowing them to have a choice according to their wish would receive the best results. What I had conducted were small group interviews, which were different from focus group interviews. A focus group interview is a semi-structured discussion of a given topic by a homogeneous group of 6-10 individuals during which participants exchange ideas and opinions (Aubel, 1994;

Sinagub, Vaughn, & Schumm, 1996). There were only 1-3 people in each of my groups and the purpose of the interview was not for them to discuss a certain topic. My group interviews are indeed semi-structured individual interviews but conducted in pairs or at most in threes. There are three main types of interviews in research: structured, semi-structured and in-depth (Britten, 1995). In structured interviews, the interviewer follows questionnaire-like questions and answers are usually in fixed choice, for example, good, fair or poor. Semi structured interviews consist of open ended questions and the interviewer may deviate in order to pursue an idea in more detail. Lastly, in depth interviews are less structured and may cover only one or two issues but in far greater detail. However, although small group interview is different from focus group interview, having interviews in groups may also benefit from the advantages of a focus group interview by allowing interactions among the students which may generate richer data. According to Frey and Fontana (1991:178) group interviews "allow opinions to bounce back and forth and be modified by the group, rather than being the definitive statement of a single respondent".

The interviews lasted around 45 minutes each. Students were also allowed to choose if they want to talk in Cantonese or in English as it was logical to assume that students might be able to express themselves better in Cantonese. Local Hong Kong students would all switch to talk in Cantonese outside classroom as Cantonese is their mother tongue. Before the start of the recording, I asked each of the students if they prefer to talk in English or Cantonese and they all indicated their preference to talk in Cantonese. All participants were given a HKD 20 McDonald's coupon as a token of appreciation of their time. The list of the interview questions is provided in Appendix III.

3.4.2 Interviewees

As mentioned in the previous section, the 40 students who took the DELTA in my institution in the second semester of 2011/2012 were invited to participate in the interviews. They were the year one students in the Institute of Textiles and Clothing who enrolled in the English for University Studies course under the English Language Centre, where DELTA was used as a subject activity. They had also completed the questionnaire survey at the beginning of the

semester. I prepared a signup sheet which I asked the two instructors of the course to pass around for their students to sign up before the end of their class. In order to encourage participation, the interviews were scheduled in the second last week of the semester when they had submitted their course assignment and still have a few weeks before the exam. Timeslots were opened for signup throughout the whole week and students were free to choose to come alone or form into groups of 2-3. In the end there were 17 volunteers (in 8 sessions with 3 groups of 3, 3 groups of 2 and 2 coming alone) of which 16 showed up in the interviews (including the student who came in for the pilot session).

3.4.3 Methods of analysis

The interviews were conducted in Cantonese and were audio-recorded, transcribed and translated. All recordings were first transcribed by myself, within three days after the interviews were conducted when my memory was still fresh and then translated into English again by myself within another three days to ensure consistency of the translation (a sample of the transcription and translation can be found in Appendix IV).

Constant comparative analysis (Miles & Huberman, 1994; Ryan & Bernard, 2003) was used to help identify, create, and see the relationships among parts of the data. Morse and Field (1996) state that in constant comparative analysis, each piece of data must be compared with every other piece of relevant data. During my transcription and translation and comparing of the translated text, it was not difficult for me to start seeing patterns emerging. It was also because my group interviews were quite structured and students basically followed the flow of my questions when they answered and not a lot of unexpected discussions appeared.

I started my coding process using a grounded approach, i.e. by using the open coding method (Cohen & Manion, 2007; Strauss & Corbin, 1998). In the process of open coding, every passage of the interview is studied to determine what exactly has been said and to label each passage with an adequate code (Boeije, 2002). NVivo9 software was used to assist the coding process and the constant comparisons. After I coded the first interview, I compared the codes within the

interview to check if there were any problems or inconsistencies, or if there were any duplication so that some codes can be combined.

Then I coded the other interviews using the same approach and after that I did axial coding by comparing the same or similar codes in different interviews to see if they can be grouped under categories and themes (Strauss & Corbin, 1990). Table 3.5 shows the list of themes and categories that I generated.

Table 3.5 List of themes and categories

1. Preactional phase Expectancy - confidence level - DELTA difficulty level - previous English results Preparation Set target Value - Importance of English - Importance of DELTA
2. Actional phase Test preparation - Action control - Subtask generation and implementation
3. Postactional Evaluation Motivation after test
4. Dynamic test motivation
5. Motivation taking the DELTA High Medium Low
6. Retake the DELTA
7. Understanding of SOS
8. Usefulness DELTA DELTA report

The SOS measures of the 16 students as well as their report usefulness measures were plotted in a scale using Winsteps software to show the profiles of the students. Then according to their performance (their rankings in the SOS scale and report usefulness scale), I looked into their

interview data (the themes and categories in NVivo) to see if any relationships between their SOS and report usefulness measures and their interview data exist. Finally selective coding (Strauss & Corbin, 1990) was used which means integrating the data around a central theme, hypothesis, or story to generate a theory (Walker & Myrick, 2006), in order to answer the research questions. Profiles of the students and their stories were built.

3.5 Ethical issues

This study is in compliance of the *Revised Ethical Guidelines For Educational Research* (2004) published by the British Educational Research Association (BERA) and has taken the below recommendations by Creswell (2009:89-92)):

General ethical issues in data collection have been covered in my study, specifically:

i. Identification of the researcher and the sponsoring institution
   In the email invitation for the participation in the questionnaire as well as on the first page of the questionnaire itself, I duly informed the students about the purpose of the survey as part of my research project and also as reference for the English Language Centre in its decision making process. When inviting volunteers to come for the interviews, and immediately before the commencement of the audio recording of the interviews, I have also fully identified and introduced myself as the researcher of the project.

ii. Identification of the purpose of the research
   I explained the research objective to the participating students before the interviews and made it clear to them that this study would be fulfilling part of my EdD requirement and might also be used as a reference for the English Language Centre in its future tests and activities planning.

iii. Identification of the benefits for participating

I stated clearly in the focus group invitation that this discussion would be of value to my research study and might be used as a reference for the English Language Centre in its future tests and activities planning; and upon completing the focus group discussion, each participant can get a McDonald cash coupon of HKD20 as a token of appreciation for their time and efforts.

iv. Guarantee of confidentiality to the participants

In the email invitation for the participation in the questionnaire, on the first page of the questionnaire itself as well as before the interviews were recorded, I assured the students that their identities would be kept anonymous.

The following ethical considerations have also been paid due regard to in my data analysis and research paper writing:

i. Protection of the anonymity of an individual

Similar to the process of data collection, the identities of all participants in the questionnaire survey have been kept anonymous and only pseudonyms were used in the data analysis for the interviews.

ii. Data interpretation

I have endeavored to provide an accurate account of the data gathered. I have ensured that the participants in the interviews were fully informed that they could request copies of the relevant transcripts and written interpretation and reports from me once I have finalized them. My Institution does not require an ethical procedure for situations which involve only the recruitment of students to facilitate an outside project (in this case my doctoral research with an overseas university) but according to the guidelines by the PolyU Human Subjects subcommittee, I need to make sure that no pressure are imposed on the students to participate. Students can only be informed of the call for volunteers and have to decide to join or not with their free and independent consent. Written approval from the director of the DELTA project had also been sought.

3.6 Chapter Summary

This chapter set out the three research questions of the study. It then describes the method of data collection and analysis. The pros and cons of the three major ways of measuring examinees' effort, i.e. post-test self-report scale questionnaire, person-fit statistics under the Item Response Theory and the Response Time Effort are discussed and explanations are given as to why post-test self-report scale questionnaire was the most practicable and useful in the DELTA context. The second section of the chapter talks about the construction of the questionnaire, how the questionnaire was piloted and the method of analysis. The third part of the chapter carries on to describe the method of analysis of the qualitative data of the study. Interviews had been conducted with 16 students and the discussions were audio-recorded and transcribed. Nivivo software was used to assist the coding process. Finally, the chapter ends with a discussion of how ethical issues have been looked after in the whole study.

Chapter 4 Findings

4.1 Questionnaire surveys

As discussed in Chapter 3, about 500 students in the 3 partner institutions took the DELTA in the second semester of 2011/2012. Invitations to complete the online survey were sent to this batch of students via email and 141 responses were collected. The response rate is only 28% which is quite low. Baruch and Holtom (2008) analysed 1607 studies published in the years 2000 and 2005 and found out that the average response rate for studies which utilized data collected from individuals is 52.7% with a standard deviation of 20.4. Apparent reasons for the low response rate in this study are that students do not have the incentive to complete online surveys as we do not offer any monetary incentives such as cash and vouchers which other researches declared as being effective in increasing the response rate in offline and online surveys (Church, 1993; Fox, Crask, & Kim, 1988) nor other rewards, and more importantly, the saliency of the survey content should have played a part in the low response rate as well (Groves, Presser, & Dipko, 2004), the DELTA is not high stakes to the students and neither is the survey. Same as in all other volunteered surveys, the sample is a self selected one and in terms of the construct of motivation, it is quite likely that the sample is also a biased sample. I am aware of this limitation of the study but as discussed in section 3.2, given the practical conditions such as the nature of the DELTA, questionnaire survey is the most feasible method of data collection.

The survey, which was in 2 sections, was analyzed using different methods. Firstly, the 'internal structure' of the SOS was examined by using factor analysis and Cronbach's Alpha and then Winsteps analysis (version 3.72) was run. Secondly, the same factoral analysis, Cronbach's Alpha and Winsteps analysis of the report usefulness scale was conducted. After that, the correlation between test motivation and the perceived usefulness of the DELTA diagnostic report was investigated to gather validation evidence on 'relations to other variables'. Finally, based on the findings above, I propose a process model of test motivation of diagnostic test.

4.1.1 The Student Opinion Scale (SOS) as an instrument of measuring motivation

4.1.1.1 Internal structure evidence of the SOS

The first section of the questionnaire survey, i.e. the SOS, contains 10 questions. The questions in the SOS are in fact statements but in this study the term 'question' will be used which is easier to understand in the context of a questionnaire. Sundre and Moore (2002) constructed 5 questions on effort and the other 5 on importance based on the Expectancy-Value Theory. Running the 141 student data using the bivariate correlation analysis (correlation between two variables) in SPSS shows that the questions under the effort factor have a significant correlation of .38 (one-tailed, $p< 0.01$) with the questions under the importance factor (Table 1). 'One-tailed' is chosen here because the directional hypothesis is that the more important the test is to a student, the more effort they should put in it. According to Field (2009:173), a correlation coefficient of +/- .1 represents a small effect, +/- .3 represents a medium effect and +/- .5 is a large effect. The significance value of $p<0.01$ tells us that the probability of getting a correlation coefficient of .38 in a sample of 141 people if the null hypothesis (there was no relationship between these variables) was true is very low (Field, 2009).

In order to test whether the SOS actually is comprised of 2 factors only, i.e. the importance factor and the effort factor, an exploratory descriptive factor analysis (principal components analysis) using SPSS was run. Factor analysis reduces a large number of interrelated variables to a smaller number of latent or hidden dimensions (Tinsley & Tinsley, 1987). There are no prior hypotheses about the results in exploratory analyses; and in descriptive methods, we assume that the subjects and variables represent the population of interest. In principal components analysis, 100% of the variance is to be factor analyzed. Eigenvalues are used to condense the variance in a correlation matrix. "The factor with the largest eigenvalue has the most variance and so on, down to factors with small or negative eigenvalues that are usually omitted from solutions" (Tabachnick and Fidell, 1996, p. 646). Traditionally only variables with eigenvalues of 1.00 or higher are considered worth analyzing (Kaiser, 1960). The Scree Plot, which visualizes the importance of the factors in descending order of magnitude of eigenvalue, is shown below. On the plot, each circle on the curve represents a factor. The curve flattens out at the third factor so only two factors should be extracted in this Scale. This is in compliance with the constructs of

the SOS, which is intended to represent two factors of motivation: the perception of importance of the test and the amount of effort exerted on the test.

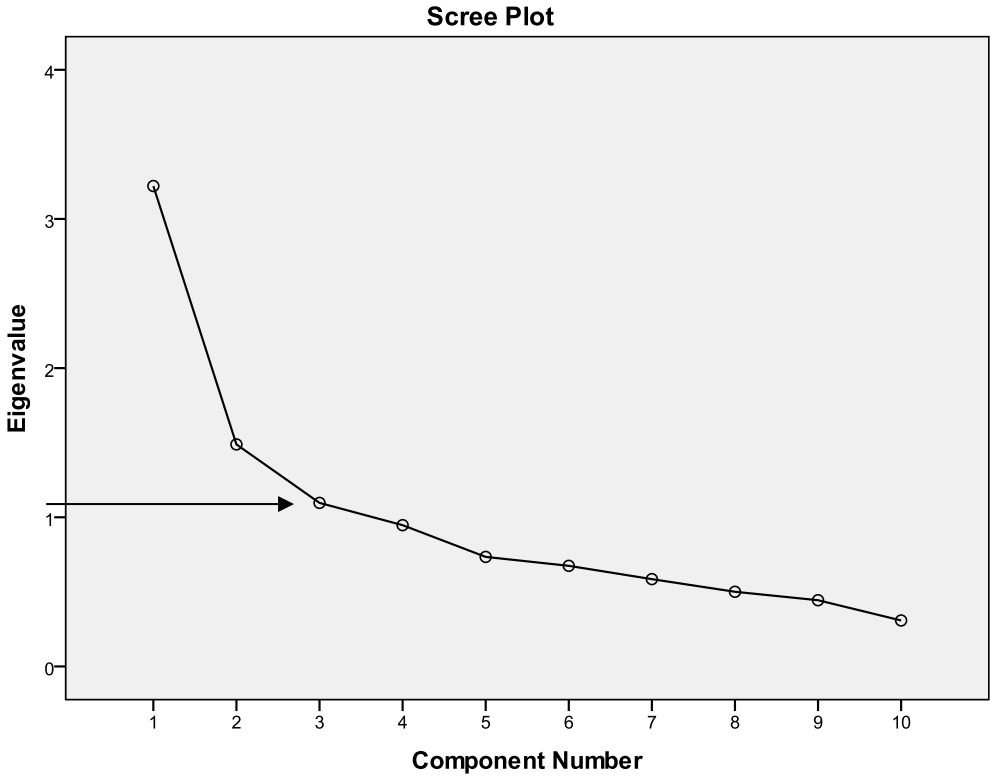


Figure 4.1 Scree Plot of SOS Factor Analysis

The scree plot shows that the third factor has an eigenvalue of about 1 in this analysis. Therefore, it is a marginal case and in order to confirm, a fixed 3-factor analysis was run to check if a third factor was contributing to the data. Factor rotation was used to calculate to what degree variables load onto these factors. Field (2009) suggests that varimax should be used as this is a good general approach that simplifies the interpretation of factors.

Table 4.1 Rotated Component Matrix of SOS

Rotated Component Matrix[a]

	Component		
	1	2	3
1. Doing well on the DELTA was important to me.	.739		
2. I engaged in good effort throughout the test.		.719	
3. I am not curious about how I did on the DELTA relative to other students.	-.584		
4. I am not concerned about the scores I receive on the DELTA.	-.734		
5. The DELTA was an important test to me.	.708		
6. I gave my best effort on the test.		.872	
7. While taking the DELTA, I could have worked harder on the questions.			.734
8. I would like to know how well I did on the DELTA.	.702		
9. I did not give the DELTA my full attention while completing it.		-.837	
10. While taking the DELTA, I was able to persist until I had completed all the questions.	.521	.323	.454

a. Rotation converged in 3 iterations.

According to Sundre and Thelk (2007), questions 1, 3, 4, 5, 8 should be testing the 'importance' factor while questions 2,6,7,9,10 should be on 'effort'. Stevens (2002) recommends that for a sample size of 100, a loading of 0.512 or above can be considered as significant. From the Rotated Component Matrix above, most of the questions fit the constructs except questions 7 and 10. One possible explanation for the fact that these two questions do not fit the constructs may be that the SOS was presented to the students in its original language, i.e. English. The SOS was conducted in English with an aim to keep its originality and also due to the fact that English is the medium of instruction and the official language of the University. All written documents are supposed to be in English and students' ability of reading in English should not be a problem. However, even though students in Hong Kong can manage everyday and academic English, my speculation is that among the four reversely written questions (Q.3 I am not curious about how I did on the DELTA relative to other students, Q. 4 I am not concerned about the scores I receive

on the DELTA, Q.7 While taking the DELTA, I could have worked harder on the questions and Q.9 I did not give the DELTA my full attention while completing it), question 7 is the only one without a 'NOT' structure. 'Could have' + past tense is used to talk about something somebody was capable of doing but did not do. This grammatical structure, for some students especially those of lowerproficiencies, may be difficult to understand. As for question number 10, according to the constructs, it should belong to factor 2 "effort". The figures are not differentiating the factors well enough (0.521, 0.323 and 0.454). My hypothesis on why this happened to question 10 is again because of how this question is phrased, 'able to persist' is not a very common expression for lower ability students. In section 4.2, I discuss my findings from student interviews that should help explain and confirm my assumptions and the data in this section.

Apart from factor analysis, the internal consistency of the SOS was also examined. The Cronbach's Alpha of the Importance scale of the SOS is .687. According to Kline (2000), the generally accepted Alpha value is .8 for cognitive tests, .7 for ability tests and for psychological constructs values of below .7 can also be expected. Cortina (1993) also mentions that the value of Alpha depends also on the number of items. Therefore .687 for a scale with only 5 items can be said to be acceptable. As for the Effort scale of the SOS, its Cronbach's Alpha is .594. This is not very satisfactory and so I looked at the item-total statistics.

Table 4.2 SOS-Effort: Item-Total Statistics

	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
2. I engaged in good effort throughout the test.	4.009	.494	.323	.468
6. I gave my best effort on the test.	3.604	.505	.390	.445
7. While taking the DELTA, I could have worked harder on the questions.	5.603	-.055	.053	.714
9. I did not give the DELTA my full attention while completing it.	3.158	.586	.376	.378
10. While taking the DELTA, I was able to persist until I had completed all the questions.	4.517	.281	.131	.573

The column Cronbach's Alpha if Item is Deleted shows the overall Alpha if that item is not included in the analysis. Field (2009) points out that we should check the 'Cronbach's Alpha if Item Deleted' column to see if any of the values is greater than the overall Alpha, since the deletion of such an item would increase the Alpha and hence the deletion improves internal consistency; in this case, any value in the column which is higher than .594. Question 7 has a value of .714 and is therefore problematic. This is in line with the results of the factor and Rasch analysis and so the item was taken out from the analysis. Further explanation is in the next section.

4.1.1.2 Validating the SOS using the Rasch model

Unlike the conventional way of analyzing the SOS as suggested by Sundre and Thelk in the SOS user manual (Sundre & Thelk, 2007), the Rasch model provides a different perspective as to validating and applying the SOS. As explained in Chapter 2 (section 2.4.3), the Rasch model is a theory of how probabilities of response should be, in order to comply with fundamental requirements of measurement. The observed frequencies of response are then compared with the expectations (Tesio, 2003). In essence, Rasch makes items (i.e. the questions in the questionnaire) and students scalable and puts them on the same 'ruler'.

Likert attitude or opinion items are likely to vary in terms of their difficulty, agreeability, or endorsability (Bond and Fox, 2007; p.112). Using the Rasch model, we would be able to tell which items are more endorsable in the SOS for this group of students, or which items are more difficult for the students to agree. The Rasch model, based on how a group of respondents responded to each item stem, arranges the items according to their respective 'difficulty'; or the term 'endorsability' should be more appropriate in this context.

Figure 4.2 below shows the distributions of item endorsability and person ability estimates of the SOS on the same logit (log odds unit) scale. According to the Winsteps menu, logit is a unit of additive measurement which is well-defined within the context of a single homogeneous test. The logit scale is an interval scale, it is from -2 to 6 in this analysis. Although the SOS is

comprised of two factors: importance and effort, we can still keep the items of the two factors in one scale. As Linacre (2009) points out, we know that the two factors are conceptually different, but they are both part of motivation and neither can be omitted.

The column with the hashes '#' on the right locates the person difficulty measures along the scale. Each '#' in this figure represents 2 students and each '.' is 1 student; on the right is the list of items arranged according to their level of endorsability. For easier identification, 'I' which stands for importance and 'E' which stands for effort was added to the coding and the question numbers are in brackets. Item 8 'I would like to know how well I did on this test' is the easiest item for this group of students to endorse; while question 7 'While taking this test, I could have worked harder on it' is the most difficult to endorse. <more> equals to higher person ability whereas <rare> indicates higher item difficulty. 'M' represents the mean of the person or item distribution and 'S' and 'T' equals to one and two standard deviations from the person or item mean respectively. In this analysis, the mean of the person is half a logit higher than that of the items, meaning that on average, the students under survey show comparatively higher motivation than the motivation level represented by the items.

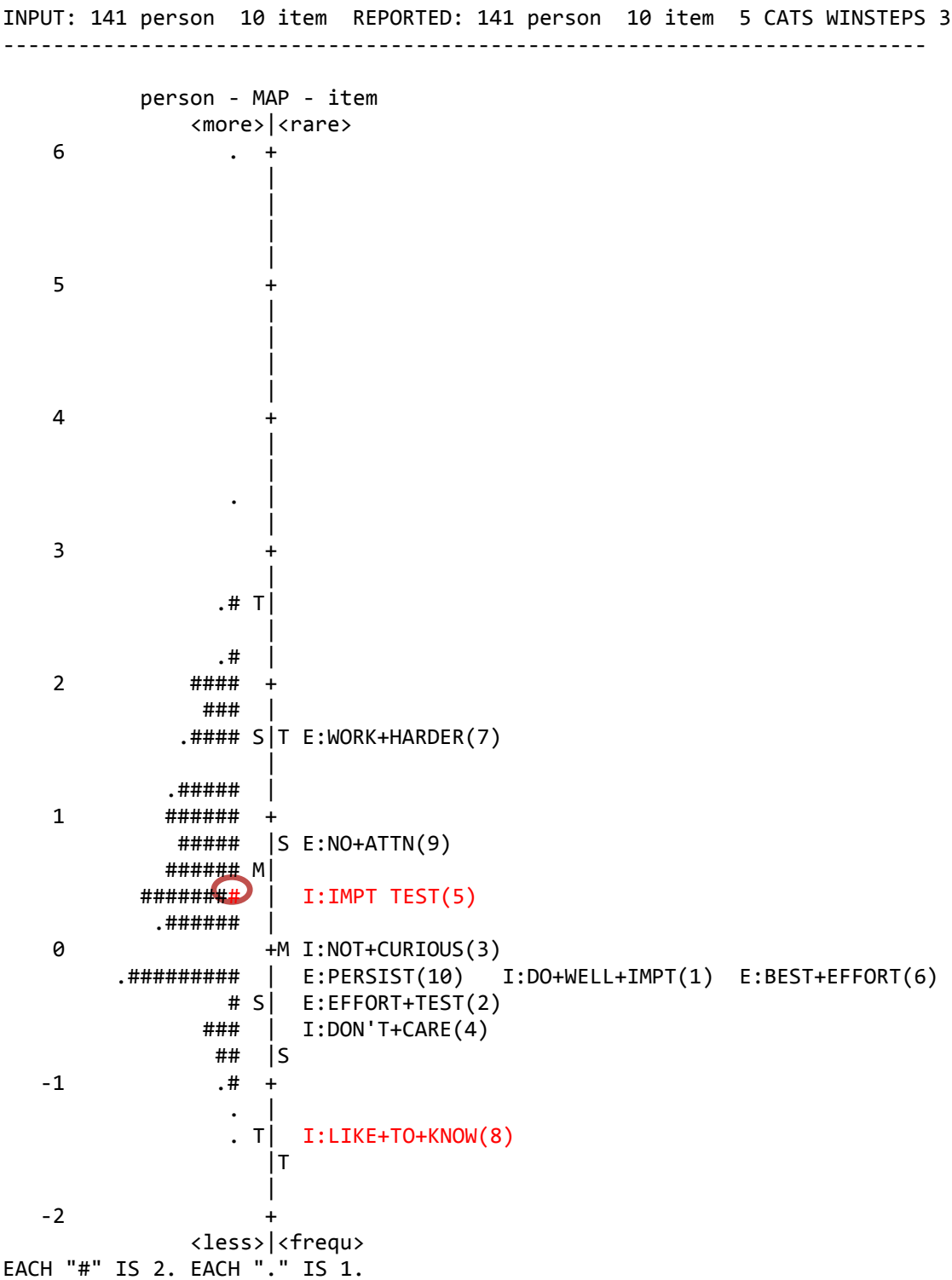


Figure 4.2 Person-Item Map of SOS

The Rasch model routinely sets at 50% the probability for any person to endorse an item located at the same point on the item-person logit scale (Bond & Fox, 2007:38). Abilities of 0, 1, 2 and 3 logits correspond to 50%, 73%, 88% and 95% probabilities of endorsing an item with 0 logit difficulty (Tesio, 2003). Take the two students represented by the '#' (in circle) in Figure 4.2 as an example, they are on the same level as question 5 (Q.5 The DELTA was an important test to me), that is to say their person ability estimates are the same as the item endorsability of this question, i.e. these two students have a chance of 50% in endorsing question 5. Under the same rule, these two students have a chance of 73% in endorsing a question one logit of endorsability below (Q.2 I engaged in good effort throughout the test) and a chance of 88% in endorsing a question two logits of endorsability below (Q.8 I would like to know how well I did on the DELTA).

Apart from providing a visual presentation of the item-person distribution as in Figure 4.2, the Rasch model also provides ways in checking the quality of the items. From the bubble chart generated by Winsteps below (Figure 4.3), items 1 and 8 are 'overfitting' meaning that the responses to the 2 items fit the Rasch model too perfectly and are too predictable from the Rasch-model perspective. That is to say, nearly all students with a motivation level higher than the endorsability estimates of the two items endorse them while those with a motivation level lower than the endorsability estimates of the items reject them. The pattern is too good to be true. Item 7, on the other hand, is 'underfitting', indicating that erratic response pattern is found in this item. Based on the Rasch analysis, students of a motivation measure above the endorsability level of Q.7 should likely endorse the question but the analysis shows that this is not the case, students with a motivation measure lower than the endorsability level of Q.7 are those who endorse the item instead. The responses are too unpredictable from the Rasch model's perspective.

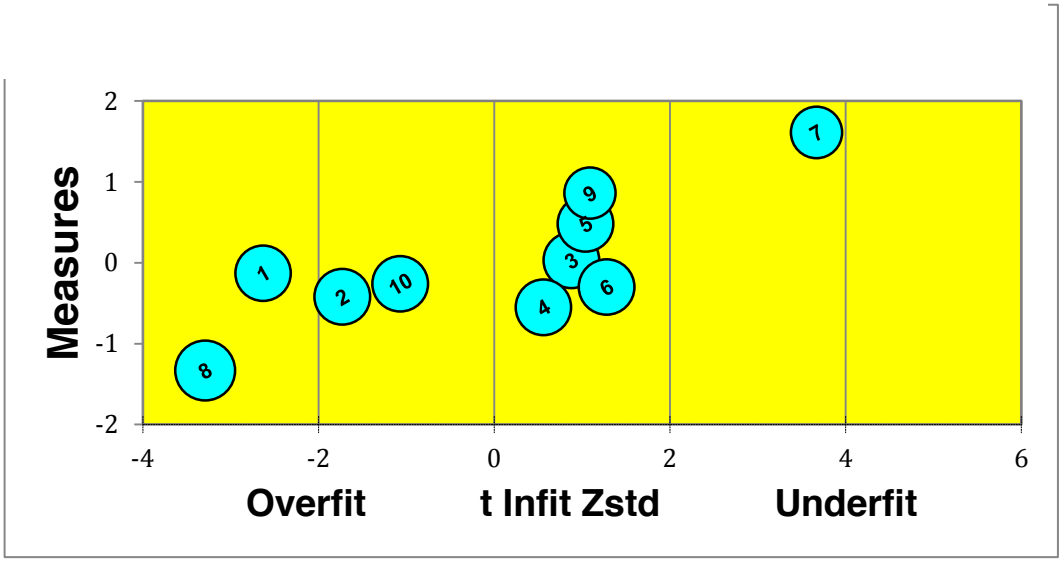


Figure 4.3 Bubble Chart of SOS

Apart from looking at the bubble chart above, which shows the infit zstd, Rasch also provides two other types of fit information, i.e. the column of infit msq and outfit msq in Table 4.3. According to Bond and Fox (2007), the general acceptance level of the mean squares is <1.3 (Bond & Fox, 2007:240). Again, question 7 does not fulfill the fit requirement.

Table 4.3 Fit Statistics of SOS

ENTRY	COUNT	ERROR	IN.MSQ	OUT.MSQ	NAME
7	141	0.12	1.4316	1.4751	E(7):WORK+HARDER
6	141	0.13	1.1546	1.1569	E(6):BEST+EFFORT
9	141	0.12	1.1317	1.1265	E(9):NO+ATTN
5	141	0.13	1.1202	1.121	I(5):IMPT TEST
4	141	0.13	1.0994	1.065	I(4):DON'T+CARE
3	141	0.13	1.0985	1.1029	I(3):NOT+CURIOUS
10	141	0.13	0.8626	0.8721	E(10):PERSIST
2	141	0.13	0.7094	0.8004	E(2):EFFORT+TEST
1	141	0.13	0.7086	0.7139	I(1):DO+WELL+IMPT
8	141	0.14	0.6429	0.634	I(8):LIKE+TO+KNOW

The results from factor analysis, Cronbach's Alpha and Rasch analysis show item 7 is problematic. Therefore one of the aims of my interview with students was to find out how they interpreted this question (see Section 4.2.1).

Since question 7 proved to be problematic, it was taken out from the analysis and the data was rerun using factor analysis which shows only 2 factors this time (Table 4.4). The scree plot (Figure 4.4) shows clearly that the third factor has an eigenvalue of less than 1 meaning that only two variables are worth analyzing. Question 10 does not seem to fall into either factor still.

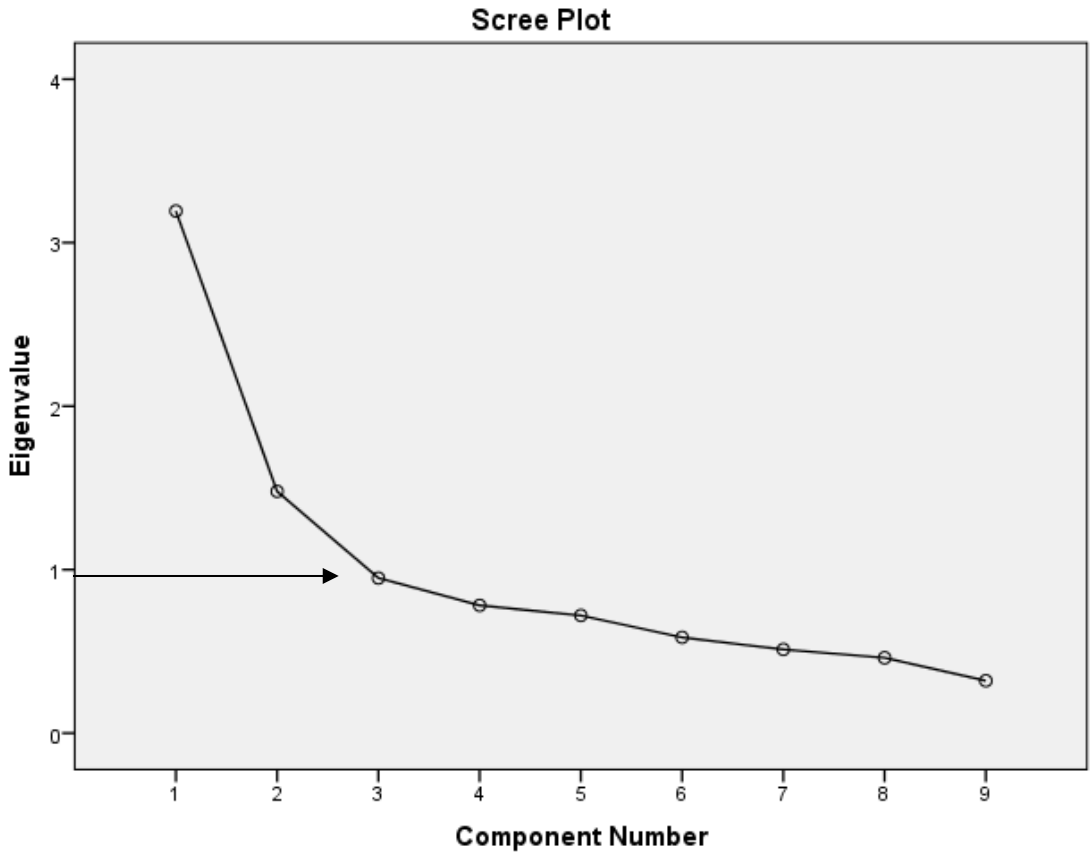


Figure 4.4 Scree Plot of SOS with 9 items

Table 4.4 Rotated Component Matrix of SOS (Q7 deleted)

Rotated Component Matrix[a]

	Component	
	1	2
1. Doing well on the DELTA was important to me.		.544
2. I engaged in good effort throughout the test.	.729	
3. I am not curious about how I did on the DELTA relative to other students.		-.710
4. I am not concerned about the scores I receive on the DELTA.		-.766
5. The DELTA was an important test to me.		.590
6. I gave my best effort on the test.	.854	
8. I would like to know how well I did on the DELTA.		.583
9. I did not give the DELTA my full attention while completing it.	-.757	
10. While taking the DELTA, I was able to persist until I had completed all the questions.	.400	.428

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Winsteps was rerun after taking out question 7 (Figure 4.5). We can see from the figure that the item on importance 'Q.8 I would like to know how well I did on the DELTA' is the easiest to endorse. The most 'difficult' to endorse item is on effort, 'Q.9 I did not give the DELTA my full attention while completing it.' We can see from the figure that there are a number of students on the scale who show higher test motivation than all the SOS items which means that those students will probably endorse 'strongly agree' or 'agree' in most if not all of the items. (the '#' above the line in Figure 4.5, showing higher 'ability' than the 'difficulty' level of the SOS items).

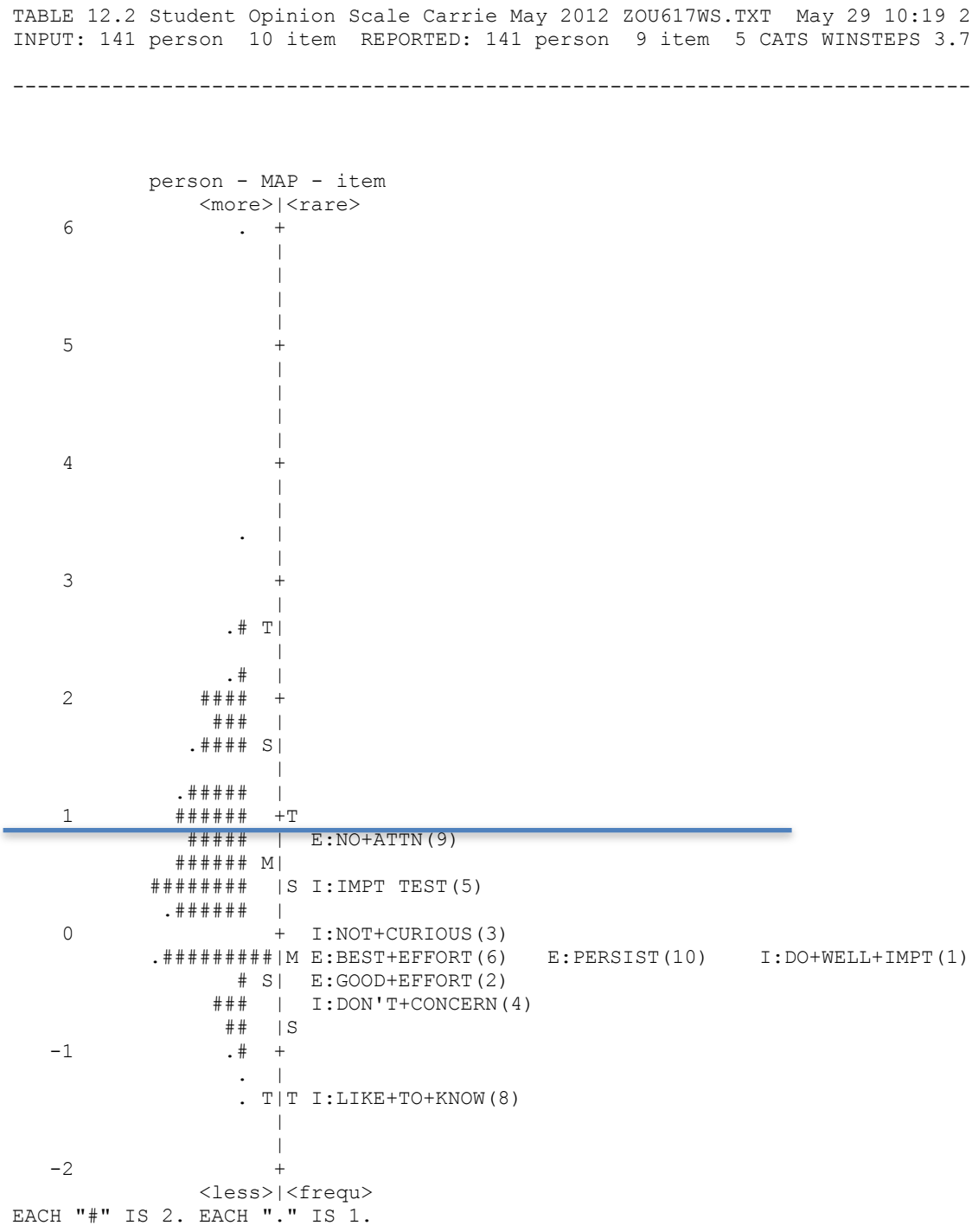


Figure 4.5 Person-Item Map of SOS (Q7 deleted)

As discussed in Section 3.3.4, under the Rasch model, each item has a different difficulty (endorsability) estimate. The Rasch model can transform the counts of the endorsements of the Likert categories into relative difficulty (endorsability) of each item and each item will be given

a difficulty estimate based on the actual empirical evidence. Each of the items in the SOS does not carry the same relative value in motivation. For example, Figure 4.5 shows that the difficulty level of item 9 is at logit 0.8 which is more difficult to endorse than item 5 which has a difficulty level at logit 0.4.

Moreover, a Likert scale of SD to SA has different relative values in the Rasch model which may lead to interpretations of results quite different from the traditional type of analysis. Sundre and Thelk's SOS manual (2007) suggests the use of the traditional method, i.e. raw scores are simply added up to get an overall score. According to the manual, in a Likert scale of SD to SA, SD would be = 1 and SA would be =5. If scores are added up as such, it is presumed that the ratio of the data, or the relative value of each response in the scale, is the same. That is, SA has a value five times greater than that of SD. However, a Likert-scale is not an interval scale. Therefore, just looking at Figure 4.5 is not enough, another Winsteps output is needed to provide a better picture of the student-item distribution.

Figure 4.6 below shows the location of the thresholds that mark the boundaries between the five Likert categories: SA, A, N, D and SD. For example, Responses of SA on items 9 (threshold 4.5) indicates higher levels of the underlying motivation than SAs on the other items; and 'neutral' in item 9 (threshold 2.5) actually indicates similar degree of motivation to 'Agree' in item 8. A threshold is a level at which the likelihood of being observed in a given response category (below the threshold) is exceeded by the likelihood of being observed in the next higher category (above the threshold) (Bond & Fox, 2007). Therefore at threshold .45, a person is more likely to indicate 'strongly agree' than 'agree'. Moreover, thresholds are not spread equidistantly, but the structure (the pattern of the threshold across items) is identical for every item in the Likert scale (Bond & Fox, 2007). Although it is not very obviously shown in this case, for example, moving from threshold 2.5 to 3.5 requires a smaller increase in motivation than moving from threshold 3.5 to 4.5.

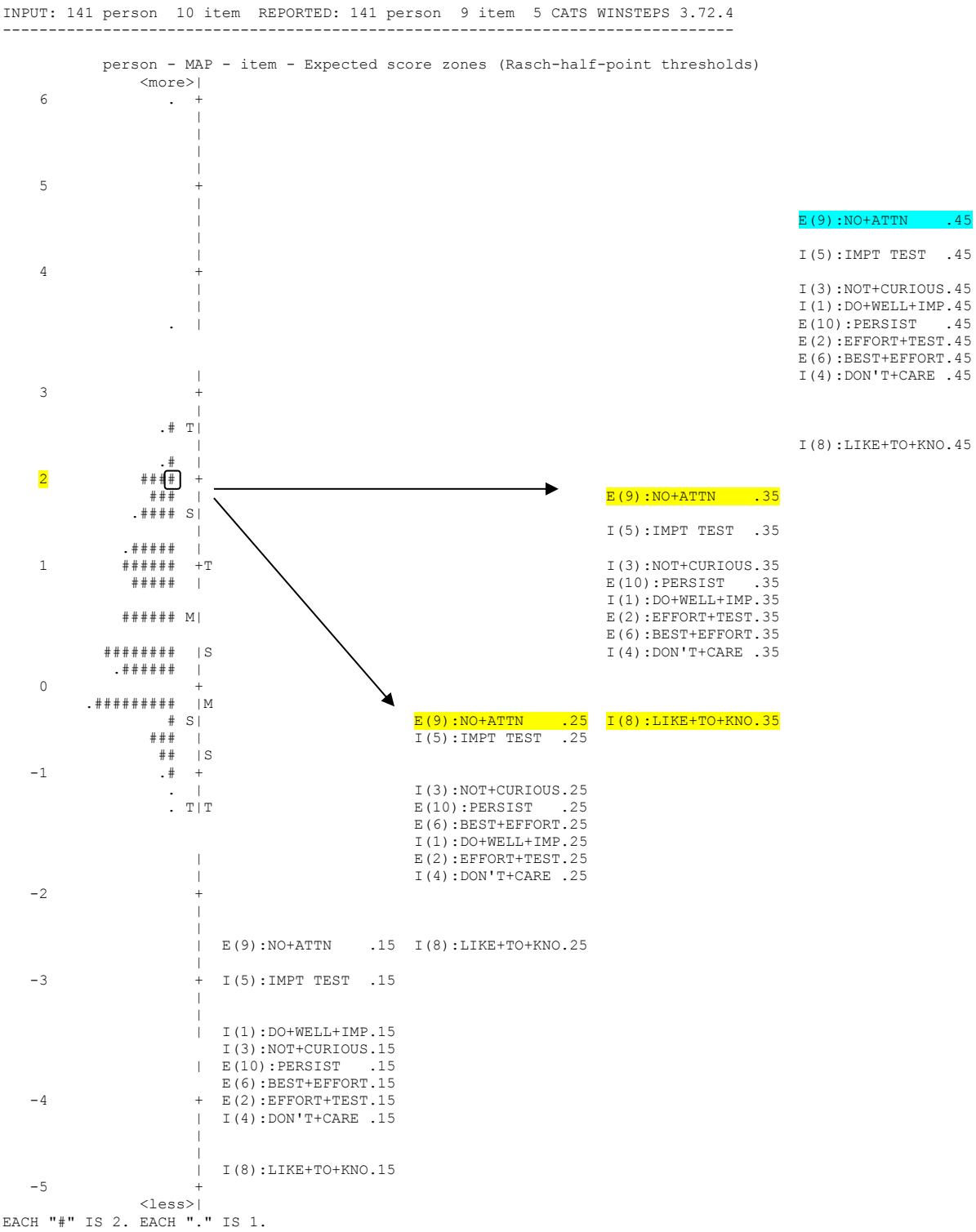


Figure 4.6 Person-Item Map of SOS with Thresholds

In Rasch models, logit measures a difference, a local distance (e.g. between subjects, between items or between ability and difficulty). Zero is conventionally assigned to the average difficulty of the items, so that one number only is sufficient to represent a measure. As mentioned before, abilities of 0, 1, 2 and 3 logits correspond to 50%, 73%, 88% and 95% probabilities of endorsing an item with 0 logit difficulty (Tesio, 2003). For example, a student at logit 2 (the '#" in square) has a probability of about 88% to endorse 'neutral' in item 9 and 'agree' in item 8; and a probability of about 50% to endorse 'agree' in item 9 (as pointed out by the two arrows respectively).

According to Linacre (1999), in order to prove if a set of Likert-type items is functioning well, the recommended minimal number of responses per category (per Likert scale) is 10. Table 4.5 shows that the current sample fulfilled this requirement (the observed count for all of the 5 category labels is greater than 10).

Table 4.5 Summary of Category Structure of SOS

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	OUTFIT MNSQ	STRUCTURE CALIBRATION
1	1	18	1	-.80	1.36	NONE
2	2	196	14	-.38	1.12	-3.31
3	3	575	41	.21	.85	-1.15
4	4	533	38	1.33	.90	.87
5	5	88	6	2.53	1.09	3.59

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

The observed average is the average of the ability estimates for all persons in the sample who chose that particular response category. For category 1, the observed average is -.80, this is the ability estimate or logit score for the persons who chose category 1 on any item in the questionnaire. Linacre (1999) states that the observed average is expected to increase as the variable increases, which means that those with higher ability (stronger attitudes) endorse the higher categories, whereas those with lower ability (weaker attitudes) endorse the lower categories. Again, the figure below shows the SOS is functioning well for this group of students (the observed average increases consistently across the rating scale). Fit statistics provide

another criterion for assessing the quality of rating scales. Outfit mean square greater than 2 indicates more misinformation than information (Linacre, 1999), meaning that the particular category is introducing "noise" into the measurement process. The column of 'OUTFIT MNSQ' above has values all below 2 which shows that each of the rating scales in the SOS meets the criterion. The last column of Table 4.5, 'STRUCTURE CALIBRATION', shows the difficulties estimated for choosing one response category over another (the thresholds) (e.g. how difficult it is to endorse 'strongly agree' over 'agree'). Threshold distances should indicate that each step defines a distinct position on the variable. That is, the estimates should be neither too close together nor too far apart on the logit scale. Guidelines recommend that threshold should increase by at least 1.4 logits, to show distinction between categories, but not more than 5 logits, so as to avoid large gaps in the variable (Linacre, 1999). The figures in Table 4.5 under structure calibration show that the 4 thresholds are 3.31, 2.16 (-3.31- -1.15), 2.02 (0.87- -1.15) and 2.72 (3.59- 0.87) logits apart respectively. Once again, the SOS meets the criteria of a sound scale.

Another way of inspecting the distinction between thresholds is to look at the probability curves. Probability curves provide a visual method of inspecting the distinction between thresholds. The probability of each response is shown across the measurement continuum. According to the Winsteps manual, the curves show how probable is the observation of each category relative to the item measure. Figure 4.7 below shows the probability curves of the SOS. Logits on the x-axes correspond to the item measure, and we can see that response 3, meaning neutral, is the most probable choice for the item at 0 logit (showing a probability of 0.5+). The points of intersection of adjacent categories are the thresholds. The probability curves plot of a well functioning scale should look like a range of hills. Each curve should have a distinct peak illustrating that each rating scale is the most probable response category for some portion of the measured variable. The probability curves below show distinct points on the measure variable. So each response category is the most probable category for some part of the continuum. Once again, the SOS is proved to be functioning well for this target group.

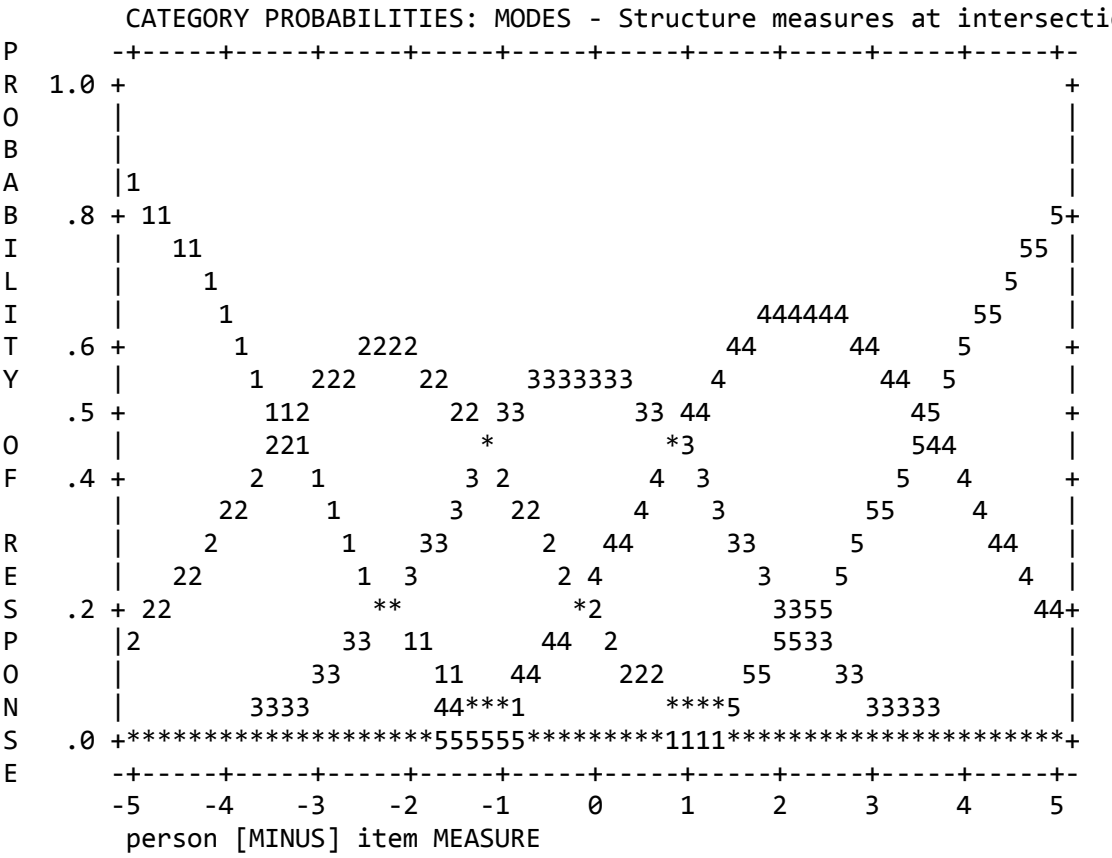


Figure 4.7 Probability Curves of SOS

All of the above Rasch analyses show that the SOS is a valid instrument for measuring test motivation in the context of a low stakes test in Hong Kong.

According to the SOS manual, SOS items are scored on a scale of 1 to 5. The highest score that a person can receive on either subscale is 25. A score of 25 on the Effort subscale means that the student perceived that they fully engaged in effortful behavior on the test and a score of 25 on the Importance subscale means that the student thought it important to do well on the test. Since under the Rasch model, each item is given a difficulty estimate based on the actual empirical evidence, each item does not carry the same relative value in the indication of motivation. Tables 4.6 and 4.7 contain the Rasch measures and model standard errors corresponding to every possible score in the SOS. For easier explanation and comparison, the tables show the SOS scores with all ten items. The measures are supposed to give a more precise estimate of students' motivation in the SOS than by just adding up the raw scores.

In this study, by adding up the raw scores, the average score of the 141 students in the Importance and Effort scale is 17.33 and 16.09 respectively. According to tables 4.6 and 4.7, a score of 17 in the Importance scale and a score of 16 in the Effort scale should be 15.98 and 15.55 under the Rasch analysis. Taking into account the standard errors, both the addition of raw scores and Rasch analysis can be said to be resulting in similar results.

Table 4.6 Table of Measures of Items on Importance

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
5	5.57	2.61	12	12.20	.93	19	17.82	1.16
6	7.45	1.53	13	12.85	.96	20	18.82	1.20
7	8.72	1.18	14	13.54	1.00	21	19.88	1.22
8	9.59	1.04	15	14.30	1.04	22	20.95	1.23
9	10.31	.97	16	15.12	1.08	23	22.09	1.30
10	10.96	.93	**17**	**15.98**	1.10	24	23.53	1.59
11	11.58	.92	18	16.88	1.13	25	25.49	2.63

Table 4.7 Table of Measures of Items on Effort

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
5	5.00	2.62	12	12.79	1.00	19	17.77	1.05
6	6.94	1.58	13	13.50	.98	20	18.59	1.08
7	8.34	1.28	14	14.19	.97	21	19.44	1.11
8	9.42	1.18	15	14.87	.97	22	20.37	1.17
9	10.37	1.12	**16**	**15.55**	.98	23	21.47	1.30
10	11.24	1.07	17	16.25	1.00	24	22.96	1.63
11	12.04	1.03	18	16.99	1.03	25	25.00	2.67

The SOS manual does not provide suggestions as to the cut off scores of low, medium or high motivation. It is up to the users of the scale to interpret. For example if we decide to set 20 as the cut off score of the Importance scale and anything above it indicates high motivation due to the importance factor, then according to the measures by Winsteps, a person should need to score 22 (true score=20.95) instead to be regarded as highly motivated. Winsteps analysis is providing a more accurate estimation with the difficulty level of each item as well as the difficulty level of each response in the Likert scale being accounted for. Nevertheless, given the standard errors, the data in this study does not show significant differences between the two methods of analysis.

4.1.1.3 Findings of SOS

The following tables show the statistics of students' responses to the SOS:

Table 4.8 Responses to the SOS-Importance questions (n=141)

	Strongly Agree		Agree		Neutral		Disagree		Strongly Disagree	
1. Doing well on the DELTA was important to me.	4	2.8%	61	43.3%	68	48.2%	8	5.7%	0	0.0%
3. I am not curious about how I did on the DELTA relative to the other students.*	2	1.4%	22	15.6%	57	40.4%	48	34.0%	12	8.5%
4. I am not concerned about the scores I receive on the DELTA.*	2	1.4%	13	9.2%	40	2.8%	71	50.4%	15	10.6%
5. The DELTA was an important test to me.	4	2.8%	15	10.6%	81	57.4%	29	27.0%	2	21.3%
8. I would like to know how well I did on the DELTA.	21	14.9%	89	63.1%	29	20.6%	2	1.4%	0	0.0%

Table 4.9 Responses to the SOS Effort questions (n=141)

	Strongly Agree		Agree		Neutral		Disagree		Strongly Disagree	
2. I engaged in good effort throughout the test	7	5.0%	75	53.2%	49	34.8%	9	6.4%	1	0.7%
6. I gave my best effort on the test.	17	12.1%	55	39.0%	54	38.3%	14	9.9%	1	0.7%
7. While taking the DELTA, I could have worked harder on the questions.*	5	3.5%	60	42.6%	65	46.1%	9	6.4%	2	1.4%
9. I did not give the DELTA my full attention while completing it.*	3	2.1%	51	36.2%	38	27.0%	45	31.9%	4	2.8%
10. While taking the DELTA, I was able to persist until I had completed all the questions.	8	5.7%	65	46.1%	57	40.4%	10	7.1%	1	0.7%

Table 4.8 clearly indicates that the DELTA test was not an important test for the students: 27% and 21.3% of the students disagreed and strongly disagreed that the DELTA was an important test to them (question 5) (with 57.4% choosing neutral in this question). However, 14.9% strongly agreed and a high of 63.1% agreed that they would like to know how well they did on

the DELTA (question 8); and 10.6% + 50.4% signified that they were concerned about their DELTA scores (question 4). These results indicate that even when the test was not very important to them, they still wanted to know how they had performed. These descriptive statistics would lead us to believe that the DELTA is at least moderately important to the students.

In question 2 of the Effort scale, slightly over half of the students (53.2%) agree that they engaged in good effort throughout the test and 39% believed that they gave their best effort on the test (question 6). Question 9 has a conflicting result with 36.2% of the students choosing agree and 31.9% choosing disagree. Another clear observation in this scale is that a large proportion of students chose neutral in these 5 questions. From these descriptive statistics, we would think that students' effort on the test was moderate to medium high.

Running the analysis of the SOS using the RASCH model provided us with a more sophisticated way of interpreting the survey results:

Table 4.10 Summary statistics of SOS

```
Student Opinion Scale May 2012  ZOU704WS.TXT  Dec  5 10:36 2012
INPUT: 141 person  10 item  REPORTED: 141 person  10 item  5 CATS WINSTEPS 3.74.0
-------------------------------------------------------------------------------

     SUMMARY OF 141 MEASURED person
-------------------------------------------------------------------------------
|          TOTAL                          MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      33.4      10.0         .69     .49      1.01    -.2   1.01    -.2 |
| S.D.       4.0         .0        1.00     .05       .77    1.5    .78    1.5 |
| MAX.      49.0      10.0        6.06    1.08      5.68    5.7   5.99    5.9 |
| MIN.      24.0      10.0       -1.41     .46       .13   -3.3    .13   -3.3 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .56 TRUE SD     .83  SEPARATION  1.50  person RELIABILITY  .69 |
|MODEL RMSE    .49 TRUE SD     .87  SEPARATION  1.78  person RELIABILITY  .76 |
| S.E. OF person MEAN = .08                                                   |
-------------------------------------------------------------------------------
person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) person RAW SCORE "TEST" RELIABILITY = .71
```

Table 4.10 shows the summary statistics of the SOS of the 141 measured persons. The mean measure of the persons is 0.69. The least able student had an ability estimate of -1.41 and the most able student had an ability estimate of 6.06.

The spread of person is determined by calculating the person strata. This statistic indicates the number of statistically distinct levels of person difficulty separated by at least three errors of measurement (Wright & Masters, 2002). Person strata is calculated using the following formula:

$$\text{Person strata} = \frac{(4 \times \text{item separation}) + 1}{3}$$

The person strata of the SOS is 2.71 and this indicates that the students can be grouped into about three statistically distinct levels of motivation.

$$\text{Person strata} = \frac{(4 \times 1.78) + 1}{3} = 2.71$$

The Winsteps manual states that strata=3 means very high, middle, and very low measures can be statistically distinguished. Figure 4.8 is the item-person map which illustrates the distributions of the 3 motivational levels of students. The numbers on the right are the students who responded to the questionnaire and the crosses on the left column are the 10 actual items. They are plotted on the same scale. Student(s) on the top band are those who indicated high motivation in the SOS and students at the bottom are those who indicated that they are not really motivated. In this study, only 1 student can be regarded as highly motivated.

Although Figure 4.5 shows that the person mean is higher than the item mean and also a large proportion of students is ranking higher than the items along the scale, still only 1 student (0.7%) is considered as highly motivated and 41 students (29%) moderately motivated by Winsteps. The reason for this is because the item 'endorsability' is low, meaning that the items carry low motivational values. Even when a student endorses most of the items, it does not mean that the

student is really highly motivated. This highlights a major difference between the traditional analysis and the analysis using the Rasch model.

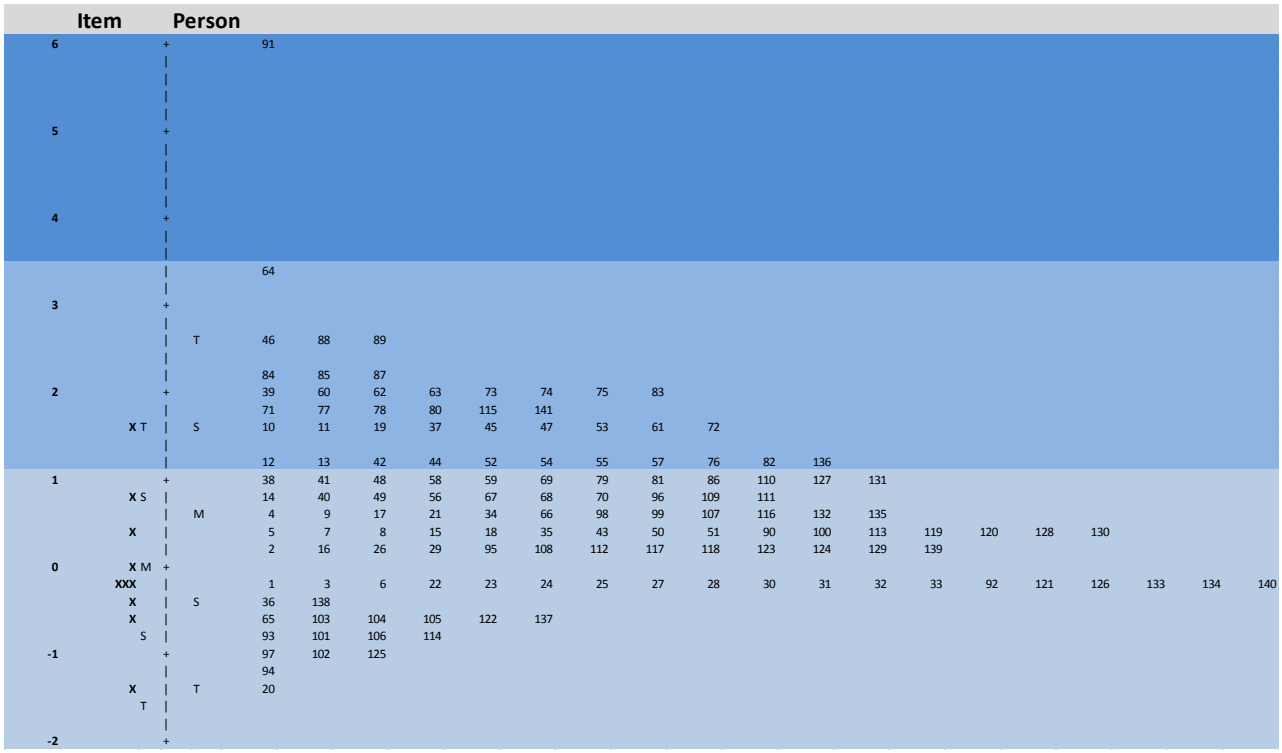


Figure 4. 8 Item-Person Map in 3 motivation levels

In order to provide a more general and descriptive explanation of the results, I grouped the 10 SOS questions according to their literal meaning. I took out the misfitting item 7 "while taking the DELTA, I could have worked harder on the questions" and the problematic item 10 "while taking the DELTA, I was able to persist until I had completed all the questions" in the Rasch analysis. Sundre and Finney (2002) conducted a validation study of the SOS and concluded that since item 1 and 5 are redundant in wordings and item 5 has a better item fit, they suggested removing item 1; therefore I also took out item 1 (doing well on the DELTA was important to me) and renamed item 5 (the DELTA was an important test to me) as 'Test has personal significance'.

Based on the Item-person map in Figure 4.8 and also with reference to the Person-Item map (Figure 4.5) earlier on, I came up with the following characteristics of the students and the number of students in the respective categories.

Student(s) with the lowest level of motivation are those who only want to know the score. In this study, there is only 1 student showing this low level of motivation (0.7%). As motivation increases, students are also concerned about their scores (14 out of 141 students = 9.9%). Those who have higher level of motivation in the test will give best/good effort in the test and they also want to compare score with others (34 students = 24.1%). 28 of the students (19.9%) apart from agreeing with all of the above statements, they also find that the test has personal significance to them. Students with the highest level of motivation as detected by the SOS are those who can be characterized by all the features listed in Table 4.11 (64 out of 141 students= 45.4%). These students endorsed all the items. The difference between this group of students and the rest is that they concentrated during the test.

Table 4.11 Characteristics of students based on SOS Winsteps analysis

	Characteristics	No. of students	(n=141)
Higher test motivation	Concentrates during test (Q.9)	64	45.4%
	Test has personal significance (Q.5)	28	19.9%
	Wants to compare his/her score with others (Q.3)	34	24.1%
	Give best/good effort in test to perform well (Q.6 & Q.2))	14	9.9%
	Concerned about score (Q.4)	1	0.7%
	Wants to know the score (Q.8)		

4.1.2 The Report Usefulness Scale as an instrument of measuring the perceived usefulness of diagnostic report

4.1.2.1 Internal structure evidence of the Report Usefulness Scale

Before a person finds a diagnostic report useful, they have to understand or know how to read it first. In the second section of the report, Questions 11 and 12 are about whether students understand the report. The majority of the students agree that the report as a whole is easy to understand (45.8%) and that they understand the meanings of the subkills listed under each of the components (81-84/141).

The DELTA Report as a whole is easy to understand.		
Answer Options	Response Percent	Response Count
Strongly agree	4.2%	6
Agree	45.8%	66
Neutral	34.7%	50
Disagree	11.1%	16
Strongly disagree	2.1%	3
answered question		141

In the DELTA report, under each component (Listening, Reading, Vocabulary and Grammar) shows a breakdown of the subskills (e.g. Listening: identifying specific information). I understand the meanings of the subskills listed in each of the components.						
Answer Options	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Response Count
Listening	5	84	41	10	1	141
Reading	5	84	44	8	0	141
Vocabulary	4	81	50	6	0	141
Grammar	5	81	50	5	0	141
answered question						141

As explained in the Methodology chapter, question 13 of my questionnaire consists of 8 questions which form the Report Usefulness Scale. Same as for the SOS, exploratory factor analysis (principal components analysis) was run and the curve in the scree plot flattens out at the second factor so only one factor should be extracted in this Scale. The component matrix below also shows that there is only one factor in this series of questions, i.e. perceived usefulness of DELTA report.

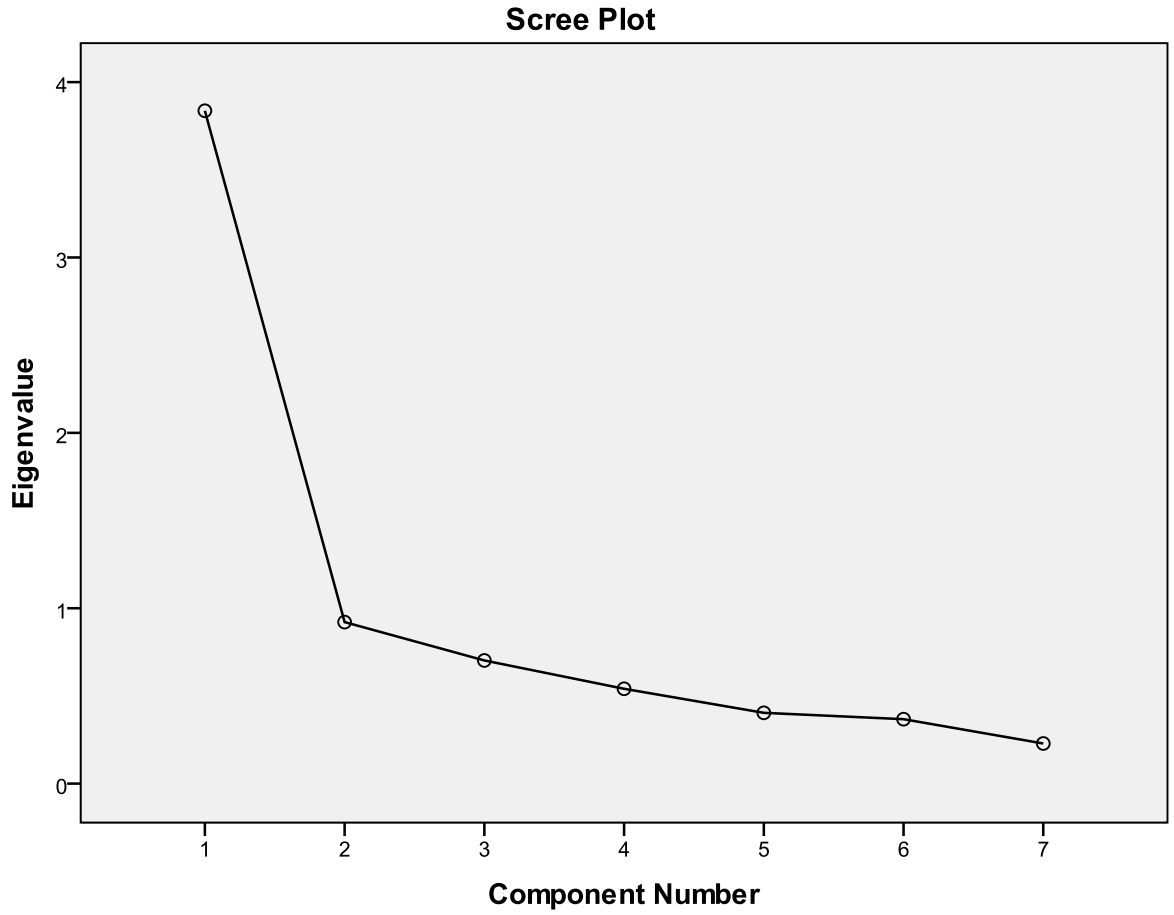


Figure 4.9 Scree Plot of Report Usefulness Scale

Table 4.12 Component Matrix of Report Usefulness Scale (8Qs)

Component Matrix[a]

The DELTA report can:	Component
	1
(Q.1) allow me to refer to it if I want to do self-study on English.	.727
(Q.2) encourage me to seek help from an English Teacher.	.617
(Q.3) tell me if I am making any progress	.801
(Q.4)tell me my English proficiency	.780
(Q.5) tell me my strengths and weaknesses.	.809
(Q.6) allow me to compare my results with other students	.557
(Q.7) guide me in how to prioritize my English learning.	.756
(Q.8) motivate me to improve my English.	.742

Extraction Method: Principal Component Analysis.

Apart from factor analysis, the internal consistency of the Report Usefulness Scale was also examined. The Cronbach's Alpha of the Report Usefulness Scale is .862 which is generally accepted as satisfactory.

4.1.2.2 Validating the Report Usefulness Scale using the Rasch model

Figure 4.10 below shows the distributions of item endorsability and person ability estimates of the Report Usefulness Scale on the same logit scale.

The figure shows that the question 'The DELTA report can tell me my strengths and weaknesses' is the easiest to endorse. The most 'difficult' to endorse question is, 'The DELTA report can encourage me to seek help from an English teacher'. We can see from the figure that again there are a number of students higher on the scale who show higher test motivation than all the Report Usefulness items which means that those students will probably endorse 'strongly agree' or 'agree' in most if not all of the questions. (the # above the line, showing higher 'ability' than the 'difficulty' level of the items).

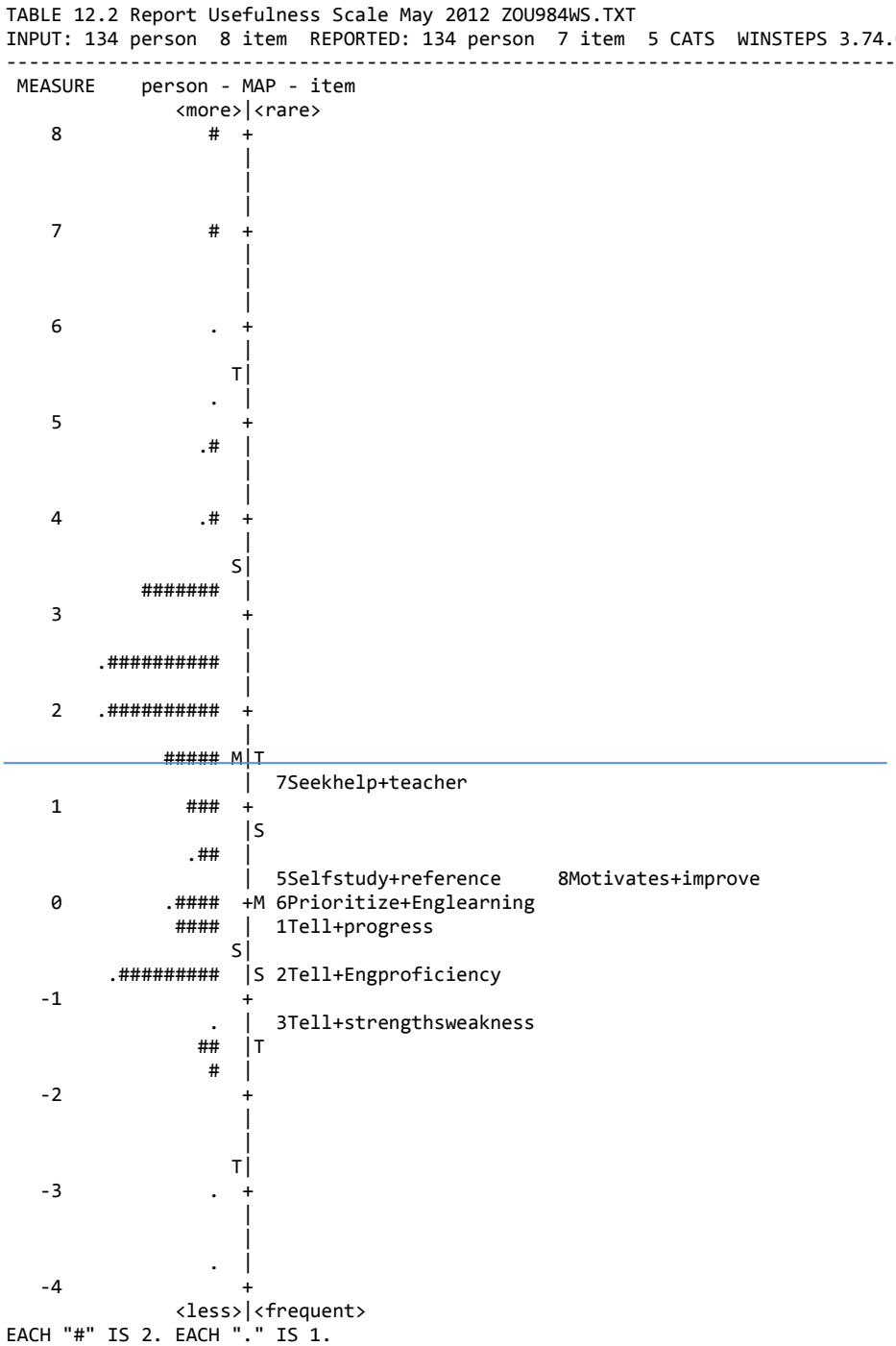


Figure 4.10 Person-Item Map of Report Usefulness Scale

In order to validate whether the 8 questions in the Scale are functioning properly, the same Rasch analysis as with the SOS is done.

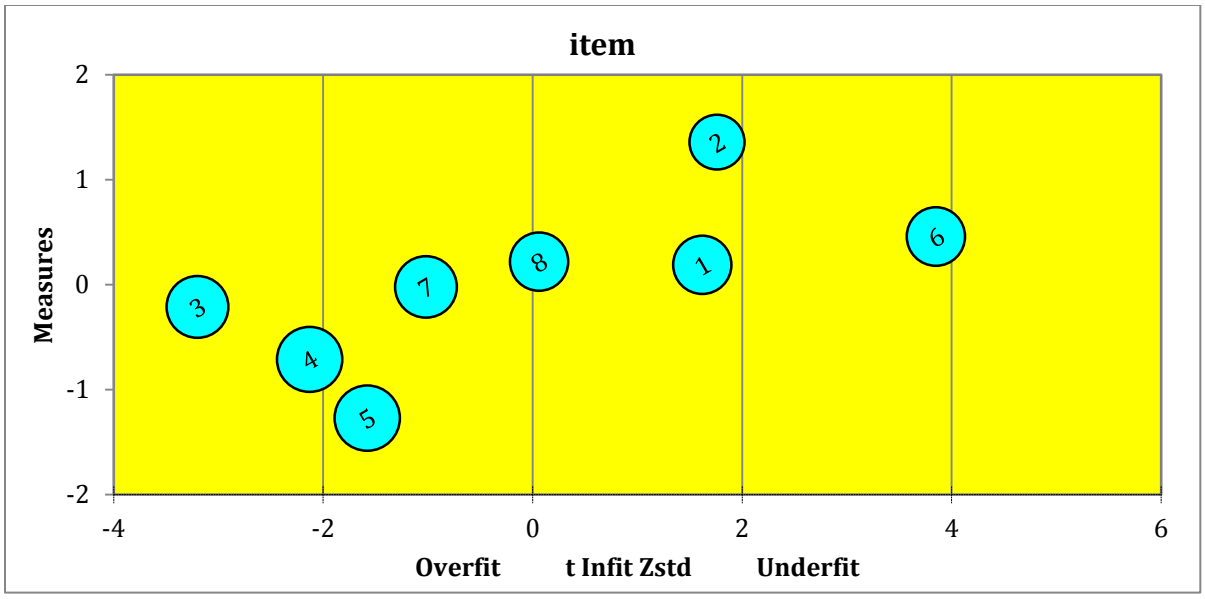


Figure 4.11 Bubble Chart of Report Usefulness Scale

The above bubble chart generated by Winsteps shows that question number 6 'The DELTA report allows me to compare my results with other students' is 'underfitting' (Infit Zstd >2), indicating erratic response pattern is found in this item. The column of infit msq and outfit msq in Table 4.13 also shows that question 6 is underfitting (mean squares >1.3).

Table 4.13 Fit Statistics of Report Usefulness Scale

ENTRY	COUNT	ERROR	IN.MSQ	OUT.MSQ	NAME
6	134	0.28	1.6334	1.6326	6Compareresults
1	134	0.29	1.1996	1.2475	1Selfstudy+reference
2	134	0.26	1.1468	1.224	2Seekhelp+teacher
8	134	0.29	0.9835	1.0017	8Motivates+improve
7	134	0.29	0.8155	0.8459	7Prioritize+Englearning
5	134	0.31	0.7286	0.7495	5Tell+strengthsweakness
4	134	0.31	0.6955	0.6841	4Tell+Engproficiency
3	134	0.3	0.6393	0.5756	3Tell+progress

Reasons why students do not agree that they can use the DELTA measure to compare scores with other students may be because DELTA is a new test and not as widely recognized yet as tests like IELTS. More evidence would be revealed from the in-depth interviews. In fact although Q.6 fits the factor analysis as shown in Table 4.12 (a loading of 0.512 or above can be considered as significant according to Stevens (2002)), Q.6 has the least loading (0.557)

compared with the other 7 questions. For the purpose of higher data analysis accuracy, question 6 was taken out from the analysis and the data was rerun using factor analysis. With Q.6 being taken out, most of the loadings of the other questions improved further.

Table 4.14 Component Matrix of Report Usefulness Scale (7Qs)

Component Matrix[a]

The DELTA report can:	Component
	1
(Q.1) allow me to refer to it if I want to do self-study on English.	.749
(Q.2) encourage me to seek help from an English Teacher.	.634
(Q.3) tell me if I am making any progress	.793
(Q.4)tell me my English proficiency	.781
(Q.5) tell me my strengths and weaknesses.	.818
(Q.7) guide me in how to prioritize my English learning.	.744
(Q.8) motivate me to improve my English.	.753

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

Table 4.15 Summary of Category Structure of Report Usefulness Scale

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	OUTFIT MSQ	STRUCTURE CALIBRATION
1	1	8	1	-2.56	1.46	NONE
2	2	66	6	-1.03	1.46	-4.26
3	3	373	35	-.02	.86	-2.40
4	4	561	52	2.33	.84	.77
5	5	64	6	5.58	1.03	5.88

OBSERVED AVERAGE is mean of measures in category.
It is not a parameter estimate.

From the summary of category structure in Table 4.15, we can see that the 7 report usefulness items are functioning well. As explained in Section 4.1.1.2 during the validation of the SOS, the observed average is expected to increase as the variable increases as this means that those with higher ability (stronger attitudes) endorse the higher categories. Here we can see that the

observed average increases monotonically across the rating scale. Structure calibrations show that the thresholds increased by at least 1.4 logits but no more than 5 logits. Fit statistics (outfit mean squares) shows no misfit in each of the rating scale thresholds (misfit = bigger than 2).

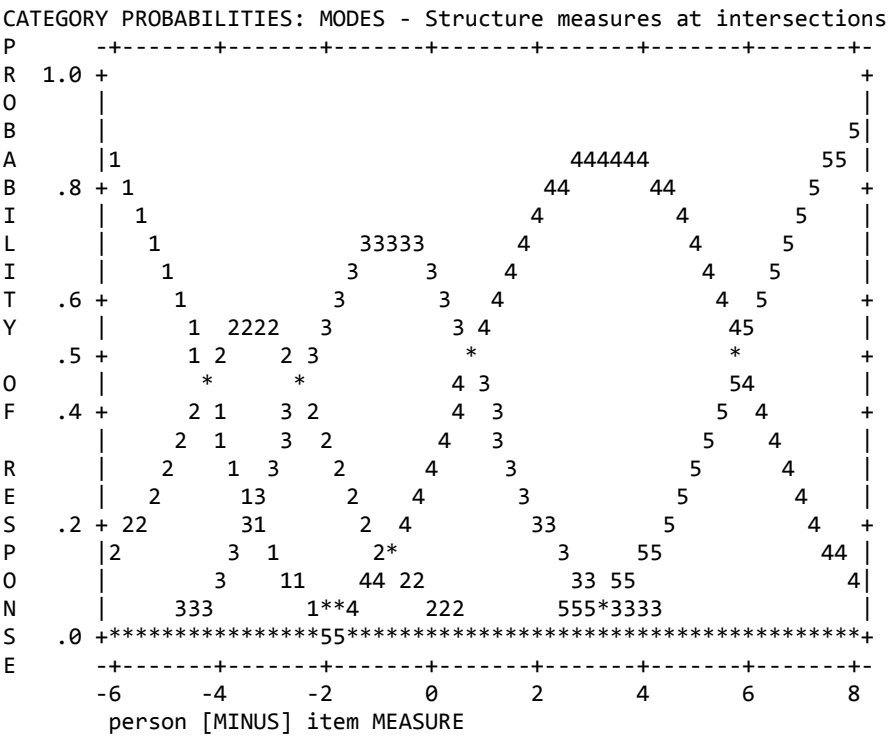


Figure 4.12 Probabilty Curves of Report Usefulness Scale

Similar to the previous analysis of the SOS, the probability curves presented in Figure 4.12 provide a visual method of inspecting the distinction between thresholds. Logits on the x-axes correspond to item-measure. As explained, probability curves of a well functioning scale are in a range of hills with distinct peak for each curve. Once again, the 7 report usefulness items are proved to be well functioning for this target group.

Under the Rasch model, each question in the report usefulness scale is given a difficulty estimate based on the actual empirical evidence. Each question does not carry the same relative value in indicating report usefulness. Table 4.14 lists the Rasch measures and model standard errors corresponding to every possible score in the scale. The measures provide a more precise estimate of students' perceived report usefulness. In this study, by adding up the raw scores, the average score of the students in the Report Usefulness Scale is 31.7 while the total score is 40. According to Table 4.16, a score of 32 in the Report Usefulness Scale should be 25.28 under the Rasch

analysis. Even if the standard errors are taken into account, the values resulted from pure adding up of the scores are quite different from those from the Rasch analysis.

Table 4.16 Table of Measures of Report Usefulness Scale

SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.	SCORE	MEASURE	S.E.
8	6.89	3.09	19	15.65	.91	30	22.94	1.27
9	9.04	1.77	20	16.16	.93	31	24.01	1.40
10	10.41	1.32	21	16.70	.95	32	25.28	1.49
11	11.31	1.14	22	17.27	.98	33	26.57	1.42
12	12.02	1.04	23	17.86	1.01	34	27.69	1.30
13	12.63	.97	24	18.50	1.03	35	28.65	1.23
14	13.18	.94	25	19.15	1.05	36	29.55	1.22
15	13.69	.91	26	19.83	1.07	37	30.48	1.27
16	14.19	.90	27	20.53	1.09	38	31.55	1.41
17	14.67	.90	28	21.26	1.12	39	33.06	1.82
18	15.16	.90	29	22.04	1.17	40	35.29	3.12

4.1.2.3 Findings of the Report Usefulness Scale

The following two tables show the statistics of 134 students' responses (there were 141 respondents to the SOS questions and 7 students did not complete the questions of the Report Usefulness Scale) to the Report Usefulness Scale:

Table 4.17 Responses to the Report Usefulness Scale-Action (n=134)

	Strongly agree		**Agree**		**Neutral**		**Disagree**		**Strongly disagree**	
1. I will refer to my Report if I want to do self-study on English.	10	7.5%	65	48.5%	48	35.8%	9	6.7%	2	1.5%
2. Having a DELTA report will encourage me to seek help from an English teacher.	5	3.7%	44	32.8%	65	48.5%	18	13.4%	2	1.5%
6. The DELTA report allows me to compare my results with other students.	8	6.0%	65	48.5%	46	34.3%	13	9.7%	2	1.5%
8. The DELTA report motivates me to improve my English.	8	6.0%	66	49.3%	48	35.8%	10	7.5%	1	0.7%

Table 4.18 Responses to the Report Usefulness Scale-Information (n=134)

	Strongly agree		Agree		Neutral		Disagree		Strongly disagree	
3. The DELTA report can tell me if I am making any progress.	6	4.5%	74	55.2%	50	37.3%	4	3.0%	0	0.0%
4. The DELTA report can tell me my English proficiency.	7	5.2%	86	64.1%	39	29.1%	2	1.5%	0	0.0%
5. The DELTA report can tell me my strengths and weaknesses.	16	11.9%	86	64.1%	30	22.4%	2	1.5%	0	0.0%
7. The DELTA report can guide me in how to prioritize my English learning.	7	5.2%	74	55.2%	46	34.3%	6	4.5%	1	0.7%

In table 4.19, a large proportion of the students agreed or was neutral when they were asked if the DELTA report promoted their incentive in taking action in improving their English while from Table 4.18, we can see that most of the students agreed that the DELTA report provided them with useful information.

Again, running the results of the Report Usefulness Scale using the Rasch model provided us with a more sophisticated way of interpreting the survey results. Table 4.19 shows the summary statistics of the 134 measured students.

Table 4.17 Summary statistics of Report Usefulness Scale

```
Report Usefulness Scale Carrie May 2012  ZOU648WS.TXT
INPUT: 134 person  8 item  REPORTED: 134 person  8 item  5 CATS  WINSTEPS 3.74.0
-------------------------------------------------------------------------------
      SUMMARY OF 134 MEASURED person
-------------------------------------------------------------------------------
|          TOTAL                        MODEL         INFIT        OUTFIT     |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      28.5       8.0       22.18    1.22      1.01    -.2   1.00    -.2 |
| S.D.       4.0        .0        3.43     .18      1.02    1.5   1.06    1.5 |
| MAX.      39.0       8.0       33.05    1.82      6.59    4.4   7.12    4.5 |
| MIN.      15.0       8.0       13.70     .90       .08   -2.3    .07   -2.3 |
|-----------------------------------------------------------------------------|
| REAL RMSE   1.44 TRUE SD    3.11  SEPARATION  2.16  person RELIABILITY  .82 |
|MODEL RMSE   1.24 TRUE SD    3.20  SEPARATION  2.58  person RELIABILITY  .87 |
| S.E. OF person MEAN = .30                                                   |
-------------------------------------------------------------------------------
person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) person RAW SCORE "TEST" RELIABILITY = .86
```

The person strata of the Report Usefulness Scale is 3.77 and this indicates that the persons can be grouped into about four statistically distinct levels.

$$\text{Person strata} = \frac{(4 \times 1.78) + 1}{3} = 3.77$$

Figure 4.13 is the item-person map which illustrates the distributions of the 4 levels of students' perceived usefulness of the DELTA report. The numbers on the right are the students who responded to the questionnaire and the crosses on the left column are the actual questions. They are plotted on the same scale. Student(s) on the top band are those who found the report more useful and students at the bottom are those who indicated that they found the report not as useful.

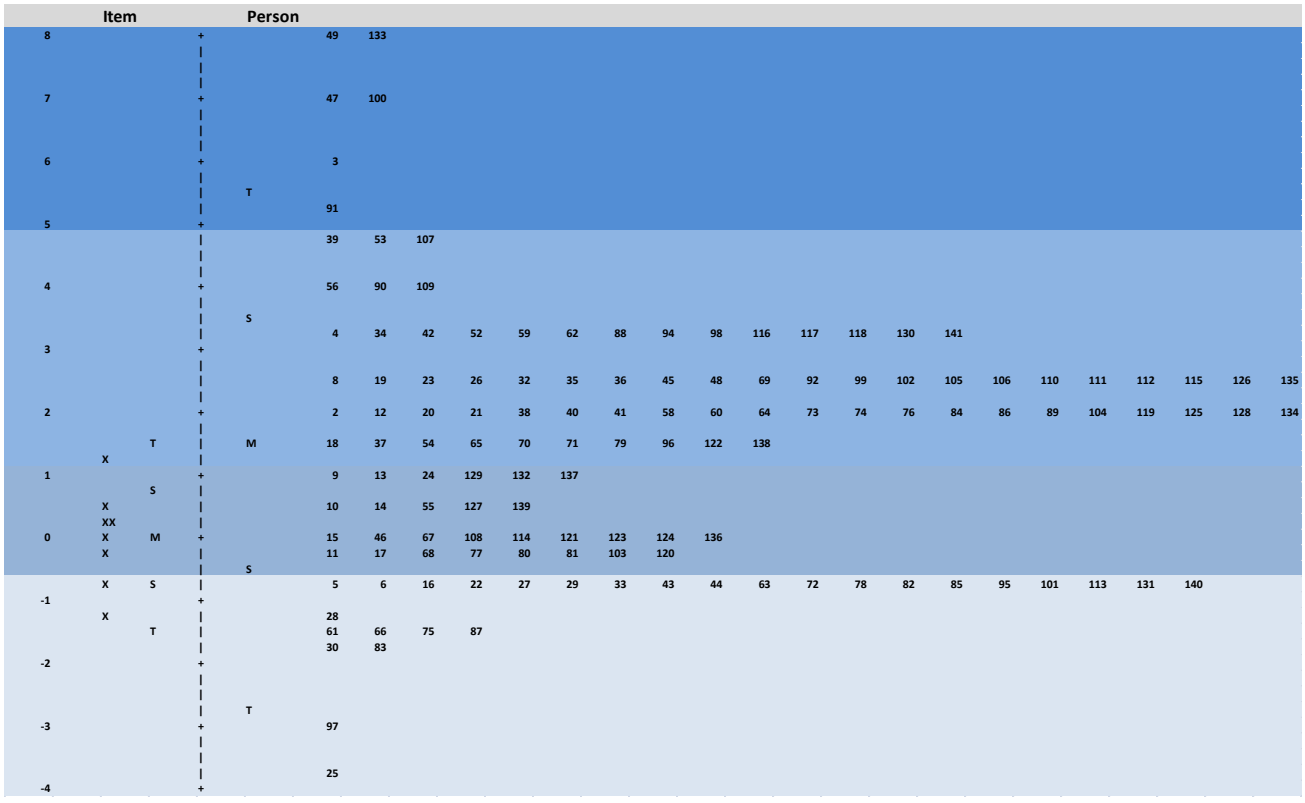


Figure 4.13 Item- Person map of 134 students on Report Usefulness Scale

Similar to the SOS analysis, although Figure 4.13 shows that the person mean is higher than the item mean and also a large proportion of students is ranking higher than the items along the scale,

according to Winsteps, only 6 and 72 students find the report very useful and useful respectively (55.3% for the two groups together). The reason for this is again because of the relatively low item 'endorsability', meaning that the items carry low values in the indication of report usefulness. Even a student who endorses most of the items in the Report Usefulness Scale does not necessarily find the report highly useful. The results, however, are better than those of the SOS (only 1 student can be regarded as highly motivated according to the Winsteps analysis of the SOS), meaning that although students are not highly motivated in the DELTA, the majority of them find the DELTA report useful. The traditional analysis of the Report Usefulness Scale at the beginning of this section can said to be in line with the results of the Rasch analysis.

In order to provide a more general and descriptive explanation of the results of the Report Usefulness Scale, I grouped the 8 items according to their literal meaning. I took out the misfitting item 6 'The DELTA report allows me to compare my results with other students' and grouped the three 'tell' items: 'Q1 The DELTA report can tell if I am making any progress', 'Q2 The DELTA report can tell me my English proficiency' and 'Q3 The DELTA report can tell me my strengths and weaknesses' into 'student satisfied with basic information of report'. Then I put 'Q5 The DELTA report can allow me to refer to it if I want to do self-study on English', 'Q6 The DELTA report can guide me in how to prioritize my English learning' and 'Q8 The DELTA report can motivate me to improve my English' together as students are being 'reflective about their English learning'. The last item 'Q7 The DELTA report can encourage me to seek help from an English teacher' remains as 'seek help from teacher'.

Based on the Item-person map in Figure 4.13 and also with reference to the Person-Item map (Figure 4.10) earlier on, I came up with the following characteristics of the students and the number of students in the respective categories. Students who find the report less useful (36 out of 134= 26.9%) are those who believe that it can only provide basic information such as their strengths and weaknesses, and their English proficiency.

For those who find the report more useful (20 students =14.9%), apart from the basic information, they think that the report can also inform them of their progress, help them to reflect upon their own English learning, such as they can refer to the report when they do self-study as well as

guiding them in prioritizing their English learning. They also agree that the report motivates them to improve their English. Those who find the report very useful (78 students, 58.2%) are the ones who would not only reflect upon their own English learning but may use the report to take a step further, i.e. to seek help from an English teacher.

Table 4.20 Students perceived usefulness of the DELTA report

Report Usefulness ↑

Characteristics	No. of students	(n=134)
Seek help from teacher	78	58.2%
Reflective about their English learning	20	14.9%
Satisfied with basic information of report	36	26.9%

All of the above findings show that most of the students find the DELTA report useful. In order to have a better idea of how useful the report is to the students high on the scale, more 'difficult items' are needed in future studies.

4.1.2.4 Relations to other variables: Correlations between the SOS and the Report Usefulness Scale

To find out the 'relations to other variables' as stated in the *Standards for Educational and Psychological Testing* (1999) of the SOS and the Report Usefulness Scale, correlation analysis between the two scales (9 questions from the SOS and 7 questions from the Report Usefulness Scale) was run using SPSS and the results are listed in Table 4.21 below.

There is a significant correlation between the Report Usefulness and the SOS importance scores, meaning that those who find the DELTA important to a certain extent also find the report useful or vice versa. However, the correlation is quite low. One apparent explanation for this is that the importance of a test to a student not only depends on whether the diagnostic report is useful or not. Whether there are other factors that make a test important to the students is further discussed

in the section on students' interviews. The correlation results also show that report usefulness has no relation to the effort score. As the questionnaire survey was anonymous, I do not know the respective DELTA measure of the 141 respondents and so correlations between their test performance and their SOS and Report Usefulness Scale cannot be done.

Table 4.21 Correlations between Importance, Effort and Report Usefulness
Correlations

		Importance score	Effort score	Report usefulness score
Importance score	Pearson Correlation		.341**	.157*
	Sig. (1-tailed)		.000	.032
	N		141	141
Effort score	Pearson Correlation	.341**		.019
	Sig. (1-tailed)	.000		.413
	N	141		141
Report usefulness score	Pearson Correlation	.157*	.019	
	Sig. (1-tailed)	.032	.413	
	N	141	141	

**. Correlation is significant at the 0.01 level (1-tailed).

*. Correlation is significant at the 0.05 level (1-tailed).

4.2 Findings of interviews

The findings of the interviews are in 3 parts. The first section of the interview explored the reasons for having two misfitting items in the SOS based on the previous factor and Rasch analyses. The second section examined students' expectancy and value believes. The third section looked into students' motivation before and after taking the DELTA.

Among the 16 students who came for the interview, 11 were female and 5 were male. They were year one students in the Institute of Textiles and Clothings at the Hong Kong Polytechnic University who enrolled in the English for University Course where the DELTA was used as one of the subject activities.

Figure 4.14 is the item-person map from the Winsteps analysis of the SOS in the previous section. The distributions of the 16 students who came for the interview are highlighted.

```
TABLE 16.3 Student Opinion Scale Carrie May 2012 ZOU909WS.TXT  May 23 17:00 2012
INPUT: 141 person  10 item  REPORTED: 141 person  10 item  5 CATS WINSTEPS 3.72.4
-------------------------------------------------------------------------------

       item - MAP - person
         <rare>|<more>
    6          +  091
               |
               |
               |
               |
    5          +
               |
               |
               |
               |
    4          +
               |
               |
               |  064
               |
    3          +
               |
               |T 046  088  089
               |
               |  084  085  087
    2          +  039  060  062  063  073  074  075  083
               |  071  077  078  080  115  141
          X T|S 010  011  019  037  045  047  053  061  072
               |
               |  012  013  042  044  052  054  055  057  076  082  136
    1          +  038  041  048  058  059  069  079  081  086  110  127  131
          X S|  014  040  049  056  067  068  070  096  109  111
              |M 004  009  017  021  034  066  098  099  107  116  132  135
          X   |  005  007  008  015  018  035  043  050  051  090  100  113 119  120  128
130
               |  002  016  026  029  095  108  112  117  118  123  124  129 139
    0       X M+
          XXX  |  001  003  006  022  023  024  025  027  028  030  031  032 033  092  121
126  133  134  140
            X  |S 036  138
            X  |  065  103  104  105  122  137
              S|  093  101  106  114
   -1          +  097  102  125
               |  094
            X  |T 020
              T|
               |
   -2          +
        <frequ>|<less>
```

Figure 4.14 Distribution of SOS measures of the 16 students

To find out if there are any correlations between their DELTA measure, SOS score and Report Usefulness Score, a correlation analysis was run using SPSS and the results are listed in Table 4.22 below.

Similar to the bigger pool (the 141 students), there is a significant correlation between their SOS importance score and effort score (0.639**) as well as a significant correlation between their Report Usefulness and SOS importance score (0.498*).

Table 4.22 Correlations between DELTA measure, SOS and Report Usefulness (16 students)

Correlations

		Importance score	Effort score	DELTA measure	Report Usefulness
Importance score	Pearson Correlation		.639**	-.222	.498*
	Sig. (2-tailed)		.008	.408	.049
	N		16	16	16
Effort score	Pearson Correlation	.639**		.234	.439
	Sig. (2-tailed)	.008		.382	.089
	N	16		16	16
DELTA measure	Pearson Correlation	-.222	.234		.120
	Sig. (2-tailed)	.408	.382		.658
	N	16	16		16
Report Usefulness	Pearson Correlation	.498*	.439	.120	
	Sig. (2-tailed)	.049	.089	.658	
	N	16	16	16	

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

*. Correlation is significant at the 0.05 level (2-tailed).

4.2.1 Reasons for misfitting items

Students were asked to interpret the two misfitting items in the SOS in their own words in the language which they find comfortable. All of them chose to explain their interpretations in Cantonese.

Question 7 in the SOS is 'while taking the DELTA, I could have worked harder on the questions'. Below is a summary table of the interpretations of the statement by the 16 interviewees.

	Interpretation of the question "while taking the DELTA, I could have worked harder on the questions'.	Count
1	When I was doing the test, I could have done better.	6
2	I should have done better on the test. (A feeling of regret)	2
3	When I was taking the DELTA, I could have done better.	1
4	If it is the DELTA test, I am able to work harder on the questions.	1
5	When I was doing the test, I worked harder on the questions than I normally would.	1
6	When I was doing the DELTA, I should have been able to work harder in reading the questions.	1
7	When I was doing the DELTA, I was exhausted and struggling hard during the test.	1
8	While taking the DELTA, I have paid enough effort in answering the questions.	1
9	While taking the DELTA, I will pay more attention and work harder on the questions than in a non-test situation.	1
10	When I was doing the test, I was doing it with the best of my ability.	1

From the summary table above, we can see that none of the 16 students was able to interpret the statement correctly. The closest interpretation is number 1 'when I was doing the test, I could have done better' with 6 students interpreted the statement as such. This may explain why the item is misfitting in both the Winsteps analysis and factor analysis.

Question 10 in the SOS is 'while taking the DELTA, I was able to persist until I had completed all the questions'. Below is a summary table of the interpretations of the statement by the 16 interviewees.

	Interpretation of the question 'while taking the DELTA, I was able to persist until I had completed all the questions'.	Count
1	When I was taking the DELTA, I was able to persist until I had completed all the questions.	7
2	When I was doing DELTA, I could keep doing until I finished the test.	2
3	When I was taking the DELTA, I was able to persist till the last minute.	1
4	Because this is the DELTA test, I was able to persist and completed all the questions.	1
5	While I was doing this test, I can…erm… I don't know what the word (persist) means…	1
6	When I was taking the DELTA, I could have done better.	1

7	When I was taking the DELTA, I could…hm…I don't know.. I don't know what this means.	1
8	When I took the DELTA, I should have tried my best to complete all the questions.	1
9	When I was doing the test, I tried my best until I finished all the questions.	1

From the summary table above, we can see that interpretations 1-3 are quite near to the original and 10 students were able to interpret the statement correctly or quite close enough. However, there are still 6 out of 16 of them whose interpretations deviate quite a lot from the original meaning. This may explain why the item although is not misfitting in Winsteps analysis, does not fall entirely into the expected factor in the factor analysis.

4.2.2 Constructing the test motivation process model

As discussed in section 2.2.4, based on Dörnyei and Ottó's process model of L2 motivation (1998), I proposed in my IFS that with slight modifications, the same process model could be applied to test motivation. In this study, I tried to look at how this model works on students in low stakes diagnostic test with the help from the results of the students interviews.

The major difference between Dörnyei and Ottó's process model, the revised models in my IFS (one for compulsory exit tests and the other for voluntary exit tests with the major difference of the two in the preactional phase where 'Goal Setting' is absent in the model for compulsory tests) and the new model that I proposed here in this study, lies in the actional phase.

I provide the three sections of the respective models below for easier explanation. Figure 4.15 is the actional phase by Dörnyei and Ottó and Figure 4.16 shows the revised model in my IFS in which Dörnyei and Ottó's general action is renamed as 'Test Preparation' with an additional 'Test Taking' stage. The flow in Figure 4.16 indicates that 'Test Preparation' stage is the major focus of a compulsory exit test with 'Test Taking' as the inevitable outcome.

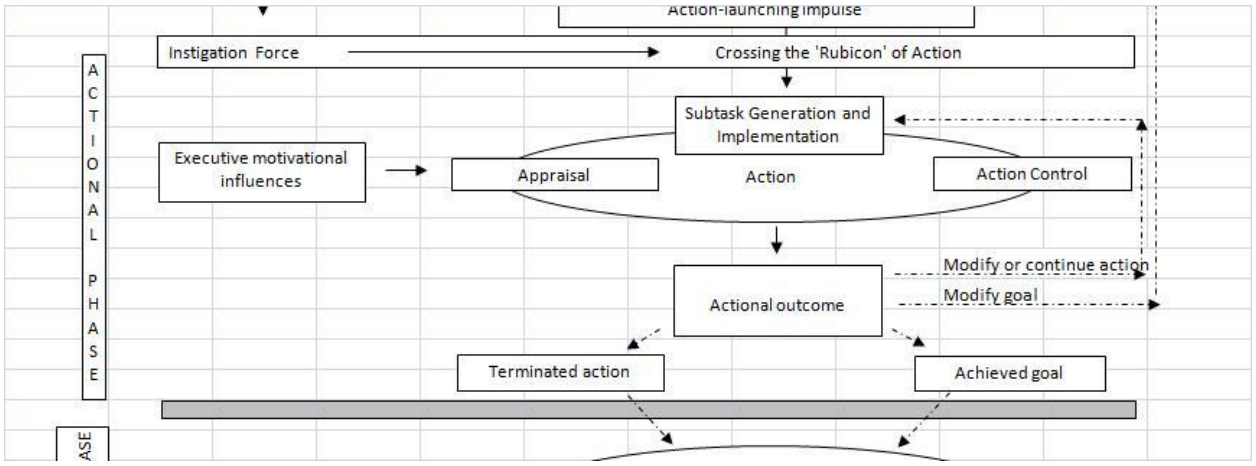


Figure 4.15 Dörnyei and Ottó's (1998) actional phase

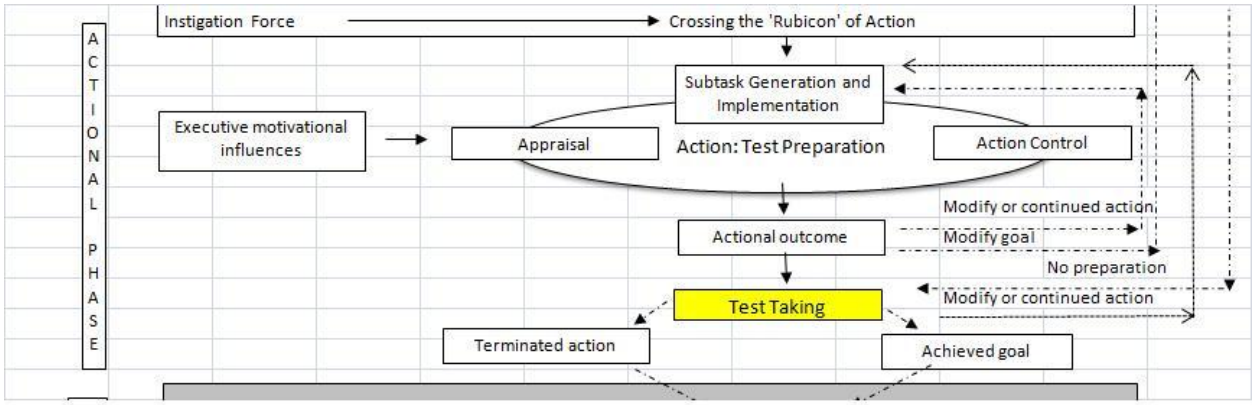


Figure 4.16 Tsang's (2011) actional phase of compulsory exit test

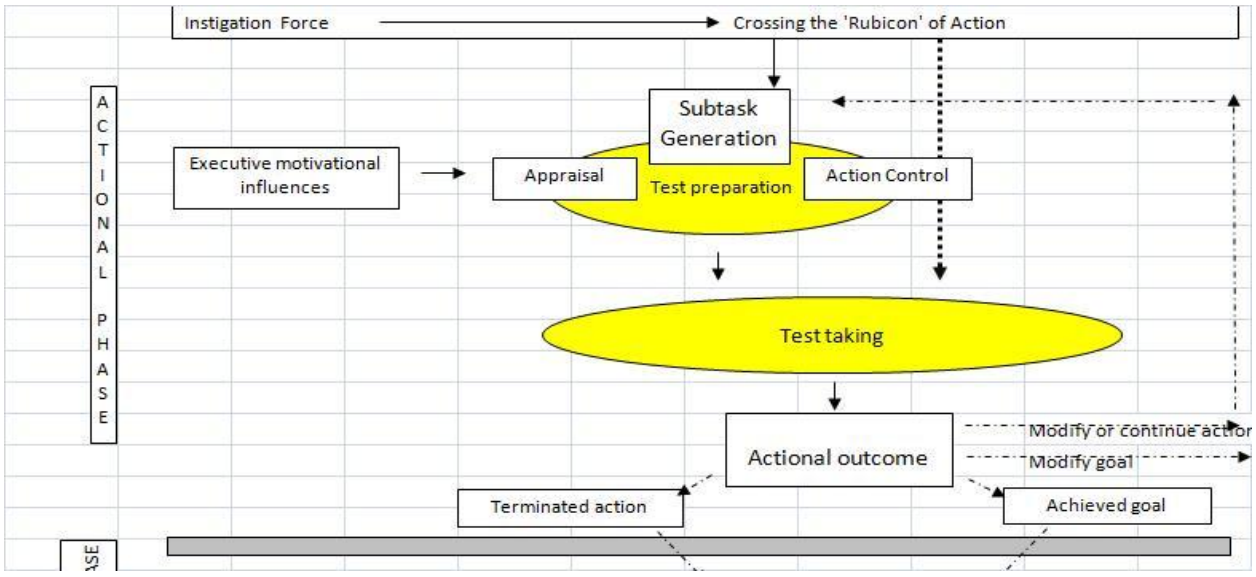


Figure 4.17 Actional phase of diagnostic and tracking test

My revised actional phase emerged from this current study on a diagnostic test shows a slightly different flow (Figure 4.17). The actual action is now in 2 tiers: test preparation and test taking. I also suggest a new dotted line which describes those students whose motivation in the preactional stage is so weak that no planning or actual preparation work would be carried out and they would go and take the test directly. The 'Actional outcome' is also moved to after 'Test taking'. The reason for this move is that for a diagnostic test which tracks students' language progress throughout the years of study, the focus should be on what happens after taking the test. The actional outcome should be after the 'Test taking' stage. The test taking experience together with the diagnostic report as a whole contributes to the actional outcome. Although in Figure 4.16 there is also a dotted line showing the flow of those students who do not prepare for the test, my new model is showing this more clearly.

Figure 4.18 below is the full version of my new process model of test motivation emerged from this study. Since the test taking act in the actional phase has been examined by the SOS, and the major purpose of a low stakes diagnostic test is to impact students in the preactional and postactional phase, my interview and analysis mainly focus on examining the motivational forces of the students in the goal setting stage of the preactional phase as well as in the postactional phase, i.e. if students are motivated before and after taking the DELTA.

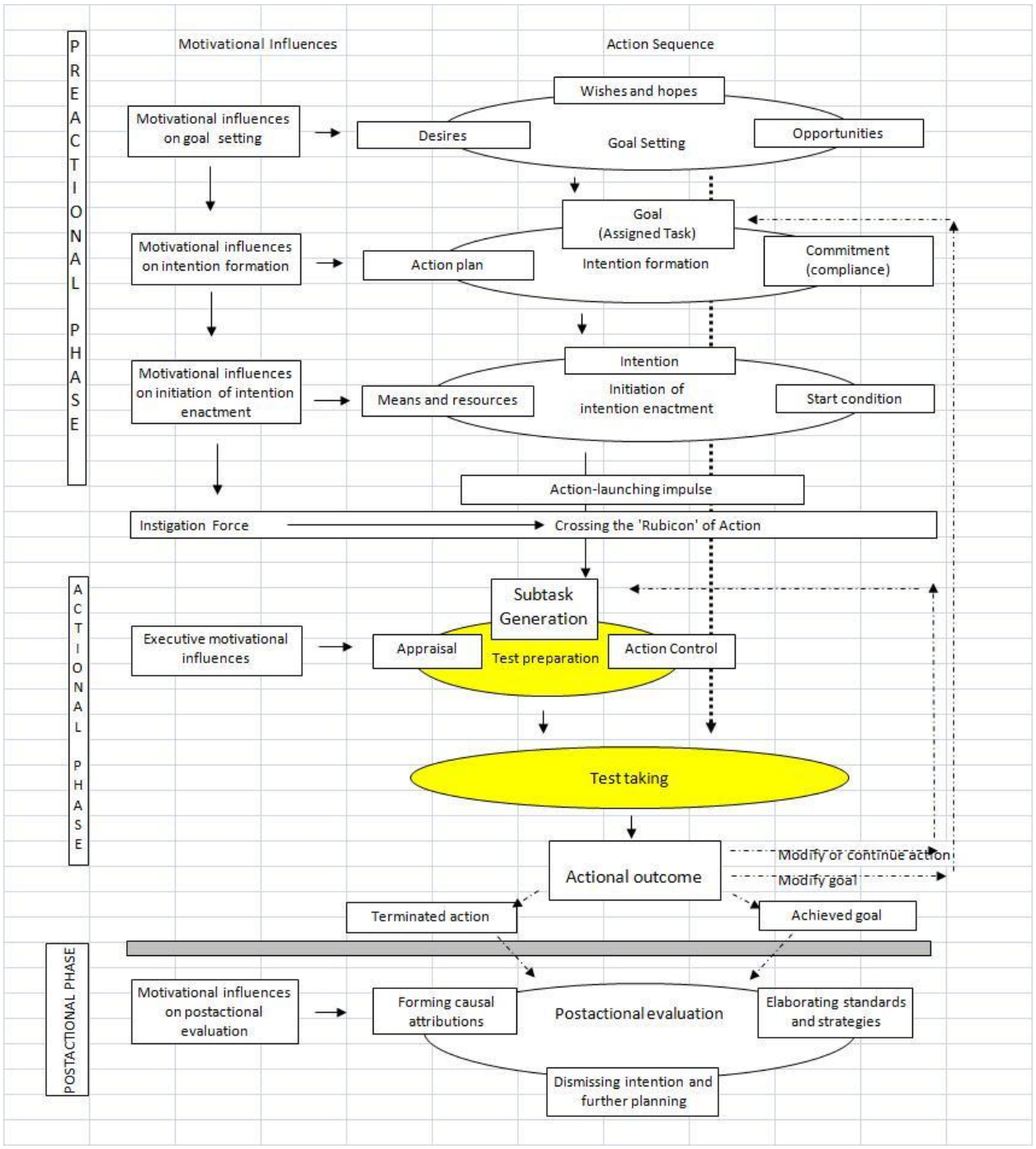


Figure 4.18 Test motivation process model

4.2.2.1 Understanding the preactional phase of the process model—motivation before taking the DELTA

According to the process model of test motivation, those students who reported having high motivation level from the SOS, reflecting high motivational influences in the actional phase, should presumably have higher motivational influences in the preactional phase.

Dörnyei listed a number of factors which would affect the motivational influences on goal setting of L2 learning (2001; 93). By modifying these factors (the originals in brackets) to suit the testing situation, motivational influences which affect goal setting of test preparation and taking should be:

1. Test-related (language-related) subjective values and norms (integrativeness)
2. Incentive values associated with test (L2 learning/proficiency): intrinsic pleasure and instrumental benefits
3. Perceived potency of potential goal
4. Environmental effects; expectations of family members and teachers; school climate

An individual's subjective values and norms in terms of a test are the basic beliefs and feelings about the significance of doing well in the test. The DELTA is a low stakes test without any consequences for the students. All 16 interviewees reported that the DELTA is not important to them. Therefore points 1 and 4 above are not applicable in this circumstance (test with low stakes). According to their SOS scores and using the cut-offs by Winsteps, only 1 student is considered to have high test motivation and 6 with medium level of motivation (Figure 4.19 below). The rest all reported comparatively low motivation. Based on the codings of the interviews (as explained in Chapter 3 Table 3.5), the themes 'expectancy' and 'value' help demonstrate Dörnyei's point 2 and 3. With regards to point 3 above, i.e., the perceived potency of potential goal; of the 7 students (1 high + 6 medium) who showed relatively higher motivation in the SOS, 4 of them reported having high confidence on their own English proficiency. 5 out of the 7 students also indicated that they were satisfied with their past English results. The 7 students all pointed out that the DELTA is of easy to medium difficulty to them.

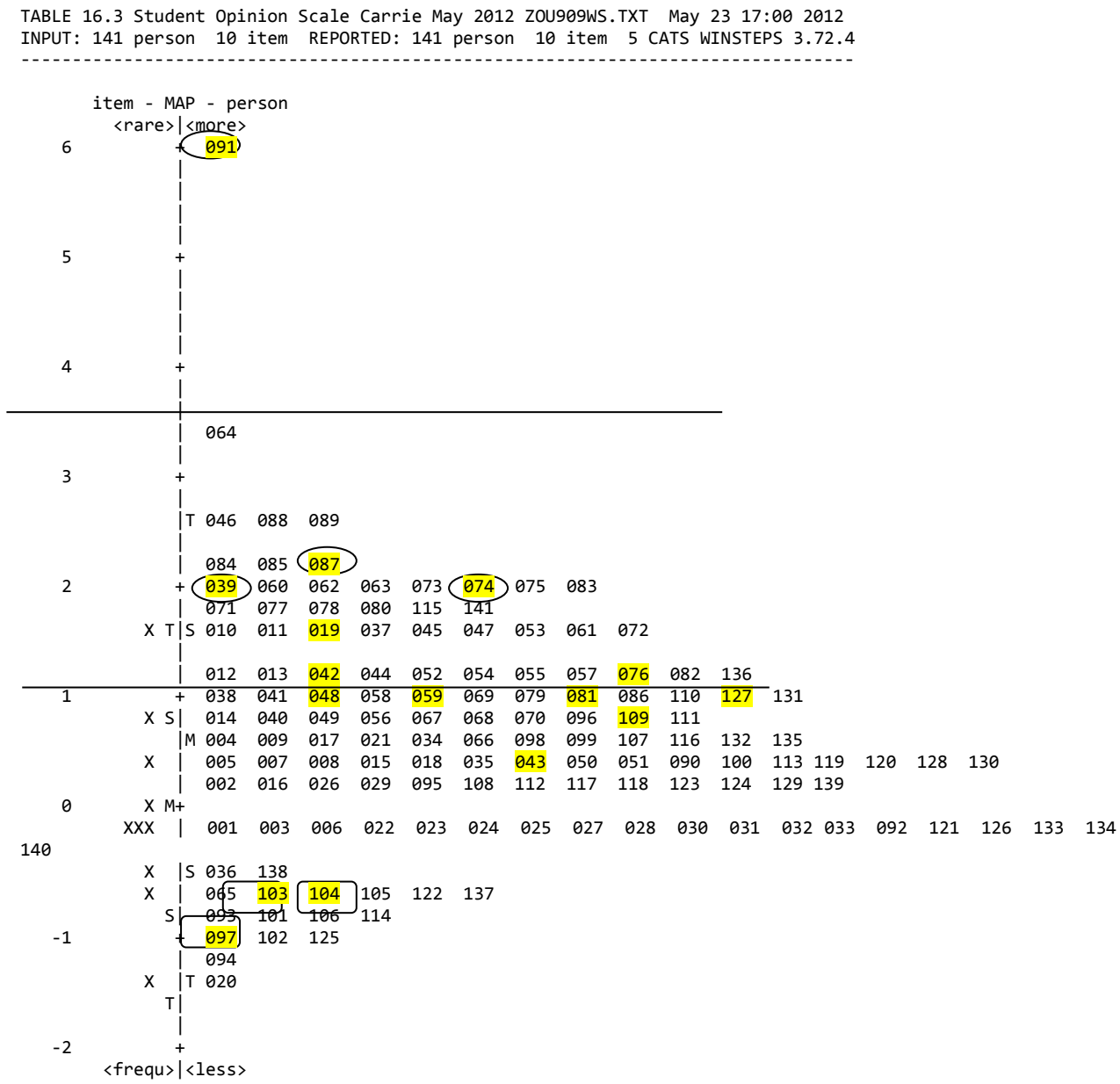


Figure 4.19 Location of students with comparatively higher and lower SOS measure

As for the incentive values of the test to these 7 students (point 2), below are the excerpts of what the highest 4 students (highlighted and in circles in Figure 4.19) said when they were asked if they want to do well in the DELTA:

> Ivan (091): Although it is not important and won't count towards my GPA, I quite like it as I think it is able to show my strengths and weaknesses. I think it is good and I wanted to do well in it and I did do it seriously.
>
> Candy (087): I think the test is quite good and it does not take a lot of time to do… I wanted to do well and I was quite motivated as I'm always interested in knowing my own English proficiency. I love languages no matter Chinese or English.
>
> Michael (039): I was very motivated when taking the DELTA. Personally I like learning English so whatever means that can test my English ability, I will try my best to do and I hope I will have slight improvement at least next time when I take it again.
>
> Shirley (074): If the DELTA is just for once, then I don't think it is very useful; but if we are doing it every year so that we know our progress, then I think it will be helpful. Of course I want to do well in any English test.

From the above excerpts, we can tell that the DELTA, as in any kind of test, has incentive values for Ivan, Candy and Michael; they have genuine interest in English/ testing their English ability. As for Shirley, she thinks that the tracking function of DELTA brings instrumental benefits—to keep track of her own English learning progress.

Wendy (097), David (103) and Ceci (104) are the ones who reported lowest motivation in the SOS (highlighted and in squares in Figure 4.19) and for them, the DELTA has low incentive values:

> Wendy: The DELTA is just a test to test our own ability, without any consequences…I can't think of any reason why it can be important to me so it doesn't matter to me at all if I do well in the test or not… if it counts towards my GPA, it will be very important.
>
> David: I didn't really care if I did well in the test or not. I was not motivated to do the test at all; it has no relevance to us at all now… unless it counts towards our GPA.

> Ceci: I was better in the first few questions, but motivation started to drop as I went along. It was like going down the slope. I don't like doing a test online and I think the test was too long and I was tired.

For the three who have low incentive value and low motivation in the test, as expected, these students expressed no wish or plans to formulate any action plan and their interviews proved that this assumption is correct. So for the students who have low motivational influences in the preactional stage, they will simply skip this stage and their process model starts at the actional phase instead, i.e. they will go straight to the test:

> Wendy (097): If it counts towards my GPA, then I will try my best to prepare for it and get higher marks. If not, then I will not pay much effort and will not care whether I perform better or worse than last year. I will not prepare for it.

> David (103): As it doesn't count towards my GPA, I will not set any target or goal or have any expectation. Will just go and take it. I will not do any preparation. If it counts towards my GPA, maybe I will look at some grammar books before the test. As for when I will start preparing, it depends on how busy I am and how urgent my other tasks are during that period.

> Ceci (104): I won't set any target or do any preparation even if it counts towards my GPA as I don't know what preparation I can do for this test.

As discussed, Ivan (091), Candy (087), Michael (039) and Shirley (074) indicated high motivation in the test and have the wish or desire to perform well in the test. They were asked if they would formulate any action plan to prepare for the test:

> Ivan: I will not set myself any preparation plan; if after taking the test, there is something tailor made for me to improve on my weaknesses, then I may work on them; but my feeling is what this test can tell me is my own strengths and weaknesses. As for the subskills in each component, even I can see that I could not answer that particular

question, I would not go and do something on this particular point just because of the DELTA test. (Interviewer: in the online version of the report, there are actually links to learning materials, have you/will you look at them?) I didn't notice the links. If it counts towards my GPA, then I may look at them.

Candy: Although I quite like this test and will try to do well during the test, I will not have the motivation to prepare for this test beforehand. I would expect myself to perform better in the second round but I will not work harder in my English because of this test… I don't know what I can do or prepare.

Michael: I will probably look at last year's report again and work on my weaker areas before the next test. If I have time, I will take out my report and look at it from time to time and maybe do something about it bit by bit during the year.

Shirley: I know I am weak in Listening skills. I will remind myself to work harder and do better in Listening in the next test, but actually I don't know what I can do to improve my Listening skills.

From the excerpts, Michael is the only student who indicated that he will lay down plans and carry out real action to study for the test. While the other 3 students, although they have the motivation to do well in the test initially, this motivational influences were not strong enough to lead them to the second phase of the preactional stage, i.e. the intention formation stage. They are the ones who will go from the goal setting stage, through the thick dotted line in Figure 4.20 and reach the test taking stage direct.

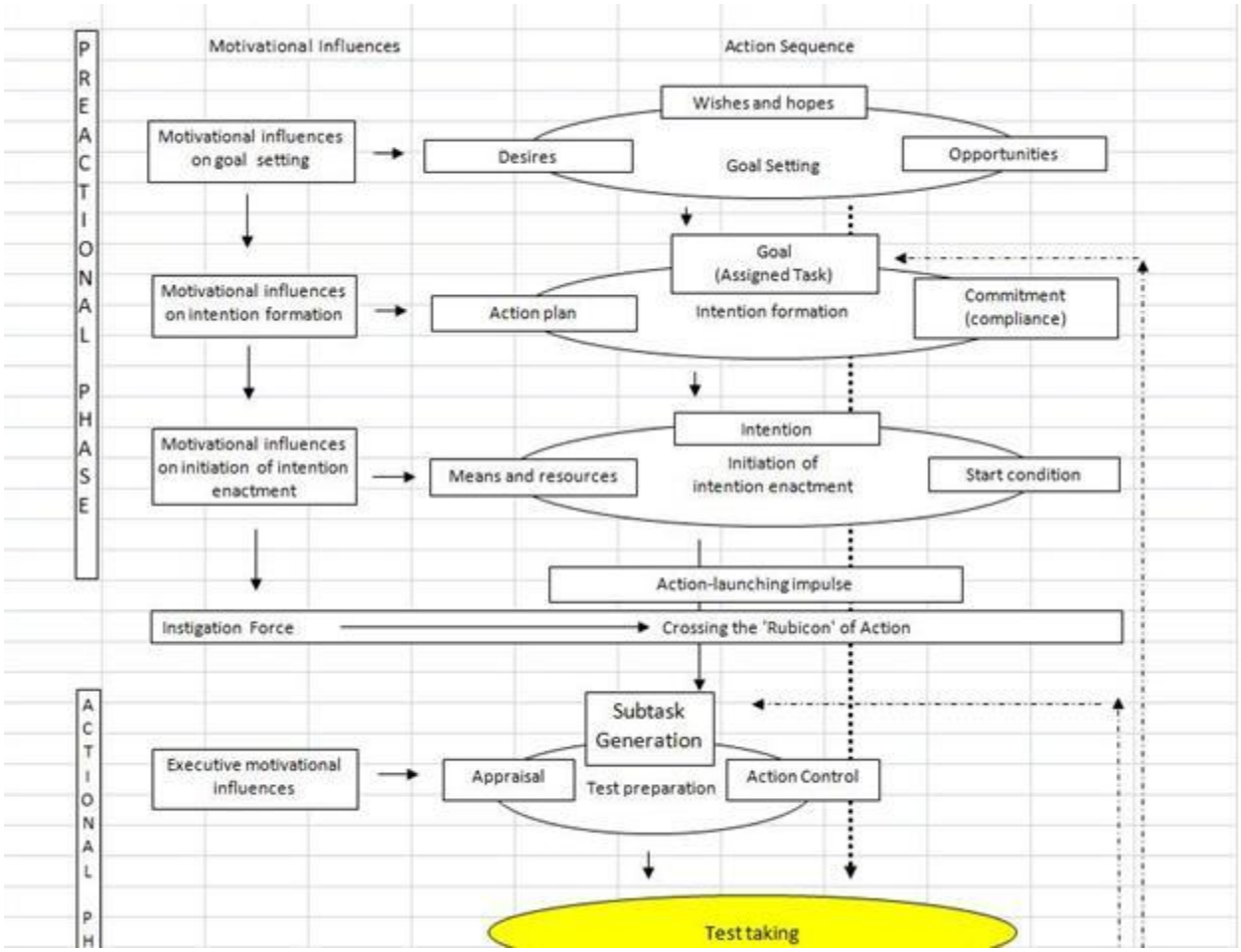


Figure 4.20 Process model of diagnostic test- Preactional stage

What causes the weak motivation in intention formation and makes the students to go straight to the test taking stage directly? According to Dörnyei (2001: 94), motivational influences on intention formation includes:

- Expectancy of success
- Perceived relevance of goal
- Need for achievement
- Degree of self-determination
- Goal properties
- Availability of task opportunities and options
- Leaner beliefs about L2 learning [test preparation]; knowledge of learning [testing] strategies; domain-specific knowledge

- Urgency, external demands; unique opportunity

The more the above factors are fulfilled, the higher the chance an intention can be formed. Ivan (091) does not have a strong **need for achievement** because he said the DELTA is low stakes and his motivation is not strong enough to care about improving the subskills of individual components. However, there is a discrepancy between his statements and his SOS score which is the highest in the scale. This discrepancy can have two indications: i) one can have low motivational influences in the preactional phase but still be highly motivated in the actional phase and so the motivational forces in both phases have no direct relation; ii) the SOS scores can be unreliable as it depends on the honesty of the students at the time when they responded to the questionnaire. It is a limitation of my research design that I did not analyze the students SOS scores before they came for the interview. If so I could have asked Ivan about the reasons for the discrepancies in his responses during the interview. Both Candy (087) and Shirley (074) commented that they do not know what they can do to prepare in order to improve in the test, i.e. they lack the knowledge of what action they can carry out and they do not have the access to task (test preparation) opportunities and options. Dörnyei (2001) explains that the development of an action plan is an imperative to forming a fully operational intention; when students lack the knowledge of what can be done or are not provided with opportunities and options to practice tasks, intention to prepare for the test would not be formed. This corresponds to the absence of Dörnyei's '**availability of task opportunities and options**.' Since there are two actions in the actional phase, i.e. test preparation and test taking, task opportunities and options here may refer to the opportunities and options of test preparation as well as the opportunities and options of taking the test.

4.2.2.2 Understanding the postactional phase of the process model-motivation after taking the DELTA

One major purpose of a diagnostic test is to impact students in the postactional phase by providing feedback, i.e. students are given a DELTA report which includes information on their language proficiency and diagnostic data of the components that they did after the test. In

Dörnyei's (2001: 100) framework, there are three major motivational influences in the postactional phase, namely, attributional factors, **self-concept beliefs** and the **quality and quantity of evaluational/attributional cues and feedback**. Satisfying these three criteria provide motivational influences in the postactional stage and students will then elaborate standards and strategies, and dismiss intention and further planning (as illustrated in the postactional stage in Figure 4.18).

The person-item map of the Winsteps analysis of the Report Usefulness Scale (Figure 4.21) shows the distribution of the 16 students:

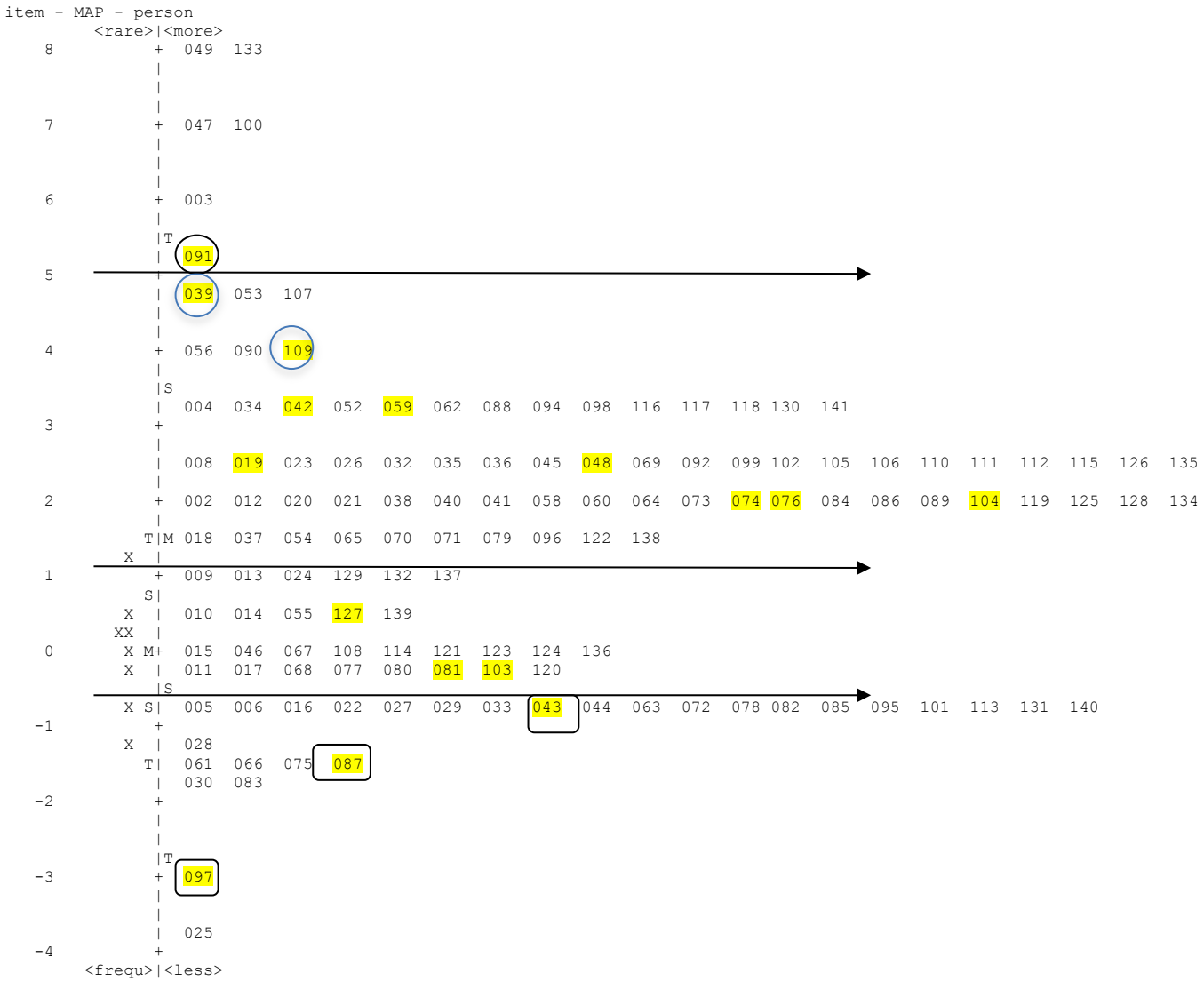


Figure 4.21 Location of students who find report more or less useful

Michael (039) and Ivan (091) (who also reported having medium to high motivation from their SOS score), and Karen (109) (all highlighted and in circles in Figure 4.21) reported in the questionnaire that they find the DELTA report useful. Below are the excerpts of what they said when they were asked to talk about the DELTA report:

> Ivan: I think the report can really show the comparative strengths and weaknesses of a student. Before the test I did have some ideas about my own ability in different areas of English and these have been reflected correctly on my DELTA report.
>
> Michael: I like the report which tells us the areas in which we have done right and areas which we have done wrong. It would be better if the report can also show us what exactly the right answers are and what my wrong answers are. I notice that there are suggested links for us to visit. I think it is quite good.
>
> Karen: The report is useful in helping me to understand my strengths and weaknesses and whether they [my strengths and weaknesses as shown on the report] match with my understandings.

The three of them who find the report useful have strong motivational influences in the postactional stage due to the **quality and quantity of evaluational/attributional cues and feedback**. The quality and quantity of evaluational/attributional cues and feedback in Dörnyei 's process model means feedback from teachers in general in the L2 learning process; in this research context, it is referred to as the quality of the DELTA report. The three students all indicated that they had evaluated their own performance and strategies.

On the other hand, Kate (043), Windy (097), Candy (087) (highlighted and in squares in Figure 4.21) found the DELTA report less useful and below is what they said about the report:

> Kate: I won't take the test and the report seriously because it is not widely recognized unlike tests like IELTS. I am not particularly interested to know my performance comparing with other students because A-level already performed that function.

> Windy: I don't actually understand some parts of it [the DELTA report] but I am not motivated to read them carefully to try to understand what they mean...If I knew there are such links [independent learning tips] and if the DELTA counts towards our GPA, then I will prepare for the test by clicking the links for sure.
>
> Candy: I looked through the report quickly but not quite understand what it means, and then I just leave it there.

Both Kate and Windy replied 'no' when they were asked if they have evaluated their test performance and strategies after the test or if they are more motivated in studying English. This proves that as the DELTA report failed to generate motivational influences on them and so no post-action was conducted. In the questionnaire survey, we found that the statement 'The DELTA report allows me to compare my results with other students' in the Report Usefulness Scale is misfitting. My assumption was that why students do not agree that they can use the DELTA measure to compare scores with other students may be because DELTA is a new test and not widely recognized yet as tests like IELTS. There is also no mechanism for them to compare their scores unless they do it informally because unlike public exams including IELTS, there are no published results which enable students to compare their performance with those of the whole population. Kate's response above is a good example confirming my assumption.

Although Candy did not find the DELTA report useful, she said,

> Candy: I am quite happy with my results in this test and I am happy to know that comparing with my classmates, I am not bad at all, so I think my English proficiency is probably ok. So it does provide me with some motivation to further improve. I am more confident now and more motivated to do even better.

Candy is a good example of having motivational influences in the postactional phase due to her better **self-concept beliefs** after the test. Self-concept beliefs in Dörnyei's model refer to one's self-efficacy under the Expectancy-value theories, i.e. people's judgement of their capabilities to carry out certain specific tasks (Dörnyei 2001:22).

Finally, the interview data above are all in line with the students' SOS and Report Usefulness Scale results; therefore it can be used as the 'Response processes evidence' as pointed out by the *Standards for educational and psychological testing* (1999).

4.3 Chapter Summary

This chapter presents the findings of the study. Firstly, the questionnaire survey which is divided into 2 main sections was analyzed using different methods. Results from factor analysis, Cronbach's Alpha, correlations check and Rasch model analysis prove that both the SOS and the perceived Report Usefulness Scale are well functioning for this particular group of students and context. Rasch model analysis of the SOS reveals that students were not motivated when they were taking the DELTA although they found the DELTA report quite useful. Secondly, the findings from the students' interviews are provided. The first section of the interview gave evidence to justify the occurrence of the two misfitting items in the SOS. Then Dörnyei and Ottó's process model of L2 motivation was applied in explaining whether students were motivated before and after taking the DELTA. Useful quotations from students showing both higher and lower motivation in the SOS and students who found the DELTA report more useful or not as useful are presented.

Chapter 5 Discussion

5.1 Summary of results

As set out in the methodology chapter, the aim of this study was to investigate test motivation in a low stakes diagnostic English test, the perceived usefulness of the diagnostic report, the motivation before and after taking the test and how the process model of test motivation in a low stakes diagnostic test can be conceptualized. The findings from the study provide answers to my research questions and offer useful insights on students' perceptions of test stakes and test value so as to understand what role the DELTA plays in motivating students in their [dynamic] L2 learning process.

5.1.1 Research question 1: Are students motivated to perform to the best of their ability whist sitting the DELTA?

Factor analysis, Rasch analysis and correlations studies, as well as students' interview data provided the internal structure evidence, response processes evidence and relations to other variables evidence as set out by the *Standards for educational and psychological testing* (1999). We can confirm that the Student Opinion Scale (SOS) by Sundre and Moore (2002) is a valid instrument in measuring students' test taking motivation in a low stakes test under the Hong Kong context. However, the SOS, with 5 questions under the 'importance' factor and the other 5 under the 'effort' factor, would need some slight amendments in the wordings as 2 of the questions show abnormality in the analysis:

7. While taking the DELTA, I could have worked harder on the questions.
10. While taking the DELTA, I was able to persist until I had completed all the questions.

Factor analysis shows that question 7 belongs to a third factor which is different from the others and question 10 does not fall into any of the two factors (importance and effort) as well. Rasch analysis also shows that question 7 is misfitting.

The SOS is used in its original language, i.e. English, in this study in which the subjects are tertiary students in Hong Kong. One possible explanation of the abnormality of the 2 questions is that students have difficulty in understanding the statements and so when I interviewed them I asked them to interpret the 2 statements in Cantonese, which is their mother tongue. Results from the interview shows that none of the students that I interviewed could interpret statement 7 correctly while only a few of them get close for statement 10.

Figure 4.5 in the Findings Chapter shows the person-item map of the SOS. After taking out question 7 which is misfitting under Rasch, the 9 SOS items are having a comparatively lower measure than the students' motivation (meaning that the items are on the whole more endorsable) as there are quite a number of '#' above the most difficult item (question 9) and the mean of the person measure higher than that of item measure. The Rasch analysis results proved that although the SOS is a valid instrument in measuring students test taking motivation in low stakes test in Hong Kong, it failed to differentiate highly motivated students as the items are too easy to endorse for most of the students. If we did not analyze the SOS using Winsteps but just by the traditional way of adding up the raw scores; first, we would not know that question 7 is problematic for this group of students in Hong Kong, second, we would not be able to have a more accurate motivation level of the students with item 'endorsability' taken into account.

As for the answer to the research question '**Are students motivated when they are taking the DELTA?'**, the results of the SOS demonstrate that they are not motivated. As explained in Section 4.1.1.3, only 1 student (0.7%) is considered as highly motivated and 41 students (29%) moderately motivated according to the Winsteps analysis.

Results from interviews with students shed more light on the characteristics of the relatively more motivated students and also the less motivated ones. The 16 students that I interviewed are in a wide range in their profile of test motivation, from extremely motivated to unmotivated. What the students actually said conforms with the Expectancy-value theories especially in terms of how to value 'value': of the 4 students who reported high and moderately high motivation, they are intrinsically interested in English and performing well in an English test is intrinsically valuable to them. While for the 3 who reported low motivation, 2 of them reported that the

DELTA has no extrinsic utility value to them as it would not count towards their GPA and 1 reported negative 'cost'. The one who reported negative cost felt that the test was too long and she felt tired doing it online. What she said falls under 'cost' in the value component which is a negative value component including factors such as expended effort and time or emotional costs such as anxiety.

5.1.2 Research question 2: Do they find the DELTA report useful?

Before students can make use of any kind of feedback, they should be able to make sense of and understand the feedback first. 50% of the students responded agree or strongly agree that they understand the DELTA report as a whole, 35% reported neutral; while about 60% of them agreed that they understand the meanings of the subskills listed in each of the 4 test components.

I constructed a Report Usefulness Scale with 8 questions in which 4 of the questions are about what kind of information students find useful and the other 4 on whether the DELTA report provides incentive for students to carry out further action. The results from this Scale can help answering my second research question, i.e. do they find the DELTA report useful?

In order to validate whether the 8 questions are functioning properly, the same factor, Cronbach's Alpha and Rasch analysis were done and the outcome is positive except for question 6 which is misfitting: 'The DELTA report allows me to compare my results with other students'.

One of the reasons why students do not agree that they can use the DELTA measure to compare scores with other students may be because the DELTA is a new test and not widely recognized yet and also there are no published results so they would not use the DELTA test result as a means for comparison. My assumption was proved to be correct from the students' interview.

Similar to the SOS, the Report Usefulness Scale failed to differentiate students who find the report very useful, i.e. there are not enough 'difficult to endorse' items. In Figure 4.10 of the

Findings Chapter, we can see that the mean of the students is much higher than that of the items and there are a lot of students whose measure is above the most difficult item on the scale.

Results of the Report Usefulness Scale demonstrate that students do find the report useful to a different degree. As explained in Section 4.1.2.3, 6 and 72 students (55.3% altogether) find the report very useful and useful respectively. The results, are better than that of the SOS (only 1 student can be regarded as highly motivated according to the Winsteps analysis of the SOS), meaning that although students are not highly motivated in the DELTA, the majority of them find the DELTA report useful. According to my summary of students' characteristics based on the Winsteps item-person map as well as the statements of the Report Usefulness Scale, over half of the students (58.2%) not only find the information provided by the report useful and reflect on their own English learning, but they would also carry out further action such as seeking help from the teacher to improve their English with reference to the report.

Correlations analysis between the SOS and the Report Usefulness scale was also run which shows a low but significant correlation between the Report Usefulness Scale and the Importance Scale under the SOS, meaning that the DELTA is important to those who find the report useful. This is an important finding which fits the expectancy-value theories because it proves that the more useful students find the diagnostic report, the higher the extrinsic utility value they have for the test and thus the higher the students' motivation in taking the test.

Although the DELTA report can said to be quite well received, some students do feel that it is not as useful. In the student interviews, two of them who scored low in the Report usefulness Scale reported that they do not understand how to read the report while one of them expressed that she did not take the report seriously because the test itself is not widely recognized. Their responses provide answer to the not highly satisfactory results of the two questions on understanding of the report in the questionnaire. It is either due to the fact that the report is indeed not easy enough to understand or it is because students do not bother to understand or read it carefully as they do not see the test itself as important. This again indicates that if the DELTA has a low attainment value to the student, not only would they be unmotivated when they take the test, they would also be unmotivated to read the DELTA report. Therefore

recommendations are provided in the later section of the chapter on how to increase students' test motivation so as to help increasing the degree of perceived report usefulness.

5.1.3 Research question 3 and 4: Are students motivated before and after taking the DELTA? How can Dörnyei and Ottó's (1998) process model of L2 motivation be applied in explaining the process of test motivation in a low stakes diagnostic test?

In my IFS, based on Dörnyei and Ottó's process model of L2 motivation (1998), I proposed that with slight modifications, the same process model could be applied to test motivation in summative tests. In this study, I look at how this model works on students in low stakes diagnostic test. The modifications that I have made on Dörnyei and Ottó's process model are renaming the original term from 'action' to 'test preparation' and adding 1 more phase in the actional phase called 'test taking' (below is my proposed test motivation process model, which is the same as Figure 4.18 and is reproduced here again for easier explanation). I suggest adding a new dotted line starting from Goal Setting that describes the flow of those students whose motivation in the preactional stage is too weak and therefore no planning or actual preparation work would be carried out and they would go and take the test direct. The results from the students' interview answered my research question 4, i.e. Dörnyei and Ottó's process model can be applied in explaining students' motivation process in low stakes diagnostic tests with minor changes.

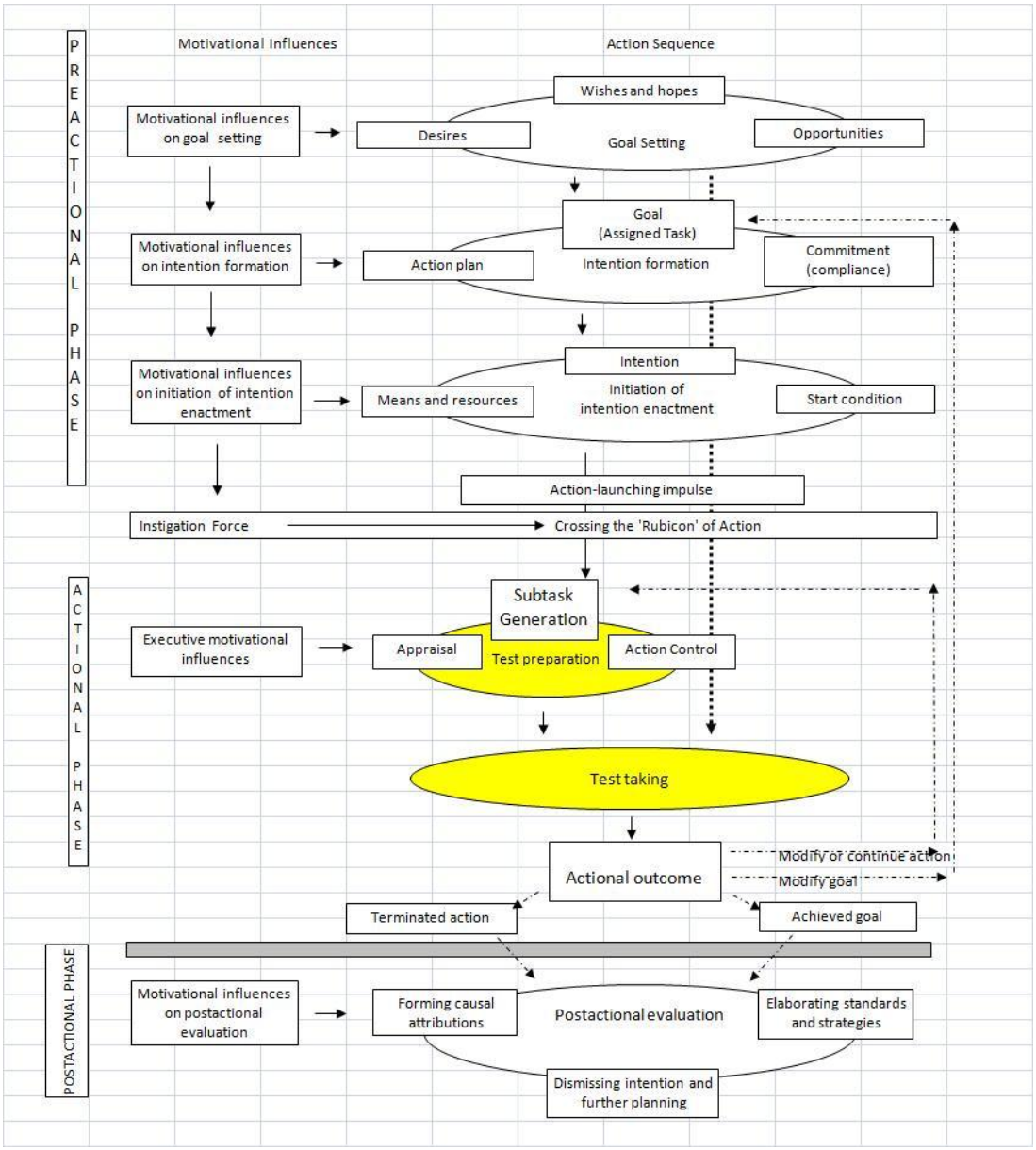


Figure 5.1 Test motivation process model (reproduction of Figure 4.18)

There are two possibilities when students go to the test-taking stage direct, the first option is that the test is compulsory for them (like the subjects of this study) but since it is low stakes, they do not have the motivation to go through test preparation. They may still have set up goals and intentions in the preactional phase; however, their motivational influences are not strong enough to push them through the 'Rubicon of action' and therefore they fail to reach the test preparation stage in the actional phase. The second possibility is that students take the test voluntarily. They are motivated in the preactional stage though the test is low stakes. They may take the tests for

many reasons such as wanting to get a diagnostic report or due to peer influence etc. and so their motivational forces are strong enough to push them forward to test-taking in the actional phase. However, they may still skip the test preparation stage (some students expressed in the interview that they did not know what and how to revise for the DELTA); or they lack action control (self control) in forcing themselves to do test preparation as the test is low stakes.

The act of test preparation in the context of the DELTA is only something optional. It is not a curriculum-embedded summative achievement test which students would be expected to study for. The intention of having the DELTA is to provide a diagnostic report which the students can refer to when they want to do self study in improving their English during the year and keep track of their proficiency throughout the years in university. As mentioned in Chapter 1, due to the huge class size and low teacher-student ratio as well as the less class contact time with teacher, it is the universities' wish to promote the habit and skill in doing self-learning among students. Therefore, the major purpose of a low stakes diagnostic test is to motivate and help students in their preactional and postactional phase. My interview with students therefore focused on examining the motivational forces of the students in the goal setting stage of the preactional phase as well as the impact of the DELTA in the postactional phase.

Dörnyei lists a number of factors which would affect the motivational influences on goal setting of L2 learning (2001; 93). For low stakes tests like the DELTA, the motivational forces in students' preactional phases mainly come from the perceived potency of potential goal and incentive values associated with test: intrinsic pleasure and instrumental benefits; which match with my interview results.

The students who showed high motivation in the SOS indicated that the DELTA has high incentive values to them while the ones low in SOS, as expected, indicated low incentive values on the DELTA. However, most of the students, no matter high or low in the SOS scale, pointed out that they were not motivated to do any preparation before taking the DELTA. In other words, students' motivational influences in the preactional stage of the test taking process are low. Therefore, they may at most set up goals but their motivation is not big enough to form intention in laying down any actual plan, not to say crossing the 'Rubicon' of action and carry out any test

preparation as what my new bold dotted line shows. Students' motivation may also fade out in any of the stages throughout the 3 phases which affect their respective action in the phases.

According to Dörnyei (2001: 100), there are three major motivational influences in the postactional phase, namely, attributional factors, self-concept beliefs and the quality and quantity of evaluational/attributional cues and feedback. What makes a diagnostic test different from the other types of test is its feedback function, i.e. students can login to the DELTA system and view their DELTA report usually 3 days after the test. From the Report Usefulness Scale and my student interviews, results revealed that for those students who find the DELTA report useful, they also feel more motivated in their postactional stage, they tend to evaluate their performances and strategies and are more motivated to improve their English after the test after the receiving of the report (Table 4.18 in the findings chapter shows that 14.9% of them will reflect about their English learning and 58.2% of them apart from reflection, will also seek help from teacher).

For the students who reported that they would not do any evaluation or are not motivated in carrying out any further action after the test, the cause of their low motivational influences in the postactional phase are once again due to the fact that the DELTA is a low stakes test in which their performances bear no consequences to their study or work; and also because they find the diagnostic report not that useful in helping them to lay down further study plans.

In conclusion, after reviewing what the students shared in the interviews, the answer to research question 3 is that students are not motivated before taking the DELTA. As for the motivation after taking the DELTA, it is largely dependent on if students perceive the DELTA report as useful or not.

5.2 The role of DELTA in L2 learning

I proposed in my IFS (Tsang, 2011) that the motivation of each of the activities under L2 learning can be described by a separate but similar process model. For example, the motivation process model of taking a language class, the process model of reading a grammar book, etc. All

these different process models may or may not overlap with each other during the main L2 learning process and will likely contribute to the main L2 learning process in different ways.

Figure 5.2 shows the points of interaction of the process models of different language learning activities and their effects on the whole L2 learning process over time. Different language learning activities may appear at different times in one's life or appear concurrently with each other. Each particular activity is affected by one's experience in their L2 learning process which may have been shaped as a result of the previous activities; and vice versa. All the individual activities will contribute to the bigger picture of an individual's L2 learning process.

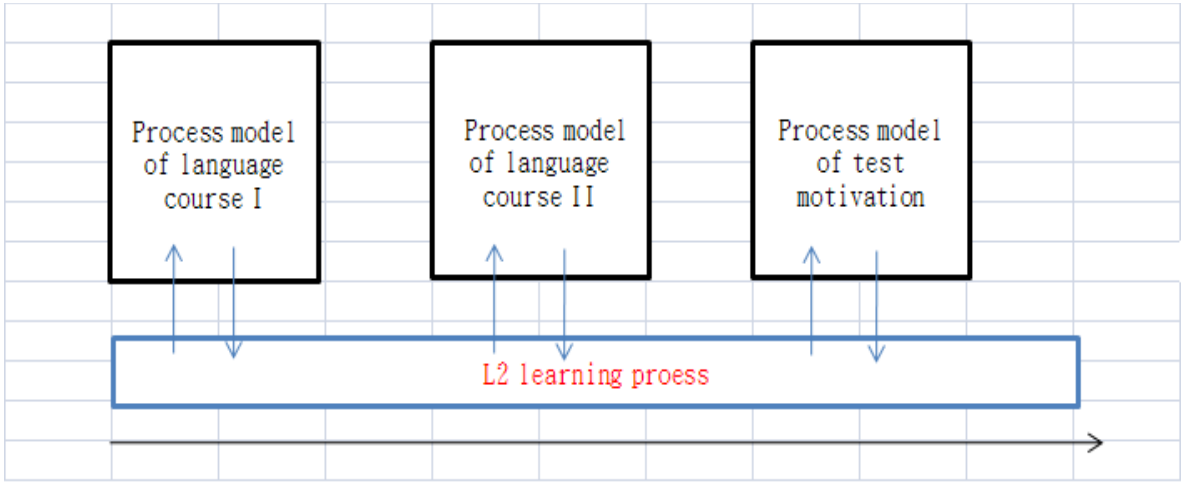


Figure 5.2 Interactions of different process models with the L2 learning process (Tsang, 2011)

The DELTA is a diagnostic and tracking test and is different from an exit test which appears only in the end of a learning process, nor other summative assessments which appear in the end of a learning phase/activity. Figure 5.3 is an illustration of the interactions of the DELTA with language learning activities in the L2 learning process. It has a direct influence in other English learning activities (students are more aware of their strengths and weaknesses and can do self study with the help of the diagnostic report) and it is also expected to be recurrent (once in each year of the 4-year university study).

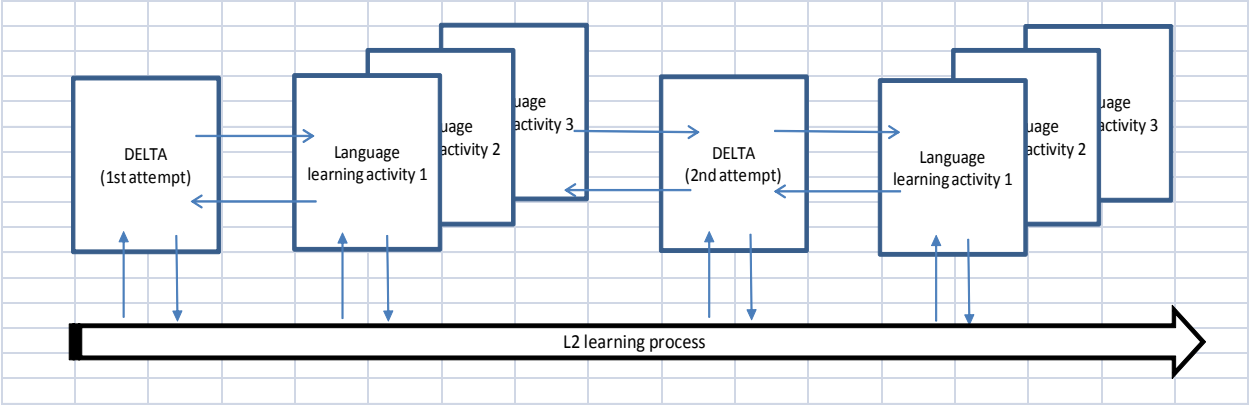


Figure 5.3 DELTA and the L2 learning process

From Figure 5.3, we can see that the DELTA, or any diagnostic test, can be playing a much more interactive role in students' L2 learning compared to other types of test. Its effectiveness can be enhanced if students are taking it more seriously and are more motivated in improving their English after the test with the help of the diagnostic report. Recommendations on how to better achieve this are in Section 5.4.

5.3 Limitations of the study

During the write up of the findings, I noticed a few technical oversights in the course of my data collection. The first shortfall of my research design is that I did not analyze the students' SOS scores before they came for the interview. If I had done this beforehand, I could have challenged some of their responses or found out the reasons behind, as some of them were scoring high in the SOS which did not exactly match with what they said in the interview. The second omission is that as the questionnaire survey was anonymous, I do not know the respective DELTA measure of the 141 respondents and so correlations between their test performance and their SOS and Report Usefulness Scale cannot be done. I could have asked the respondents to provide their DELTA measure in the survey.

Other than the technical flaw, due to the constraints in time and scope of this thesis, there are a number of limitations in this study that I was aware of right at the start of the project:

1. As mentioned in the methodology section already, both questionnaire survey and interview are a kind of self-report and the validity of responses is reliant on the truthfulness of the examinees.

   However as I have already explained, questionnaire survey and interview are the most practicable research methods in this case given the resources and time available. Moreover, due to the setting of the DELTA, methods like the measuring of response time are not possible.

2. Same as in many other studies, the students who responded to the questionnaire survey or came for an interview are very likely to be the ones who are more motivated or more willing to express themselves. The method of data collection in this study may not be able to take into account the points of view from the more passive students.

   In this study, whether the 16 interviewees can only represent the outspoken ones and whether the more passive students may have other opinions, I am not able to know. However, their profiles show a range in their SOS scores as well as their Report Usefulness scores and even lower motivated students in the DELTA volunteered to come for the interview. I believe that the sample is representative.

3. Both questionnaire survey and interview have only been conducted at one time in the longitudinal test preparation and taking process. Therefore the results can only review students' thoughts in a particular moment in time.

4. The survey on 'how students make use of the DELTA report" is just a prediction of their possible action. Not until we survey students who have taken their 2nd attempt on how they have applied the report in the previous year, we would not be able to tell.

5.4 Implications of the study

Although I noticed that there are insufficiencies in this study, I believe that the data I have collected and presented here is appropriate and can contribute to the field under the constraints that I faced. It is the first study to have analyzed the motivation scale (SOS) and developed and analyzed the Report Usefulness Scale using the Rasch model. There is no published research that I am aware of looking at the way in which diagnostic testing is perceived and used by the stakeholders.

The implication of this study is threefold. Firstly, both the Student Opinion Scale by Sundre and Moore (2002) and the Report Usefulness Scale is shown to be a valid tool in measuring test motivation in low stakes diagnostic tests in Hong Kong although more difficult items are needed in both scales in order to better differentiate the students with higher motivation/who find the report very useful. It is therefore likely that the SOS can be applied in other kind of tests or in other countries as well provided that different adjustments would be made after piloting. In the context of this study, question 7 and 10 in the SOS are problematic mainly due to language issues. I would recommend that in the future if the SOS is used in similar context, a bilingual version should be used (as in all official government documents) so that students can refer to the Chinese version if in doubt or to rewording the SOS by avoiding reversely written statements. The word 'persist' which a lot of students did not know can be replaced by terms such as 'carry on' and 'keep on'. As for the 8 questions in the Report Usefulness Scale, question 6 "The DELTA report allows me to compare my results with other students' is misffiting. I would suggest taking this question out in future studies because the purpose of the DELTA is diagnostic instead of comparing students across cohorts; students are not marked against each other but against themselves (meaning that one's growth is not at another's expense). I would recommend to replace this question with 'The DELTA report can tell me what I can expect to achieve in the future' as there is a predicted growth track on the DELTA report as well as the list of items which the students failed to answer this round. After adding this question, there will be 5 questions on 'information' and 3 questions on 'action'. More 'difficult to endorse' items should be added in future studies for both the SOS and the Report Usefulness Scale.

Secondly, the results of this study revealed that there is a significant (although not high) correlation between the SOS (the importance scale) and the Report Usefulness Scale meaning that those who view the test important would probably also find the report more useful and vice versa. Therefore in order to increase the perceived usefulness of the diagnostic report, we have to make the students feel that the DELTA is an important test to them. Recommendations of how to achieve this is in the coming section. On the other hand, in order for the students to be more motivated in taking the DELTA, we have to improve the diagnostic report so that students would find it more useful. The DELTA team understands this and has made improvement of the DELTA report one of their top priority in their research agenda in the coming year.

Thirdly, students' interview data supported that the process model of L2 motivation by Dörnyei and Ottó (1998) can be slightly modified and used in explaining students' test motivation in low stakes diagnostic tests. While the SOS can measure students' degree of motivation in the actional phase, the Report usefulness Scale can be used in illuminating the postactional phase as well as in the preactional phase because if students find the diagnostic report useful, their incentive values associated with the test will also increase. Together with the results from the students' interview, we can say that students are not motivated before taking the DELTA but whether they are motivated after the DELTA very much depends on if they perceive the DELTA report useful. However, interview results also shown that the stakes of a test is the most important factor affecting one's motivation in the whole dynamic process.

Further studies that may complement what I have presented above are in development or about to begin at the time I am rounding up this thesis. I will present my recommendations and our plans on the way forward in the next section.

5.5 The way forward

As discussed in the research rationale in Chapter 1, motivation influences the validity of the interpretation of test results. Apart from this, with reference to the findings of this study, motivation level also determines what one does before, during and after taking a test. As this

study has revealed, students are not motivated in the case of the DELTA. Raising students' motivation in the DELTA is paramount as it is also for the own good of the students as it is assumed that the DELTA report can help with their English learning and low motivation is a cause of them to pay less attention to reading the report. In order to boost students' motivation, a few things could be done in the future:

1. Increase the stakes of the DELTA

   According to all motivation theories, coupled with Chinese learners' pragmatic reasons in learning English and taking tests and well supported by the results from the students' interview in this study, increasing the stakes of the DELTA would definitely raise students' motivation in preparing and taking the test. University or Language Centre administrators can increase the stakes of the DELTA by ways such as making the DELTA a requirement, performance in DELTA contributes to part of the students' GPA, or set pass/fail mark in order to fulfil the requirement etc. 27.8% of the students indicated that they would not or probably not take the DELTA again next time round. Although the percentage is not very high, students' desire to take the DELTA again cannot be claimed as satisfactory. So making DELTA a requirement (without setting pass/fell benchmark since increasing the stakes of the test too much may also causes issues such as negative washback (Wall, 1997) or test anxieties from students) is also a way to increase students' participation rate in the 2nd and subsequent attempts.

2. Promotion of DELTA

   One of the 'shortcomings' of DELTA according to the students is that the test is not widely and highly recognized such as TOEFL and IELTS. Yet they are not diagnostic in nature with an aim of helping students to improve. However, students are not aware of the differences between the different types of tests, be it summative proficiency test or formative diagnostic test nor do they realize the potential bias of the two big tests in the world (Khan, 2009). Promotion of the DELTA, at least within Hong Kong, is needed so that more tertiary institutions are using it in an attempt to increase the face validity of the test and in turn increase the extrinsic utility value of the test to the students. It would also be advantageous if the DELTA can be used in some other countries outside Hong Kong

and more research output can be generated in conferences and academic journals. And work is ongoing to achieve this.

Promotion of the test within an institution is also needed again for face validity purposes as there are students saying that 'if my classmates are doing it and paying effort in it, I am more motivated.'

3. Increase the 'relevance' and 'usefulness' of DELTA

Some low motivated students find the DELTA of no relevance to them: "I didn't really care if I did well in the test or not. I was not motivated to do the test at all; it has no relevance to us at all now" one said. Apart from considering the raising of test stakes, another solution is to show students the 'relevance' and 'usefulness' of DELTA. They need to understand that studying and preparing for the DELTA and working according to the suggestions in the diagnostic report can help improving their English. As stated in the first chapter, the examination-led and highly competitive education system in Hong Kong has built a learning model in the secondary education where the teacher is in full control of the learning process; while it is the hope of tertiary institutions to promote independent learning among students especially in language learning due to the higher teacher-student ratio and less contact hours in English language classes. It is therefore important to educate students the importance of independent language learning and how the DELTA can help them in their independent language learning. For those of the students who believed English has a personal importance to them, increasing their beliefs in DELTA's relevance and usefulness in improving their English can hopefully lift their 'attainment value' under the expectancy-value theories. This can be done with the help from English teachers if they can show the link of the DELTA and the curriculum, and help students in understanding and making better use of the report.

4. Enhance teachers' test literacy

Follow up on the last part of point 3 above, i.e. teachers can show students the link of the DELTA and the curriculum and help students in making better use of the report, it is a prerequisite that the teachers themselves can buy into the DELTA or it is important for

them to at least understand clearly the functioning of the test themselves, i.e. teachers have to be test literate in the first place. According to Inbar-Lourie (2008), being literate in assessment means having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results. Therefore, providing teacher training on assessment in particular how the DELTA works is very important. After they have the knowledge, i.e. being test literate, the DELTA team can do promotion of the benefits of the test. It is when the teachers can truly appreciate and believe in the test before they can persuade their students to.

5.6 Potentials for researchers and administrators

Apart from promoting higher motivation from students in taking the DELTA, diagnostic and tracking test is a relatively new area with great potential of further development for both researchers and university/language centers administrators:-

i. For researchers- Survey of students of second and subsequent attempts

As discussed in the limitation section, this study was only targeted at students who have taken the DELTA the first time. Both questionnaire survey and interview can only be conducted at one time in the longitudinal test preparation and taking process and the results can therefore, only reveal students' thoughts in a particular moment in time. Moreover, the survey on 'how students make use of the DELTA report" is just a prediction of their possible action. As mentioned before, more 'difficult to endorse' items should be added in future studies; for students of second and subsequent attempts in particular, an extra section on what actually had been done in the previous year can be added. In that additional section, I suggest putting in items concerning real action taken, for e.g. 'Based on the DELTA report, I laid down my language learning plan' or 'I did self-study in my weaker area as identified by the DELTA report.'.

The DELTA is a tracking test which aims at helping the students throughout their 4-year university studies. Therefore the study on motivation in the DELTA should also be a longitudinal study. Survey of students of second and subsequent attempts is essential in order to track the possible changes in their motivation level throughout the years and also to confirm with the students their claimed usage of the DELTA report in their L2 study. It would also be very useful to find out if motivation level during the test and perceived usefulness of the DELTA report has anything to do with their DELTA measure in their subsequent DELTA attempts.

As the DELTA coordinator in my institution, this is going to be one of my research targets in the coming years.

ii. For ELCs and administrators

As discussed in sections 5.4, in order to better motivate the students in taking the DELTA, it is important to raise the attractiveness/importance of the test and the usefulness of the DELTA report. There are a number of on-going or planned projects with an aim to make the DELTA a better test.

- Research project on DELTA for teaching and independent learning

According to the expectancy-value theories and also confirmed by the results in this study, apart from the fact that the DELTA is a low-stakes test, students feel less motivated in taking it mainly because they feel that the test is separated from the curriculum and has no relation to what they are learning in class. The DELTA team has therefore proposed to conduct an action research project on applying the DELTA report in classroom learning and will begin the project in 2013.

The project aims at creating classroom teaching materials to help students in applying the results from the DELTA to identify specific learning needs and create independent learning plans. For example, teachers will go through teaching materials with their students on how to understand the DELTA report and how to make use of the DELTA report. Based on the strengths and

weaknesses as identified by the DELTA report, students have to write a learning plan at the beginning of the semester on what area(s) of English they would like to work harder at in order to improve, what are they going to do in order to improve (they will need to search for specific materials in the University's Self Access Learning Centre for example) as well as the timeline in carrying out their proposed plan. Throughout the semester, they are reminded and encouraged to follow their plan. It is also a course requirement for students to submit a reflection at the end of the semester on what they have done and their comments on this activity.

The project team would then evaluate the effectiveness of the DELTA report and the associated purpose-designed teaching materials to see if the report and materials can aid student learning. This project would be the first step in bridging the DELTA with the curriculum. If the feedback from this project is positive, English Language Centres of the participating institutions and also the DELTA administrators can consider the further development of classroom learning materials and make the DELTA better linked to the syllabus.

- Inter-University Collaborative Online Self-Access (ICOSA) project

Students revealed in the interviews that after receiving the DELTA report, they do not know what they can do with it in order to improve their language skills or to have a better DELTA result next time. No matter how informative the DELTA report can be, if students are unable to find resources related to their weaknesses laid out by the report, it would be of little practical use to them.

Starting in 2012, five participating Hong Kong tertiary institutions collaborate in the ICOSA project with an aim to share and develop online self-access English language learning materials. One of their targets is to provide learning materials for the DELTA. The project team plans to survey existing materials to find out what are already available among the institutions and what materials could be modified or developed which can help students in improving the individual subskills as shown in the DELTA report. All project materials will be shared in a repository and an indexing system will be used in order to help participating institutions to find materials in the

repository. Links to relevant materials would be provided in the DELTA report so that students can do independent learning by visiting the sites and work on their weaker areas.

The project will be completed in 2 years' time and at the moment only generic learning sites are provided in the DELTA report. Upon completion of the project, it is hoped that students would find the DELTA report more useful as they can find related learning materials in a particular area which they would like to work on. With an improved perceived usefulness of the DELTA, it is expected that motivation in taking the DELTA would be increased simultaneously.

5.7 Chapter Summary

This chapter provides a summary of the research results and the answers to the research questions: firstly, students were not very motivated during the DELTA; secondly, they found the DELTA report quite useful; thirdly the process model of L2 motivation could be modified and help in explaining students' test-taking motivation; lastly, students were not motivated before taking the DELTA and their motivation after taking the DELTA largely depends on whether they perceive the DELTA report as useful. Suggestion on the role of the DELTA in the L2 learning process is made. The chapter then goes on to discuss the limitations of the study and the possible means in lifting students' motivation in taking the DELTA by ways of increasing the stakes of the test, better promotion of the test, increasing the 'relevance' and 'usefulness' of DELTA in classroom learning and also by enhancing teachers' test literacy. Lastly, recommendations on the way forward and also other on-going or under-planning related projects are mentioned. It is believed that this study is an innovative and valuable one as it is the first study to have validated the SOS and the Report Usefulness Scale using Rasch analysis. The use of Rasch analysis in compiling students' profile in order to help with the qualitative analysis (reporting on students' interview data) is also novel. There is little (if any) research on the way in which diagnostic testing is perceived and used by the stakeholders.

References

Alderson, J. C. (2000). Technology in testing: The present and the future. *System, 28*(4), 593-603.

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London; New York, NY: Continuum

Alderson, J. C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring. In J. Fox (Ed.), *Language testing reconsidered* (pp. 21-39). Ottawa: Univeristy of Ottawa Press.

Alderson, J. C., & Clapham, C. (1995). Assessing student performance in the ESL classroom. *TESOL Quarterly, 29*(1), 184-187.

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology, 61*(2), 207-215.

Association of Language Testers in Europe (1999). *Multilingual glossary of language testing terms*. Cambridge [England]: Cambridge University Press.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review; Psychological Review, 64*(6p1), 359.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton: Van Nostrand.

Atkinson, J. W., & Birch, D. (1978). *An introduction to motivation*. New York: Van Nostrand.

Aubel, J. (1994). *Guidelines for studies using the group interview technique*. Geneva: International Labour Organization.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (2010). Language Enhancement in Hong Kong Universities: Some observations and Recommendations. Consultant's Report to the University Grants Committee.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice : designing and developing useful language tests*. Oxford [England]: Oxford University Press.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J.: Prentice Hall.

Bandura, A. (1986). Fearful Expectations and Avoidant Actions as Coeffects of Perceived Self-Inefficacy. *American psychologist, 41*(12), 1389-1391.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. (1997). *Self-efficacy: The exercise of control*: New York: Freeman.

Barkoukis, V., Tsorbatzoudis, H., Grouios, G., & Sideridis, G. (2008). The assessment of intrinsic and extrinsic motivation and amotivation: Validity and reliability of the Greek version of the Academic Motivation Scale. *Assessment in Education: Principles, Policy & Practice, 15*(1), 39-55.

Barry, C., & Finney, S. (2009). Does it Matter How Data are Collected? A Comparison of Testing Conditions and the Implications for Validity. *Research & Practice in Assessment, 3*(1), 1-15.

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations, 61*(8), 1139-1160.

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*(3), 441-462.

B.E.R.A. (2004). *Revised ethical guidelines for educational research.* British Educational Research Association.

Biggs, J. B. (1987). *Student approaches to learning and studying*. Melbourne: Australian Council for Educational Research.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education, 5*(1), 7-74.

Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality & quantity, 36*(4), 391-409.

Boekaerts, M. (1987). Situation-specific judgments of a learning task versus overall measures of motivational orientation. In E. L. De Corte, H.; Parmentier, R.; Span, P. (Ed.), *Learning and instruction: European research in an international context* (pp. 169-179). Oxford/Leuven: Pergamon Press/Leuven University Press.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum.

Boyle, J. (1997). Native-speaker teachers of English in Hong Kong. *Language and Education, 11*(3), 163-181.

Britten, N. (1995). Qualitative interviews in medical research. *BMJ: British Medical Journal, 311*(6999), 251.

Brown, E., & Glover, C. (2006). 6 Evaluating written feedback. *Innovative assessment in higher education*, 81.

Brown, J. D. (2001). *Using surveys in language programs*: Cambridge Univ Pr.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in higher education, 31*(2), 219-233.

Chan, V. (2001). Readiness for learner autonomy: What do our learners tell us? *Teaching in Higher Education, 6*(4), 505-518.

Chan, V., Spratt, M., & Humphreys, G. (2002). Autonomous language learning: Hong Kong tertiary students' attitudes and behaviours. *Evaluation & Research in Education, 16*(1), 1-18.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education, 11*(1), 38-54.

Cheng, X. (2000). Asian students' reticence revisited. *System*, *28*(3), 435-446.

Choi, P. K. (2003). 'The best students will learn English': ultra-utilitarianism and linguistic imperialism in education in post-1997 Hong Kong. *J. Education Policy, 18*(6), 673-694.

Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly, 57*(1), 62-79.

Clark, R., & Gieve, S. N. (2006). On the discursive construction of 'the Chinese learner'. *Language, Culture and Curriculum*, *19*(1), 54-73.

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. *Language: Social psychological perspectives, 147154*.

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, Self-confidence, and Group Cohesion in the Foreign Language Classroom. *Language learning, 44*(3), 417-448.

Clément, R., & Kruidenier, B. G. (1983). Orientations in second language acquisition: The effects of ethicity, miliey and target language on their emergence. *Language learning, 33*(3), 273-291.

Cohen, L. M., & Manion, L. (2007). L. & Morrison, K.(2000) Research Methods in Education. *Routledge Falmer, London: ISBN: 0, 415*(19541), 1.

Cole, J., Bergin, D., & Whittaker, T. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*(4), 609-624.

Coniam, D., & Falvey, P. (2002). Selecting models and setting standards for teachers of English in Hong Kong. *Journal of Asian Pacific Communication, 12*(1), 13-37.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology, 78*, 98-98.

Creswell, J. W. (2009). *Research design : qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology, 70*(4), 519-538.

Dörnyei, Z. (2001). *Teaching and researching: motivation*. Pearson Education.

Dörnyei, Z. (2002). The motivational basis of language learning tasks. *Individual differences and instructed language learning, 2*, 137-158.

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. *Language learning, 53*(S1), 3-32.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *Aila Review, 19*(1), 42-68.

Dörnyei, Z., & Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics*, 43-69.

Drasgow, F., & Hulin, C. L. (1990). Item response theory. *Handbook of industrial and organizational psychology, 1*, 577-636.

Dweck, C. (1986). Motivational processes affecting learning. *American psychologist, 41*(10), 1040.

Eccles, J., Adler, T., Futterman, R., Goff, S., Kaczala, C., Meece, J., & Midgley, C. (1983). Expectancies, values, and academic behaviors. *Achievement and achievement motives*, 75-146.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality & Social Psychology Bulletin, 21*(3), 215.

Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement, 66*(4), 643-656.

Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*(3), 311-326.

Eklöf, H. (2010). Skill and will: test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice, 17*(4), 345-356.

Eklöf, S. (2005). *The return of piracy : decoloization and international relations in a maritime border region (the Sulu Sea), 1959-63*. Lund: Centre for East and South-East Asian Studies, Lund University.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58*(3), 357-381.

Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315-339.

Ferris, D., & Tagg, T. (1996). Academic Listening/Speaking Tasks for ESL Students: Problems, Suggestions, and Implications. *TESOL Quarterly, 30*(2), 297-320.

Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE Publications.

Forlin, C. (2007). A collaborative, collegial and more cohesive approach to supporting educational reform for inclusion in Hong Kong. *Asia Pacific Education Review, 8*(2), 276-287.

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate a meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*(4), 467-491.

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal, 28*(2), 175-187.

Fullilove, J. (1992). The tail that wags. *Institute of Language in Education Journal, 9*, 131-147.

Gao, X. (2006). Understanding changes in Chinese students' uses of learning strategies in China and Britain: A socio-cultural re-interpretation. *System, 34*(1), 55-67.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold

Gardner, R. C. (2006). The socio-educational model of second language acquisition: A research paradigm. *Eurosla Yearbook, 6*(1), 237-260.

Gardner, R. C. (2007). Motivation and second language acquisition. *Porta Linguarum 8*, 9-20.

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second-language acquisition. *Canadian Journal of Psychology/Revue canadienne de psychologie, 13*(4), 266.

Gardner, R. C., & Lambert, W. E. (1972). Attitudes and motivation in second language learning. *Rowley, Mass*.

Gardner, R. C., Smythe, P., & Clément, R. (1979). Intensive second language study in a bicultural milieu: An investigation of attitudes, motivation and language proficiency. *Language Learning, 29*(2), 305-320.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education, 1*(1), 3-31.

Gieve, S., & Clark, R. (2005). The chinese approach to learning: Cultural trait or situated response? The case of a self-directed learning programme. *System, 33*, 261-276.

Glas, C. A. W., & Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement, 27*(3), 217-233.

Glover, C., & Brown, E. (2006). Written feedback for students: too much, too detailed or too incomprehensible to be effective?. *Bioscience Education*, (7).

Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2-31.

Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice, 10*(2), 169-207.

Hartley, J., & Chesworth, K. (2000). Qualitative and quantitative methods in research on essay writing: no one way. *Journal of Further and Higher Education, 24*(1), 15-24.

Heckhausen, H., & Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. *Goal directed behavior: The concept of action in psychology, 10*, 134-159.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in higher education, 27*(1), 53-64.

Hill, P. W., & Wan, T. (2006). *Assessment Reforms in Hong Kong.* Paper presented at the 32nd IAEA Conference, Singapore.

Hirvela, A., & Law, E. (1991). A survey of local English teachers' attitudes towards English and ELT. *Institute of Language in Education Journal, 8*, 25-38.

Huhta, A. (2008). Diagnostic and Formative Assessment. *The Handbook of Educational Linguistics*, 467-482

Huhta, A., Luoma, S., Oscarson, M., Sajavaara, K., Takala, S., & Teasdale, A. (2002). DIALANG–a diagnostic language assessment system for learners. *Case Studies of the Use of the Common European Framework*, 130-145.

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185-212.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385-402.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement, 20*(1), 141-151.

Kennedy, K. J., Fok, P. K., & Chan, K. S. J. (2006). Reforming the Curriculum in a Post-Colonial Society: The Case of Hong Kong. *Planning and Changing, 37*(1), 111-130.

Kennedy, P. (2002). Learning cultures and learning styles: myth-understandings about adult (Hong Kong) Chinese learners. *International Journal of Lifelong Education, 21*(5), 430-445.

Khan, S. Z. (2009). Imperialism of international tests: An EIL perspective. *English as an international language. Perspectives and pedagogical issues*, 190-205.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). New York: Routledge.

Lado, R. (1961). *Language testing : the construction and use of foreign language tests : A teacher's book*. London: Longmans.

Lai, M. (2000). Motivation of English language learners in Hong Kong-use of case histories. *Asia Pacific Journal of Language in Education 3*, 43-60.

Lai, M. (2001). Hong Kong students' attitudes towards Cantonese, Putonghua and English after the change of sovereignty. *Journal of Multilingual and Multicultural Development, 22*(2), 112-133.

Lai, M. (2009). 'I love Cantonese but I want English'- A qualitative account of Hong Kong students' language attitudes. *The Asia-Pacific Education Researcher, 18*(1), 79-92.

Lee, C. (1999). Learning through tutorial discussion and learners' culture: Some preliminary observations from the views of Hong Kong Chinese tertiary students. *Language Culture and Curriculum, 12*(3), 255-264.

Lee, M. H., & Gopinathan, S. (2003). Reforming university education in Hong Kong and Singapore. *Higher Education Research and Development, 22*(2), 167-182.

Lee, W. O. (1996). The cultural context for Chinese learners: conceptions of learning in the confucian tradition. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese Learner: Cultural, Psychological, and Contextual Influences* (Vol. 1, pp. 25-41). Hong Kong; Melbourne, Australia: Comparative Education Research Centre; Australia Council for Educational Research.

Lin, A. (2005). Critical, transdisciplinary perspectives on language-in-education policy and practice in postcolonial contexts: The case of Hong Kong. *Decolonisation, globalisation: Language-ineducation policy and practice*, 38-54.

Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of outcome measurement, 3*(4), 382-405.

Linacre, J. M., & Wright, B. D. (2000). WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis [Computer software]. Chicago: MESA Press.

Littlewood, W. (1996). "Autonomy": An anatomy and a framework. *System, 24*(4), 427-435.

Littlewood, W. (2000). Do Asian students really want to listen and obey?. *ELT journal*, *54*(1), 31-36.

Littlewood, W. (2001). Students' attitudes to classroom English learning: a cross-cultural study. *Language Teaching Research, 5*(1), 3-28.

Liu, N. F., & Littlewood, W. (1997). Why do many students appear reluctant to participate in classroom learning discourse?. *System*, *25*(3), 371-384.

Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education, 33*(3), 263-275.

Lu, D. (2002). English medium teaching at crisis: Towards bilingual education in Hong Kong. *Gema: Online Journal of Language Studies, 2*(1).

Lumley, T., & Qian, D. (2003). Assessing English for employment in Hong Kong. *Assessment practices: Case studies in TESOL. Practice Series*, 135-147.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman

Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25*(2), 107-135.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45*(1), 35-44.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative analysis*. Thousand Oaks, CA: Sage.

Morrison, B. (Ed.). (2011). *Independent Language Learning: Building on Experience, Seeking New Perspectives*: Hong Kong University Press.

Morse, J. M., & Field, P.-A. (1996). *Nursing research : the application of qualitative approaches* (2nd ed.). London: Chapman & Hall.

Mousavi, S. A. (2002). *An encyclopaedic dictionary of language testing*. n.p.: Rahnama Publications.

Murphy, D. (1987). Offshore Education: A Hong Kong Perspective. *Australian Universities' Review, 30*(2), 43-44.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education, 31*(2), 199-218.

Noels, K. A., & Clément, R. (1996). Communicating across cultures: Social determinants and acculturative consequences. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 28*(3), 214.

Noels, K. A., Pelletier, L. G., Clement, R., & Vallerand, R. J. (2003). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*(1), 57-85.

Oxford, R. L. (1996). *Language learning motivation : pathways to the new century*. Mānoa: Second Language Teaching & Curriculum Center, University of Hawaiì at Mānoa.

Pennington, M. C., & Yue, F. (2007). English and Chinese in Hong Kong: pre-1997 language attitudes. *World Englishes, 13*(1), 1-20.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan, 48109*, 1259.

Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: the students' perspective. *Assessment & Evaluation in Higher Education, 33*(2), 143-154.

Qian, D. D. (2007). Assessing University Students Searching for an English Language Exit Test. *RELC journal, 38*(1), 18-37.

Qian, D. D. (2008). English language assessment in Hong Kong: A survey of practices, developments and issues. *Language Testing, 25*(1), 85-110.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. n.p.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field methods, 15*(1), 85-109.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55*(1), 68-78.

Sadler, D. R. (1998). Formative assessment: revisiting the territory. *Assessment in education, 5*(1), 77-84.

Sanchez, R., Truxillo, D., & Bauer, T. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *The Journal of applied psychology, 85*(5), 739.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education : theory, research, and applications* (3rd ed.). Upper Saddle River, N.J.: Merrill.

Sinagub, J. M., Vaughn, S., & Schumm, J. S. (1996). *Focus group interviews in education and psychology*: Sage Publications, Incorporated.

Solberg, V., O'Brien, K., Villareal, P., Kennel, R., & Davis, B. (1993). Self-efficacy and Hispanic college students: Validation of the college self-efficacy instrument. *Hispanic Journal of Behavioral Sciences, 15*(1), 80.

Squires, J. E., Estabrooks, C. A., Newburn-Cook, C. V., & Gierl, M. (2011). Validation of the conceptual research utilization scale: an application of the standards for educational and psychological testing in healthcare. *BMC health services research, 11*(1), 107.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, N.J.: L. Erlbaum.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research : techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, Calif.: Sage Publications.

Sundre, D. L., & Finney, S. J. (2002). *Enhancing the validity and value of learning assessment: Furthering the development of a motivation scale.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*(1), 6-26.

Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14*(1), 8-9.

Sundre, D. L., & Thelk, A. D. (2007). The Student Opinion Scale (SOS) A Measure of Examinee Motivation-Test Manual: James Madison University.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics*. New York: Harper Collins.

Tang, C., & Biggs, J. (1996). How Hong Kong students cope with assessment. *The Chinese learner: Cultural, psychological and contextual influences*, 159-182.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478.

Tesio, L. (2003). Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. *Journal of Rehabilitation Medicine, 35*(3), 105-115.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *The Journal of General Education*, 129-151.

Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of counseling psychology, 34*(4), 414.

Tremblay, P. F., & Gardner, R. C. (1995). Expanding the motivation construct in language learning. *The Modern Language Journal, 79*(4), 505-518.

Tsang, C. (2011). *A study of two English exit tests on students' motivation to learn English at a Hong Kong university*. Institution Focus Study. Institute of Education, University of London.

Tsui, A. B. M. (1996). Reticence and anxiety in second language learning. *Voices from the language classroom*, 145-167.

Tung, P., Lam, R., & Tsang, W. K. (1997). English as a medium of instruction in post-1997 Hong Kong: What students, teachers, and parents think. *Journal of Pragmatics, 28*(4), 441-459.

UGC. (2012). *UGC to launch New Arrangement to Sponsor Collaborative Language Enhancement Projects*. Hong Kong.

Ushioda, E. (1996). Developing a dynamic concept of L2 motivation. In T. Hickey & J. Williams (Eds.), *Language, education and society in a changing world* (pp. 239-245). Dublin, Ireland: IRAAL/Multilingual Matters.

Ushioda, E. (1998). *Effective motivational thinking: A cognitive theoretical approach to the study of language learning motivation*: Publicacions de la Universitat Jaume I.

Ushioda, E. (2001). Language learning at university: Exploring the role of motivational thinking. *Motivation and second language acquisition, 23*, 93-125.

Vallerand, R., Pelletier, L., Blais, M., Briere, N., Senecal, C., & Vallieres, E. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*(4), 1003.

Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 21*(3), 323.

Walker, D., & Myrick, F. (2006). Grounded theory: An exploration of process and procedure. *Qualitative Health Research, 16*(4), 547-559.

Wall, D. (1997). Impact and washback in language testing. *Encyclopedia of language and education, 7*, 291-302.

Watkins, D. A. (2009). Motivation and competition in Hong Kong secondary schools: The students' perspective. In C. K. K. Chan & N. Rao (Eds.), *Revisiting The Chinese Learner- Changing Contexts, Changing Education* (pp. 71-88). Hong Kong: Springer: Comparative Education Research Centre, The University of Hong Kong.

Watkins, D. A., & Biggs, J. B. (1996). *The Chinese Learner: Cultural, Psychological, and Contextual Influences*: ERIC.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education, 31*(3), 379-394.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, N.J.: General Learning Press.

Weiner, B. (1986). Attribution, emotion, and action. *Handbook of motivation and cognition: Foundations of social behavior, 1*, 281-312.

Weiner, B. (1992). *Human motivation : metaphors, theories, and research*. Newbury Park, Calif.: Sage.

Weinstein, C. E. A., Taylor W.; Jung, Jaehak; Dearman, Jeremy K. (2011). Strategic and self-regulated learning for the 21st century: The merging of skill, will and self-regulation. In B. Morrison (Ed.), *Indepedent Language Learning: Building on Experience, Seeking New Perspectives* (pp. 41-54). Hong Kong: Hong Kong University Press.

White, C. (2011). Inside independent learning: Old and new perspectives. In B. Morrison (Ed.), *Independent Language Learning: Building on Experience, Seeking New Perspectives* (pp. 13-24). Hong Kong: Hong Kong University Press.

Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of motivation. *Contemporary Educational Psychology, 25*, 68-81.

Wigfield, A., & Eccles, J. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. *Development of achievement motivation*, 91¡V120.

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J. A., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology; Journal of Educational Psychology, 89*(3), 451.

Wise, S., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1-17.

Wise, S. L., & DeMars, C. E. (2006). An Application of Item Response Time: The Effort Moderated IRT Model. *Journal of Educational Measurement, 43*(1), 19-38.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*(3), 227-242.

Wright, B. D. (1995). 3PL or Rasch? *Rasch Measurement Transactions, 9*(1), 408-409.

Wright, B. D., & Masters, G. N. (2002). Number of person or item strata. *Rasch Measurement Transactions, 16*(3), 888.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education, 45*(4), 477-501.

Young, J. W. (2007). Predicting college grades: the value of achievement goals in supplementing ability measures. *Assessment in Education: Principles, Policy & Practice, 14*(2), 233-249.

Zhang, S., & Thompson, N. (2004). DIALANG: A Diagnostic Language Assessment System (review). *The Canadian Modern Language Review/La revue canadienne des langues vivantes, 61*(2), 290-293.

Appendix I

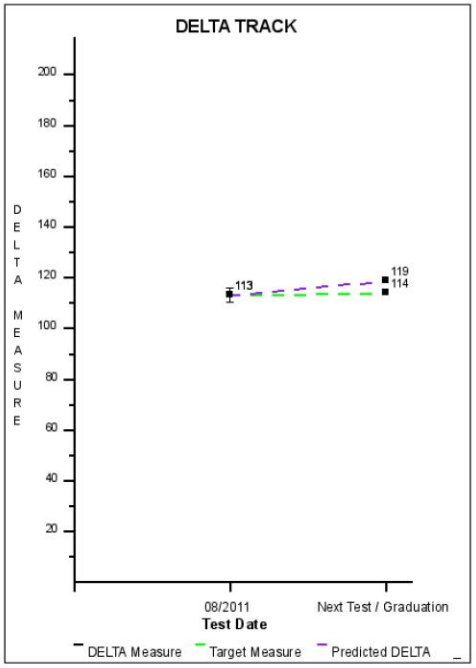
THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

DELTA
Diagnostic English Language Tracking Assessment

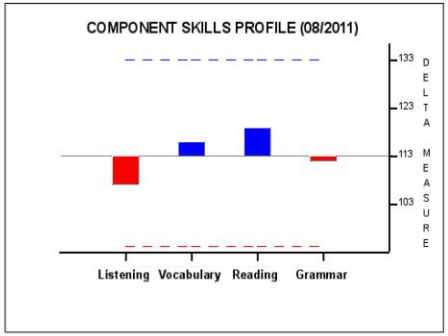
Diagnostic English Language Tracking Assessment
Candidate Report

Name: Demo Student **Student No:** 11000001D 4 September 2012

This is a report of your performance in the Diagnostic English Language Tracking Assessment (DELTA). Each time you take the DELTA, you will receive a DELTA Measure and a diagnostic report. The reports are cumulative, so that you can track your progress in improving your English.



The DELTA Track shows your English proficiency calculated from your performance on the DELTA tests. Each time you take the tests, your Delta Measure is plotted to show your progress.



The Component Skills Profile above shows the contribution that the scores that you attained on each component has made to your DELTA Measure. Components below the line indicate areas of relative weakness.

Last updated: 08/09/2011 3:10:21 PM

Appendix I

Component Diagnostic Report

The four reports below show your performance on each of the four tests in the DELTA. They show a description of the subskills tested by each of the items you did, in descending order of difficulty. Your proficiency level as indicated by your DELTA Measure is also shown. Items below the line of your proficiency level are those that you would be expected to answer correctly. The items that are highlighted indicate the subskills that you should focus on in your English language learning.

Listening

	Subskills tested	Text Type	Theme
More ↑	× Understanding information and making an inference	TV/Radio interviews	Media and communication
	× Understanding main ideas and supporting ideas	TV/Radio interviews	Business and marketing
	√ Identifying specific information		
			DELTA 113
	× Interpreting a word or phrase as used by the speaker	TV/Radio interviews	Business and marketing
	× Identifying specific information	TV/Radio interviews	Business and marketing
	× Understanding main ideas and supporting ideas	TV/Radio interviews	Business and marketing
	× Inferring a speakerâ€™s reasoning	TV/Radio interviews	Media and communication
	√ Interpreting an attitude or intention of the speaker		
	√ Interpreting a word or phrase as used by the speaker		
	× Identifying specific information	TV/Radio interviews	Business and marketing
	× Identifying specific information	TV/Radio interviews	Media and communication
	√ Identifying specific information		
	√ Understanding main ideas and supporting ideas		
	√ Inferring a speaker's reasoning		
	√ Understanding main ideas and supporting ideas		
	× Identifying specific information	Dialogues and conversations	Travel
	√ Identifying specific information		
	× Identifying specific information	TV/Radio interviews	Media and communication
	√ Identifying specific information		
	√ Identifying specific information		
	√ Identifying specific information		
	× Interpreting a word or phrase as used by the speaker	TV/Radio interviews	Media and communication
	√ Identifying specific information		
	× Identifying specific information	Dialogues and conversations	Travel

Difficulty

Last updated: 08/09/2011 3:10:21 PM

Less

√ Understanding main ideas and supporting ideas
√ Identifying specific information

The report for Listening above indicates subskills to work on. You can find resources for improving your listening skills in general at http://elc.polyu.edu.hk/CILL/listenin.htm. Please view the online version of this report to access more resources.

Last updated: 08/09/2011 3:10:21 PM

Appendix I

Vocabulary

		Academic Word Sublist	Words for revision
More ↑	√	AWL Sublist 8	
	×	AWL Sublist 10	intrinsic
	×	AWL Sublist 7	inferred
	×	AWL Sublist 8	ambiguity
	√	AWL Sublist 9	
	√	AWL Sublist 9	
			DELTA 113
	×	AWL Sublist 5	sustain
	√	AWL Sublist 10	
	√	AWL Sublist 9	
	×	AWL Sublist 2	distinction
	√	AWL Sublist 6	
Difficulty	×	AWL Sublist 1	variables
	√	AWL Sublist 8	
	√	AWL Sublist 9	
	√	AWL Sublist 7	
	√	AWL Sublist 8	
	√	AWL Sublist 3	
	√	AWL Sublist 5	
	√	AWL Sublist 3	
	√	AWL Sublist 6	
	×	AWL Sublist 3	minor
	√	AWL Sublist 5	
Less	√	AWL Sublist 2	
	√	AWL Sublist 2	

In the report for Vocabulary above, the items are drawn from the Academic Word List (AWL). The AWL is divided into 10 sublists of words used in academic English. Sublist 1 consists of the most frequent words. Sublist 2 contains the next most frequent and so on. You can find more information about the AWL at http://elc.polyu.edu.hk/CILL/eap/wordlists.htm. Please view the online version of this report to access more resources.

Last updated: 08/09/2011 3:10:21 PM

Appendix I

Reading

		Subskills tested	Text type	Theme
More ↑	√	Interpreting an attitude or intention of the writer		
	×	Interpreting a word or phrase as used by the writer	Feature articles	Health and bodycare
	√	Interpreting a word or phrase as used by the writer		
	√	Interpreting a word or phrase as used by the writer		
	×	Interpreting a word or phrase as used by the writer	Feature articles	Language
				DELTA 113
	×	Understanding main ideas and supporting ideas	Feature articles	Language
	×	Interpreting an attitude or intention of the writer	Feature articles	Language
	√	Interpreting a word or phrase as used by the writer		
	√	Understanding information and making an inference		
Difficulty	√	Interpreting a word or phrase as used by the writer		
	√	Identifying specific information		
	√	Interpreting an attitude or intention of the writer		
	√	Understanding main ideas and supporting ideas		
	√	Identifying specific information		
	×	Identifying specific information	Feature articles	Language
	√	Interpreting a word or phrase as used by the writer		
	√	Identifying specific information		
	√	Understanding main ideas and supporting ideas		
	√	Understanding main ideas and supporting ideas		
	√	Understanding main ideas and supporting ideas		
	√	Identifying specific information		
Less	√	Understanding main ideas and supporting ideas		
	√	Identifying specific information		

The report for Reading above indicates subskills to work on. You can find resources for improving your reading skills in general at http://elc.polyu.edu.hk/CILL/reading.htm. Please view the online version of this report to access more resources.

Last updated: 08/09/2011 3:10:21 PM

Grammar

		Subskills tested	Example
More ↑	×	Voice	The law should (amend) BE AMENDED immediately.
	√	Determiner	
			DELTA 113
	√	Present perfect tense	
	×	Determiner	She had a long holiday in (the Asia) ASIA last year.
	×	Preposition	It was late so he came home (on) IN a taxi.
	√	Voice	
	×	Simple present tense	It always (rained) RAINS at this time of year.
	√	Subject and verb agreement	
	√	Gerund	
	√	Present perfect tense	
Difficulty	×	Conditional	If I (will have) HAVE a day off from work, I will go to the beach.
	√	Relative pronoun	
	×	Possessive	It is important to abide by the (universities) UNIVERSITYâ€™S regulations.
	√	Phrasal verb	
	√	Word order	
	√	Gerund	
	√	Determiner	
	√	Gerund	
	√	Verb formation	
	√	Connective	
	√	Adjective	
	√	Modal	
	√	Word form	
Less	√	Voice	
	√	Word form	

The report for Grammar above indicates subskills to work on. In the second column, examples are provided for the incorrect items. The error is in brackets and the correct form is in capital letters. You can find resources for improving your grammar skills in general at http://elc.polyu.edu.hk/CILL/grammar.htm. Please view the online version of this report to access more resources.

Last updated: 08/09/2011 3:10:21 PM

Appendix I

Overall Performance

Your Component Skills Profile suggests that you should priorities your English language learning as follows:

1. Listening
2. Grammar
3. Vocabulary
4. Reading

You should study the diagnostic information relating to your performance on each test component to gain an idea of your areas of strength and weakness. You should aim to improve your proficiency by focusing in particular on those areas in which you have shown weakness.

The next time you take the DELTA, the tests will be targeted to your proficiency level, enabling you to demonstrate the progress that you have made.

Thank you for taking the DELTA and good luck in your English language learning.

Language Testing Unit
English Language Centre

Last updated: 08/09/2011 3:10:21 PM

Appendix II

DELTA student motivation survey (2011-2012)

1. Introduction

Thank you for taking the Diagnostic English Language Tracking Assessment (DELTA). As part of our on-going research and development, we are interested in your perceptions of the DELTA test, your motivation in taking the DELTA and the effectiveness of the DELTA and the DELTA Report. We would be grateful if you could spare a few minutes to complete this survey and answer all the questions as completely as possible. In order to help you to answer the questions, please have your DELTA Report with you.

The results of this questionnaire will provide us with valuable information for improvements. All responses and identities will be kept strictly confidential.

We appreciate your time. Thank you.

Appendix II

DELTA motivation survey (2011-2012)

2. Test motivation

This section is a measure of students' motivation and is taken from the Student Opinion Scale (Sundre and Moore 2002).

***1. Doing well on the DELTA was important to me.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

***2. I engaged in good effort throughout the test.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

***3. I am not curious about how I did on the DELTA relative to other students.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly disagree

***4. I am not concerned about the scores I receive on the DELTA.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

Page 2

Appendix II

DELTA motivation survey (2011-2012)

***5. The DELTA was an important test to me.**

- ◯ Strongly Disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

***6. I gave my best effort on the test.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

***7. While taking the DELTA, I could have worked harder on the questions.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

***8. I would like to know how well I did on the DELTA.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

***9. I did not give the DELTA my full attention while completing it.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

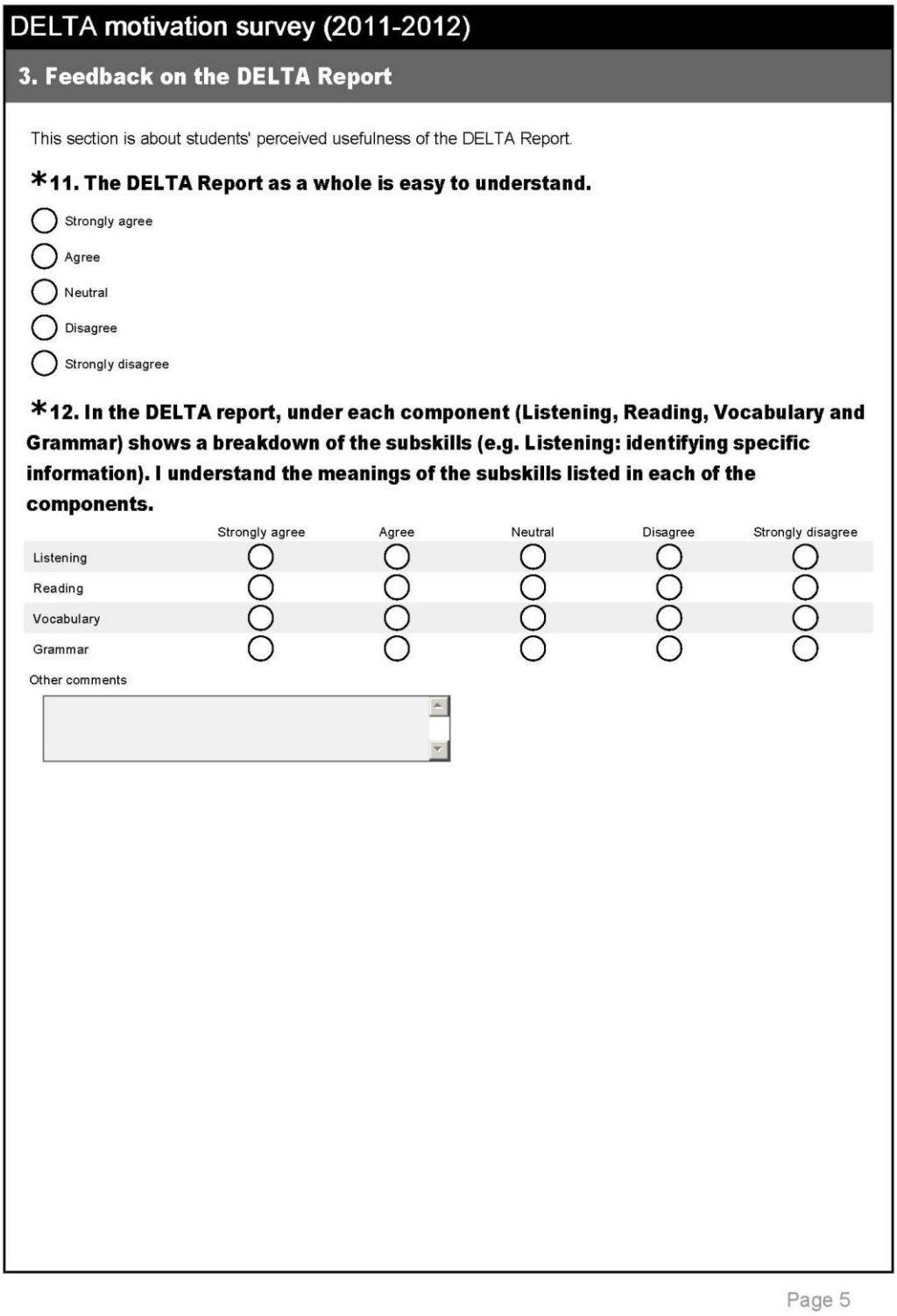
DELTA motivation survey (2011-2012)

***10. While taking the DELTA, I was able to persist until I had completed all the questions.**

- ◯ Strongly Agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly Disagree

Page 4

Appendix II



DELTA motivation survey (2011-2012)

3. Feedback on the DELTA Report

This section is about students' perceived usefulness of the DELTA Report.

***11. The DELTA Report as a whole is easy to understand.**

- ◯ Strongly agree
- ◯ Agree
- ◯ Neutral
- ◯ Disagree
- ◯ Strongly disagree

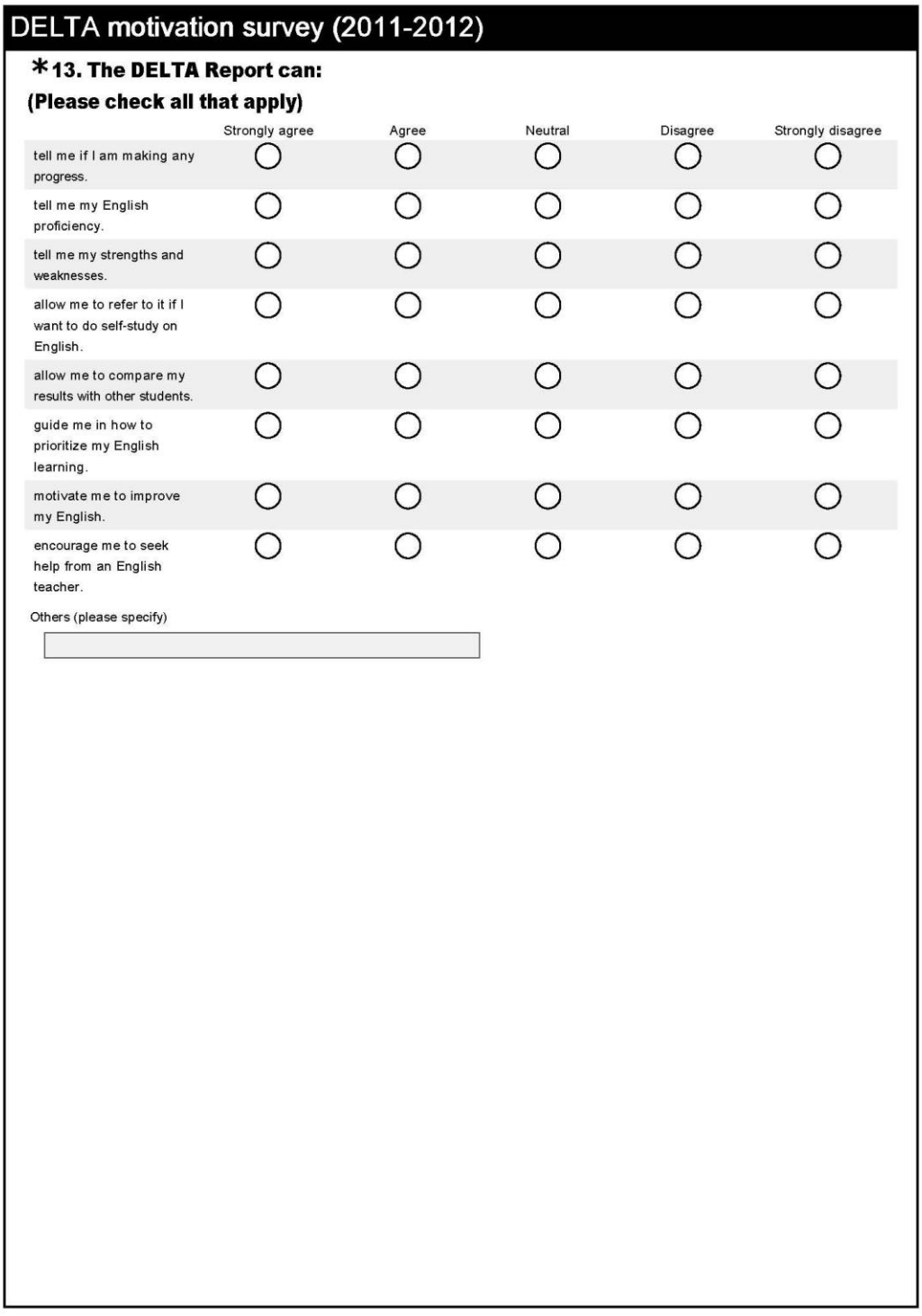
***12. In the DELTA report, under each component (Listening, Reading, Vocabulary and Grammar) shows a breakdown of the subskills (e.g. Listening: identifying specific information). I understand the meanings of the subskills listed in each of the components.**

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Listening	◯	◯	◯	◯	◯
Reading	◯	◯	◯	◯	◯
Vocabulary	◯	◯	◯	◯	◯
Grammar	◯	◯	◯	◯	◯

Other comments

Page 5

Appendix II



DELTA motivation survey (2011-2012)

*13. The DELTA Report can:

(Please check all that apply)

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
tell me if I am making any progress.	○	○	○	○	○
tell me my English proficiency.	○	○	○	○	○
tell me my strengths and weaknesses.	○	○	○	○	○
allow me to refer to it if I want to do self-study on English.	○	○	○	○	○
allow me to compare my results with other students.	○	○	○	○	○
guide me in how to prioritize my English learning.	○	○	○	○	○
motivate me to improve my English.	○	○	○	○	○
encourage me to seek help from an English teacher.	○	○	○	○	○

Others (please specify)

Page 6

Appendix II

DELTA motivation survey (2011-2012)

4.

***14. I want to take the DELTA again because (Please check all that apply)**

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I enjoy taking tests.	◯	◯	◯	◯	◯
I can know my English proficiency.	◯	◯	◯	◯	◯
I can see if I am making any progress.	◯	◯	◯	◯	◯
I want to compare my results with other students.	◯	◯	◯	◯	◯
The DELTA can tell me my strengths and weaknesses.	◯	◯	◯	◯	◯
The DELTA can guide me to prioritize my English learning.	◯	◯	◯	◯	◯
The DELTA can motivates me to improve my English.	◯	◯	◯	◯	◯

Others (please specify)

***15. I do not want to take the DELTA again next year because (Please check all that apply)**

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I hate tests.	◯	◯	◯	◯	◯
I have no time.	◯	◯	◯	◯	◯
the test is too long.	◯	◯	◯	◯	◯
the test is not a requirement for my studies.	◯	◯	◯	◯	◯
I do not think the test is a useful aid to language learning.	◯	◯	◯	◯	◯

Others (please specify)

***16. We would like to gain a deeper understanding of students' view later through face to face or telephone interviews. Please fill in your information below if you would like to take part.**

Name:

Student Number:

Department:

Email Address:

Page 7

Appendix III

List of interview questions

Part 1: The DELTA motivation questionnaire

In the DELTA motivation questionnaire that you completed after the test, there's a question "while taking the DELTA, I could have worked harder on the questions". How do you interpret this statement?
Another question in the questionnaire was "while taking the DELTA, I was able to persist until I had completed all the questions". How do you interpret this statement?

Part 2 DELTA and the expectancy-value model

Questions on expectancy: Are you satisfied with your previous A-level English results? Do you find the DELTA easy or difficult? What do you think about your English proficiency?
Questions on value: How important is English proficiency to you? Is the DELTA an important test to you? Why? Do you want to have a DELTA report each year? Will you make/Have you made use of the DELTA report? If so, in what way?

Part 3.1 Motivation as a dynamic process –-Preactional phase

Do you want to do well in the DELTA?
If you have to retake the DELTA again next year, will you set yourself a target? Will you have any expectation? What expectation do you have?
Will you do anything to prepare for it? When will you start to prepare? How much time will you spend on preparing? Why did you say X time?

Appendix III

Part 3.2 Motivation as a dynamic process –-Actional phase

Were you motivated in taking the DELTA? Why?
Can you describe your motivation level during the DELTA test?

Part 3.3 Motivation as a dynamic process –-Postactional phase

After taking the DELTA, have you evaluated your own performance?
After taking the DELTA, are you motivated in improving your English?
Can the DELTA report help you understand your strengths and weaknesses? And thereby help you in formulating strategies on how to improve/study English?
Will you retake the DELTA next year? Why?
Do you have the same level of motivation throughout the whole test preparation and test taking process?
Do you have the same level of motivation throughout the whole process of English study?
What are the things which will affect your motivation during your process of preparing and taking DELTA?
What are the things which will affect your motivation during your process of English study?
Do you think a diagnostic and tracking test like the DELTA is one of the effective ways in motivating students to study English?

Appendix III

How long does this motivation after the DELTA test and reading of the DELTA report last?
Finally are there any other things you want to say about the DELTA?

Appendix IV

Sample interview transcription and translation

Interviewer	Thank you for coming to the interview today. My name is Carrie. I am an instructor in the English Language Center, the Hong Kong Polytechnic University. I would like to talk to you today about your feelings and experience of the DELTA. Today's discussion will help not only my own doctoral study but provides valuable information for the ELC on its future planning.	
	I would like to mention a few points before we begin: - Feel free to express your opinions, whether you agree or disagree and whatever you have to say is fine. I want to know your opinions based on your own personal experience. - You are free to choose whether you want the interview to be conducted in English or in Cantonese. - I will treat our conversation as confidential. I am not asking for anything that could identify you and I will only use first names during the discussion. - I am audio recording the interview today because I don't want to miss any of your comments. I will not include your names or any other information that could identity you in any reports I write. Are you ok with this? - Finally this interview is going to take about 45 minutes. At the end of the interview, I will give you a $20 coupon to thank you for your participation. Do you have any questions before we start?	
Interviewer	Do you prefer to talk in English or in Cantonese?	
Candy	I prefer to talk in Cantonese.	
Interviewer	Okay, then we will switch to Cantonese from now on.	

Appendix IV

(Below is the transcription and translation of the interview from Cantonese to English.)		
Interviewer	I would like you to tell me your first name, your major and when you took the DELTA.	
Candy	My name is Candy. I am a year 1 student in the Institute of Textiles and Clothing. I took the DELTA at the beginning of this semester.	
Interviewer	In the DELTA motivation questionnaire that you completed after the test. There's a question "while taking the DELTA, I could have worked harder on the questions" (in English). How do you interpret this statement?	
Candy	I think it means "when I was doing the test, I worked harder on the questions than I normally would (in Cantonese)."	
Interviewer	Another question in the questionnaire was "while taking the DELTA, I was able to persist until I had completed all the questions" (in English). How do you interpret this statement?	
Candy	I think it means "I was able to persist until I finished with all the questions (in Cantonese)."	
Interviewer	Are you satisfied with your previous A-level English results?	
Candy	Hm… I was expecting a slightly better grade but it was ok. I am satisfied.	
Interviewer	Do you find the DELTA easy or difficult?	

Candy	I think it's normal, about right.	
Interviewer	What do you think about your English proficiency?	
Candy	It depends on which area of English. I am not confident in writing and grammar and I am stronger with reading, listening and oral. But in general, I think my English is good in the class.	
Interviewer	How important is English proficiency to you?	
Candy	It is very important and I love languages, no matter Chinese or English, so I think it is important.	
Interviewer	Is the DELTA an important test to you? Why?	
Candy	Hm…no…it has nothing to do with my study. It won't affect my GPA.	
Interviewer	Do you want to have a DELTA report each year?	
Candy	Yes, I think it will be good to have a report which shows us our strengths and weaknesses so that we know if we are making any progress in English each year.	
Interviewer	Will you make/Have you made use of the DELTA report? If so, in what way?	
Candy	Apart from knowing my grade and my strengths and weaknesses, I did not read the rest of the report in detail. I looked through the	

	report quickly but did not quite understand what it means, and then I just leave it there.	
Interviewer	Do you want to do well in the DELTA?	
Candy	I think the test is quite good and it does not take a lot of time to do… I wanted to do well and I was quite motivated as I'm always interested in knowing my own English proficiency. I love languages no matter Chinese or English.	
Interviewer	If you have to retake the DELTA again next year, will you set yourself a target? Will you have any expectation? What expectation do you have?	
Candy	I will not have the motivation to prepare for this test beforehand, so I will not set myself a target.	
Interviewer	Will you do anything to prepare for it? When will you start to prepare? How much time will you spend on preparing? Why did you say X time?	
Candy	Although I quite like this test and will try to do well during the test, I will not have the motivation to prepare for this test beforehand. I would expect myself to perform better in the second round but I will not work harder in my English because of this test… I don't know what I can do or prepare.	
Interviewer	Were you motivated in taking the DELTA? Why?	
Candy	I was quite motivated as I am always interested in knowing my	

	own English proficiency.	
Interviewer	Can you describe your motivation level during the DELTA test?	
Candy	Yes, I think my motivation has 7 out of 10 too.	
Interviewer	After taking the DELTA, have you evaluated your own performance?	
Candy	Yes I did think about why I performed like this in each component. I would like to know what the right answers are and what my wrong answers are. Especially for vocab, it would be much more useful if we can compare our wrong answers with the right ones. I am quite happy with my results in this test and I am happy to know that comparing with my classmates, I am not bad at all, so I think my English proficiency is probably ok. So it does provide me with some motivation to further improve.	
Interviewer	After taking the DELTA, are you motivated in improving your English?	
Candy	Yes, I am more confident now and more motivated to do even better.	
Interviewer	Can the DELTA report help you understand your strengths and weaknesses? And thereby help you in formulating strategies on how to improve/study English?	

Candy	I think that the test is quite good and I like the report which tells us what we have done right and what we have done wrong. And I also find the difficulty level of the test just right. And I notice that there's suggested links for us to visit. I think it is quite good.	
Interviewer	Will you retake the DELTA next year? Why?	
Candy	It depends if I have the time. Maybe. I quite like it and I think it does not need a lot of time to do the test. But it's hard to push myself to do it if it is voluntary. I may be lazy.	
Interviewer	Do you have the same level of motivation throughout the whole test preparation and test taking process?	
Candy	No, my motivation fluctuates. It rises when the test is getting nearer.	
Interviewer	Do you have the same level of motivation throughout the whole process of English study?	
Candy	No. For me it depends on how interactive are the English lessons. I like discussions with teachers and classmates.	
Interviewer	What are the things which will affect your motivation during your process of preparing and taking DELTA?	
Candy	I think the content of the test will affect my motivation. I am more interested in daily communicative English.	
Interviewer	What are the things which will affect your motivation during your	

	process of English study?	
Candy	Same as my motivation for the test. I think the content of the English lessons, whether they are interactive will affect my interests in studying English.	
Interviewer	Do you think a diagnostic and tracking test like the DELTA is one of the effective ways in motivating students to study English?	
Candy	Yes, I think it will be better than not having one no matter what. If the test is compulsory, we will at least have a way of tracking our English and force ourselves to do well during the test.	
Interviewer	How long does this motivation after the DELTA test and reading of the DELTA report last?	
Candy	It can only last about a week I think…	
Interviewer	Finally are there any other things you want to say about the DELTA?	
Candy	I don't like it online, I prefer pen and paper.	
Interviewer	That's all for the interview today. Thank you once again for your participation. You can contact me anytime via email if you have any questions about the test, the research project or the interview today.	