Exploring the Washback of a Large-scale High-stakes Chinese Test, the
Hanyu Shuiping Kaoshi, on Learner Factors

Shujiao Wang

Department of Integrated Studies in Education
McGill University, Montreal

The thesis is submitted in June, 2013.

A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of Master of Arts

© Shujiao Wang 2013
Abstract

Washback research has tended to focus on whether washback exists and whether there is intended washback brought about by examinations in English as a second/foreign language (ESL/EFL contexts). This study, on the other hand, investigated how learner factors, such as learning strategies and beliefs, related to the washback of a large-scale high-stakes Chinese second language proficiency test, the Hanyu Shuiping Kaoshi (HSK). Using a mixed-methods research (MMR) approach, quantitative data were collected from Chinese as a second/foreign/heritage language (CSL) learner survey responses (n = 60) and qualitative data were elicited from 8 interviews and HSK related documents. Findings revealed that, an increasing number of people wanted to learn Chinese and take the HSK because they were interested not only in the Chinese language and culture, but also hoped to study, work or travel in China. They felt that becoming HSK-certified, helped them feel more motivated to learn Chinese and increased job opportunities. Thus, there were significant washback effects of the HSK on learning Chinese. Similar to previous research on washback, however, this study showed that the HSK had positive impact on some learners but negative effects on others. The finding showed statistically significant differences between regular Chinese learning (e.g., CSL courses, distance learning, or self-learning) and HSK test-specific learning among the four skills - speaking, listening, reading and writing. Significant predictors for the change in learning strategies between regular and test-specific learning were nationality, language proficiency, HSK performance and motivation. Finally, although it is claimed that the validity and reliability of the HSK content are high, this study points to some aspects that could be improved, such as including more subjective question forms and emphasizing the output (speaking and writing) measurements. This study has implications for HSK test developers, CSL teachers and CSL learners.
Résumé

Les recherches sur le washback, qu’on définit comme l’impact des examens déterminants sur l’enseignement et l’apprentissage, tentent de se questionner sur l’existence du washback en général ainsi que sa présence prévue par des épreuves d’anglais comme langue seconde ou étrangère. Cependant, cette étude s’est plutôt penchée sur la manière dont le profil de l’apprenant, soit ses stratégies d’apprentissage ainsi que ses croyances, influe sur le washback (l’effet en retour) d’une épreuve rigoureuse de langue chinoise à grande échelle appelée le Hanyu Shuiping Kaoshi (HSK). À l’aide d’une approche de méthodes mixtes de recherche, nous avons recensé des données quantitatives issues d’un sondage auprès des apprenants du chinois comme langue seconde/étrangère/maternelle. Pour ce qui est des données qualitatives, elles ont été tirées de huit entrevues et de plusieurs documents faisant référence au HSK. Les résultats ont révélé que le nombre d’apprenants du chinois ne cesse d’accroître et qu’ils ont choisi de faire le HSK parce qu’ils avaient l’intention d’étudier, de travailler ou de voyager en Chine, en plus de leur vif intérêt pour la langue et la culture chinoises. Ces étudiants se sont rendu compte que le fait de réussir le HSK les a aidés à gagner davantage de motivation à apprendre le chinois et à élargir leur champ de débouchés. Ainsi, l’impact du HSK s’avérait significatif sur l’apprentissage de la langue chinoise. Bien que cette étude soit analogue à des recherches antérieures sur le washback, elle a démontré que l’impact du HSK s’avérait positif chez certains apprenants, mais négatif chez d’autres. D’un point de vue statistique, les conclusions de la présente recherche ont souligné des différences significatives entre les cours réguliers du chinois et ceux basés sur le HSK en ce qui concerne quatre compétences : l’oral, l’écoute, la lecture et l’écriture. Parmi les principaux facteurs jouant un rôle sur la variété des stratégies d’apprentissage, on retrouvait la nationalité, la maîtrise de la langue, la performance du HSK et la motivation. Finalement, même si la validité et la fiabilité du contenu du HSK apparaissent élevées, cette étude révèle certains aspects qui pourraient être améliorés, soit l’ajout de structures interrogatives subjectives et l’emphase sur les outils d’évaluation du sortant, c’est-à-dire l’expression orale et l’écriture. Cette recherche constitue donc une référence incontournable pour les concepteurs du HSK, pour les enseignants et pour les apprenants du chinois langue seconde.
Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor, Dr. Carolyn. E. Turner, for her continuous support, guidance and constructive feedback throughout all the time of research and writing of this thesis. I would like to thank the second reader, Dr. Mela Sarkar, for her encouragement and mentorship.

I am also truly appreciative towards the Chinese as a second/foreign/heritage language learners, whose agreement to participate in the study enabled this research to take place. Their voluntary participation in this research enabled me to finish this study and produce meaningful findings. I sincerely thank the directors, teachers in Chinese institutions/schools/programs, including Confucius Institutes at the University of Waterloo, Montreal and Vancouver, Chinese programs in McGill University, Shandong University and Dalian Foreign Language University. I thank many other friends and people, whom I did not mention here, but provided me with help through various ways seen or unseen. They will always be remembered.

Lastly, I extend my most sincere gratitude to my family, my parents Liyan Chen and Dexin Wang, for giving birth to me at the first place and supporting me spiritually throughout my life. I would like to thank my husband Zhenhuan Wei for his love, support and tolerance in my life.
Table of Contents

Abstract ........................................................................................................................................ ii
Résumé .......................................................................................................................................... iii
Acknowledgements ................................................................................................................... iv
Table of Contents ........................................................................................................................... v
List of Tables .................................................................................................................................... ix
List of Figures .................................................................................................................................... xi
List of Abbreviations and Acronyms ............................................................................................ xii

Chapter 1: Introduction .................................................................................................................. 1

1.1 The Rationale and Purpose of the study ............................................................................... 1
1.2 Research Questions .............................................................................................................. 3
1.3 The Significance of This Study ........................................................................................... 4
1.4 List of Definitions of Terms ............................................................................................... 5
1.5 Structure of the Thesis ......................................................................................................... 7
1.6 Chapter Summary ................................................................................................................ 8

Chapter 2: Literature Review ....................................................................................................... 9

2.1 Definitions of Washback ..................................................................................................... 9
2.2 Nature of Washback ............................................................................................................ 11
2.3 Theoretical Frameworks and Models for Washback Studies ............................................ 13
   2.3.1 Alderson & Wall’s washback hypotheses ..................................................................... 13
   2.3.2 Hughes and Bailey’s “PPP” basic model of washback .............................................. 14
   2.3.3 Cheng’s explanatory model of washback ................................................................. 15
   2.3.4 Gu’s basic model of CET washback ......................................................................... 17
   2.3.5 Hungerland’s contextual washback model ............................................................... 19
2.4 Washback on Teaching and Learning ................................................................................ 20
EXPLORING THE WASHBACK OF THE HSK ON LEARNER FACTORS

2.5 HSK Test and Washback Studies in the Chinese Context ........................................ 25
  2.5.1 Chinese proficiency test (HSK)........................................................................... 25
  2.5.2 The impact of the HSK and washback studies in the Chinese context........... 28

2.6 Learning Beliefs and Strategies.............................................................................. 31
  2.6.1 Language learning beliefs, attitudes and perceptions........................................ 31
  2.6.2 Language learning strategies and methods....................................................... 32
  2.6.3 Relationships among beliefs, strategies, motivation and learning outcomes... 34

2.7 Chapter Summary.................................................................................................. 35

Chapter 3: Methodology ............................................................................................... 37

3.1 Research Questions................................................................................................. 37
3.2 Mixed Methods Sequential Explanatory Design..................................................... 37
3.3 Participants............................................................................................................. 40
3.4 Instruments............................................................................................................ 41
3.5 Procedures............................................................................................................. 44
  3.5.1 Ethical issues.................................................................................................... 44
  3.5.2 Data collection.................................................................................................. 45
3.5.3 Data Analysis.................................................................................................... 48
  3.6.1 Quantitative data analysis process ................................................................. 49
  3.6.2 Qualitative data analysis.................................................................................. 51
3.7 Chapter Summary.................................................................................................. 53

Chapter 4: Presentation of the Results......................................................................... 54

4.1 Analysis of the HSK related documents................................................................. 54
  4.1.1 The reliability and validity of the HSK............................................................ 54
  4.1.2 Test-takers of the HSK................................................................................... 55
  4.1.3 Testing score.................................................................................................... 56
  4.1.4 HSK and Scales.............................................................................................. 57
EXPLORING THE WASHBACK OF THE HSK ON LEARNER FACTORS  

4.2 Questionnaire Data.............................................................................................................................. 59
  4.2.1 Why does the learner take the HSK test.......................................................................................... 59
  4.2.2 The washback effects of the HSK on learner factors................................................................. 62
  4.2.3 What factors contribute to washback effects? .............................................................................. 68
  4.2.5 What is the participant’s view of the HSK and the effects of the HSK? ............................... 70

4.3 Interview Results........................................................................................................................................ 72
  4.3.1 Profile of the participants............................................................................................................... 72
  4.3.2 Why does the learner take the HSK test?...................................................................................... 73
  4.3.3 The impact of the HSK on learning............................................................................................... 76
  4.3.4 What is the participant’s view of the HSK?................................................................................... 79

4.4 Chapter Summary....................................................................................................................................... 82

Chapter 5: Discussion of Results .................................................................................................................. 85
  5.1 Research Question 1: Learner’s Interest in Chinese and the HSK ............................................... 85
  5.2 Research Question 2: the Washback Effects on Learner Factors............................................. 88
  5.3 Research Question 3: How does the HSK Affect Learning?......................................................... 90
  5.4 Research Question 4: the Impact of the HSK............................................................................... 93
  5.5 Chapter Summary.............................................................................................................................. 95

Chapter 6: Conclusion................................................................................................................................. 97
  6.1 Research Summary............................................................................................................................ 97
  6.2 Limitations of the Study................................................................................................................... 98
  6.3 Future Research Directions............................................................................................................ 100
  6.4 Implications of the Study................................................................................................................ 101
    6.4.1 Implications for test-developers............................................................................................... 101
    6.4.2 Implications for students and teachers................................................................................... 102
    6.4.3 Implications for institutions................................................................................................... 103
  6.5 Chapter Summary............................................................................................................................ 105
List of Tables

Table 2.1 Summarized Washback Studies on Teaching and Learning …… 22
Table 2.2 The New HSK Test Structure ……………………………………… 27
Table 3.1 Summary of data collection ………………………………………… 47
Table 3.2 Summary of the procedures and methods of data analysis ……… 53
Table 4.1 Summary of Participants in Each Level of HSK Tests ………….. 55
Table 4.2 Summary of Participants of Each Level from Different Continents … 56
Table 4.3 Relationship among the New HSK Tests and Scales ……………… 58
Table 4.4 Summary of Descriptive Statistics of Responses to the reason for taking the HSK. …………………………………………………………….. 60
Table 4.5 Summary of Descriptive Statistics of Responses to Q47, Q48, Q51 and Q59 ………………………………………………………………………. 61
Table 4.6 Summary of the t Test Analyses of Asian and Western Groups …… 62
Table 4.7 Summary Results for Paired t Tests of Washback Effects on Learning 63
Table 4.8 Summary Results of t-tests for Speaking Strategies ………………… 64
Table 4.9 Summary Results of t-tests for Listening Strategies ………………… 65
Table 4.10 Summary Results of t-tests for Reading Strategies ………………… 66
Table 4.11 Summary Results of t-test for Writing Strategies ………………….. 67
Table 4.12 Summary Results of t-test for classroom activities ………………… 68
Table 4.13 Summary Results for Multiple Regression Model for the Predictor Variables ……………………………………………………………… 70
Table 4.14 Summary of Descriptive Statistics of Responses to Q54, Q55, Q56, Q57 and Q65 ……………………………………………………………… 71
Table 4.15 Summary of Descriptive Statistics of Responses to Q60, Q61, Q62 and Q63

Table 4.16 General Information of the Interview Participants
List of Figures

Figure 2.1 Basic model of washback - “PPP” .................................15
Figure 2.2 Model of the effect of washback within the HONG KONG educational context........................................................................................................17
Figure 2.3 Basis model of CET washback........................................18
Figure 2.4 Contextual Washback Model..........................................20
Figure 2.5 Washback Framework of GEPT on Students’ Learning........24
Figure 3.1 Visual diagram of the MMSE research design of the study........39
Figure 3.2 Proceeding of qualitative content analysis of the pertinent documents...........................................................................................................49
Figure 4.1 Description of Chinese Language Proficiency..................58
List of Abbreviations and Acronyms

CE: College English

CEFR: Common European Framework of Reference for Languages: Learning, Teaching, and Assessment

CET: College English Test

CLB: Canadian Language Benchmarks

CSL: Chinese as a Second/Foreign/Heritage Language

EFL: English as a foreign language

ESL: English as a second language

HSK: Hanyu Shuiping Kaoshi (Chinese Proficiency Test)

L2: Second Language

MMR: Mixed Methods Research

MMSE: Mixed Methods Sequential Explanatory

REB: McGill Research Ethics Board
Chapter 1: Introduction

1.1 The Rationale and Purpose of the study

Washback, a term now commonly used in educational measurement, language testing and applied linguistics, has become a prevalent phenomenon in educational research. It is generally understood as the influence of an examination on teaching and learning (Alderson and Wall, 1993; Cheng, 2004; Hughes, 2003). In recent decades, an increasing number of research studies have investigated the multifaceted meanings and mechanisms of washback effects on language teaching and learning. There is convincing evidence to suggest that language examinations, especially high-stakes large-scale tests, have powerful washback effects on teaching and learning within different educational contexts (Andrew, Fullilove and Wong, 1997; Cheng and Curtis, 2004).

The rapid development of China's economy has boosted the popularity of Chinese language and culture globally. *Hanyu Shuiping Kaoshi* (HSK) which translates literally to “Chinese Proficiency Test”, also known as the “Chinese TOEFL”, has become a widely accepted standardized test for Chinese as a second/foreign/heritage language (CSL) learners and it plays a vital role in certifying Chinese language proficiency for higher education and professional purposes. By 2010, over 170 countries had participated as regular hosts of the HSK and it had been taken by approximately 100 million people (including domestic ethnic minority candidates) (Luo, Zhang, Xie, and Huang, 2011). Despite the popularity of this test, very few empirical studies have been conducted
on the washback effects from the old version (before 2009) or the revised version (after 2009)\textsuperscript{1}.

My experience as a CSL and CFL teacher in China and North America have given me insider’s knowledge of the HSK test and CSL teaching and learning. In order to further understand the washback effects of a large-scale high-stakes test, it is important to look at the phenomenon in a specific educational context by investigating different aspects of teaching and learning. Only by piecing together the different tenets of such an investigation can understandings of the complex relationships between the language test and learner factors be understood. Compared to the widely researched washback effects from English as a Second/Foreign language (ESL/EFL) tests (e.g., washback from the International English Language Testing System, IELTS), studies on the washback of other language tests seem to be sparse. In addition, there is little research that has examined washback effects of tests on learning, in contrast to the considerable attention paid to washback effects on teaching (Wall, 2000; Watanabe, 2004).

Bailey (1999) observed that although language learners are key participants in the testing process, there is relatively little research that documents their points of view and their related learning behavior/strategies. The present study aims to address these gaps.

Specifically, this study intends to investigate the impact of the HSK test on

\footnotetext{\textsuperscript{1} The China National Office for Teaching Chinese as a Foreign Language is responsible for developing the HSK exams and introduced a new format in November 2009. The emphasis of the new format is comprehensive language and communication ability. Most notable is the inclusion of spoken and written segments at all levels, a reformation of the ranking system and use of new question structures. The new ranking system reduces the three test formats to Elementary, Intermediate and Advanced, compared to the 11 levels in the original test.}
learner factors from the test-takers’ view, taking into account the international trend in testing of combining assessment of learning\(^2\), assessment for learning\(^3\), with assessment as learning\(^4\). The “learner factors” in this study include motivation for learning, learning attitudes/beliefs, learning strategies and learning activities.

### 1.2 Research Questions

As discussed above, there is a lack of research that investigates the washback effects of the HSK on learner factors. The present study aims to explore a global question: What are the washback effects of the Chinese Proficiency Test, Hanyu Shuiping Kaoshi (HSK), on CSL learner factors?

There are four secondary questions:

1) Why are learners interested in learning Chinese and taking the HSK?

2) What is the evidence of washback effects on the learner factors?

3) How does the HSK affect the way learners learn Chinese? In other words, what factors contribute to the change of learning strategies/practice used in regular Chinese learning as opposed to HSK specific learning\(^5\)?

4) What are the views of the CSL learners towards the HSK and its impact?

---

\(^2\) Assessment of learning occurs at end of year or at key stages. It is summative and for grading and Report cards.

\(^3\) Assessment for learning is ongoing, diagnostic, and formative. It is for ongoing planning.

\(^4\) Assessment as learning actively involves students. It is ongoing, and it involves self and peer assessment. It provides students with the opportunity to use the feedback to improve learning.

\(^5\) I define regular Chinese learning to include CSL courses, distance learning, or self-learning on a daily basis. On the other hand, test-specific learning reflects participants’ learning largely on the basis of the test’s impact and includes all the learning beliefs, strategies and processes that affect success on the HSK, such as learning test-taking strategies and practicing simulated exam papers.
1.3 The Significance of This Study

The need to study washback effects on learning has been elucidated by a number of researchers (Shih, 2007; Wall, 2000; Watanabe, 2004). Despite the common concern shared among these researchers about L2 learners, no further studies have explored the washback effects on learning behaviors and learning beliefs with respect to the HSK. Therefore, a more systematic study of these factors is called for. As Watanabe (2004) pointed out, the study of learner behaviors is a promising area for further research. The present study addresses the area of washback on learning and in this respect, this study will not only help fill the gap in this research area, but the findings may also help shed light on further research, for example the assessment of CSL learning.

This research also has social significance for Chinese language learning and/or teaching. China is rising in terms of global status and an increasing amount of people are interested in learning the Chinese language and culture. This change has evolved since China opened its doors to the world thirty years ago. Nowadays 40 million people around the world are learning Chinese as a foreign language (Custer, 2010). That number is expected to multiply in the coming years.

In addition, this study also has pedagogical and practical value. The insights gained can help CSL learners reflect on their learning attitudes, beliefs, strategies and methods, and trigger a deeper understanding of their learning processes. Such a reflection may not only help learners to increase their awareness of how they learn, but also help them change their perceptions towards tests. It also may influence CSL teachers to adjust their teaching approaches. As a result,
with a more positive perception of the washback effects from tests, learners may be encouraged to promote their learning outcomes through improving their learning methods/strategies.

In sum, there are several potentially salient implications of this study. Although the issue of washback has been touched upon across a range of subjects, it has not been dealt with in depth from test-takers’ perspectives. This study will contribute a new perspective to the existing washback literature, and contribute to the ongoing call in the field for more intensive investigation of washback on learning (Alderson, 2004; Andrews et al., 2002; Turner, 2009). Furthermore, the findings of this study may not only provide findings that are applicable to pedagogical and methodological issues of CSL teaching, but contribute to the development of second language acquisition and second language assessment theories in general.

1.4 List of Definitions of Terms

Below I introduce a list of definitions of terms and concepts that will be used frequently in this thesis. These definitions are provided to guide the reader. They are interpreted within the context of CSL learning.

**High-stakes tests**: This term is used to describe tests that have major consequences for students, teachers and schools or are the basis of a major decision, such as for admission purposes into a university. High-stakes tests are likely to be influential on the teaching and learning behaviors of those involved in the tests. Thus, even a minor change in the test can cause strong washback effects on the stake-holders (Shohamy, Donitsa & Ferman, 1996).
**Washback:** This term is used to refer to the influence on teaching and learning generated by a test (usually a high-stakes test). The influence can be either positive or negative (Bailey, 1996). Washback is further defined in the next chapter.

**HSK:** *Hanyu Shuiping Kaoshi* (HSK), which is translated as Chinese Proficiency Test, is a national standardized test designed to evaluate Chinese proficiency of non-native Chinese speakers (including foreigners, overseas Chinese and Chinese national minorities).

**Chinese as a Second/Foreign/Heritage Language (CSL):** Increased interest in China from those outside China has led to more interest in studying Standard Chinese as a foreign language, the official language of mainland China and Taiwan. However the teaching of Chinese both within and outside China is not a recent phenomenon. Westerners started to learn Chinese language in the 16th century. In recent decades, China has helped 60,000 teachers promote CSL internationally, and an estimated 40 million people were studying Chinese as a second language around the world at the end of 2008 (Custer, 2010). Since the HSK is designed for non-native Chinese speakers as well as overseas Chinese, I include Chinese as a Heritage Language in CSL.

**Chinese Language Proficiency Scales for Speakers of Other Languages** (Office of Chinese Language Proficiency Scales for Speakers of Other Languages, 2009), abbreviated as *Scales*, is an official document with guidelines for CSL teaching and learning. It was created to meet the needs of Chinese language teaching and learning worldwide. It was developed by language education and
testing experts from over 80 universities in China and abroad. Designed for learners of Chinese as a foreign language, the Scales provide a five-band all around description of learner ability to use their knowledge and skills of the Chinese language for communication. It is an important basis on which the language proficiency of learners of Chinese can be measured.

1.5 Structure of the Thesis

This thesis consists of six chapters. In this chapter, I introduced the current study. The review of the literature in Chapter 2 presents an overview of theoretical and empirical washback studies in relation to language assessment. Its first section includes washback definitions, investigation of the teacher factor and the learner factor in washback studies, and washback models. The second section introduces the HSK test and discusses the washback studies in CSL context. The third section discusses language learners’ beliefs/attitude/perceptions, strategies/methods, and the relationship of them with the tests’ impacts. Chapter 3 addresses the mixed methods research methodology of the present study and describes the participants, instruments, data collection and analyses. Chapter 4 presents qualitative and quantitative results of HSK related documents, participants’ perceptions, behaviors, strategies and beliefs toward the HSK, the washback effects of the HSK and the factors contributing to the washback effects. In Chapter 5, the discussion section follows the research questions and builds on the analyses of the findings described in the previous chapter. Finally, concluding remarks are found in Chapter 6. Implications, suggestions and limitations of the current study are presented.
1.6 Chapter Summary

This chapter opened with the rationale and the purpose of the study. Then I introduced the research questions that the thesis aims to answer and described the significance of the research. Finally, the chapter outlined definitions of key terms used in this thesis and provided a brief overview of the structure of the thesis. In the next chapter, I review the literature of washback studies, the HSK test, and learning theories.
Chapter 2: Literature Review

The purpose of this chapter is to gain insight into the complex dimensions of washback and illuminate the impact of high-stakes tests on learning, especially in the context of CSL teaching and learning. To provide a theoretical basis for this research, first of all, I review relevant literature on washback in second language (L2)/foreign language (FL) education. This includes definitions of washback; the nature of washback – negative and positive, either or neither; washback studies from teachers’ and learners’ perspectives; and theoretical frameworks/models for studying washback. Then I give detailed information about the background of the context of the research, and I introduce the HSK test and related test impact studies. In the last part of this chapter, I address language learners’ beliefs, attitudes and knowledge, as well as their learning strategies and methods, and comment on the relationships among these factors.

2.1 Definitions of Washback

In the last decade, there has been an increasing interest in the issue of washback and several different definitions have been proposed in the field of applied linguistics, language testing/assessment and educational measurement. Washback, sometimes called backwash, refers to the impact of a test on teaching and it also refers to the extent to which the test influences language teachers and learners to do things “they would not necessarily otherwise do because of the test” (Alderson & Wall, 1993, p.117). Prodromou (1995) further stated that, “the backwash effect can be defined as the direct or indirect effect of examinations on teaching methods” (p. 13). Messick (1996) viewed washback as an “instance of
the consequent aspect of construct validity” (p. 242), which covers elements of
test use, the impact of testing on test-takers and educators, the interpretation of
results by decision-makers, and any possible misuses, abuses, and unintentional
effects of tests.

Wall (1997) distinguished test impact and test washback in terms of the
scope of the effects. According to Wall, impact encompasses a broader meaning
than washback, and includes any of the effects that a test may have on individuals,
policies or practices, within the classroom, the school, the educational system or
society as a whole. On the other hand, she saw washback as the effect of tests on
teaching and learning. Hamp-Lyons (1997) criticized the term washback as being
too narrow and pointed out that general education and educational measurement
tend to employ the more general term of impact, which includes effects beyond
the classroom as well, such as effects on the educational system and society as a
whole. Turner (2001) stated that “‘washback’ in second language education,
generally refers to the influence of a test or other evaluation procedures on
teaching and learning; and the effect could be positive or negative” (p.138).

In summary, although different researchers prefer different terms, they all
refer to facets of the same phenomenon: the influence of testing on teaching and
learning. The washback of the HSK in this study is defined as the effects or
influences of the test on CSL teaching and learning. In addition, since the
intention of the current study is to explore test takers’ perceptions and their
learning methods/strategies under the influence of HSK, this research mainly
examines the washback effects of the HSK on learning.
2.2 Nature of Washback

Although there is general agreement in the field of language assessment on the existence and importance of the washback phenomenon, there still is considerable variety in opinions about how washback functions (Bailey, 1996). The way it is perceived depends on the theoretical standpoint researchers take as well as the educational context that they are associated with. A case in point is that no consensus can be reached in the language testing community as to whether washback effects are positive or negative, or both. Consistent with this view, Cheng and Curtis (2004) stated that the use of tests has been subject to criticism and pointing out positive or negative washback effects could be seen as a form of criticism.

In the past, most discussions of the influence of examinations emphasized their supposed negative effects (Andrews, 2004). Undertaking the review of World Bank research carried out in 14 African countries on the effects of examinations over the curriculum and on the quality of teaching and learning, Kellaghan and Greaney (1992) concluded that most of the focus is on the negative effects that high-stakes examinations can have on the classroom (e.g., the neglect of subjects and skills which are not examined, the excessive use of past exam papers as teaching material, frequent practice in examination-taking techniques) and on individuals (e.g., the promotion of a passive concept of learning).

There are other researchers who view tests as having an impact in a more positive way. Alderson (1986), for example, argued for innovations in the
language curriculum through innovations in language testing. Noble and Smith (1994) put forward that “the goal of current measurement-driven reforms in assessment is to build better tests that will drive schools toward more ambitious goals and reform them toward a curriculum and pedagogy geared more toward thinking and away from rote memory and isolated skills” (as cited in Cheng & Curtis, 2004, p.5). Crooks (1988) emphasized that tests can have a positive effect on learning if teachers stress the need for deep learning rather than surface learning.

Meanwhile, many researchers have argued that it is difficult to determine whether effects of tests are positive or negative. For instance, Alderson & Wall (1993) stated that there is no simple relationship between tests, good or bad, and their effects on the classroom. Messick (1996) highlighted that, “a poor test may be associated with positive effects and a good test with negative effects because of the other things that are done or not done in the education system” (p. 242). Brown and Hudson (1998) claimed that when the assessment procedures correspond to the course goals and objectives, a positive washback effect occurs. In contrast, if the assessment procedures do not reflect the curriculum objectives, they are likely to generate a negative washback.

In summary, washback could be either positive or negative, or both. Its influence on teaching and learning is a complex phenomenon. Hence, more empirical studies are needed to explore the nature of washback, and how it functions to influence teaching and learning. That is one of the reasons for carrying out the present study.
2.3 Theoretical Frameworks and Models for Washback Studies

Since 1993, there has been a shift from making assertions about washback to establishing theoretical frameworks for washback studies. In this section, five influential theoretical frameworks are introduced. They are Alderson & Wall’s (1993) washback hypotheses, Hughes and Bailey’s (1996) “PPP” basic model of washback, Cheng’s (2005) explanatory model of washback, Gu’s (2007) College English Test (CET) washback model and Hungerland’s (2005) contextual washback model.

2.3.1 Alderson & Wall’s washback hypotheses

As the pioneers in the research on washback, Alderson & Wall (1993) provided the most thorough and comprehensive explanation of the term washback. They came up with 15 hypotheses regarding washback to illustrate tests’ impacts on teachers and learners, teaching and learning, and attitudes concerning method and content of teaching and learning. The hypotheses are:

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

(Alderson & Wall, 1993, p. 120-121).

Since the current study concentrates on the influence of HSK on CSL learners, hypotheses (2, 5, 6, 8, 10, and 11) provide the guidance for investigating students’ perceptions of the HSK and the washback effects of the test.

2.3.2 Hughes and Bailey’s “PPP” basic model of washback

Hughes (1993) discussed the mechanisms by which washback works. He made a distinction of participants, processes and products in teaching and learning, recognizing that these three may be affected by the nature of a test. Bailey (1996) synthesized Hughes’ ideas and proposed a basic model of washback to explore the complicated mechanisms of washback (see Figure 2.1). She categorized the impact of a test into two categories: washback to the learners, or the direct impact of the test on test-takers; and washback to the program, or the impact of the test on teachers, administrators, curriculum developers and researchers.
(Bailey, 1996, p. 264.)

Figure 2.1: Basic model of washback - “PPP”

From this model, we can see that a test has direct influences on the participants who are involved in various processes. These processes result in products specific to each type of participant. In addition, unlike the washback hypotheses by Alderson & Wall (1993) that only demonstrates the linear relationship between tests and teaching/learning, this model also shows the interaction among different components. The current study adapted this basic model of washback in a simple way and I will describe it in detail in the following chapter.

2.3.3 Cheng’s explanatory model of washback

Drawing from all sources in the classroom at the macro and micro levels, Cheng (2005) proposed a working model to study the complex and multidimensional nature of washback (see Figure 2.2). This model served as a theoretical and methodological guide for Cheng’s washback study on the revised
Hong Kong Certificate of Education Examination in English (HKCEE). The curriculum as planned at Level 1 refers to the intended changes behind the new examination syllabus; the curriculum in action and curriculum as outcome elements were studied in Level 2 and 3. In addition, Cheng (2005) put forward a new notion for studying washback - washback intensity. This refers to the degree of washback in an area, or in a number of areas of teaching and learning affected by examination. The new concept supplements the concept of washback and enriches the 15 Washback Hypotheses by Alderson & Wall (1993). The current study adapted this explanatory model of washback in terms of test reform and I describe it in detail in the following chapter.
<table>
<thead>
<tr>
<th>Level 1</th>
<th>Participants</th>
<th>Processes</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision – making level</td>
<td>ED</td>
<td>Syllabus design</td>
<td>new HKCEE</td>
</tr>
<tr>
<td></td>
<td>HKEA &amp; CDC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 2</th>
<th>Participants</th>
<th>Processes</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervening level</td>
<td>Textbook publisher</td>
<td>Material development bring about changes in teaching and learning through teacher education</td>
<td>new materials</td>
</tr>
<tr>
<td></td>
<td>Teacher educators</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level 3</th>
<th>Participants</th>
<th>Processes</th>
<th>Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>School level</td>
<td>Participants</td>
<td>Changes in attitudes towards teaching and learning changes to teaching and learning activities</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Panel chairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teachers &amp; Students</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**New 1996 HKCEE**

**Improved learning**

Note: ED = Education Department
HKEA = Hong Kong Examinations Authority
CDC = Curriculum Development Committee
(Cheng, 2005, p.242)

**Figure 2.2: Model of the effect of washback within the HONGKONG educational context**

**2.3.4 Gu’s basic model of CET washback**

Gu (2007) conducted an empirical study of the College English Test (CET)\(^6\) in China and developed a basic model of CET washback (see Figure 2.3).

---

\(^6\) CET is a national English as a foreign language test in mainland China. The purpose of the CET is to examine the English proficiency of undergraduate students and postgraduate students and
According to Gu, the CET exerts influence on CET participants, including CET administrators, teachers and students. The impact manifested on three aspects: the perceptions of the test and its washback; the processes of both regular classroom teaching and learning and test-oriented practice; and finally the products of candidates’ learning outcomes presented in the CET (Gu, 2007).

(Gu, 2007, p.7)

**Figure 2.3: Basic model of CET washback**

Gu found that each group of participants (administrators, teachers and students) is related to major factors impacting college English (CE) teaching and learning. The group of administrators is related to administrative measures, such as CE course planning, measures to cope with teaching staff shortage, and material incentives to teachers. Teacher factors include the composition of CE
teaching staff, teacher training, and teachers’ views on teaching and practice. The group of students is related to learner factors, such as students’ educational background, time investment in English learning and input they receive in English. The present study considers this model as it explores students’ perceptions of the HSK.

2.3.5 Hungerland’s contextual washback model

Hungerland (2005) contributed the Contextual Washback Model (see Figure 2.4), which illustrated the test-specific factors, and contextual factors that affect the washback on an individual learner within a high-stakes language assessment and learning environment. In particular, test-specific factors appeared to reflect participants’ perceptions largely on the basis of specific properties of the test, while contextual factors appeared to reflect participants’ perceptions largely on the basis of extra-test factors. This model emphasizes that extra-test factors, such as participants’ affective experiences and immediate surroundings, appear to play an important role in shaping washback on learners. The author also stated it is important to keep in mind that the individual factors represented in the Contextual Washback Model are to be interpreted as surrounded by the wider context of the educational system, culture and society. She suggested that researchers should focus greater attention on contextual factors, and more precisely on learner perceptions that may be shaped by both test-specific and contextual factors, rather than on the impact of the test alone. The current study considers this model due to its focus on the individual learner.
Overall, these 5 models of washback all offer different resources for my study. With these considerations, this study makes use of these 5 models. They served as a baselines and theoretical framework to design the research methodology.

2.4 Washback on Teaching and Learning

In the field of second or foreign language education, testing is often used as one of the most widely favored instruments to encourage innovations in teaching and learning. The washback theoretical framework and models discussed above have sparked empirical research that has aimed to discover the effects of washback on language teaching and learning. Some representative studies are summarized in Table 2.1. They are discussed chronologically and their

**Figure 2.4: Contextual Washback Model**

(Hungerland, 2004, p.101)
methodological considerations as to how to measure various aspects of washback from different stakeholders are examined.
Table 2.1: Summary of Washback Studies on Teaching and Learning

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exams studied</td>
<td>2 EFL oral tests</td>
<td>TOEFL</td>
<td>Old &amp; New HKCEE</td>
<td>GEPT</td>
<td>IELTS writing test</td>
</tr>
<tr>
<td>Purpose</td>
<td>To examine the two national tests’ impact of the oral test on teaching</td>
<td>To ascertain influence of the TOEFL on class teaching</td>
<td>To compare teachers’ perceptions toward both exams</td>
<td>To explore the stakeholders’ perceptions of the GEPT and its washback on teaching, learning and school’s policies</td>
<td>To examine how preparation classes impact score gains</td>
</tr>
<tr>
<td>Methodologies</td>
<td>1) Student questionnaires 2) Structured interviews with teachers and inspectors 3) Analysis of inspectorate bulletins</td>
<td>1) Interviews with teachers and students 2) Classroom observations</td>
<td>1) Teacher/student questionnaires 2) Structured interviews with teachers 3) Classroom observations</td>
<td>1) Interviews with department heads, teachers, students, and family members 2) Classroom observations</td>
<td>1) Two IELTS writing tests 2) Two questionnaires consisting of participant and process variables respectively</td>
</tr>
<tr>
<td>Collected Evidence</td>
<td>1) More positive washback found in ESL 2) More negative washback found in ESL</td>
<td>1) More occurrences of teacher talk, the use of meta language in non-TOEFL classes 2) Fewer opportunities for pair work, laughter, and turn-taking in TOEFL classes</td>
<td>1) An increased change in teaching content and activities 2) A lack of change in teaching methodologies</td>
<td>1) Small but varied aspects of washback found in students at both schools with and without exit requirements 2) External, intrinsic and test factors explain GEPT's minor impact on students' learning</td>
<td>An improvement in test scores for learners in test-preparation or academic-oriented classes, but those in the former progressed no more than those in the latter</td>
</tr>
<tr>
<td>Conclusions</td>
<td>Washback changes over time because of factors including language status and test uses.</td>
<td>TOEFL affects both what and how teachers teach, but the effect varies with teachers.</td>
<td>The change on teaching content rather than methodology was attributed to inadequate training and qualifications of secondary English teachers.</td>
<td>The current washback theory didn't account for GEPT washback, so a new learning washback model has been developed.</td>
<td>Test preparation classes have no apparent benefit to improve test scores.</td>
</tr>
</tbody>
</table>
It can be seen that most research findings showed that high-stakes tests had a significant impact on L2 teaching in that the tests altered teachers’ methods (e.g., Alderson & Hamp-Lyons, 1996; Cheng, 1997, 2005). On the other hand, other studies indicated that washback happened to some teachers, but not to others, and Cheng (1997, 1999) showed that washback happened in some specific contexts, but had limited impact on others. In addition, results also suggested that tests had an impact on how teachers administered tests (Wall & Alderson, 1993) and imposed anxiety and fear on teachers due to their high-stakes status (Ferman, 2004).

Whereas a majority of the research has investigated washback on teaching, little research has examined the effects on learning, with the exception of the afore-mentioned research conducted on the GEPT and HKCEE. In Cheng’s early study (Cheng, 1997), she examined the washback of the HKCEE on learning through a survey. Results indicated that the test was the most significant factor involved in motivating students to learn English, more so than future career plans. In her later study, Cheng (1998) found that although students changed their learning beliefs after the content of the test had been changed, they reported retaining their original learning processes, learning strategies and individual motivations to learn English.

As Watanabe (2004) stated, "relatively well explored is the area of washback to the program, while less emphasis has been given to learners" (p. 22). One notable exception and significant contribution was Shih’s study. Shih (2007) investigated the washback of the GEPT on English learning in Taiwan and found
that existing theories did not fully explain the washback of this test on educational context, so a new tentative washback framework was proposed (see Figure 2.5).

(Shih, 2007, p. 151)

**Figure 2.5: Washback Framework of GEPT on Students’ Learning**

In summary, washback studies have predominantly investigated those examinations that were tailored to bring about top-down educational reform. The
empirical research of washback effects on learning process and learning beliefs seems sparse. While some researchers have pointed out this oversight (e.g., Shih, 2007; Wall, 2000, Watanabe, 2004), more research on this topic is still needed. In addition, since most studies concerning washback effects on learning only employed qualitative research, or relied on questionnaires, I would argue that more in-depth information will emerge if diverse research methods are involved, such as mixed methods.

2.5 HSK Test and Washback Studies in the Chinese Context

2.5.1 Chinese proficiency test (HSK)

2.5.1.1 Introduction

The Chinese Proficiency Test, also known as the HSK or Chinese TOEFL, is a national standardized test designed to evaluate Chinese proficiency of non-native Chinese speakers (including foreigners, overseas Chinese and Chinese national minorities). Development of the HSK began in 1984 at Beijing Language and Culture University and in 1992 the HSK became the official national standardized test. Hanban (中国国家对外汉语教学领导小组办公室7) is currently responsible for the HSK exams and introduced a new format in November 2009. The HSK test was launched by Hanban after experts from

7 Hanban is the colloquial abbreviation for the Chinese National Office for Teaching Chinese as a Foreign Language. It is a non-government and non-profit organization affiliated with the Ministry of Education of the People’s Republic of China. Hanban is most notable for the Confucius Institute program, but it also sponsors the Chinese Bridge competition, which is a competition in Chinese proficiency for non-native speakers. According to the mission statement: Hanban is committed to developing Chinese language and culture teaching resources and making its services available worldwide, meeting the demands of overseas Chinese learners to the utmost degree, and to contributing to global cultural diversity and harmony. Generally, Hanban is charged with cultivating knowledge and interest in the Chinese language and culture in nations around the world that are not native speakers of Chinese.
various disciplines, such as Chinese language teaching, linguistics, psychology and educational measurement, conducted in-depth surveys on CSL teaching at home and abroad. The new HSK is an international standardized Chinese language ability test based on the Chinese Language Proficiency Scales for Speakers of Other Languages. Test developers took into consideration the differences between professional learners and non-professional learners living in China and learners residing overseas, and also incorporated aspects of the previous version of the HSK, and drew on the latest findings in global language testing.

2.5.1.2 Test structure

The new HSK assesses non-native Chinese speakers’ abilities in using the Chinese language in their daily, academic and professional lives. The new HSK consists of a writing test and a speaking test, which are independent of each other. There are six levels of the writing tests, namely the HSK Levels 1 to 6. There are three levels of the speaking test, namely the HSK Speaking Test, beginner, intermediate and advanced. The current test structure is presented in Table 2.2. During the speaking test, test takers’ speech is recorded. According to regulations by China’s Ministry of Education, foreign students can enroll in degree-granting major courses in a Chinese university/college based on their HSK test results. The HSK test results can also serve as a reference tool for employers when making recruiting decisions.
Table 2.2: The New HSK Test Structure

<table>
<thead>
<tr>
<th>Level</th>
<th>Words</th>
<th>Characters (cumulative/new)</th>
<th>Written test</th>
<th>Oral test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Listening</td>
<td>Reading</td>
<td>Writing</td>
</tr>
<tr>
<td>1</td>
<td>150</td>
<td>178</td>
<td>20 questions</td>
<td>20</td>
<td>(Not tested)</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>349</td>
<td>35 questions</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>600</td>
<td>623</td>
<td>40 questions</td>
<td>30</td>
<td>10 items</td>
</tr>
<tr>
<td>4</td>
<td>1200</td>
<td>1074</td>
<td>45 questions</td>
<td>40</td>
<td>15 items</td>
</tr>
<tr>
<td>5</td>
<td>2500</td>
<td>1710</td>
<td>45 questions</td>
<td>45</td>
<td>10 items</td>
</tr>
<tr>
<td>6</td>
<td>5000</td>
<td>2633</td>
<td>50 questions</td>
<td>50</td>
<td>1 composition</td>
</tr>
</tbody>
</table>


2.5.1.3 Test principles

The new HSK follows the principle of test-teaching alignment, is based on the current trends in international Chinese language instruction, and is closely related to CSL textbooks. The purpose of the test is to promote instruction and learning. This test claims to measure objectively and accurately learners’ Chinese
language abilities. The new HSK sets clear test objectives to allow the test takers to be able to improve their Chinese language abilities in a systematic and efficient way.

2.5.1.4 Testing purposes

The new HSK retains the former HSK’s orientation as a general (or universal) Chinese language ability test for adult learners. According to Chinese Language Proficiency Test Levels 1-6 (Office of Chinese Language Council International, 2010), the results of the test can serve several purposes:

1) A reference for an educational institution’s decision-making concerning recruiting students, assigning students to different classes, allowing students to skip certain courses and granting students academic credits;

2) A reference for employers’ decision-making concerning the recruitment, training and promotion of test takers;

3) A method for Chinese language learners to assess and improve their proficiency in Chinese; and

4) A method for Chinese language training institutions to evaluate training results.

2.5.2 The impact of the HSK and washback studies in the Chinese context

Inspired by washback research conducted in the ESL/EFL context, an increasing number of washback studies have been conducted in the Chinese context over the past decade mainly concerning EFL teaching (Gu, 2007; Qi, 2004, 2007; Wang, 2010). Both Gu (2007) and Qi (2007) demonstrated that
Despite the test designers’ efforts to reduce the weighting on linguistic knowledge in high-stakes tests, teachers continue to focus on teaching linguistic knowledge rather than communication skills. Similarly, Liu and Dai’s (2003) study revealed that more than 90% of Chinese college English instructors maintained that the CET cannot objectively reflect students’ communicative competence, and 85% would like to improve their teaching methods according to the test reform.\(^8\)

The development of the HSK speeded up the development of CSL teaching at home and abroad (Zhang, 2001). Compared to the abundant research on EFL tests in the Chinese context, it is surprising that there is very little washback research on Chinese language tests. This is not commensurate with its important status. Yang and Liao (2000) examined a case study of the old HSK’s impact on 67 Thai college-level students whose major was Chinese. The results revealed that the students with positive beliefs about learning and testing got better scores than those with negative beliefs. They also indicated that the HSK provided negative influence on CSL teaching, such as paying more attention to linguistic knowledge but ignoring the communicative skills (including speaking and writing).

Huang and Li (2010) surveyed 150 CSL teachers from 8 Chinese universities and found that both positive and negative washback of the old HSK version exists in current CSL education. On the one hand, teachers reported that

\(^8\) In order to raise the efficiency of college English (CE) teaching in China, a revised curriculum, the College English Curriculum Requirements, was officially announced in 2004. It was developed and introduced as guidelines for CE instruction. Following the revision of the curriculum, the CET was revised from its previous focus on reading comprehension to an emphasis on assessing students’ aural aspects of English, namely, listening and language use.
the HSK objectively and fairly reflects students’ language proficiency. Since “the HSK is the baton of CSL teaching and learning” (p.26), the test can inform teachers about the students’ strengths and weaknesses, and help them to adjust their teaching objective and teaching methodologies. It can motivate students’ self-learning as well. On the other hand, they criticized that the HSK overemphasized the memorization of vocabulary and grammar rules instead of the use of real language abilities. They also pointed out that, to some extent, it is not fair for western test-takers compared to the ones from Asia. Since several Asian languages have been written with a form of Chinese characters, such as Vietnamese, Korean and Japanese, it is easier for people from these countries to read and write Chinese. It is crucial to note that most washback studies of the HSK are based on the old HSK and the washback effect of the new version on learning cannot be taken for granted without empirical evidence. In order to get a comprehensive understanding of the washback effects, more research on all the stakeholders of the HSK, particularly learners, is called for.

Some researchers have focused on investigating the HSK’s impact on other aspects beyond teaching and learning, for example the CSL textbooks and HSK preparation materials. Under the influence of structural linguistics theories, the old HSK and most currently used CSL textbooks were designed to focus on the normalization and accuracy of the language, such as emphasizing grammar and sentence structure practice. To address this issue, Li and Zhang (2011) pointed out that the new HSK has been revised to measure the test-takers’ language output capacity, thus the CSL textbooks should reflect this transfer of
linguistic knowledge to language competency development. In addition, they stated that the CSL textbooks/materials should correspond to different learners’ characteristics, such as the nationality factor. Since it was reported that half of the HSK candidates are from South Korea, they suggested that CSL textbooks and the HSK preparation materials for the South Korean market need to be further developed.

In summary, this section introduced the HSK test and washback studies in the Chinese context. In the next section, I will focus on reviewing the literature on L2 learner’s learning beliefs and strategies.

2.6 Learning Beliefs and Strategies

Learning beliefs and strategies are two very important and well-developed concepts in linguistic and educational research. In this section, I will only address those that have guided my study.

2.6.1 Language learning beliefs, attitudes and perceptions

The inherently complex nature of learner belief systems, coupled with researchers’ different agendas, has led to the proliferation of terms used to describe learner beliefs, attitude, and perceptions (henceforth referred to as beliefs) in language learning. Generally, learner beliefs have been addressed in the literature in relation to two targets: “learners” beliefs about themselves and their beliefs about learning contexts. The former have been commonly defined as how students understand and make sense of their own learning (Williams & Burden, 1999). Beliefs about the learning contexts have included how students experience and understand aspects of the classroom (e.g., teaching methods) (Brown, 2009).
With respect to beliefs about themselves, Yashima (2009) suggested that international posture is an important aspect of defining context. This relates to how students see themselves as “connected to the international community, concerned for international affairs” (p. 146). Drawing on the literature, for the current study I propose that the language learners’ beliefs be defined within these parameters: views concerning general facts, perspectives that guide the students’ learning behaviors and thoughts that determine the ultimate success of their learning in the target context.

From the 1970s to 1980s, various instruments such as the Foreign Language Attitude Scale (Bartley, 1970), the Attitude/Motivation Test Battery (Gardner, 1985), and the Foreign Language Classroom Anxiety Scale (Horwitz, 1988) were created and used to explore learners’ beliefs. Although the survey research has continued to dominate this area, high-profile qualitative and mixed method research has been published (e.g. Lyons, 2009; Yoshida, 2009). These studies have provided evidence that beliefs are dynamic, socially constructed, responsive to context and susceptible to change (Zhong, 2012). More research is needed to explore the nature of and the factors concerning the beliefs of the L2 learners towards high-stakes exams.

2.6.2 Language learning strategies and methods

Influenced by the development of cognitive psychology, research on language learning strategies started in the 1960s. There has been controversy regarding the appropriate ways of defining language learning strategies. Early on, Wenden and Robin (1987) defined learning strategies as “sets of operations, steps,
plans, routines used by the learner to facilitate the obtaining, storage, retrieval and use of information” (p.19). More recently, Dörnyei (2005), Gu (2010) and Oxford (2011) argued that the term learning strategies in L2 can be compared to the versatile concept of self-regulation in educational psychology. In these studies, strategic, self-regulated learning refers to “ways of tracking the learning task at hand and managing the self in overseeing the learning process…under the constraints of the learning situation and learning context for the purpose of learning success” (p. 2). It emphasizes self-learning strategies and methods (referred as strategies) to achieve their goals taking into account specific beliefs, behaviors and processes as well as the learning environment.

Much research has demonstrated that choosing learning strategies is affected by various factors, which include gender, age, nationality, learning style, motivation, learning context and language proficiency. For example, studies have shown significant gender differences between male and female learners. Female learners have been shown to adopt a wider range of learning strategies than males (O’Malley & Chamot, 1990; Oxford, 2011). Moreover, research has demonstrated that people with higher language proficiency use a greater number of learning strategies than those whose proficiency level is lower (Griffiths, 2003). Findings also indicated that learners with high motivation use a significantly greater number of strategies than less motivated learners (Oxford, 2011). These studies help provide an understanding of strategy use of L2 learners and highlight that strategy use is a complex phenomenon that interacts with numerous variables.
2.6.3 Relationships among beliefs, strategies, motivation and learning outcomes

Findings of language learning beliefs and strategy studies have repeatedly demonstrated that positive associations between beliefs стратегии and proficiency in the target L2 (O’Malley & Chamot, 1995; Griffiths, 2003). Park (1995) investigated 332 Korean university EFL students’ beliefs about language learning, their language learning strategies, and the relationships among their beliefs, strategy use, and L2 proficiency (based on TOEFL scores). The results of Pearson correlation indicated that self-efficacy and beliefs about social interactions were related to L2 proficiency and also to the reported use of learning strategies. It was also found that those who held positive beliefs about spoken English tended to use more communication strategies which were significantly related to L2 proficiency. A multiple regression analysis with the TOEFL scores as a criterion variable and belief and strategy variables as predictor variables showed that belief and strategy variables shared a total variance of 16% with the TOEFL scores. Based on his findings, Park concluded that there were significant correlations among beliefs, strategy use and L2 language proficiency.

Research studies also demonstrated that motivation levels significantly correlate with test performance. Sundre and Kitsantas (2004) examined the predictive power of self-regulated strategies and test-taking motivation on achievement performances under consequential and non-consequential test conditions. Sixty-two undergraduate students were asked to take two parallel classroom tests: one that counted towards their class grade (consequential) and
one that did not (non-consequential). Each test consisted of a multiple-choice section and an essay. The results showed that self-regulation strategy use significantly predicted both multiple-choice conditions.

The studies of the relationship between beliefs, strategies, motivation and learning outcomes are illuminating. They help to explain the interactions of different variables during the L2 learning process. However, studies of learning beliefs and strategies in the field of CSL are sparse. More research is needed to explore the impact that beliefs have on the processes and outcomes of language learning, and the interactions of beliefs and language learning strategies with other learner factors in relation to washback effects.

2.7 Chapter Summary

An overview of the washback research in second language education and in the Chinese context suggests that despite the increasing number of studies carried out in this area, there are remarkably few empirical studies concerning CSL tests, and particularly those that investigate and establish the washback effect of the new HSK.

The above review of the literature shows three limitations concerning the HSK and washback studies in the worldwide CSL educational context. The first limitation relates to the research methods employed by the researchers. A number of studies only rely on survey instruments such as questionnaires, rather than other quantitative, qualitative or mixed methods research methodologies. The second limitation is associated with defining or conceptualizing the term “washback”. Several studies address the issue of washback, but they have not
achieved a thorough understanding of the theoretical underpinnings of this concept, especially from the learners’ perspective concerning learning behaviors and beliefs in the learning process. Furthermore, in some cases (e.g., Yang & Liao, 2000) researchers seem to not provide sufficient data to back up their claims. Consequently, in light of the above limitations, more rigorous research is called for in the area of CSL and Chinese language tests. The present study aims to help fill some of these gaps, focusing on the impact of the HSK test on L2 learners’ learning. In the next chapter, I will present the methodology used in this study.
Chapter 3: Methodology

This section provides an overview of the methodology employed in this study. It starts by describing the research questions and the rationale for the research methodology. It then discusses the research design and gives a detailed description of the research setting, the participants and the instruments. Finally, the qualitative and quantitative data collection procedures and the data analysis process, including how the data were prepared and organized, are explained.

3.1 Research Questions

The following research questions guided the present study. The global research question is: What are the washback effects of the HSK on CSL learner factors?" There are four secondary questions:

1) Why are learners interested in learning Chinese and taking the HSK?
2) What evidence of washback effects is there on learner factors?
3) How does the HSK affect the way they learn? In other words, what factors contribute to the change of learning strategies and practice used in regular learning as opposed to HSK specific learning?
4) What are the views of CSL learners towards the HSK and its impact?

3.2 Mixed Methods Sequential Explanatory Design

Drawing on the literature review in Chapter 2 and on the basis of a prior pilot study, this study was conducted in the framework of mixed methods research (MMR) which combines two different methodological approaches, quantitative and qualitative, to provide a more holistic picture of a research problem (Creswell & Plano Clark, 2011). As a methodology, mixed methods research involves
philosophical assumptions that guide the direction of the collection and analysis of data and the mixing of qualitative and quantitative approaches in some or all of the phases in the research process. MMR focuses on collecting, analyzing, and mixing both quantitative and qualitative data in a single study or series of studies. Its central premise is that the use of both approaches in combination provides a better understanding of research problems than either approach alone.

As Greene (2007) noted, the primary purpose of a study conducted with an MMR focus is to better understand the complexity of the social phenomena being studied. The literature review section has demonstrated that there are multiple facets of change and washback that can occur in a systemic environment as well as within an individual learning context. The discussion shows that MMR is well suited to capture the complexity of the washback processes inherent in an educational context.

According to Creswell (2009), Creswell and Plano Clark (2011), Greene (2007) and Teddlie and Tashakkori (2009), by using quantitative and qualitative methods within one study, MMR can incorporate the strengths of both approaches and reduce some problems associated with single methods. In the view of these strengths, in order to understand the phenomenon of washback of a large-scale, high-stakes exam in a worldwide educational context, a mixed-methods sequential explanatory design (MMSE) was deemed an appropriate methodological approach for addressing the research questions (Creswell & Plano Clark, 2011). First of all, I reviewed HSK related documents such as the Scales and HSK exam papers and conducted a qualitative content analysis. In the second phase, I collected and
analyzed the quantitative data from a questionnaire I developed. This quantitative data makes up the primary data set. Its results were analyzed and then probed further though one-on-one interviews. These qualitative data were collected to help explain and build on the statistical results by exploring participants’ views in more depth (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009). In the end, both kinds of data were cross-examined and integrated. Figure 3.1 provides a visual diagram of the research design.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Procedure</th>
<th>Product</th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qual Data Collection</td>
<td>• Documents review</td>
<td>• Field notes</td>
<td>July - Sept, 2012</td>
</tr>
<tr>
<td></td>
<td>• Content analysis</td>
<td>• Thematic analysis</td>
<td></td>
</tr>
<tr>
<td>QUAN Data Collection</td>
<td>• N=60 CSL learners</td>
<td>• Numeric data</td>
<td>Sept - Nov, 2012</td>
</tr>
<tr>
<td></td>
<td>• Questionnaire</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Descriptive Statistics</td>
<td>• Means, SDs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Hypothesis Testing</td>
<td>• T-tests</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Multiple Regression Analysis</td>
<td>• Predict variables</td>
<td></td>
</tr>
<tr>
<td>QUAN Data Analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qual Data Collection</td>
<td>• N=8</td>
<td>• Transcripts</td>
<td>Dec, 2012 - Jan, 2013</td>
</tr>
<tr>
<td></td>
<td>• Individual in-depth interview</td>
<td>• Field notes</td>
<td></td>
</tr>
<tr>
<td>Qual Data Analysis</td>
<td>• Coding</td>
<td>• Coded text</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Thematic development</td>
<td>• Themes</td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>• Interpretation and explanation of the</td>
<td>• Discussion</td>
<td>Feb - Mar, 2013</td>
</tr>
<tr>
<td></td>
<td>quan and qual results</td>
<td>• Implication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Future research</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.1: Visual diagram of the MMSE research design of the study
As shown in the figure above, either qualitative (qual) or quantitative (quan) data or both forms of data (quan + qual) were collected and analyzed. With respect to the overall process of data collection, the phases proceed in a specific sequence over time. The research design is further explained below.

3.3 Participants

When selecting participants, Patton’s (2002) maximum variation sampling technique was used. Based on Patton (2002) and Maxwell (1996), purposive sampling involves selecting particular settings and individuals/groups deliberately not only based on the research questions, but also based on information available about these individuals/groups. In order to get a representative sample, the participants were chosen according to three parameters: 1) the type of CSL program they have taken; 2) their Chinese language ability; and 3) geographical factors. As Strauss and Corbin (1998) stated, a researcher’s experience and knowledge are what sensitizes the researcher to significant problems and issues in the data and allows him/her to seek explanations for alternatives and to theorize emergent concepts. My previous experience as a CSL instructor in China and North America allowed me to have an insider’s view which helped in my initial sample selection.

All participants volunteered and met the specific criteria (the 3 parameters set out above). The population of the focus group (used in the pilot study and discussed in “Procedure” below) included 4 CSL learners, 1 HSK test developer and 2 experienced CSL teachers. The population in the quantitative data collection included 60 CSL learners who had previously taken the HSK test in
China or outside of China. Fifty-six participants completed the new version of the HSK in the year 2012, while the other four students took the old version during the year 2009-2010. The 60 participants (Female = 27, Male = 33) came from the university language programs, weekend language schools, HSK preparation classes and Confucius Institute programs and had varying language proficiencies. They were from 15 countries and areas, including Canada, Belgium, Austria, Vietnam, Mexico, Korea, France, Italy, USA, Pakistan, Malaysia, Australia, Spain, Japan, Indonesia, and Greece. Considering their language background and cultural contexts, they were divided into 2 groups (Western=26, Asian=34). The variety of school settings and proficiency levels were of considerable importance, because tests might have different amounts and types of washback on some learners than on other learners.

In the last section of the questionnaire, students were asked if they were willing to be interviewed. If they were, they were asked to fill out the information section, including their contact details. There were 21 interview participants who indicated their willingness to be interviewed. Of these, I chose 8 interviewees who were representative of the differences in gender (4 females and 4 males), in age group (3 from age group of 18 to 24, 3 from age group of 24 to 34, 2 from age group of 35+), and their self-reported Chinese competence (2 of low level, 3 of intermediate level and 3 of high level).

3.4 Instruments

As indicated above, the literature on washback research contributed to my decision on what types of instruments to use for this research. There are three
types of instruments which were used: HSK related documents analysis, structured questionnaires, and one-on-one interviews.

3.4.1 HSK-related documents and reports

The HSK-related documents and reports issued by Hanban were taken as official sources reflecting the test constructors’ intentions. The purpose of the review and qualitative content analysis of documents was to identify the characteristics of the HSK and to find out what the HSK claims to measures (e.g., linguistic knowledge or language use ability), what is the test-taker component, and whether or not the HSK represents the Chinese Language Proficiency Scales for Speakers of Other Languages.


3.4.2 Structured questionnaire

I developed an online structured questionnaire based on the learning strategy and learning beliefs theories, washback theories and washback models, which were mentioned in Chapter Two. The questionnaire was piloted with my focus group. Online questionnaires offered several advantages: 1) they are the
most economical tool if the researcher has limited time and resources; 2) they are especially suitable for large samples to investigate and compare general trends; 3) they are less threatening for respondents than observations can be; and 4) they allow access to participants across different time zones and locations.

The questionnaire was structured in three parts (See Appendices A & B). Part 1 asks for background information. Part 2 concerns participants’ learning strategies used for regular learning and test-specific learning. As I mentioned in the previous chapter, “regular Chinese learning” includes CSL courses, distance learning, or self-learning on a daily basis; “test-specific learning” reflects participants’ learning largely on the basis of the test’s impact and includes all the learning beliefs, strategies and processes that affect success on the HSK, such as learning test-taking strategies and practicing simulated exam papers. Part 3 is related to participants’ beliefs of the HSK (e.g., expectations of the HSK, whether the HSK was a valid test, etc.). Both Part 2 and 3 use Likert-scale items. The use of Likert-type scale questions to measure opinions, beliefs, and attitudes is well documented in the behavioral sciences and in L2 research specifically (Busch, 1993). In order to test the reliability of the questionnaire, I tested the Cronbach’s Alpha of the quantitative data. The result showed that Cronbach’s $\alpha = .90$, which indicated the questionnaire has high reliability.

3.4.3 One-on-one interview

In-depth individual interviews were undertaken with the participants’ consent after the questionnaires were completed. The purpose of the interviews was to qualify the quantitative data; that is, to help contribute to a more
comprehensive and nuanced understanding of the examined relationships (i.e.,
regular learning of a language or learning in order to pass a specific proficiency
test, test-specific learning). Qualitative interviewing was chosen because it is
presumed to be an appropriate and effective approach to inquire about specific
social processes or particular individuals’ perspectives through direct contact with
those involved in natural contexts (Locke, Spirduso, & Silverman, 2000). In
addition, the qualitative interviews served as a complementary tool to verify
findings from the survey questionnaire (Phakiti, 2003). The design of the
interview protocol was informed by findings from the students’ questionnaire and
HSK related document review. Interviews were conducted in Chinese or English
depending on the participants’ preference (See Appendices C & D).

3.5 Procedures

3.5.1 Ethical issues

In compliance with the McGill Research Ethics Board (REB) rules, I
obtained access to the research site and consent from the participants before
conducting this study. Since there were no sensitive or personal questions in the
questionnaire or interviews, no ethical concerns were identified in completing
either the survey questionnaires or the interviews. No psychological, emotional,
economic, cultural, and/or social risk was foreseen for the participants.

The online questionnaire has a lead-in page, which consists of a consent
form where participants must click on an agreement button to participate in the
study before proceeding to the actual questionnaire. Considering that the CSL
learners who participated in the current research were from different countries and
with different language proficiency levels in Chinese, I prepared the Letter of Consent (see Appendix E and F) and the online questionnaire (see Appendix A and B) in four language versions: English, Chinese, Japanese and Korean. All of the questionnaires were translated by one or two native speakers of the target language with language assessment knowledge. It is important in this process that the translation produces a questionnaire which is not only comparable in terms of language (linguistic equivalence), but, in particular, is conceptually comparable (conceptual equivalence). In addition, translating the questionnaire ensured that the participants properly understood the nature and purpose of the study and this made the results more reliable. Interview participants were given a second Letter of Consent specifically designed for the interviews in English and Chinese versions (see Appendix C and D). Furthermore, student participation was voluntary. The participating students were also told that they could skip any question and were free to withdraw even after starting to fill out the questionnaire or starting with the interview.

The participants of the survey and the interview were assured that their responses to the questionnaire or the interview would not be released to anyone, including the university authorities, without their consent and that their responses would be used only for the stated research purposes. They were assured that the issue of confidentiality was taken seriously by McGill REB rules.

3.5.2 Data collection

First of all, the initial version of the questionnaire was piloted on a focus group, which consisted of 4 CSL learners in Canada and Korea, 1 HSK test
developer and 1 experienced CSL teacher, during the summer of 2012. The results indicated that the first draft of the questionnaire combined too many formats such as closed and open-ended items, questions or statements in Likert-scale, etc. It was time-consuming and hard to analyze. Therefore, based on the feedback from the pilot survey, the questionnaire was refined, reframed and revised. The revised questionnaire includes only 2 types of statements, which are multiple choice and 4-point Likert-scale items, and there are 73 questions in the questionnaire.

After piloting the questionnaire, I conducted an intensive review and content analysis of the documents pertaining to the HSK test papers, the CSL teaching materials, Language Proficiency Scales for Speakers of Other Languages, HSK research reports and official HSK guidelines. Content analysis of qualitative data is a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Stembler, 2001).

The following stage involved the administration of the questionnaire. First of all, I posted my recruitment flyer in the admission offices of the university CSL programs (McGill University, University of Waterloo and Nanjing University), Confucius Institutes (Vancouver, Montreal, Waterloo and Atlanta), Chinese language weekend schools (Suzhou and Melbourne) and HSK test centers (Dalian, Montreal, Toronto, and Vancouver). I also posted the recruitment flyer on

http://bbs.chinese.cn/forum and HSK group on www.facebook.com. The CSL learners who were interested in participating in my research contacted me individually to have access to the online survey. In addition, I sent emails to my previous students who had taken the HSK and asked them if they were interested
in participating. By December 1\textsuperscript{st}, 2012, I had received 65 questionnaires, of which 60 were complete with clear data.

The last part of data collection was the in-depth one-on-one interview. This round of qualitative data collection was conducted from December 2012 to January 2013. For the interviewees in Montreal and Toronto, I conducted face-to-face interviews. The participants signed the consent form and gave permission to audiotape the conversation. For the interviewees overseas, I conducted the interview via Skype. They signed the consent form and gave permission to audiotape the conversation online after the questionnaire. The interviews lasted between 40 and 60 minutes. The interviewees were given the choice of conducting the interview in either English or Chinese or a combination of both. Out of the 8 participants interviewed, two of the interviewees chose to conduct the whole interview in Chinese, 2 of them chose English, while 4 of them started with Chinese and switched to English as the interviews went on. At the end of this stage, all data resources were cross-examined to finally develop theories to explain the findings.

The summary of data collection is presented in Table 3, and the survey and interview instruments are attached in the Appendices B, C, D and E.
Table 3.1: Summary of data collection

<table>
<thead>
<tr>
<th>Source of data</th>
<th>Participants</th>
<th>Instruments</th>
<th>Data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSK Documents Review</td>
<td>Researcher</td>
<td>Official HSK research reports</td>
<td>Qualitative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HSK exam paper</td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td>60 CSL learners</td>
<td>Structured-questionnaire</td>
<td>Quantitative</td>
</tr>
<tr>
<td>One-on-one Interview</td>
<td>8 CSL learners</td>
<td>Interview guide</td>
<td>Qualitative</td>
</tr>
</tbody>
</table>

3.5.3 Data Analysis

As mentioned above, a mixed methods sequential explanatory design combining both qualitative and quantitative methods of data analysis was adopted for this study. Quantitative analyses of survey data, which provided the main data set, involved descriptive statistical analyses and inferential analyses. These were applied to analyze interrelationships among numbers of variables and to explain these variables in terms of their common underlying dimensions (Kline, 2005; Purpura, 1998, 1999). In addition, a close examination and intensive content analysis of the pertinent documents (sample test of the HSK, textbooks, etc.) were conducted. Based on Glaser & Laudel’s (2004) content analysis process circle, Figure 3.2 shows the proceeding of qualitative content analysis from the theory to the final analysis and interpretation. Furthermore, qualitative analyses of the data of the individual interviews were conducted to provide deeper insights into the quantitative findings. Analyses involved the constant comparative method (Bogdan & Biklen, 1998) in which the data is classified into categories. Finally,
the different types of data sources were synthesized. The qualitative analyses findings helped to explain and expand the quantitative findings.

Figure 3.2: Proceeding of qualitative content analysis of the pertinent documents

3.6.1 Quantitative data analysis process

3.6.1.1 Descriptive statistics

When dealing with the questionnaire data involving various components of the learning beliefs/attitudes towards the HSK content, impact and score use, I first relied on frequency counts to identify percentages of learners’ responses by category, and I also examined the mean and standard deviation (SD) of each question. In addition, I used Confidence Interval (CI) to determine whether the mean of each question contained the true population mean.

3.6.1.2 Inferential statistical analysis

A very common question is whether it is legitimate to use Likert scale data in parametric statistical procedures that require continuous/interval data, such as
linear t-test, Regression, ANOVA, and Factor Analysis. One camp maintains that as ordered categories, the intervals between the scale values are not equal. Any mean, correlation, or other numerical operation applied to them is invalid. Only non-parametric statistics should be used on Likert scale data (Jamieson, 2004) or chi-square could be used to analysis categorical data. The other group maintains that while technically the Likert scale item is ordered, using it in parametric tests is valid in some situations. For example, Lubke and Muthen (2004) found that it is possible to find true parametric values in factor analysis with Likert scale data, if assumptions about skewness and number of categories are met. Likewise, Glass and Sanders (1972) found that F tests in ANOVA could return accurate p-values on Likert items under certain conditions. With these considerations, in this study, I first examined the questionnaire items for skewness (a measure of distribution symmetry) and kurtosis (a measure of whether the distribution is peaked or flat relative to a normal distribution). I found that most of the items were normally distributed with skewness and kurtosis values close to zero. For the non-normal distributed cases, no extreme skew distribution was found. The assumption of the chi-square requires the data are independent; however, my data from regular learning and test specific learning are from the same group of participants, that is, they are dependent data. Hence, I could not conduct a chi-square analysis. Based on the above reasons, I could treat that the 4 grade Likert scale data from the questionnaire as continuous data, and statistical analyses for continuous data were conducted.

The second part of the questionnaire consists of five parts, the four L2
skills (speaking, listening, reading and writing) and classroom activities. In order
to answer the research question “what are the washback effects of the HSK on
learning,” I asked the three sub-item questions twice about learning practices
(strategies and methods) on regular learning and test-specific learning in the
questionnaire. These three sub-item questions included a two-factor model of
cognitive strategy use, involving comprehending, storing/memory, and
using/retrieval processes, and a one-factor model of metacognitive strategy. Since
the results of regular learning versus test-specific learning are from the same
group of people, I performed paired-sample t tests to investigate if there is a
difference in the participants’ learning strategies and methods between the two
kinds of learning.

My hypothesis states there is no difference between the learners’ strategy
and method use on regular learning and test-specific learning, thus, \(H_0: \mu_1 = \mu_2\).
The alternative hypothesis states there is a statistically significant difference
between them, thus, \(H_1: \mu_1 \neq \mu_2\). That is to say, washback is present. Series
comparisons were followed to find out where the differences are. Moreover,
multiple regression analysis was conducted to explore which factors (gender, age,
nationality, language proficiency, the course type, motivation, etc.) contribute to
the washback effects. The computer program Statistical Package of the Social
Sciences (SPSS, 20.0) for Windows was used to perform descriptive statistics and
inferential statistics.

3.6.2 Qualitative data analysis

When reviewing the HSK related documents, the content was reduced and
synthesized using focused summaries pertaining to the HSK construction and implementation. Several steps were followed in analyzing the interview data. First, I read over each transcript to gain a general sense of the information they provided. Next, I used open coding (Strauss & Corbin, 1998) to help build ideas inductively and remain more attentive to what the interviewees had to say, rather than pre-imposing my ideas or the findings from the statistical analyses on this part of the data analysis. Open coding is an analytic process through which concepts are identified and their properties and dimensions are discovered in data. Third, I proceeded with axial coding (i.e., a process of relating categories to their subcategories, whereby coding occurs around the axis of a category) by focusing mainly on selecting, categorizing, synthesizing, and interpreting the interview data to provide explanations of the research questions addressed (McMillan & Schumacher, 2001). I assigned units of data to four categories: background, motivation, test content, and test impact, and analyzed the data logically. Each category was further broken down into several subcategories followed by interviewees’ quotes to represent those subcategories. In the next step, I related the data to the 4 categories to build theoretical explanations for the research questions. Finally, I reported the outcomes from an individual perspective or holistic perspective depending on the context. A description of the procedures and methods of data analysis is shown in Table 3.2.
Table 3.2: Summary of the procedures and methods of data analysis

<table>
<thead>
<tr>
<th>Analysis of Documents</th>
<th>Questionnaire</th>
<th>Interview</th>
<th>Integration of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Goals</td>
<td>• SPSS</td>
<td></td>
<td>• Questionnaires</td>
</tr>
<tr>
<td>• Contents</td>
<td>• Descriptive stats</td>
<td>• Organizing data</td>
<td></td>
</tr>
<tr>
<td>• Skills</td>
<td>(frequency counts, SD)</td>
<td>• Categorization</td>
<td>• Interview</td>
</tr>
<tr>
<td></td>
<td>• Inferential stats</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(t-tests, multiple regression)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.7 Chapter Summary

The MMR methodology used in this study combined both quantitative and qualitative approaches. The documents analysis related to the HSK exam involved official HSK research reports, CSL teaching materials, and HSK exam papers. The survey participants were 60 CSL learners from 15 countries who had taken the HSK test. Eight learners volunteered to provide their opinions in the structured interview. Having described the research methods and procedures, I will turn to present the major research findings emerging from the data in the following chapter.
Chapter 4: Presentation of the Results

This chapter presents the qualitative and quantitative findings of the study. It begins by examining the characteristics of the HSK from HSK related documents. Then the quantitative results obtained from questionnaires are reported and summarized. After that, the qualitative findings derived mainly from interviews are reported. The findings from the interview are presented and summarized by categories and themes.

4.1 Analysis of the HSK related documents

In order to identify the characteristics of the HSK, I conducted an intensive content analysis of HSK related documents, which include *Scales* (2009), *Report on Overseas Enforcement of New Chinese Proficiency Test (New HSK)* (2011), *Report of Researching and Producing the New HSK* (2010), and *Chinese Language Proficiency Test Level 1 to 6* (2010). In this section, I address the validity and reliability of the HSK, the implementation of the HSK outside of China, as well as the relationship of the HSK and CSL teaching official guidelines.

4.1.1 The reliability and validity of the HSK

Reliability and validity are generally considered as the most important features when designing a test or evaluating its potential usefulness. Validity refers to the extent to which we can interpret a given test score as an indicator of a test takers’ language ability, while reliability concerns the consistency of measurement. In the *Report on Overseas Enforcement of New Chinese Proficiency Test (New HSK)* (2011) (abbreviated as *Report*), the degree of difficulty, differentiation, validity and reliability of 36 HSK exam papers were
analyzed by the Assessment Systems Corporation Item (ASCI) and Test Analysis Package (TAP). The *Report* indicated that most levels of the HSK tests have relatively ideal validity and reliability, except for Level 1. It revealed that the new HSK has reached the anticipated target basically through proof, test and operation. However, all these data were only based on the analysis of the objective task, and did not include the subjective tasks, such as writing task in the Written HSK and Oral HSK. In order to get a holistic, comprehensive and objective understanding of the HSK tests, data from the subjective tasks need to be taken into account.

### 4.1.2 Test-takers of the HSK

A growing number of people have been interested in learning Chinese and have taken the HSK test. Taking the year 2010 as an example, there were 76,088 participants from 174 test centers in 65 countries and areas taking the tests. Table 4.1 below shows that the number of test-takers who took the HSK Written Exams is 16 times greater than those who took the Oral Exam.

<table>
<thead>
<tr>
<th>Category</th>
<th>Level</th>
<th>N</th>
<th>%</th>
<th>Category</th>
<th>Level</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written</td>
<td>1</td>
<td>4817</td>
<td>5.18%</td>
<td>Oral</td>
<td>Ele.</td>
<td>840</td>
<td>16.1%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>8120</td>
<td>8.72%</td>
<td></td>
<td>Int.</td>
<td>1789</td>
<td>34.36%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>10859</td>
<td>11.67%</td>
<td></td>
<td>Adv.</td>
<td>2577</td>
<td>49.50%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>18458</td>
<td>19.83%</td>
<td></td>
<td>Total</td>
<td>5206</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>31473</td>
<td>33.81%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>19358</td>
<td>20.80%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>93085</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In 2010, among all the test-takers, 60% were female and 40% were male. Most of them (89.5%) are in the 16 to 35 age group. From the test-takers distribution around the world (see Table 4.2 below), Asia is the best represented continent and Koreans contributed 54.37% to all the participants. The HSK clearly has a great influence on Asian countries.

<table>
<thead>
<tr>
<th>Category</th>
<th>Level</th>
<th>Asia N</th>
<th>%</th>
<th>Europe N</th>
<th>%</th>
<th>America N</th>
<th>%</th>
<th>Oceania N</th>
<th>%</th>
<th>Africa N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written HSK</td>
<td>1</td>
<td>2424</td>
<td>2.96%</td>
<td>1567</td>
<td>21.49%</td>
<td>509</td>
<td>20.27%</td>
<td>125</td>
<td>17.03%</td>
<td>192</td>
<td>30.70%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>5509</td>
<td>6.72%</td>
<td>1771</td>
<td>24.29%</td>
<td>520</td>
<td>20.71%</td>
<td>178</td>
<td>24.25%</td>
<td>142</td>
<td>22.72%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8178</td>
<td>9.98%</td>
<td>1808</td>
<td>24.80%</td>
<td>552</td>
<td>21.98%</td>
<td>139</td>
<td>18.94%</td>
<td>182</td>
<td>29.12%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>16510</td>
<td>20.15%</td>
<td>1239</td>
<td>16.99%</td>
<td>458</td>
<td>15.24%</td>
<td>175</td>
<td>23.84%</td>
<td>76</td>
<td>12.16%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>30323</td>
<td>37.01%</td>
<td>672</td>
<td>9.22%</td>
<td>375</td>
<td>14.93%</td>
<td>75</td>
<td>10.22%</td>
<td>28</td>
<td>4.48%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>18980</td>
<td>23.17%</td>
<td>234</td>
<td>3.21%</td>
<td>97</td>
<td>3.86%</td>
<td>42</td>
<td>5.72%</td>
<td>5</td>
<td>0.80%</td>
</tr>
<tr>
<td>Oral HSK</td>
<td>Ele.</td>
<td>599</td>
<td>13.08%</td>
<td>166</td>
<td>41.50%</td>
<td>75</td>
<td>37.50%</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Int.</td>
<td>1530</td>
<td>33.41%</td>
<td>173</td>
<td>43.25%</td>
<td>73</td>
<td>36.50%</td>
<td>13</td>
<td>48.15%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Adv.</td>
<td>2450</td>
<td>53.51%</td>
<td>61</td>
<td>15.25%</td>
<td>52</td>
<td>26.00%</td>
<td>14</td>
<td>51.85%</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4.1.3 Testing score

According to *Chinese Language Proficiency Test Level 1 to 6* (2010), Levels 1 and 2 of the HSK consist of listening comprehension and reading comprehension sections. There are a total of 200 points possible, with each section accounting for 100 points. A passing score is a score of 120 or above. Levels 3, 4, 5 and 6 consist of listening comprehension, reading comprehension
and writing sections. There are a total of 300 points possible, with each section accounting for 100 points. A passing score is 180 or above. The HSK Speaking Test (Beginner, Intermediate and Advanced Level) has 100 points possible. A passing score is 60 or above.

The database of HSK participation in 2010 revealed that the passing rates of HSK Level 1, 2 and 3 reached 85%, and the passing rate of the advanced levels (5 and 6) exceeded 60%. The Report (2011) also confirmed that the degree of difficulty of the HSK (all six levels) seems to be low. In contrast, the passing rate of the speaking tests is much lower than the written tests: beginner level (67.74%), intermediate level (51.76%) and advanced level (47.07%).

4.1.4 HSK and Scales

The *Chinese Language Proficiency Scales For Speakers of Other Languages* (Office of Chinese Language Council International, 2009), which is a guideline document for CSL, has drawn on the results of research on international language proficiency scales such as the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) and the Canadian Language Benchmarks (CLB) (*Chinese Language Proficiency Test Levels 1 to 6, 2010*). It provides a five-band all-round description of their ability to use their knowledge and skills of the Chinese language for communication. The Scales serve as a reference standard for drawing up a syllabus of teaching Chinese for speakers of other languages, for compiling Chinese textbooks, and for assessing the language proficiency of CSL learners. From the general description
of Chinese Language Proficiency (see Figure 4.1), it is clear that communicative ability is the core focus.

Figure 4.1: Description of Chinese Language Proficiency

Table 4.3 below shows the corresponding relationship between each level of the New HSK tests and the Scales.

Table 4.3: Relationship among the New HSK Tests and Scales

<table>
<thead>
<tr>
<th>New HSK</th>
<th>Vocabulary</th>
<th>Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSK Level 6</td>
<td>5000</td>
<td>Band 5</td>
</tr>
<tr>
<td>HSK Level 5</td>
<td>2500</td>
<td></td>
</tr>
<tr>
<td>HSK Level 4</td>
<td>1200</td>
<td>Band 4</td>
</tr>
<tr>
<td>HSK Level 3</td>
<td>600</td>
<td>Band 3</td>
</tr>
<tr>
<td>HSK Level 2</td>
<td>300</td>
<td>Band 2</td>
</tr>
<tr>
<td>HSK Level 1</td>
<td>150</td>
<td>Band 1</td>
</tr>
</tbody>
</table>

4.2 Questionnaire Data

To help me comprehend the mechanisms of washback, I first used a questionnaire to investigate the purpose of learning Chinese, the reason for taking the HSK and if the students set passing the HSK as their goal of learning Chinese (See Appendices A & B).

4.2.1 Why does the learner take the HSK test?

In order to understand the students’ motivation for taking the HSK, student respondents were first asked to indicate the main reason for learning Chinese.

The results showed that one third of the learners would like to study, work, travel in China; another one-third of the learners are interested in learning a new language and culture. Twenty percent of the participants were learning Chinese because it is required by their academic programs and/or professional projects. Less than 8% of the participants learned Chinese due to encouragement of their friends or relatives.

The results indicating the main reason for the learners taking the HSK are displayed in Table 4.4. Thirty-eight percent of the participants take the test to identify their level of language proficiency. The majority of participants took the HSK because they wanted to pursue higher education (32%) or increase job-hunting/career opportunities in China.
Table 4.4: Summary of Descriptive Statistics of Responses to the reason for taking the HSK

<table>
<thead>
<tr>
<th>Reason</th>
<th>Frequency</th>
<th>Valid</th>
<th>Cumulative Percent</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>To have an indicator of your language proficiency</td>
<td>23</td>
<td>38.3</td>
<td>38.3</td>
<td>38.3</td>
</tr>
<tr>
<td>To be able to pursue higher education in China</td>
<td>19</td>
<td>31.7</td>
<td>70.0</td>
<td></td>
</tr>
<tr>
<td>To help job hunting</td>
<td>14</td>
<td>23.3</td>
<td>93.3</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>4</td>
<td>6.7</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

Four questions on the questionnaire (Q47, Q48, Q51 and Q59) are about motivation for learning Chinese and taking the HSK. The results displayed in Table 4.5 indicate that when they began to learn Chinese, only 15% of the learners were not familiar with the HSK test (Q47) and 40% of them had a clear goal – passing the HSK (Q48). Most learners (Mean=1.92, SD = .79) believed that taking the HSK benefited their motivation for learning Chinese (e.g., the test motivated them and helped them understand their own learning needs) (Q51). After a close look at the data, I found the people (N=4) who failed the HSK had much weaker motivation than those who passed (Q59).
Table 4.5: Summary of Descriptive Statistics of Responses to Q47, Q48, Q51 and Q59

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q47</td>
<td>55.0%</td>
<td>30.0%</td>
<td>13.3%</td>
<td>1.7%</td>
<td>1.62</td>
<td>.78</td>
</tr>
<tr>
<td>Q48</td>
<td>13.3%</td>
<td>26.7%</td>
<td>16.7%</td>
<td>43.3%</td>
<td>2.90</td>
<td>1.12</td>
</tr>
<tr>
<td>Q51</td>
<td>30.0%</td>
<td>53.3%</td>
<td>11.7%</td>
<td>5%</td>
<td>1.92</td>
<td>.79</td>
</tr>
<tr>
<td>Q59</td>
<td>23.3%</td>
<td>61.7%</td>
<td>11.7%</td>
<td>3.3%</td>
<td>1.95</td>
<td>.70</td>
</tr>
</tbody>
</table>

Notes:
Q47: When you began to learn Chinese, you did not know the HSK.
Q48: When you began to learn Chinese, you set passing HSK as goal.
Q51: Taking the HSK benefits your motivation of learning Chinese.
Q59: It motivates students to enhance their proficiency in Chinese.

In order to investigate the group difference between test-takers in the West and those in Asia on passing HSK as their goal when they begin to learn Chinese, a case two t-test for independent samples was conducted (see Table 4.5).

According to the Levene’s test, equal variance was assumed: $F (1, 58) = 3.25$, $p > .05$. Since it is not significant, the assumption has been met. The results of the t-test suggest that there is a significant difference between groups, $t (58) = -4.10$, $p < .001$. The population means of the two groups are different. The means of Western and Asian groups showed that more learners from the Asian group ($M =$
2.44, $SD = 1.08$) set passing the HSK as their goal for learning Chinese than those in the Western group ($M = 3.50, SD = .86$).

Table 4.6: Summary of the t Test Analyses of Asian and Western Groups

<table>
<thead>
<tr>
<th></th>
<th>Asian</th>
<th></th>
<th>Western</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Q 48</td>
<td>2.44</td>
<td>1.08</td>
<td>3.50</td>
<td>.86</td>
</tr>
<tr>
<td></td>
<td>-4.10</td>
<td>.000*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$p < .001$ *

4.2.2 The washback effects of the HSK on learner factors.

The second part of the questionnaire was used to measure the learners’ learning strategies applied in regular learning and HSK test-specific learning in the four skills (speaking, listening, reading and writing). Under each skill, there are three sub-items questions/statements. All the questions/statements used a four-point Likert-scale format (1 = always, 2 = often, 3 = rarely and 4 = never). The differences between the learners’ learning practices (strategies and methods) used in the two learning types in the four learning skills were tested with a paired-sample $t$-test. Results are shown in Table 4.7. An examination of the data results showed statistically-significant differences between “regular learning” and “HSK test-specific learning” in speaking ($t (59) = -2.38, p = .02$), listening ($t (59) = -3.28, p = .002$), reading ($t (59) = -4.27, p < .001$), writing ($t (59) = -2.95, p = .005$) and in total ($t (59) = -3.32, p = .002$). The washback effect in classroom activities is not significant. Therefore, overall, it can be said that there are washback effects of the HSK on learning.
Table 4.7: Summary Results for Paired t Tests of Washback Effects on Learners Learning

<table>
<thead>
<tr>
<th>Variable</th>
<th>95%CI</th>
<th>Mean</th>
<th>Mean</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Test-specific</td>
<td>Regular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speaking</td>
<td>-.38</td>
<td>-.03</td>
<td>2.29</td>
<td>2.48</td>
<td>59</td>
<td>-2.38</td>
<td>.020*</td>
</tr>
<tr>
<td>Listening</td>
<td>-.44</td>
<td>-.11</td>
<td>2.31</td>
<td>2.58</td>
<td>59</td>
<td>-3.28</td>
<td>.002*</td>
</tr>
<tr>
<td>Reading</td>
<td>-.55</td>
<td>-.19</td>
<td>2.07</td>
<td>2.44</td>
<td>59</td>
<td>-4.27</td>
<td>.000**</td>
</tr>
<tr>
<td>Writing</td>
<td>-.31</td>
<td>-.01</td>
<td>2.61</td>
<td>2.78</td>
<td>59</td>
<td>-2.95</td>
<td>.005*</td>
</tr>
<tr>
<td>Total</td>
<td>-.31</td>
<td>-.8</td>
<td>2.32</td>
<td>2.52</td>
<td>59</td>
<td>-3.32</td>
<td>.002*</td>
</tr>
<tr>
<td>Classroom</td>
<td>-.17</td>
<td>-.01</td>
<td>.53</td>
<td>.54</td>
<td>59</td>
<td>-1.88</td>
<td>.07</td>
</tr>
</tbody>
</table>

p < .05 *, p < .001 **

As mentioned above, no significant difference between regular learning and test-specific learning in speaking practices was found. However, in the three sub-items of the speaking practices, there was a significant effect with a large effect size on spending time on speaking exercises in the HSK example papers or simulated exam, \( t(59) = -3.56, p = .001 \) (see Table 11). The learners paid less attention to the test preparation materials in their regular learning \( (M = 3.12, SD = .88) \) than in test-specific learning \( (M = 2.65, SD = 1.07) \) (Q13/ Q16), but they kept the same learning strategies of the other two cognitive sub-items (Q11/ Q14 and Q12/ Q15). This is shown in Table 4.8 below.
Table 4.8: Summary Results of t Tests for Speaking Strategies

<table>
<thead>
<tr>
<th>Variable</th>
<th>95%CI</th>
<th>Mean</th>
<th>Mean</th>
<th>Lower</th>
<th>Upper</th>
<th>Test- specific</th>
<th>Regular</th>
<th>df</th>
<th>t</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q 11 vs. Q 14</td>
<td>-.24</td>
<td>.14</td>
<td>2.12</td>
<td>2.12</td>
<td>59</td>
<td>-.52</td>
<td>.61</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q 12 vs. Q 15</td>
<td>-.21</td>
<td>.14</td>
<td>2.12</td>
<td>2.15</td>
<td>59</td>
<td>-.38</td>
<td>.71</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q 13 vs. Q 16</td>
<td>-.73</td>
<td>-.20</td>
<td>2.65</td>
<td>3.12</td>
<td>59</td>
<td>-3.56</td>
<td>.001*</td>
<td>.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: \( p \leq .001 * \)
Q11/ Q14: Repeating what you have learned/heard.
Q12/ Q15: Spending time with Chinese speaking people.
Q13/ Q16: Spending time on speaking exercises in the HSK example papers or simulated exam papers.

Table 4.9 shows that all three sub-items of the listening strategies used in the two learning types had significant differences (Q17/20, \( p = .02 \); Q18/21, \( p = .01 \); Q19/22, \( p < .001 \)). In other words, the HSK test has a significant influence on test-takers in learning listening. Not only did the strategies concentrating on test-specific preparation change, but also the regular learning behaviors (e.g., watching Chinese TV, listening to Chinese radio or listening to textbook CDs) changed due to the test.
### Table 4.9: Summary Results of $t$ Tests for Listening Strategies

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lower</th>
<th>Upper</th>
<th>Test-specific Mean</th>
<th>Regular Mean</th>
<th>df</th>
<th>$t$</th>
<th>$p$</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q17 vs.</td>
<td>0.03</td>
<td>0.40</td>
<td>2.48</td>
<td>2.27</td>
<td>59</td>
<td>2.35</td>
<td>0.02*</td>
<td>0.23</td>
</tr>
<tr>
<td>Q20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q18 vs.</td>
<td>-0.63</td>
<td>-0.07</td>
<td>2.27</td>
<td>2.61</td>
<td>59</td>
<td>-2.53</td>
<td>0.01*</td>
<td>0.32</td>
</tr>
<tr>
<td>Q21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q19 vs.</td>
<td>-0.97</td>
<td>-0.40</td>
<td>2.18</td>
<td>2.87</td>
<td>59</td>
<td>-4.76</td>
<td>0.000**</td>
<td>0.68</td>
</tr>
<tr>
<td>Q22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $p < .05*; p < 0.001**

Q17/ Q20: Watching TV programs, movies or listening to radio in Chinese.
Q18/Q21: Listening to textbooks CDs.
Q19/Q22: Spending time on listening exercises in the HSK example papers or simulated exam papers.

In learning reading (see Table 4.10), the learners reported paying less attention to learning new vocabulary, grammar rules and spending time on preparation of the test related materials during regular learning ($p = .003$, $p < .001$) than HSK test-specific learning. However, there were no significant differences between regular and test-specific learning strategies in the case of reading Chinese books and magazines.
Table 4.10: Summary Results of t Tests for Reading Strategies

<table>
<thead>
<tr>
<th>Variable</th>
<th>95%CI</th>
<th>Mean</th>
<th>Mean</th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower</td>
<td>Upper</td>
<td>Test-specific</td>
<td>Regular</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q23 vs. Q26</td>
<td>-0.07</td>
<td>0.37</td>
<td>2.45</td>
<td>2.30</td>
<td>59</td>
<td>1.35</td>
<td>1.82</td>
</tr>
<tr>
<td>Q24 vs. Q27</td>
<td>-0.49</td>
<td>-0.10</td>
<td>1.75</td>
<td>2.03</td>
<td>59</td>
<td>-3.07</td>
<td>.003*</td>
</tr>
<tr>
<td>Q25 vs. Q28</td>
<td>-1.30</td>
<td>-0.67</td>
<td>2.00</td>
<td>2.98</td>
<td>59</td>
<td>-6.27</td>
<td>.000**</td>
</tr>
</tbody>
</table>

Note: $p < .05*$; $p < 0.001**$
Q23/ Q26: Reading Chinese books and magazines.
Q24/ Q27: Learning new vocabulary and grammar rules.
Q25/ Q28: Spending time on reading exercises in the HSK example papers or simulated exam papers.

Similar to the speaking practices, there is also a significant effect with a large effect size on spending time on writing exercises in the HSK example papers or simulated exam ($p = .001$) (see in Table 4.11 below). The learners paid less attention to the test preparation materials in regular learning ($M = 2.38$, $SD = 1.15$) than in test-specific learning ($M = 3.07$, $SD = .92$), but they kept the same learning strategies and methods of the other two sub-items on both learning types (keeping a diary or writing essays in Chinese and reading writing samples, analyzing the structure and expression and using familiar words and structures to make new sentences).
Table 4.11: Summary Results of t Tests for Writing Strategies

<table>
<thead>
<tr>
<th>Variable</th>
<th>95%CI</th>
<th>Mean</th>
<th>Mean</th>
<th>Lower</th>
<th>Upper</th>
<th>Test-specific Mean</th>
<th>Regular Mean</th>
<th>N</th>
<th>t</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q29 vs. Q32</td>
<td>- .28</td>
<td>.11</td>
<td>2.82</td>
<td>2.90</td>
<td>60</td>
<td>-.843</td>
<td>.403</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q30 vs. Q33</td>
<td>- .10</td>
<td>-.22</td>
<td>2.62</td>
<td>2.57</td>
<td>60</td>
<td>.731</td>
<td>.468</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q31 vs. Q34</td>
<td>- .99</td>
<td>-.38</td>
<td>2.38</td>
<td>3.07</td>
<td>60</td>
<td>-4.518</td>
<td>.000**</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05; **p < 0.001

Q29/ Q32: Keeping a diary/ writing essays in Chinese.
Q30/ Q33: Using familiar words and sentence structures to make new sentences and reading writing samples, analyzing how other works.
Q31/ Q34: Spending time on writing exercises in the HSK example papers or simulated exam paper.

The results of the classroom activities (see Table 4.12) suggest that there was little difference in students’ preference for the classroom activities, which include teaching methods, involvement of the classroom activities, teaching content and classroom assessments format. The only sub-item showing a significant difference between regular and test-specific is the textbook. For the test-specific learning, students hoped their textbook would relate to the HSK test (M = 2.33, SD = 1.00), for example, having exercises or content in the textbook similar to the HSK tasks; while in their regular learning, students did not show as much desire for their textbooks to be related to the test (M = 2.10, SD = .99).
### Table 4.12: Summary Results of *t* Tests for Classroom Activities

<table>
<thead>
<tr>
<th>Variable</th>
<th>95%CI</th>
<th>Mean</th>
<th>Mean</th>
<th>N</th>
<th><em>t</em></th>
<th><em>p</em></th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q35 vs. Q41</td>
<td>-0.07</td>
<td>.21</td>
<td>1.83</td>
<td>1.90</td>
<td>59</td>
<td>0.94</td>
<td>1.82</td>
</tr>
<tr>
<td>Q36 vs. Q42</td>
<td>-0.20</td>
<td>.13</td>
<td>1.87</td>
<td>1.83</td>
<td>59</td>
<td>-0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>Q37 vs. Q43</td>
<td>-0.14</td>
<td>.17</td>
<td>1.87</td>
<td>1.88</td>
<td>59</td>
<td>0.22</td>
<td>0.83</td>
</tr>
<tr>
<td>Q38 vs. Q44</td>
<td>0.36</td>
<td>.03</td>
<td>2.48</td>
<td>2.32</td>
<td>59</td>
<td>-1.70</td>
<td>0.09</td>
</tr>
<tr>
<td>Q39 vs. Q45</td>
<td>-0.44</td>
<td>-0.02</td>
<td>2.33</td>
<td>2.10</td>
<td>59</td>
<td>-2.23</td>
<td>0.03*</td>
</tr>
<tr>
<td>Q40 vs. Q46</td>
<td>-0.39</td>
<td>0.09</td>
<td>2.30</td>
<td>1.15</td>
<td>59</td>
<td>-1.27</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Note: *p* < .05*

Q35/ Q41: Structural/grammar teaching approach.
Q36/ Q42: Communicative language teaching approach.
Q37/ Q43: Being involved in all the classroom activities.
Q38/ Q44: Learning test-taking strategies.
Q39/ Q45: Textbooks related to the HSK test preparation.
Q40/ Q46: Quizzes, mid-terms and final-exams similar to the HSK.

#### 4.2.3 What factors contribute to washback effects?

Previous research has suggested that there are various factors mediating the process of washback. For example, Watanabe (2004) listed the following factors: 1) test factors (e.g., test methods, test content, skills tested, purpose of the test, and decisions that are made by the test results; 2) prestige factors (e.g., stakes of the test, status of the test, and status of the test within the entire educational system); 3) personal factors (e.g., teachers’ educational backgrounds, their beliefs about the best methods of teaching and learning); 4) micro-context factors (e.g., the school setting in which the test preparation is being carried out); and 5)
macro-context factors (e.g., social factors). Based on previous research, in the
current study nine language-learning variables were selected as potential
predictors of washback effects on HSK from the learners’ perspective. These
features are: gender, nationality, age, how long they have been learning Chinese,
purpose of learning Chinese, purpose of taking HSK, type of courses, language
proficiency, HSK performance and motivation.

In order to explore what might explain the significant overlap between the
washback effects and individual learner factors, a multiple regression analysis was
applied. By using the “enter” method, the above-mentioned nine features were
entered as predictor independent variables and the washback effect (the change of
score between regular learning and test-specific learning) as the dependent
variable.

Table 4.13 below reveals a significant model for the dependent variable, $F$
$(5, 54) = 2.64, p = .014$. It is found that nationality ($\beta = .31, p = .05$), language
proficiency ($\beta = 0.41, p = .05$), HSK performance ($\beta = .34, p = .01$) and
motivation ($\beta = -.49, p = .002$) were significant predictors. However, gender, age,
the time of learning Chinese, purpose of learning Chinese and type of Chinese
course were not significant predictors. Since only the motivation factor among the
nine factors is continuous data, it is found that the stronger a learner’s motivation
for taking the HSK, the more obvious the washback effect. The overall model fit
was adjusted $R^2 = .34$. That is to say, these 9 features contribute 34% to the
independent variable, which is the washback effect.
Table 4.13: Summary Results for Multiple Regression Model for the Predictor Variables

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>B</th>
<th>SE</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-.07</td>
<td>.12</td>
<td>-.07</td>
<td>-.53</td>
<td>.60</td>
</tr>
<tr>
<td>Nationality</td>
<td>.29</td>
<td>.15</td>
<td>.31</td>
<td>2.01</td>
<td>.05*</td>
</tr>
<tr>
<td>Age</td>
<td>.16</td>
<td>.11</td>
<td>.24</td>
<td>1.45</td>
<td>.15</td>
</tr>
<tr>
<td>How long have learned Chinese</td>
<td>-.03</td>
<td>.09</td>
<td>-.05</td>
<td>-.29</td>
<td>.77</td>
</tr>
<tr>
<td>Purpose of learning Chinese</td>
<td>.06</td>
<td>.06</td>
<td>.14</td>
<td>1.08</td>
<td>.29</td>
</tr>
<tr>
<td>Purpose of taking HSK</td>
<td>-.09</td>
<td>.06</td>
<td>-.22</td>
<td>-1.75</td>
<td>.09</td>
</tr>
<tr>
<td>Type of Chinese Course</td>
<td>.06</td>
<td>.05</td>
<td>-.18</td>
<td>-1.29</td>
<td>.20</td>
</tr>
<tr>
<td>Language proficiency</td>
<td>.18</td>
<td>.09</td>
<td>.30</td>
<td>1.98</td>
<td>.05*</td>
</tr>
<tr>
<td>HSK performance</td>
<td>.65</td>
<td>.24</td>
<td>.34</td>
<td>2.70</td>
<td>.01*</td>
</tr>
<tr>
<td>Motivation</td>
<td>-.21</td>
<td>.06</td>
<td>-.49</td>
<td>-3.34</td>
<td>.002*</td>
</tr>
</tbody>
</table>

Note: B=un-standardized beta coefficient, SE= standard error, β = standardized beta coefficient, t = t-test statistic, p = significance value, p ≤.05*

4.2.5 What is the participant’s view of the HSK and the effects of the HSK?

Q54, Q55, Q56, Q57, Q58, Q64 and Q65 concern the test-takers’ view about the nature and content of the HSK. Table 4.14 below shows that most of the participants agreed that the HSK test measures test-takers’ communicative competence and language abilities and that HSK score reports inform the test-takers’ of their strengths and weaknesses. However, 86.6% of the participants believed that the test pays too much attention to assessing test-takers’ linguistic
knowledge (e.g., vocabulary, grammar rules and sentence structures), and that the test needs some improvement. Although some of them thought the HSK test well represents the *Scales*, others expressed the opinion that they did not really know what the *Scales* is.

Table 4.14: Summary of Descriptive Statistics of Responses to Q54, Q55, Q56, Q57 and Q65

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q54</td>
<td>16.7%</td>
<td>58.3%</td>
<td>23.3%</td>
<td>1.7%</td>
<td>2.10</td>
<td>.68</td>
</tr>
<tr>
<td>Q55</td>
<td>18.3%</td>
<td>68.3%</td>
<td>11.7%</td>
<td>1.7%</td>
<td>1.97</td>
<td>1.13</td>
</tr>
<tr>
<td>Q56</td>
<td>13.3%</td>
<td>55%</td>
<td>30%</td>
<td>1.7%</td>
<td>2.20</td>
<td>.68</td>
</tr>
<tr>
<td>Q57</td>
<td>1.7%</td>
<td>31.7%</td>
<td>50%</td>
<td>16.7%</td>
<td>2.82</td>
<td>.72</td>
</tr>
<tr>
<td>Q65</td>
<td>13.3%</td>
<td>61.7%</td>
<td>21.7%</td>
<td>3.3%</td>
<td>2.15</td>
<td>.68</td>
</tr>
</tbody>
</table>

Note:
Q54: The HSK tests focus on assessing test-takers’ language abilities and communicative competence in their daily, academic and professional lives.
Q55: The HSK tests focus on assessing test-takers’ linguistic knowledge.
Q56: The HSK test well represents the “Chinese language proficiency scales for speakers of other languages”, which is the guideline of Chinese teaching.
Q57: You don’t think there are any aspects of the HSK that can be improved.
Q65: The HSK score reports inform you of your strengths and weaknesses.

Q60, Q61, Q62 and Q63 (see Table 4.15 below) are about the test-taker’s perspective on the HSK’s influence on CSL learning and teaching. The results showed that the HSK test has potential washback effects. In the matter of learning, 70% of the participants showed positive washback in reporting that the HSK test helps them adjust their learning strategies and leads to success in future Chinese
learning. In contrast, more people believed that it encourages memorization of vocabulary and grammar rules instead of the use of real language abilities. This can be seen as negative washback. Concerning teaching approaches, interviewees hold different opinions. Only, half of them thought the test would motivate teachers to improve their teaching methodology.

Table 4.15: Summary of Descriptive Statistics of Responses to Q60, Q61, Q62 and Q63

<table>
<thead>
<tr>
<th>Question</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q60</td>
<td>10.0%</td>
<td>38.3%</td>
<td>46.7%</td>
<td>5.0%</td>
<td>2.47</td>
<td>.75</td>
</tr>
<tr>
<td>Q61</td>
<td>11.7%</td>
<td>58.3%</td>
<td>25%</td>
<td>5.0%</td>
<td>2.23</td>
<td>.72</td>
</tr>
<tr>
<td>Q62</td>
<td>30%</td>
<td>43.3%</td>
<td>21.7%</td>
<td>5.0%</td>
<td>2.01</td>
<td>.85</td>
</tr>
<tr>
<td>Q63</td>
<td>23.3%</td>
<td>41.7%</td>
<td>33.0%</td>
<td>1.7%</td>
<td>2.13</td>
<td>.79</td>
</tr>
</tbody>
</table>

Note:
Q60: It motivates teachers to improve their methodology in teaching Chinese.
Q61: It helps the test-takers adjust their learning strategies and leads to success in future Chinese learning.
Q62: It encourages memorization of vocabulary and grammar rules instead of the use of real language abilities.
Q63: It forces the test-takers to focus on test-taking strategies instead of the content (e.g., you may forget most of the vocabulary and grammar rules after the test).

4.3 Interview Results

4.3.1 Profile of the participants

The 8 participants are CSL learners with different learning experiences and from different countries. Three are females and five are males. Three of them are from Korea, three of them are from Canada, one is from Japan, and one from
Italy. Four of them have experience in studying or working in China. Seven of them reported having received CSL training from professional academic programs, and the eighth one is mainly relying on self-learning. Table 4.16 below presents the general background on the participants.

**Table 4.16: General Information of the Interview Participants**

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>M</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Nationality</td>
<td>Canada</td>
<td>Korea</td>
<td>Korea</td>
<td>Japan</td>
<td>Canada</td>
<td>Canada</td>
<td>Korean</td>
<td>Italy</td>
</tr>
<tr>
<td>Age group</td>
<td>25-34</td>
<td>18-24</td>
<td>25-34</td>
<td>34-45</td>
<td>45+</td>
<td>18-24</td>
<td>25-34</td>
<td>25-34</td>
</tr>
<tr>
<td>Years learning Chinese</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2.5</td>
<td>&lt;1</td>
<td>2.5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Chinese course type</td>
<td>Confucius Institute</td>
<td>Universi ty CSL program</td>
<td>HSK preparation course</td>
<td>Weeke nd language course</td>
<td>Self-learning</td>
<td>Universi ty credit course</td>
<td>Universi ty Credit course</td>
<td>Universi ty CSL program</td>
</tr>
</tbody>
</table>


**4.3.2 Why does the learner take the HSK test?**

Three out of 8 interviewees attributed a portion or all of their HSK test-taking to self-motivation. For example, S1 identified his self-motivation as the sole influential factor. The following excerpt is taken from S1’s comments.
Interviewer: 你为什么参加 HSK 考试？

[Translation: Why did you take the HSK?]

S1: 我喜欢学习新的语言。我已经学汉语四年了，而且我还在北京工作了半年。我觉得自己应该通过一个考试反映出现在的汉语水平，它可以帮我找到我的优点和弱点。

[Translation: I love learning new languages. I have been learning Chinese for 4 years and I was working in Beijing for half a year. I thought I needed a test to examine my proficiency level and HSK is an appropriate test. I hoped that through the test I could find out my strengths and weaknesses.]

Interviewer: 你考了几级？你为何选了这一级？

[Translation: Which level of the HSK did you take? Why did you choose this level?]

S1: 我考了中级（4级）。我联系了 HSK 温哥华考试中心告诉他们我想报名考试。他们询问了我学汉语的经历和水平，建议我考这一级，并且给了我一个关于 4 级考试的单词表。

[Translation: I took the Intermediate level (level 4). I contacted the HSK center in Vancouver and told them I would like to register for the test. Based on my learning experience, they suggested I take level 4 and gave me a vocabulary list of this level.]

In addition to the above-mentioned internal factors, the interviewees were also motivated by some external influences. The influences that drove students to learn Chinese and to take the HSK were viewed from the combined perspectives
of society, school and career. As one of the most crucial reasons, CSL learners’ interest in learning Chinese was influenced by the significant role of China in world social, economic and political affairs. For example, when asked what caused her to take the HSK, S3 explained:

S3: Taking the HSK test is very popular in Korea. With regard to the impact exerted by the workplace, I had been informed that the HSK certificate would be an asset for job-hunting and professional development. My friend told me that her company (a famous trading company) adopted the HSK as the indicator of employees’ Chinese ability. In order to pursue her increased remuneration, she took the HSK.

Two interviewees from Korea and one interviewee from Italy pointed out that their academic programs require them to pass a specific level of the HSK. For example, S2, a student at a University CSL program in China, took the HSK because his goal is to pursue his higher education in a Chinese University. HSK level 5 is required as the language requirement.

Interviewer: 你为什么参加 HSK 考试？

[Translation: Why did you take the HSK?]

S2: 主要是受到了父母的鼓励，我决定到中国读本科。一般中国的大大学要求留学生达到汉语水平考试的 5 级或者 4 级才可以开始上学。3 年前我开始学汉语的时候我就设定考过 HSK 考试作为我的目标。我第一次考试不合格，但是我第二次考通过了。

[Translation: With the encouragement of my parents, I decided to pursue my undergraduate studies in China. Generally, Chinese universities set the
HSK level 5 or 4 as their language proficiency requirement for international students. When I began to learn Chinese 3 years ago, I set passing HSK level 5 as my goal. I failed the first time trying but I succeeded the second time.]

In summary, through the above findings, students pointed out the following sources of impact that motivated or compelled them to take the HSK: themselves, a program (academic/professional), society and other people like parents or friends.

**4.3.3 The impact of the HSK on learning**

In general, the HSK caused various degrees of washback (negative, positive or neutral) on learning among the participants. Some of the interviewees had spent much time and energy on preparing for the HSK, whereas, some spent no time preparing.

Interviewer: HSK 考试对你学习方面有什么影响吗，课堂和课后都包括？

[Translation: Did taking the HSK change your learning behaviors (both in the classroom and after-school contexts)]]

S6: 我认为有影响。在平时学汉语的时候，我比较喜欢练习口语，找中国人聊天儿，有时候为了练习写作，我还找了中国学伴，每周我们都交换日记，短文。后来我开始准备 HSK 考试了，我发现这个考试要求的内容和我平时学习时侧重的不一样，它要考查非常多的单词和语法知识。所以我每天花很多时间背单词和做考试模拟题，我考过
I think there is an influence of the HSK on my learning. Before I took the HSK, I used to spend much time on practicing speaking, for example, talking with Chinese speaking people. Sometimes, in order to practice my writing, I got a language exchange partner. We exchanged our dairies and essays every week. Then while preparing for the HSK, I found that the text content is to some extent different with what I put particular emphasis on. It tests a lot of vocabulary and many grammar rules. So I used to learn Chinese by memorizing word lists and doing mock/practice exams. I passed level 5 with a good score. After that, I forgot most of the vocabulary and lost the drive to learn Chinese.

Similarly, S2 holds similar views about the HSK, but from a different angle. Because of the test-driven culture in Korea and other East Asian countries, not only is most course content driven by exams, but also teachers’ and students’ attention is focused on the content that will be tested in the exams. S2 indicated that his experience of learning Chinese was greatly pressured by his Korean classmates, teachers and parents. Moreover, he expressed that it is hard to make connections between tasks and the university setting in China that they were preparing to enter.

S2: 你知道韩国人考试很厉害吧。我的同学们每天花 10 几个小时学习。
老师上课讲关于 HSK 的东西很重要，我希望他们多讲，考前我也请
了家教辅导，花钱很多。考完之后很多东西都忘了。我现在上中国大学的专业课觉得非常难，内容和 HSK 很不一样。

[Translation: You know Korean students are crazy about any exams. My classmates spent over 10 hours a day studying for the HSK every day. I think preparatory classes for both the knowledge and test-taking strategies for the HSK play a crucial role in improving my HSK result. I hoped they could cover more in class. Besides, I had a private tutor to help me as well. I have spent lots of money on test preparation. However, although I passed the level 5, I forgot a lot after the test. When I entered the Chinese university credit program, I found the courses and textbooks were so different than the test. They were too hard for me.]

S5, however, saw the effect of the HSK on learning in a more positive light. In contrast to the perceptions of other learners, S5 explained the impact of the HSK in a different way. This view showed that passing the HSK was not only a milestone in students’ CSL learning, but also for some served as an impetus for their later studies. Because S5 was a Chinese language beginning level learner, she responded in English.

S5: I have been learning Chinese for only half a year. I think it is a great experience taking the HSK. It is harder than what I thought before I took it. Because in the next level, there will be Chinese characters involved rather than just Pinyin⁹. I will devote more energy to learning the characters. It is

---

⁹ Pinyin is the official phonetic system for transcribing the sound of Chinese characters into Latin script in China, Taiwan and Singapore. It is often used to teach Standard Chinese and spell Chinese names in foreign publications and may be used as an input method to enter Chinese
good for reading and writing. It motivates to enhance the proficiency in Chinese. You know your strengths and weaknesses, and then you strengthen that aspect. It is good for students.

Compared to other learners, S1 had a neutral point of view. The HSK had little impact on his learning methods, strategies and beliefs. However, in most of the above cases, interviewees showed that they learned specifically for the exam. Only S1 held a different perspective.

S1: I still kept my original learning strategies for learning Chinese. Because my purpose for taking the test was to examine my language proficiency, I hoped to find out my weaknesses through the test and improve my Chinese. By the way, I think the word list [the one that was given by the test center before the test] is good. I have often reviewed it since the exam.

Concerning the impact of the HSK on teaching approach, activities and textbooks, students generally had two opinions. On the one hand, they prefer teachers talking about the HSK in class, such as test content, test-taking strategies; on the other hand, students do not want teachers to address anything concerning the test in their regular CSL class.

4.3.4 What is the participant’s view of the HSK?

Some interviewees believed that the HSK was valid and reliable, whereas others had neutral or negative perspectives on these issues. Participating students
believed that the HSK was seen as a credible test in the public view. However, they still pointed out several issues and problems with the test. The following comments illustrate the participants’ general opinion.

Interviewer: What do you think about the validity and reliability of the HSK? Are there any aspects of the HSK that can be improved?

S6: I think the most important thing in learning a language is the “communicative” ability. The HSK doesn’t even have a speaking task, and the majority of the test tasks are in the form of multiple-choice.

Interviewer: HSK has a separate spoken test, didn’t you know?

S6: Really? I did not know that. I might have a try sometime later.

Questioning the validity of the HSK, L6 raised the issue of multiple-choice (MC) questions. In fact, a number of researchers (e.g. Alderson, 2000; Hughes, 2003; Shih, 2007) have disapproved of MC questions in testing language skills. In addition, much evidence indicates that test-takers can be coached in answering MC questions. The test scores therefore are not commensurate with language proficiency. There was a revision of the HSK in 2009. The new HSK has a speaking test, but only a few people are aware of this change.

Interviewer: 你觉得这个考试怎么样，比如说内容，形式方面？

[Translation: What did you think about the HSK, for example in terms of its content and the structure?]

S7: 我觉得这个考试不能说全部，但是比较客观的反映了学习的水平。它考查的也大都是日常生活的内容为主，是我们需要知道的。可是，它考察很多语言点和生词。我觉得这样的考试对汉字国的人（比如日本）有利，
EXPLORING THE WASHBACK OF THE HSK ON LEARNER FACTORS

我们知道意思。但是对于字母语言的人，有时候我依靠语言分析判断能力猜测单词的意思。我觉得 HSK 可以分成几个不同的目的和类别，比如雅思，现在的是普通的考试，如果要去中国留学的可以有学术类的考试。另外，我觉得 HSK 考察 “input”很多，但是“output”不够。

[Translation: I think this test objectively reflects the test-taker’s learning ability. It examines the language people used in daily life. However, it over emphasizes the linguistic rules and vocabulary. I think people from the logographic Chinese writing context [such as Japan] will have an advantage. It is very easy for us to understand the meaning of the vocabulary and paragraphs. But for those “alphabet language” people, they sometimes rely on language analysis and judgment ability to guess the meaning of the words. In addition, I hope the HSK can have different versions for different purposes [like IELTS]. The current version could serve as a general-purpose test, and if people want to go to China to study, they could take the academic-purpose exams. Third, the HSK focuses on input but lack output. For example, there is no speaking part; also, the writing part is very easy and short.]

In responding to my question about the HSK’s validity and reliability, S7 made a concrete analysis of the specific question in the listening section, as she thought that the talks in the listening part were too long and with too much information. Although the difficulty of the test is not high, test-takers need to take notes and answer questions at the same time; as a result, they might miss some “useful” information.
In general, the participants thought that the test fairly and accurately reflected their proficiency levels. Of the eight participants, six felt that the test generally reflected their abilities, but the scores they received in the test were higher than their actual performance level. The other two learners experienced frustration about the mismatch of their usual class performance compared to their test performance, which fueled their skepticism about the ability of the test to accurately reflect their language abilities. S8 suggested that a mismatch between ability level and test result could be attributed to nervousness and stress. Some researchers have raised this concern about emotional influence on test performance (e.g. Shohamy, 1982; Fox, 2005).

Participants perceived several advantages once they were HSK-certified, ranging from increased self-confidence to securing an edge over their competitors when applying for jobs. The first advantage a majority of participants could foresee was that the HSK certificate would be the badge of their Chinese proficiency. With respect to job opportunities, as S4 claimed, his company in Japan prefers to hire employees who know the Chinese language for potential internationalization. Similarly, passing the HSK would also favor those students who applied for further studies in China. As S2 shared at the beginning of this chapter, the HSK certificate could help people be admitted to undergraduate programs in a Chinese university.

4.4 Chapter Summary

In this chapter, I have elaborated on the quantitative and qualitative data of participants’ views of the HSK and the impact on learning. First of all, based on
the documents review and the data from the learners, it appears that the validity, reliability and content of the HSK are of high quality, but some aspects still need improvement. For example, the HSK needs more alternative question forms rather than MC, and it should emphasize output skills (speaking and writing). Second, there are more examinees from Asian countries than from Western countries and South Korean test takers are the largest group. Third, the significant differences between regular learning and test-specific learning show that there were washback effects from the HSK. On the one hand, there is positive washback, such as benefiting learner’s motivation for learning Chinese and helping them to understand their strengths and weaknesses and, as a result, helping them to adjust their learning strategies. On the other hand, like other high-stake exams, the HSK inevitably has negative effects, such as overemphasizing linguistic knowledge such as vocabulary and grammar rules, and neglecting communicative skills. In addition, expanding on the previous research, nationality, language proficiency, HSK performance and motivation were significant predictors for the change of learning behaviors between regular learning and test-specific learning. However, gender, age, the time of learning Chinese, purpose of learning Chinese and type of Chinese course were not significant predictors. Moreover, participants believed that the HSK fairly and accurately reflected their proficiency level and they perceived several advantages to becoming HSK-certified, such as promoting motivation of learning Chinese and increasing job opportunities. Last but not least, though there are separate HSK speaking tests, they do not appear to have public credibility and some CSL learners do not seem to be aware of them. In this regard,
although the HSK claims to assess students’ communicative ability to use language through real-life tasks, it is unclear whether test scores actually reflect this competence. Owing to the fact that little evidence can be retrieved to demonstrate construct validity, it appears validity issues need further development and transparency. Based on the results reported in this chapter, I will discuss these findings in relation to the research questions in the next chapter.
Chapter 5: Discussion of Results

This chapter discusses the findings in light of the theoretical framework and models presented in Chapter Two. I have chosen to discuss the most salient results that correspond to the global research question and the 4 secondary research questions. It should be noted that the four secondary questions are closely related to one another and all of them serve to address the global question. Thus, questions addressed in one section might be revisited in another section.

5.1 Research Question 1: Learner’s Interest in Chinese and the HSK

Research question 1 is: Why are the learners interested in learning Chinese and taking the HSK?

The results from the document review, the questionnaires and the interviews revealed that an increasing number of people were interested in learning Chinese and taking the HSK because they were not only interested in the Chinese language and culture, but hoped to study, work or travel in China. It is notable that the importance and usefulness of learning Chinese and taking the HSK could be attributed to the significant role that China plays in the world. As L5 said, “HSK certificate is like my visa to China.”

During the past few decades, the standardization and promotion of Mandarin Chinese within China has eased the way to promote Mandarin as a global language in the world. Nowadays, Chinese has been gaining popularity around the world. The number of learners studying Mandarin as a foreign language has reached 40 million (Custer, 2010). Not surprisingly, CSL learners’ interest in the HSK has surged. From the year 2000, when only 4,500 people took
the exam, the number leaped in 2006 to 370,000 people who took HSK at nearly 200 locations throughout China and the world (Wang, 2006). The number of examinees increased to 100 million (including domestic ethnic minority candidates) by 2010 (Luo et al., 2010). In the meantime, many countries have been taking measures to promote Chinese as an alternative foreign language offered in schools, including the United States, Thailand, Indonesia, Zimbabwe, and Sweden. In Canada, Chinese is becoming the third most spoken language after English and French (Statistics Canada, 2013). Over 2000 universities in over 100 different countries have started offering Chinese language courses. The increase in learning Chinese reflects the fact that China is recognized by the world as having a promising economy and a lucrative market (Li, 2008). Being the second largest economy worldwide, China has maintained an average annual growth rate of 8% percent in the past two decades. Overall, the Chinese economy has developed considerably and China’s role in the world economy has too. As a result, more and more people are becoming interested in learning the Chinese language and taking the HSK.

Just as western countries leverage their superpower status to promote their language and culture, so too does China. Chinese lags behind English as a global or influential language, but there is no doubt about its rising appeal, especially in Asia. China is the biggest trading partner of dozens of countries, including Japan, Korea, and Vietnam. Fluent Chinese speakers from these countries are likely to create more opportunities for their career development.

In order to promote the development of world Chinese language teaching
and build an effective platform for language cultural exchanges and cooperation between China and other countries, Hanban was established in 1984. Since its establishment, Hanban has started many initiatives to promote Chinese language and culture in the world, such as developing the HSK test, providing qualified teachers from China to teach abroad, and offering scholarships to CSL students to attend Chinese programs or degree education in China.

Moreover, both the qualitative and quantitative data in this study show that students with high Chinese proficiency usually have clear objectives for learning Chinese and taking the HSK. The passing rate of the HSK of the learners who set passing HSK as their goal is remarkably higher than those who did not set this goal. That is to say, setting clear short/long term learning goals will have a positive effect on the learning outcomes.

Overall, participants indicated that they had a positive attitude toward learning Chinese and taking the HSK. Their motivation was demonstrated at two interrelated levels: personal motivators and societal motivators. Personal motivators can be categorized into interest and instrumental orientations. Individual interest in Chinese language and culture and sustained efforts in spending time on Chinese learning and resorting to different learning resources and test-preparation resources are important motivators. Participants adopted the HSK as a yardstick to assess their Chinese proficiency. For example, they were spurred by their previous failures in the test to improve their Chinese. In addition, they displayed three types of instrumental orientations: mark orientation, further-education orientation, and job orientation. From the results showed in the previous
chapter, learning motivation is a key factor in increasing performance on the test. Motivation is highly correlated with outcome and higher effort will yield better performance.

5.2 Research Question 2: the Washback Effects on Learners

Research question 2 is: What is the evidence of washback effects on the learners?

A significant finding from the survey was there were washback effects of the HSK on learning behaviors. The HSK generated various degrees of washback on different individuals. Among the participants’ responses concerning the four skills (speaking, listening, reading and writing), some statistically significant differences emerged between regular learning and HSK test-specific learning. In line with Alderson & Wall’s (1993) observation that “tests can be powerful determiners, both positive and negative” (p. 117), my results showed that the HSK had a favorable impact on some learners but negative effects on others.

In terms of positive effects, taking the HSK benefits learner’s motivation for learning Chinese and helps them understand their own learning needs. The qualitative data supported the quantitative findings as learners referred to the impact of the HSK to measure their strengths and weaknesses, justify their learning practices and motivate them to learn more. For the exam-oriented or mark-oriented students, achieving high scores on the exam or passing the exam is a driving force for learning. This point is supported by Huang & Li (2010), who revealed that preparing for the HSK is useful and helpful for adjusting learning
goals, improving weaknesses, shaping learning methods and enhancing proficiency in the long-term.

The high stakes of the HSK draws CSL learners’ attention to their own learning. This echoes the theoretical statement of Shohamy et al. (1996) that if the stakes of a test are high in the society, its influence over its stakeholders will also be strong. Furthermore, as Gardner (2012) clearly stated, “assessment is our focus but learning is our goal” (p. 2). Assessment for learning can help learners discern where they are and where they need to go in the learning. As L7 said, HSK could help learners to find out their language level and the strengths and weaknesses of their language abilities. It could be seen as a feedback tool for language learning. Being able to recognize one’s current level and also being willing to take future steps is a clear sign that learner’s agency is activated, and a test is seen as part of the learning continuum.

Nevertheless, there are some negative effects of the HSK, such as encouraging memorization of vocabulary and grammar rules instead of practicing real language abilities. It forces the test-takers to focus on test-taking strategies instead of content, thus leading students to forget vocabulary and grammar rules after the test. Huang & Li (2010) also reported this trend. This, in turn, leads CSL teachers to focus on a structural teaching approach. This approach emphasizes grammatical structures and involves memorization of rules, mechanical manipulation of language forms, such as concentrating on teaching discrete points of language, vocabulary and pronunciation. The structural teaching approach neglects communicative competence, which is a significant goal of CSL. Hence, a
combination of a structural teaching approach and communicative teaching approach is called for. Compared to the HSK old version, the New HSK Test Syllabus (abbreviated as syllabus) stated that HSK is a standardized test to measure the test candidates’ Chinese communication skills and created the optional Oral HSK. However, because the Oral HSK is not yet well known, most of the participants in this study paid much less attention to the speaking practices than the other three skills. They were more interested in the learning product than in the process. The lack of spoken tests makes the test less relevant to measuring students’ communicative language skills. The inclusion of a compulsory speaking component in the test could strengthen the development of communicative competence proficiency. Not only could it promote learners’ interest in building up oral communicative abilities in and outside the classroom, but it can also help the learners apply the skill in real-life situations.

5.3 Research Question 3: How does the HSK Affect Learning?

Research question 3 is: How does the HSK affect the way they learn? In other words, what factors contribute to the change of learning strategies and practice used in regular learning and HSK specific learning?

As discussed earlier, four significant predictors of the washback regression model were found: language proficiency, HSK performance, motivation and nationality. The interview data also showed a great difference in learning beliefs and practices between Asian CSL learners and Western CSL learners. In this section, some significant issues of influences emerging from the data set are addressed.
As Little (2007) noted, the ability to take charge of one’s own learning characterizes learning autonomy. Independent or autonomous language learning has mainly been associated with Western and tertiary educational settings; it is sometimes perceived to be more problematic in an Asian context (Lamb, 2004). While warning against stereotyping “Asian learners”, Littlewood (1999) suggested that socialization practices in family and school inevitably influence learners’ attitudes and responses to academic freedom, and therefore different types of learning autonomy may be suited to Asian educational contexts.

In order to investigate the deep reasons underlying the differences between East and West, here I highlight the characteristics of Asian students and the Asian test-driven context. The East Asian educational system is examination-driven. One typical example is that almost all Korean participants in this research were aware of the overwhelming washback effects of the HSK. Under such an enormous impact, there is no denying that most CSL teachers would teach to the test and students would learn to the test. This may be attributed to the time-honored Confucian tradition of the role played by testing in Asian countries. Confucius, who is acclaimed as the greatest thinker and educator in ancient China, has had an enduring impact on teaching and learning in East Asia for centuries. The core of his educational values in terms of learning and studying are identified in four main themes: (1) perfecting oneself morally; (2) acquiring knowledge/skills for survival, self-sufficiency and careers; (3) contributing to society by linking learners’ personal pursuit of knowledge to a higher moral and social responsibility, and (4) obtaining social respect/mobility (Lee, 1996). With
this influence, the Imperial Examination (also called the Civil Service Examination) was chosen as the unique examination to select government officers starting in the 7th century and continuing to the 20th century in China. Korea also used testing to employ government officials of high rank during the Chosun Dynasty (1392-1897). Given that it has been the test that provides opportunities for individuals to climb the ladder of social status, Koreans in general have tended to have an implicit faith in the potential usefulness of testing (Choi, 2008). Nowadays, exams play an equally important role in the society. Test scores are viewed both as a marker of students’ academic success, and as the premise to their future career. In other words, academic achievement leads to better job opportunities and higher social status.

Within this context and in this study, the learning beliefs and practices of Asian students are reflected in several aspects during test-specific learning, which is different from regular learning. First, passing the exam is the only goal for students, parents, teachers and schools and is the key to success in the future. However, in their regular learning, their purpose is mastering linguistic knowledge and communicative skills. Another typical feature of the learning beliefs and practices is that the test is knowledge-focused. It is widely believed that if a student would like to learn a language well, he or she has to accumulate or acquire the basic linguistic knowledge (grammar rules, vocabularies, etc.). Guided by this belief, external factors such as teachers, parents and schools tend to stress the importance of what they perceive as knowledge of a language. Within such a social setting and the pressure of passing the exam, learning is generally
viewed as a process of accumulating knowledge, analyzing, and memorizing, rather than a process of acquiring practical skills, such as communication. In regular learning, however, participants expressed paying more attention to practicing oral and writing skills. Third, inherent in Asian learning beliefs is that the best test-taking strategy is rote memorization of word lists and practicing the simulated exam papers. According to Gu & Johnson (1996), rote-memorization is regarded as simplistic and associated with a surface approach to learning, and this strategy seems to have received a lot of criticism. However, more than one interviewee in this study expressed the view that recitation and rote-memorization was boring but effective in preparing for the HSK. In addition, participants from the East Asian countries showed a universal test-taking strategy, which is, practicing the simulated exam papers. As L2 claimed, the more simulated exams he practiced, the better results he would receive. It is helpful for improving the test results by becoming familiar with the test form and content.

5.4 Research Question 4: the Impact of the HSK

Research question 4 is: What are the views of the CSL learners towards the HSK and its impact?

It should be noted that there was a debate among the test-takers on the appropriateness of the HSK content. The first position holds that the new HSK is well-developed in the sense of being practical, scientific, valid and reliable. It is believed that linguistic knowledge is fundamental to the development of communicative competence. From this point of view, without acquisition of linguistic knowledge or sufficient input, the enhancement of other competencies
would not be possible. Compared to the old HSK, the new HSK, with its written part of the test, does require the use of language in real-life tasks. Besides, the difficulty level of the HSK, especially for the elementary levels, is not high, so learners’ achievement can be recognized as accurate. The HSK can help boost learner’s confidence in learning Chinese and benefits their long-term learning.

The second position, however, criticized the continuous focus on vocabulary and grammar, which indicates the HSK is outdated in terms of current language proficiency tests. The vocabulary inventory provided in the HSK preparation is seen as giving the wrong message to test-takers. That is, as long as they have memorized the vocabulary list, they can pass the test. Meanwhile, critics state that there is an overwhelming reliance on MC and discrete-point tasks rather than performance-based question forms. As such, the test has exerted a considerable adverse influence on teaching and learning. This claim is supported and reinforced by Wang (2008) and Chen (2009). Some alternative types of testing tasks such as translation, spot and compound dictation, short-answer questions, and writing tasks should be added to the current format of the HSK. Such a modification could mitigate the negative washback of objective tasks of the test on CSL teaching and learning.

In addition, the participants questioned the absence of an oral HSK test. Apart from the mismatch between the Syllabus and the HSK, it is believed that the inclusion of the oral aspect of the language would result in more attention being focused on the practical use of the language by CSL learners and teachers. The chief goal for introducing the Oral HSK, according to the Syllabus, was to
increase an emphasis on measuring the students’ oral communicative competence. Based on the data described in the previous chapter, however, the passing rate of the Oral HSK is much lower than the Written HSK. This indicated that CSL students’ spoken ability needed to be improved. Moreover, what is controversial is that the Oral HSK is optional and some students do not even know about this independent speaking test. From the perspective of institutions and job markets only the Written HSK is required. Consequently, the number of the Oral HSK test-takers in this study was much lower than those who took the Written HSK and the passing rate of the Oral HSK was low as well. Thus test construction reform is needed.

When asked about the popularity of the HSK, participants attributed the HSK “whirlwind” to its established credibility, changes in the socioeconomic climate, and vigorous propaganda conducted by the government. The HSK could be a reflection of one’s language ability for the learners, a reflection of a teaching effect, or a criterion for institutions and companies selecting candidates. Drawing on the findings, it appears that in order to serve as the panacea for all the stakeholders, there need to be improvements to the HSK.

5.5 Chapter Summary

In summary, this study indicates that the HSK has significant impact on learners’ learning. The findings indicated that due to the rising status of China internationally, CSL learners are more interested in and have positive attitudes toward learning Chinese and taking the HSK. Both inside and outside motivators act on the learner’ learning processes. The HSK generated various degrees of
positive and negative washback on different individuals. Asian CSL learners were found to be quite different from learners from other countries in terms of learning beliefs and practices. I have argued that these reasons for learning are related to Confucian values and were actively promoted by individuals, families, schools and society at an overall level. Exam culture has been predominant and influential in these areas. The supremacy of learning and the likelihood of upward social mobility have motivated East Asian people for generations to strive to learn for exams in order to realize their aspirations and social status. Though some people believe that the new HSK is well developed in the sense of being practical, scientific, valid and reliable, others argue that the HSK still needs further improvement by increasing the diversity of test items and tasks, paying more attention to the Oral test and enhancing the test’s validity. In the next chapter, I describe the implications and limitations of this study.
Chapter 6: Conclusion

This chapter discusses tentative conclusions and some suggestions for further research. First of all, I summarize the findings of the study. Secondly, I acknowledge limitations of the study. Last but not least, I address the implications and suggest future research directions.

6.1 Research Summary

Traditional washback research focuses on whether washback exists and whether there is intended washback brought about by examination change. The emphasis of this body of scholarship has been on ESL/EFL tests. This study, in contrast, investigated how the intertwined learner factors fit into the washback phenomenon of a large-scale high-stakes Chinese test. As such, it is a new contribution to the field of language testing.

The present research adopted MMR methods and the findings revealed that more and more people were interested in learning Chinese and taking the HSK because they were not only interested in Chinese language and culture, but also hoped to study, work or travel in China. One of the most significant reasons was China’s national strength and the fact that its international status is advancing. Since motivation was highly correlated with learning practices and outcomes, a significant finding from the survey and interview data was that there were washback effects of the HSK on learning. This study showed that HSK generated various degrees of washback on different individuals and had a favorable impact on some learners but negative effects on others, which supports Alderson & Wall’s (1993) claim that “tests can be powerful determiners, both positive and
negative” (p. 117). There were statistically significant differences in learner’s responses regarding regular learning versus HSK test-specific learning across the four skills: speaking, listening, reading and writing. Data showed the HSK could help learners identify their strengths and weaknesses, justify their learning practices and motivate them to learn more Chinese. The Syllabus, Scales and the new version of the HSK have attached great importance to improving students’ “integrated” abilities of using Chinese. In addition, expanding on previous research, nationality, language proficiency, HSK performance and motivation were found as the significant predictors for the change of learning strategies between regular learning and test-specific learning. Moreover, participants felt that the HSK reflected their proficiency level and they perceived several advantages if they were HSK-certified, such as promoting motivation of learning Chinese and increasing their job opportunities. Although the validity, reliability and content of the HSK appear satisfactory, improvement is still needed, for example, more subjective question forms, emphasizing the output (speaking and writing) measurement.

6.2 Limitations of the Study

This study has provided detailed empirical evidence that yields insights into the complex relationships of the learning process and the HSK. With respect to the literature on washback, learner factors and HSK washback, adopting the MMR approach to both data collection and analyses is a strength of the study and it has shed light on the relationship among learning strategies, learning beliefs and the exam. However, there are some limitations that need to be acknowledged.
First, the findings of the current study must be considered in relation to certain limitations regarding the participants and nature of the investigation. The findings discussed were based on a limited number of participants who volunteered to share their learning and test experiences and might thus not fully represent the perceptions of other CSL learners.

Second, the results of the quantitative data of this research relied on students’ self-reported strategy use and the beliefs in one questionnaire. It has some inherent drawbacks. First of all, it is not clear if the reported strategies and beliefs were congruent with their overt behaviors and thought. Also, some participants took the HSK test some time ago, and their responses might not reflect their present perceptions and beliefs.

Third, learners’ test performances in this study were based on pass or fail, not a real score. Originally, the reason for employing pass or fail as the performance indicator of the HSK was that most of the participants took the new HSK while a small group of participants took the old version of the HSK. There is no official equivalence of the old and new scores, so it is difficult to compare these two different kinds of scores. However, due to the high passing rate which had been discussed in the previous chapter, especially for the elementary proficiency levels’ tests, I could not accurately compare the learners’ language proficiency only based on the two bands: pass or fail. If the HSK test scores provided by the learners were available, various correlations among learning practice, learning beliefs and learning performance could be explored.
Fourth, this study used self-reported strategies and beliefs. In the future, self-reported surveys, which are internal and mental, as well as learner behavior observation in the classroom, could be combined. Another option for documenting learning strategies and beliefs is through the use of a diary. Diary studies can be used to collect data on a learner’s strategy over a period of time (Cohen & Scott, 1996, p. 99). Although there is a debate about the usefulness of diary studies in the field of SLA, they do allow research to focus on language learning specifically from the learner’s point of view.

6.3 Future Research Directions

As already noted, this study has made some contributions to the field of language assessment and CSL teaching/learning, particularly relating to the development of the test’s impact on learners’ practices and beliefs. However, the results of the present study also suggest the need for more studies probing washback of the HSK on all stakeholders.

First of all, I found, like Bernaus, Moore, and Azevedo (2007), that students’ attitudes toward the class and the teacher were related to many affective variables. Such results suggest that it would be useful to investigate teachers and institutions to examine how their decisions, actions and beliefs relate to affective characteristics of student learning.

Secondly, as mentioned in Chapter 4, this study chose to examine only some affective aspects of the learner factors in relation to washback effects, such as psychological characteristics. In this regard, it would be beneficial for future research to conduct studies using this method to further examine the
attitudes/beliefs underpinning the learners’ practice during regular learning and test-specific learning.

Third, the HSK is a high-stakes and large-scale test, therefore further research is needed on the HSK preparation programs/courses in various contexts and on the impact of the subtests (oral and written) to explore whether test preparation is effective. In this regard, it is also important to identify what factors contribute to the changes in learning strategies/beliefs in order to deliver more effective test-preparation training programs.

6.4 Implications of the Study

This study points to implications for the HSK test developers, CSL teachers and CSL learners.

6.4.1 Implications for test-developers

One implication of this study is that in future revisions of the test, HSK developers might give serious thought to whether positive washback might be enhanced by increasing the range of task types and other performance conditions; that is to further improve both its context and construct validity. Consideration needs to be given to making answers less predictable by varying the type and format of the tasks and by broadening the range of topics covered and the text types required in the tasks. This also should add relevance to the test because test-stakeholders would be able to see the relationship between the testing activities and the potential language use in real life. Test relevance could also be improved by incorporating interactive testing methods and testing integrative skills that are used in daily communications. These changes could lead to more positive
washback effects and a beneficial cycle among learning, teaching, and testing, with testing as a pivotal mediating role in the cycle.

Furthermore, as Huang & Li (2010) stated, the interpretation of scores on the HSK test needs to be more accessible, scientific and comprehensive. Score reports are one of the washback processes and provide direct feedback of the test takers’ performance; they also serve as the medium for delivering the information to all the score users in the washback processes. Such feedback could help the learners and teachers discover their strengths and weaknesses and guide them in improving their learning or teaching methods.

6.4.2 Implications for students and teachers

As Shohamy (1993) pointed out, “external tests are currently used to force students to study, teachers to teach, and principals to modify the curriculum” (p. 513). In terms of students, my findings suggest that by taking the test, learners gain insight into their strengths and weaknesses in Chinese and gain self-confidence and other advantages by passing the exam. In spite of this, students should make further efforts to improve their communicative ability and develop all of their language skills. Although mastering vocabulary and grammar rules plays a major role in the HSK preparation, other aspects of language abilities cannot be ignored, particularly for the CSL learners who plan to pursue their higher education in China. The results of this study suggest that if CSL learners are to be adequately prepared for academic study in China, passing a specific level of the written HSK is not sufficient. Only by recognizing the limitations of the HSK in this regard, will learners be able to make better informed choices.
about how to prepare themselves for university study. The HSK should not be the only element in this preparation.

According to the positive correlation between motivation and outcomes, CSL teachers should create a supportive and dynamic learning environment by setting clear learning goals/sub-goals, fostering learner autonomy, increasing students’ self-confidence, and maximizing their free choice. In order to motivate themselves to enjoy the continual challenge of improving results, learners should set clear short term and long term learning goals, as well as maintain their positive attitudes and confidence.

In addition, due to the results of the strong impact of the job market and learning motivation, CSL teachers should encourage students to possess a global outlook, use critical thinking, and embrace multiple perspectives of languages, cultures, and identity. It is of pedagogical importance to draw lessons from the research findings concerning ESL/EFL teaching and make efforts to combine structural and communicative teaching approaches.

Moreover, a clear finding of the study is that students spent much time and effort on memorizing vocabulary lists but struggled to retain the words they have learned. Teachers should help them avoid this negative washback effect by using other methods, such as visual or audio presentation modes to increase the amount of engagement with new words.

**6.4.3 Implications for institutions**

For institutions, it may no longer be sufficient to make gate-keeping decisions based on a simple formulaic approach relating to a level or band score
of the HSK. Washback is a complicated phenomenon and no single policy can guarantee success without any side effects, especially when tests are adopted for high-stake purposes. It is important to consider the recommendation by Huang & Li (2010) that the Oral HSK certificate is needed as one of the supplementary materials for decision-making. Interpretation of test scores by institutions should be made with full reference to the local contexts. The consideration of available support mechanisms in a particular institution, from CSL program to tutorial support, should be balanced against the demands of a particular course with regard to the academic-oriented learning. The institutions should understand that the international student experience would be enhanced if ongoing language support could be provided to help students become a part of the academic discourse community.

Therefore, on the one hand, as Shohamy (1993) advocated, in order to ensure the construct of test validity and increase the match between the curriculum and the test, test developers should focus on improving test design to facilitate language teaching and learning. On the other hand, it will be beneficial if adequate attention is directed at the stakeholders. Given the involvement of the learner factors in the language learning process, consideration will need to be given to all issues and conditions concerning the observable (i.e., strategies and behaviors) and unobservable dimensions (i.e., beliefs and attitudes) of learning. Overall, this will require the joint efforts of test developers, teachers, researchers and particularly test-takers.
6.5 Chapter Summary

In conclusion, this study is unique in that it explored the relationships between CSL learners’ factors and the HSK test using an MMR design. Both quantitative and qualitative data demonstrated the complexity of the washback processes that affect learning processes. The quantitative data make generalizations about contributing factors, while the qualitative data offered more in-depth and detailed information about the learners’ perceptions of the impact of the test. The analysis showed that it is important to make future improvements to the HSK that will emphasize positive washback and minimize negative washback. This study has enriched and expanded knowledge in washback theory development in second language assessment, and specifically contributes evidence to the validation of the HSK.
REFERENCES


Hughes, A. (1993). *Backwash and TOEFL 2000, unpublished manuscript*, University of Reading, UK.


Turner, C.E. (2001). The need for impact studies of L2 performance testing and rating: Identifying areas of potential consequences at all levels of the testing cycle. In A. Brown et al. (Eds.), *Experimenting with uncertainty: Language testing essays in honour of Alan Davies* (pp.138-149). Cambridge, Cambridge University Press.


Appendix A:

Questionnaire

Dear students,

I would like to ask you to help with my thesis study by answering the following questions concerning your perceptions of Hanyu Shuiping Kaoshi (HSK) and its impact on your learning. All information will be absolutely confidential. Please give your answers thoughtfully and carefully as only this will guarantee the success of the investigation. Thank you very much for your help.

Your Background Information

Check [ ] what you think is the most appropriate.

1. Your gender: [ ] female [ ] male
2. Your nationality: ________________
3. What is your age: ________________
4. What is your first language: ________________
5. How long have you been learning Chinese: ________________
6. What is your purpose in learning Chinese:
   [ ] You are interested in learning new languages.
   [ ] You would like to study/work/travel in China or related to China.
   [ ] It is required by the academic program/professional project.
   [ ] It is encouraged by my parents or friends.
   [ ] Other (please specify: ____________________)
7. What types of Chinese courses have you taken:
   [ ] University credit course (please specify your major: ____________)
   [ ] Weekend language school
   [ ] HSK preparation course
   [ ] Confucius Institute course
   [ ] Other (please specify: ____________________)
8. Which HSK level have you taken?
   New version (after 2010): [ ] Elementary 1,2 [ ] Intermediate 3, 4 [ ] Advanced 5, 6
   Old version (before 2010): [ ] Elementary 1,2,3 [ ] Intermediate 4,5,6,7,8 [ ] Advanced 9,10,11
9. If you took the new HSK, did you succeed in it? [ ] Yes [ ] No
If you took the old HSK, did you get the HSK certificate? [ ] Yes [ ] No

10. The HSK provides an indicator of Chinese language proficiency. Why did you take it?
[ ] To have an indicator of your language proficiency
[ ] To be able to pursue higher education in China
[ ] To help with job hunting
[ ] Others (please specify: ____________________ )

In the brackets [ ], please mark the following on a four point scale as:


During HSK test-specific learning, you tried to improve your speaking by…

11. [ ] Repeating what you have learned/heard.
12. [ ] Spending time with Chinese speaking people.
13. [ ] Spending time on speaking exercises in the HSK example papers or simulated exam papers.

During regular learning, you tried to improve your speaking by…

14. [ ] Repeating what you have learned/heard.
15. [ ] Spending time with Chinese speaking people.
16. [ ] Spending time on speaking exercises in the HSK example papers or simulated exam papers.

During HSK test-specific learning, you tried to improve your listening by…

17. [ ] Watching TV programs, movies or listening to radio in Chinese.
18. [ ] Listening to textbook CDs.
19. [ ] Spending time on listening exercises in the HSK example papers or simulated exam papers.

During regular learning, you tried to improve your listening by…

20. [ ] Watching TV programs, movies or listening radio in Chinese.
21. [ ] Listening to textbook CDs.
22. [ ] Spending time on listening exercises in the HSK example papers or simulated exam papers.

During HSK test-specific learning, you tried to improve your reading in Chinese by…

23. [ ] Reading Chinese books and magazines.
24. [ ] Learning new vocabulary and grammar rules.
25. [ ] Spending time on reading exercises in the HSK example papers or simulated exam papers.

**During regular learning, you tried to improve your reading by…**

26. [ ] Reading Chinese books and magazines.
27. [ ] Learning new vocabulary and grammar rules.
28. [ ] Spending time on reading exercises in the HSK example papers or simulated exam papers.

**During HSK test-specific learning, you tried to improve your writing in Chinese by…**

29. [ ] Keeping a diary/ writing essays in Chinese.
30. [ ] Using familiar words and sentence structures to make new sentences and reading writing samples, analyzing how other works
31. [ ] Spending time on writing exercises in the HSK example papers or simulated exam papers

**During regular learning, you tried to improve your writing by…**

32. [ ] Writing diaries in Chinese.
33. [ ] Using familiar words and sentence structures to make new sentences and reading writing samples, analyzing how other works
34. [ ] Spending time on writing exercises in the HSK example papers or simulated exam papers.

*In the brackets [ ], please mark the following on a four point scale as:*


**During HSK test-specific learning, in your regular Chinese class, you found the following things helpful…**

35. [ ] structural/grammar teaching approach.
36. [ ] communicative language teaching approach.
37. [ ] being involved in all the classroom activities.
38. [ ] learning test-taking strategies.
39. [ ] textbooks related to the HSK test preparation.
40. [ ] quizzes, mid-terms and final-exams similar to the HSK.

**During regular learning, in your regular Chinese class, you found the following things helpful…**
41. [ ] structural/grammar teaching approach.
42. [ ] communicative language teaching approach.
43. [ ] being involved in all the classroom activities.
44. [ ] learning test-taking strategies in regular class.
45. [ ] textbooks related to the HSK test preparation.
46. [ ] quizzes, mid-terms and final-exams similar to the HSK.

When you began to learn Chinese…

47. [ ] You did not know anything about the HSK.
48. [ ] You set passing the HSK as your goal.

Among the four skills (listening, speaking, reading and writing),

49. [ ] You know your strengths and weaknesses.
50. [ ] You feel anxious about your weaknesses and want to spend more time on them.

From your perspective, what do you think about the HSK’s impact on learners’ learning motivation and beliefs?

51. [ ] You believe that taking the HSK benefits your motivation for learning Chinese (e.g., the test can motivate learners, help learners understand their own learning needs, etc.).
52. [ ] You believe that the more time you spend on the HSK preparation, the better the results you could achieve.
53. [ ] You believe that the preparation for the HSK is useful and helpful for improving your Chinese language proficiency in the long-term.

From your perspective, what do you think about the nature and content of the HSK?

54. [ ] The HSK tests focus on assessing test-takers’ language abilities and communicative competence in their daily, academic and professional lives.
55. [ ] The HSK tests focus on assessing test-takers’ linguistic knowledge (e.g., vocabulary, grammars rules and sentence structures.)
56. [ ] The HSK test well represents the “Chinese language proficiency scales for speakers of other languages”, which is the guideline of Chinese teaching.
57. [ ] You don’t think there are any aspects of the HSK that can be improved.
58. [ ] The HSK is a large-scale high-stakes test and has public credibility.

From your perspective, what do you think about the effects of the HSK?

59. [ ] It motivates students to enhance their proficiency in Chinese.
60. [ ] It motivates teachers to improve their methodology in teaching Chinese.
61. [ ] It helps the test-takers adjust their learning strategies and leads to success in future Chinese learning.
62. [ ] It encourages memorization of vocabulary and grammar rules instead of the use of real language abilities.
63. [ ] It forces the test-takers to focus on test-taking strategies instead of the content (e.g., you may forget most of the vocabulary and grammar rules after the test).

From your perspective, what do you think about the HSK score use?

64. [ ] The HSK scores are an appropriate indicator of Chinese ability.
65. [ ] The HSK score reports inform you of your strengths and weaknesses.
66. [ ] If passing a specific level of the HSK is the only path leading to the completion of your degree/program, you will view the HSK differently and will spend more time on preparing for the HSK.
67. [ ] The HSK certificate will help in your future job hunting.
68. [ ] The HSK certificate will help you pursue higher education in China.

Would you be willing to participate in a one-on-one interview? Yes [ ] No [ ]

Thank you very much for your time!
Appendix B:

调查问卷

麦吉尔大学

王姝娇

亲爱的同学们：

此问卷旨在调查您对汉语水平考试（HSK）的看法，以及 HSK 对您学习方面的影响。请您认真回答，答案的真实性是此次问卷调查成功的基础。您所有的信息将会绝对保密。非常感谢您的帮助。

背景信息

(请选择最适合的答案，在括号内打勾【√】)

1. 你的性别：【】女【】男
2. 你的国籍：________________
3. 你的年龄：________________
4. 你的第一语言是：________________
5. 你学习汉语多长时间了？____________
6. 你学习汉语的目的是什么：
   【】 我喜欢学习新的语言。
   【】 我打算去中国或与中国相关的地方学习、工作或者旅游。
   【】 这是我的学习专业或者工作项目所要求的。
   【】 受到父母或者朋友的鼓励。
   【】 其他（请注明：___________________________）
7. 你参加过哪类汉语培训课程：
   【】 大学学分课程(请注明你所学专业: ____________ )
   【】 周末中文学校
   【】 HSK 考试培训班
   【】 孔子学院课程
   【】 其他（请注明：___________________________）
8. 你参加过哪个等级的HSK考试？

新版（2010年以后）【】初级 1,2【】中级 3,4【】高级 5,6

旧版（2010年以前）【】基础 1,2,3【】初中等 4,5,6,7,8【】高等 9,10,11
9. 如果你参加了新版 HSK，你通过了 HSK 考试吗？ [ ] 是 [ ] 否

如果你参加了旧版 HSK，你得到了 HSK 证书吗？ [ ] 是 [ ] 否

10. HSK 考试体现了你的汉语能力水平。您为何参加 HSK 考试？
[ ] 了解汉语学习水平。
[ ] 获取在中国参加高等教育的资格。
[ ] 帮助就业和职业发展。
[ ] 其他（请注明：__________________________）


在备考 HSK 期间，你通过以下方式提高口语水平：

11. [ ] 重复你所学到的和听到的
12. [ ] 与说汉语的人进行交流
13. [ ] 做 HSK 样卷和模拟题中的口语练习

在日常学习汉语时，你通过以下方式提高口语水平：

14. [ ] 重复你所学到的和听到的
15. [ ] 与说汉语的人交流
16. [ ] 做 HSK 样卷和模拟题中的口语练习

在备考 HSK 期间，你通过以下方式提高听力水平：

17. [ ] 看汉语电视节目，电影或者听汉语广播
18. [ ] 做教科书 CD 的听力练习
19. [ ] 做 HSK 样卷和模拟题中的听力练习

在日常学习汉语时，你通过以下方式提高听力水平：

20. [ ] 看汉语电视节目，电影或者听汉语广播
21. [ ] 做教科书 CD 的听力练习
22. [ ] 做 HSK 样卷和模拟题中的听力练习
在备考HSK期间，你通过以下方式提高阅读水平：

23. [ ] 阅读汉语书刊、杂志
24. [ ] 学习新的词汇和语法点
25. [ ] 做HSK样卷和模拟题中的阅读练习

在日常学习汉语时，你通过以下方式提高阅读水平：

26. [ ] 阅读汉语书刊、杂志
27. [ ] 学习新的词汇和语法点
28. [ ] 做HSK样卷和模拟题中的阅读练习

在备考HSK期间，你通过以下方式提高写作水平：

29. [ ] 写汉语日记或作文
30. [ ] 运用不同的词语和句型结构造句，读写作例文，分析其他作品
31. [ ] 做HSK样卷和模拟题中的写作练习

在日常学习汉语时，你通过以下方式提高写作水平：

32. [ ] 写汉语日记或作文
33. [ ] 运用不同的词语和句型结构造句，读写作例文，分析其他作品
34. [ ] 做HSK样卷和模拟题中的写作练习


（如果你没有参加过任何汉语课程，请略过43题到54题。）

当你备考HSK期间，在你的日常汉语课堂中，你发现以下方式对你有所帮助：

35. [ ] 结构／语法教学法
36. [ ] 交际语言教学法
37. [ ] 融入所有的课堂活动中
38. [ ] 学习应试策略
39. [ ] 与HSK考试相关的教科书
EXPLORING THE WASHBACK OF THE HSK ON LEARNER FACTORS

40. [ ] 与 HSK 考试相似的小测验／期中／期末考试

在日常学习汉语时，在你的日常汉语课堂中，你发现以下方式对你有所帮助:

41. [ ] 结构／语法教学法
42. [ ] 交际语言教学法
43. [ ] 融入所有的课堂活动中
44. [ ] 学习应试策略
45. [ ] 与 HSK 考试相关的教科书
46. [ ] 与 HSK 考试相似的小测验／期中／期末考试

当你开始学习汉语时，

47. [ ] 你对 HSK 一无所知。
48. [ ] 你把通过 HSK 考试作为你的学习目标。

在听、说、读、写这四种技能中，

49. [ ] 你知道自己的优势和弱点。
50. [ ] 你对自己的弱点感到焦虑并花更多的时间在它上面。

你如何评价 HSK 对学习动机和信念方面的影响？

51. [ ] 你认为参加 HSK 考试对你学汉语的动机方面有积极的影响（比如，考试激励学生，帮助学生了解自己的学习需求等等）。
52. [ ] 你认为在 HSK 准备上花的时间越多在考试中的成绩就越好。
53. [ ] 你认为备考 HSK 对长时间内提高汉语水平有帮助。

你如何评价 HSK 的考试内容和属性？
54. [ ] HSK 专注于考察应试者的语言运用能力和日常生活、学习、工作中的语言交流能力。

55. [ ] HSK 专注于考察应试者语言点知识（比如，单词，语法和句子结构）。

56. [ ] HSK 考试很好的体现了《国际汉语能力测试标准》的要求。

57. [ ] 你认为 HSK 考试很完美，没有需要改进的地方。

58. [ ] 你认为 HSK 是大规模高风险考试，你认为 HSK 具有公共信度。

你如何评价 HSK 考试的影响？

59. [ ] 它激励学生提高汉语水平。

60. [ ] 它激励老师改进汉语教学方法。

61. [ ] 它帮助应试者调整学习策略并使其在未来的汉语学习中获得成功。

62. [ ] 它只鼓励单词记忆和语法规则学习，而不是真正的语言运用能力。

63. [ ] 它强迫应试者专注于应试技巧而不是考试内容本身（比如说你可能在考完试后忘掉了大部分的生词和语法）。

你如何评价 HSK 的分数应用？

64. [ ] HSK 成绩恰当的反应了学习汉语的水平。

65. [ ] HSK 成绩报告单体现出你的优势和薄弱环节。

66. [ ] 如果通过特定等级的 HSK 考试是你完成学业所必须要求的，你会对 HSK 有不同的看法并且花更多的时间和精力准备 HSK 考试。

67. [ ] HSK 证书将在你未来就业／工作中有所帮助。

68. [ ] HSK 证书将有助于你在中国接受高等教育。

您愿意参加一对一 Skype 采访吗？ 愿意[ ] 不愿意[ ]

非常感谢你的参与！
Appendix C:

**Interview Questions**

(English Version)

Shujiao Wang

**Introductory statement:** Thank you for agreeing to participate in this interview. Today I would like to ask you some questions concerning your perceptions of the HSK and its impact on your learning. Before we start, I want to assure you that what you are going to talk about here will be used for the purpose of this research project only. I will not let any other people listen to the recording or see the transcripts.

1. How long have you been learning Chinese? Describe your experience in learning Chinese? What is your purpose of taking the HSK?
2. Why did you take the HSK and not other Chinese proficiency tests? Do you think the HSK indicates your language proficiency?
3. Did the HSK influence your motivation for learning Chinese?
4. Among the four skills (speaking, listening, reading and writing), what are your strengths and weakness? What factors do you think contribute to your strengths and weakness?
5. What do you think are good strategies for learning Chinese? Did you follow what you believed when preparing for the HSK?
6. When you prepared for the HSK, how did you prepare the speaking, listening, reading and writing? Did you keep the same learning methods after you took the test?
7. Do you think that your Chinese proficiency has improved as a result of your preparation for the HSK?
8. Do you think that taking the HSK has induced long-term impact on your Chinese learning?
9. In your regular Chinese class, does your teacher mention anything about the HSK and test-taking strategies? What teaching approach do you think is the most appropriate in your regular Chinese class?
10. What kind of materials did you use in preparing for the HSK? The textbooks? The HSK samples? The HSK guidelines?
11. Did parents, friends or other people including who has passed the HSK have an impact on you taking the HSK? How do you think they can better assist you to prepare the HSK?
12. Do you think the HSK test reflects its objective - assessing test-takers’ language using abilities and communicative competence in their daily, academic and professional lives?
13. Are there any aspects of the HSK that can be improved?
14. What advantages can you perceive if you are HSK-certified? Do you think the HSK certificate will help you for your future job hunting or pursuing higher education?
15. If passing a specific level of the HSK is the only path leading to the completion of your degree, would you view the HSK differently? How does this policy influence your test preparation?
访谈问题

简介：感谢你参加这次访谈。今天，我会采访您一些关于你对 HSK 考试的态度以及 HSK 考试对你学习上的反反驳作用方面的问题。在我们开始之前，我想再次神匿名此次访谈只是用于研究目的，其他任何人都不会听到本次访谈的录音和看到记录。

1. 你学习汉语多久了？可以介绍一些你学汉语的经历吗？你考 HSK 的目的是什么？
2. 你为什么参加 HSK 考试而不是其他的汉语考试？你认为 HSK 能反映你学习的真实水平吗？
3. HSK 考试会不会影响你学习汉语的动机？
4. 在听、说、读、写四中技能中，你的优势和弱点是什么？什么因素造成了你的优势和弱点？
5. 你认为什么是学汉语的好学习策略？当你准备 HSK 考试时你也是按照这样的策略吗？
6. 当你准备 HSK 时，你是怎样分别准备听、说、读、写四方面的？你考完试后你还按照同样的方法学习吗？
7. 你觉得准备 HSK 考试对你汉语水平的提高有帮助吗？
8. 你认为参加 HSK 对你学汉语有长期的影响吗？
9. 在你日常课上，你的老师会讲 HSK 考试或者应试技巧吗？你认为什么样的教学方法是最适合日常汉语课堂的？
10. 你用什么材料来准备 HSK 考试，教科书，模拟题还是考试大纲？
11. 你的朋友，同学，家长对你参加 HSK 考试有影响吗？你认为他们应该怎样帮助你备考？
12. 你认为 HSK 考试反应了它的初衷“测评学生的语言运用能力和日常交往能力”吗？
13. HSK 哪些方面需要改进？
14. 如果你获得了 HSK 证书，你会具有哪些优势？你认为 HSK 证书对你日后就业或者继续读书有帮助吗？
15. 如果通过某一等级的 HSK 是你取得学位的必要条件，那么你对 HSK 会有不同的态度吗？这样的政策会对你备考产生影响吗？
Appendix E:

Informed Consent for Participation in Survey Research

Shujiao Wang

You are being invited to participate in a research project conducted by Shujiao Wang, a MA student in the Second Language Education program at McGill University, Department of Integrated Studies in Education. This research is being conducted to fulfill the requirements for a master’s thesis, under the supervision of Dr. Carolyn E. Turner, a professor in the above-mentioned program. Here is a brief description of the research and the reasons why I am collecting data:

**Purpose:** The purpose of this study is to investigate and arrive at an accurate description and explanation of learners’ perceptions of a high-stake large-scale standard Chinese proficiency test - *Hanyu Shuiping Kaoshi* (HSK), the washback effects of the HSK on test-takers’ learning, and the factors contributing to the way learners interpret and react to the washback.

**Procedure:** As a participant in this study, you would be invited to fill out a questionnaire which would be taken approximately 15 minutes. The questionnaire will ask you to answer questions about your background, what you think about the HSK test and its impact on your learning process, as well as your views about teaching and learning Chinese. You will have the opportunity to see the questionnaire before signing this consent form.

Any information about you that is obtained as a result of your participation in this research will be kept strictly confidential. It will not be possible to match you with your data in this report. Only my supervisor and I will have access to identifiable data. All data will be destroyed 5 years after the completion of the study. Intended dissemination of this study’s results is master’s thesis and possible later publication in article form or presentation at a refereed conference.

**Conditions of Participation:**

This survey is available both on-line and in paper-and-pencil format and in two languages: English and Chinese. If you would like to complete the on-line version, please contact me at the below email address. If you would prefer to complete the paper-and-pencil version, please sign at the bottom and return the completed survey to the admission
offices. (You will be provided with an envelope in which to place the completed survey, which they can seal to prevent the admissions office staffs from seeing your responses.)

I have read the description of the research project and hereby agree to participate.

Name: _______________________
Signature: ________________________

Telephone: __________________
E-mail: ____________________________

If you have any inquiries or need further information concerning this study, please do not hesitate to contact me by phone at (1) 514-502-2988, by email (shujiao.wang@mail.mcgill.ca) or Dr. Carolyn Turner (Carolyn.turner@mcgill.ca). If you have any questions or concerns regarding your rights or welfare as a participant in this research study, please contact the McGill Ethics Officer at 514-398-6831 or lynda.mcneil@mcgill.ca.

Thank you for your time.
Appendix F:

**Consent for Participation in Interview Research**

I volunteer to participate in a research project conducted by Ms. Shujiao Wang from McGill University. I understand that the project is designed to gather information about the impact of the HSK on learning. I will be one of approximately 15 people being interviewed for this research.

1. My participation in this project is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time without penalty. If I decline to participate or withdraw from the study, no one will be told.
2. I understand that most interviewees will find the discussion interesting and thought-provoking. If, however, I feel uncomfortable in any way during the interview session, I have the right to decline to answer any question or to end the interview.
3. The online skype interview will last approximately 30-45 minutes. Notes will be written during the interview. The interview will be audio-taped. If I don’t want to be taped, I will not be able to participate in the study.
4. I understand that the researcher will not identify me by name in any reports using information obtained from this interview, and that my confidentiality as a participant in this study will remain secure. Subsequent uses of records and data will be subject to standard data use policies which protect the anonymity of individuals and institutions.
5. Faculty and administrators from my campus will neither be present at the interview nor have access to raw notes or transcripts. This precaution will prevent my individual comments from having any negative repercussions.
6. I understand that this research study has been reviewed and approved by the Research Ethics Board (REB) for Studies Involving Human Subjects at the McGill University.
7. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.
8. I have been given a copy of this consent form.

Name: _______________________     Signature: _________________________

Telephone: __________________    E-mail: ____________________________

For further information, please contact:
Ms. Shujiao Wang,
Integrated Studies,
McGill University
Tel: 1-514-502-2988
Email: shujiao.wang@mail.mcgill.ca