Demonstrating and Evaluating Expertise in Communicating in Chinese as a Foreign Language

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Zhini Zeng, M.A.

Graduate Program in East Asian Languages and Literatures

The Ohio State University

2015

Dissertation Committee:

Professor Galal Walker, Advisor

Professor Mari Noda

Professor Xiaobin Jian
Abstract

Effective language learning entails a lifelong experience of performing the target language in meaningful contexts. Assuring the effectiveness of a language program therefore requires us to assess our students’ performances throughout their learning career. Standardized foreign language assessment such as OPI, however, is found problematic in many aspects and cannot meet the needs to recommend foreign language learners for employment. This study therefore proposes a paradigm shift by switching the focus of assessment from “proficiency” to “expertise.”

By bridging the advances in studies of expertise in the field of cognitive psychology and what is emerging in the field of foreign language pedagogy, this dissertation consists of two sub-studies.

The first study operationalized the concept of “expertise” in using Chinese as a foreign language. To identify characteristics working towards expertise, the author shadowed two Americans subjects who speak Chinese at work, interviewed them and their Chinese counterparts working with them on the subject of domain-related Chinese use. The study found non-native expert users of Chinese possessed a highly developed communicative repertoire characterized by colloquial expressions including regional and dialectical usage and by formal expressions including occupational use of jargons, literary allusions, and formulaic polite speech. Moreover, experts are able to manipulate these discourse features skillfully, in an automated and creative manner. The achieved
automaticity can be seen from the expert’s ability to hold new information in short-term memory while continuing the on-going interactions. Moreover, their ability to repurpose the linguistic forms for their own communicative intention marks the achieved flexibility in indexing socio-cultural information through language practices.

The second study explores how a domain-related performance given by a non-native expert user of Chinese is assessed by people from different areas of expertise. The study obtained both quantitative and qualitative evaluations of the two American subjects’ domain-related Chinese uses from three groups of native Chinese raters including domain insiders, non-teacher domain outsiders and Chinese instructors. The study found a substantive discrepancy between evaluations of domain insider and evaluations of Chinese instructors. However, no significant differences in evaluations were found between the two groups of domain-outsider raters. Domain insiders attend to the nature of each domain activity, only recognizing the utterances that fulfill certain functions within the domain. At the same time, a lack of domain knowledge prevents Chinese teachers from assigning meanings to lexical, grammatical, phonological and discursive structures specifically used for a certain domain activity.

Based on these empirical findings, it is recommended that Chinese language program in the United States use performance portfolio assessment with descriptive and transparent criteria to assess advanced level Chinese uses in highly contextualized domain-related performances. Moreover, native Chinese speakers with domain expertise and Chinese instructors should be recruited as raters to work cooperatively to assess foreign language learners’ expertise in communicating in Chinese within one’s domain.
Dedication

Dedicated to my parents, without whose humor and love, this dissertation would not have been possible.
Acknowledgements

Like any dissertation project, this one has been a journey, and many people have helped along the way. I would like to thank each of my committee members for motivating me to expand my research in extremely fruitful directions.

I owe my deepest gratitude to my advisor, Galal Walker, whose unconventional view of pedagogy opened many intellectual and occupational doors for me. Dr. Walker always encouraged me to look beyond teaching Chinese as a foreign language, the benefit of which allowed me to mature into an educator and trainer in every aspect of my field. Dr. Walker is also the first person to inspire me with the concept of “expertise” while at the same time, giving me freedom to explore my research interests. Through his mentorship, I have been able to turn what I love doing into a career.

I am indebted to Professor Mari Noda, who has always been a role model and inspiration, both professionally and personally. Her attentiveness to detail, insightful observations on language use, innovative approach to training and generous yet uncompromising guidance has enabled me to complete this project. Those times we spent discussing Mukoda Kunkio’s good humor and Shigematsu Kiyoshi’s profoundness in her Japanese individual sessions were tremendously enjoyable.

I sincerely thank Professor Xiaobin Jian for those enlightening discussions on the topic of training advanced learners, which has served as the inspiration for this study. He also demonstrated to me his genuineness as a scholar by frankly revealing his confusion and depression to me when he witnessed some advanced learners blindly conforming to
Chinese cultural norms. It was these concerns that inspired me to further pursue my curiosity about the current research topic.

My six years’ studying and working in DEALL have been the most precious experience. I am indebted to all the professors, teachers, and fellow graduate students for their contributions to my study. I especially thank Debbie Knicely for her professional assistance (and countless lighthearted jokes) throughout the course of my study.

Sincere thanks go to the individuals that participated in this research: the two American subjects who graciously allowed me to shadow them at work in China and donated their time to be interviewed and the16 Chinese raters who spent countless hours evaluating many performance samples and answering my interview questions.

I am also grateful to all my students whose achievements are the best motivation for me to advance in my career. Especially, I want to thank Alison Beale, Adam Dehmohseni, Kevin Slaten and Nicholas Holtkamp, for their patience, tolerance and unwavering trust in me when my life as a teacher was still in its embryonic stage.

I want to thank four special individuals: Ying Liu, Junqing Jia, Xin Zhang and Cong Li. Their contribution to this dissertation is undoubtedly significant. It is, however, for those bottles of wine we drank together and the midnight hot-tub conversation at Hocking Hills, that I want to thank them. I am fortunate to have you as my best friends.

Finally, I wish to thank members of my family for their unconditional love and support. It is impossible to name all the individuals who contributed to this project. I am extremely grateful to all of them. Its shortcomings, however, are my own responsibility.
Vita

April 1987 .......................................................... Born, Shanghai, P. R. China

2009 .......................................................... B. A. Teaching Chinese as a Foreign Language, Shanghai International Studies University, Shanghai, China

2011 .......................................................... M.A. Chinese Pedagogy, Department of East Asian Languages and Literatures, the Ohio State University, Columbus, OH

2009-2010 .......................................................... Graduate Research Assistant, Chinese Language Flagship Program, the Ohio State University, Columbus, OH

2010-2015 .......................................................... Graduate Teaching Assistant, Department of East Asian Languages and Literatures, the Ohio State University, Columbus, OH

2015 to present .......................................................... Assistant Professor of Chinese, Department of Modern Languages, University of Mississippi, Oxford, MS
Publications


Fields of Study

Major Field: East Asian Languages and Literatures
   Chinese language pedagogy track

Minor Field: Chinese linguistics
Table of Contents

Abstract .......................................................................................................................................... ii
Acknowledgements .......................................................................................................................... v
Vita ................................................................................................................................................... vii
List of Tables .................................................................................................................................... xiv
List of Figures ................................................................................................................................... xv
Chapter 1: A Paradigm Shift: From Proficiency to Expertise ........................................................... 1
1.1 A Clarification of the Terminology ............................................................................................... 1
   1.1.1 A Historical Overview of the Proficiency Movement ............................................................ 3
   1.1.2 The Development of the Language Competence Models ...................................................... 5
Linguistic theory is primarily concerned with an ideal speaker-listener, in a completely homogenous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky 1965, 3). 6
   1.1.3 Conclusion: Problems Associated with Terminology ............................................................ 10
1.2 A Critical Appraisal of Foreign Language Proficiency Assessment ........................................... 13
   1.2.1 The Undefined “Real-World”: Questioning the Content Validity of the OPI ......................... 14
   1.2.2 The Non-conversation: Questioning the Construct Validity of the OPI ............................... 16
   1.2.3 The Problem with Comparisons to the “Native Speaker”: A Norm-Referenced Test? ........ 16
   1.2.4 The Unrealistic “Global Rating”: Question the Rating Procedure of OPI .... 18
1.3 Shifting the Focus from Competence to Performance and from Proficiency to Expertise .......... 18
   1.3.1 Rampton’s Proposal: The Emergence of the Term “Expertise” in Foreign Language Education .......................................................... 18
1.3.2 Advantages of Adopting the Term Expertise in Foreign Language Education ................................................................. 22
1.3.3 Defining Expertise .................................................................................................................................................. 27
1.3.4 Defining Domain as it Relates to Learning Chinese as a Foreign Language ................................................................. 29
1.3.5 The Path from Competence to Performance: Defining Performance .......... 31

Chapter 2 Expertise in Nonverbal Domains ........................................................................................................... 33
  2.1 Experts Excel Mainly in Their Own Domains ................................................................. 33
  2.2 Experts Perceive Large Meaningful Patterns within Their Domain ......................... 36
  2.3 Experts Solve Problems Quickly and Accurately ............................................................. 39
  2.4 Experts Have Superior Short- and Long-term Memory ................................................. 41
  2.5 Experts Perceive and Represent a Problem on a Deeper Level .................................. 42
  2.6 Experts Spend Much Time Analyzing a Problem Qualitatively .................................. 44
  2.7 Experts Possess Strong Self-monitoring Skills .............................................................. 46
  2.8 Conclusion .................................................................................................................................................. 48

Chapter 3 Identifying expertise in using Chinese as a foreign language ......................... 49
  3.1 Literature Review: Describing Experts in Learning/Using a Foreign Language 49
      3.1.1 Expertise in Foreign Language Learning ................................................................. 50
      3.1.2 Expertise in Foreign Language Use ......................................................................... 53
      3.1.3 Understanding Advanced L2 Learners and Professional-level Chinese Use .................................................................................................................. 57
      3.1.4 Conclusion: Inadequacies in the Previous Literature ........................................ ..... 59
  3.2 Research Questions ......................................................................................................................... 62
  3.3 Research Methods ......................................................................................................................... 64
      3.3.1 The Subjects ......................................................................................................................... 65
      3.3.2 Data Collection ....................................................................................................................... 69
      3.3.2.1 Ethnographic Observation of Working Practices at Chinese Workplaces ......................................................................................................................... 70
3.3.2.2 Stimulated Recall (SR) ................................................................. 72
3.3.2.3 Interviews ................................................................................. 74
3.3.3 Method of Analysis ..................................................................... 76
3.4 Profiles of Two Expert Users of Chinese ........................................ 77
3.4.1 Jack Mayer .................................................................................. 77
3.4.1.1 Chinese Learning Background ................................................. 77
   3.4.1.2 Domain-related Experience .................................................... 78
   3.4.1.3 Motivation for Learning Chinese and Working in China .......... 79
3.4.2 Alan Bahar ................................................................................... 81
   3.4.2.1 Chinese Learning Background ................................................. 81
   3.4.2.2 Domain-related Experience .................................................... 82
   3.4.2.3 Motivation for Learning Chinese and Staying in China .......... 83
3.5 Identifying Expertise: An Analysis of Professional Use of Chinese at Work.... 85
   3.5.1 Experts Possess A Highly Developed Communicative Repertoire .... 85
      3.5.1.1 An Untextbookish Colloquial Style ....................................... 86
      3.5.1.2 Dialectal Expressions and Accents ....................................... 89
      3.5.1.3 Domain-specific Terminology ............................................. 92
      3.5.1.4 A Rich Store of Cultural Knowledge .................................... 94
   3.5.2 Experts Process the Communicative Repertoire in An Automated and
      Creative Manner ........................................................................... 98
      3.5.2.1 Automaticity ..................................................................... 98
      3.5.2.2 Creative Indexing ............................................................... 101
   3.5.3 Experts Assume the Responsibility of Making Improvements .......... 107
3.6 A High Expectation for Multilingual Experts ..................................... 111
Chapter 4 Assessing Expertise in Communicating in Chinese as a Foreign Language . 118
4.1 Literature Review: Alternatives in Assessment .................................... 118
   4.1.1 Portfolio Assessment in Foreign Language Education .................. 120
   4.1.2 360-degree Feedback: Evaluating Professional Performance in Human
      Resource Management (HRM) ....................................................... 124
5.3 Establish Transparent Categories and Criteria for Evaluating Expertise ...... 196
5.4 Apply Modern Technology to Implement the Assessment ........................................... 202
  5.4.1 Create an Online Platform ........................................................................................ 202
  5.4.2 You Tube-type Technology ..................................................................................... 203
  5.4.3 Video-conference Technology .................................................................................. 204
  5.4.4 Digital Data Storage ................................................................................................. 205
5.5 Administer the Assessment thorough a Collaborative Participation of Various
  Institutions .......................................................................................................................... 205
5.6 A New Look at CFL Assessment: Should We Change? Why? What? And How?
  206
  5.6.1 The Pressing Need .................................................................................................... 207
  5.6.2 The Potentiality of A Change .................................................................................. 208
  5.6.3 Areas for Future Research ...................................................................................... 211

Appendix A Recruitment Letter for American Subject ......................................................... 214
Appendix B Subject Recruitment Posting for Expatriate BBS ........................................... 215
Appendix C Recruitment Email for Chinese Interviewees (Identified by the American
  Subjects) ............................................................................................................................. 216
Appendix D Recruitment Email for Chinese Raters .............................................................. 218
Appendix E Interview Questions to the American Subjects ................................................. 220
Appendix F Stimulated Recall Interview Questions to the American Subjects ................. 221
Appendix G Interview Questions to the Chinese Counterparts ........................................... 222
Appendix H Interview Questions to the Chinese Raters ...................................................... 224
Appendix I Evaluation Form Sample ................................................................................... 225
Appendix J Alan Bahar’s Portfolio Evaluation Data ............................................................. 226
Appendix K Jack Mayer’s Portfolio Evaluation Data ............................................................ 237
Bibliography ....................................................................................................................... 248
List of Tables

Table 1. Categories of Language Difficulty ................................................................. 26
Table 2. Characteristics of Expert Learner ................................................................. 51
Table 3. Scoring Rubric Used for Portfolio Assessment in the Current Study ........ 134
Table 4. Rating Criteria Used for Portfolio Assessment in the Current Study ........ 136
Table 5. Descriptions of Performance Samples in Alan’s Portfolio ....................... 139
Table 6. Descriptions of Performance Samples in Jack’s Portfolio ....................... 140
Table 7. Aggregate Portfolio Scores for Alan Bahar .............................................. 142
Table 8. Aggregate Portfolio Scores in Various Domains, Alan Bahar .................. 142
Table 9. Average Portfolio Scores by Rater Types, Alan Bahar ...........................(1)
Table 10. Jack Mayer’s Aggregate Portfolio Scores ................................................. 144
Table 11. Jack Mayer’s Aggregate Portfolio Scores in Different Domains ............ 146
Table 12. Average Portfolio Scores by Rater types, Jack Mayer .............................. 146
Table 13. Scores on Cultural Appropriateness, Performance Sample 2 ............... 155
Table 14. Scores for Alan on Cultural Appropriateness, Performance Sample 4 158
Table 15. Average Scores on Rhetorical Effectiveness by Rater Types, Alan Bahar 163
Table 16. Average Scores on Rhetorical Effectiveness by Rater Types, Jack Mayer 163
Table 17. Selected Comments by Rater Types, Alan Bahar, Sample 2 ............... 167
Table 18. Selected Comments by Rater Types, Alan Bahar, Sample 5 ............... 170
Table 19. Selected Comments by Rater Types, Jack Mayer, Sample 4 ............. 171
Table 20. Average Scores on Domain References by Rater Types, Alan Bahar .... 176
Table 21. Average Scores on Domain References by Rater Types, Jack Mayer .... 176
Table 22. Chinese Instructors’ Comments on Domain References in Performance Sample 6, in Alan’s Portfolio ...........................................................(179
Table 23. Content Selection Framework for CFL Portfolio Assessment ............... 194
Table 24. Sample Rating Criteria for CFL Portfolio Assessment .......................... 198
Table 25. Sample Rating Criteria for CFL Portfolio Assessment (Translated Version) .............................................................................................................................. 200
List of Figures

Figure 1. Similarities and Differences between Models of Communicative Competence......................................................................................................................7
Figure 2. The “Inverted Pyramid”: Components of communicative competence (Savignon, 1983: 46)...........................................................................................................................................9
Figure 3. Components of Communicative Competence (Savignon, 2002: 8).............. 10
Figure 4. Relationships Among Measurement, Testing, and Assessment (Adapted from Bachman 1990)........................................................................................................................................119
Chapter 1: A Paradigm Shift: From Proficiency to Expertise

1.1 A Clarification of the Terminology

The notion of proficiency in foreign language learning has been determined by tests and assessments such as the Interagency Language Roundtable (ILR) Proficiency Scale, or its academic counterpart, the ACTFL proficiency guidelines. The fact that these prevailing scales have been embraced by a few federal agencies, including the U.S. Department of State’s Foreign Service Institute (FSI), the Federal Bureau of Investigation (FBI), the National Security Agency and the Central Intelligence Agency (CIA), vividly reflects the popularity of achieving this mediocre goal in the government-sponsored foreign language training and testing community. However, despite the enormous enthusiasm and funds invested in promoting proficiency standards and developing proficiency assessment tools and trained testers, many insightful scholars in the field of second language theory and testing have vehemently voiced skepticisms and penetrating criticisms of the proficiency-oriented paradigm in their field (Lantolf and Frawley 1985; Johnson 2001).

As the exclusive licensee of the ACTFL Oral Proficiency Interview (OPI), Language Testing International (LTI) defines proficiency as “one’s ability to use language for real world purposes to accomplish real world linguistic tasks, across a wide

---

1 “American Council on the Teaching of Foreign Languages”.
http://www.languagetesting.com/understanding-proficiency
range of topics and settings.” This definition has largely defined the purpose of these tests as long as a set of language descriptors expressed in the ACTFL proficiency guidelines for many years. In the spirit of expansion, Lowe (1985) offers an operational definition of proficiency as “the global rating of general language ability over a wide range of functions and topics at any given level” (p. 16), which can further explain why ACTFL, indicating a global view of language ability, uses the term proficiency rather than competence or ability.

In practice, however, the term proficiency is viewed as synonymous with the term competence. For example, Stern (1983) dedicated an entire chapter to the concept of proficiency in his book *Fundamental concepts of language teaching*, in which he uses both terms interchangeably:

> Among different learners at different stages of learning second language competence or proficiency ranges from zero to native-like proficiency … The native speaker’s ‘competence’, ‘proficiency’, or ‘knowledge of the language’ is a necessary point of reference for the second language proficiency concept used in second language theory (p. 341).

The failure to recognize a difference between the two concepts reveals the ambiguities that exist in their respective definitions and the resulting confusion in their operations. Marysia Johnson (2001) discusses the serious theoretical and practical consequences of this blurring of definitions of the two terms by examining the design of the Test of Spoken English (TSE) and the Speaking Proficiency English Assessment Kit (SPEAK), two of the most popular assessment methods of the speaking ability of international teaching assistants, students and international professionals who are...

---

nonnative speakers of English. Johnson finds that although both tests are based on the communicative language ability model on the surface, giving an impression that the construct of speaking ability has some solid theoretical foundations, in reality, they adopt one global rating that underlies the proficiency theory, without seeing any contradiction in combining it with the communicative competence theory that advocates a non-unitary concept of language ability that attends to its multidimensional feature. This state of confusion directly affects the validity of the tests, as “we do not know what these tests measure – communicative speaking language ability or language speaking proficiency” (Johnson 2001, 174). At the same time, the social consequences of neglecting the current controversy regarding proficiency versus competence can also be seen in some academic institutions’ reluctance to use either test to assess the language ability of their international teaching assistants.

In order to make a distinction between the two concepts, we start a historical review of proficiency and competence, by inspecting the former in the so-called “proficiency movement” that had begun to impact the field in the late 1970s and early 1980s while examining the latter one in its Chomskyan formulation.

1.1.1 A Historical Overview of the Proficiency Movement

Since the late 1980s, proficiency has become a movement within the second-language profession, with the intent to introduce a national metric to define achievement in second language instruction. The long history of efforts aiming at assessing second language competence in the U.S. government began in the 1950s, when the United States’ lack of preparation in foreign languages was recognized as a serious problem
during the Korean War. Therefore, both the proficiency rating scale and the proficiency tests based on it were born out of a practical necessity. In 1973, the Interagency Language Roundtable (ILR) assumed the responsibility for developing a standardized version of the scale and eventually refined the government’s definitions of proficiency in the four language skills, known as the ILR scale\(^3\). The ILR scale was soon adopted for hiring, promotion and salary bonus purposes in certain federal agencies. For example, in US Department of the Navy, the Foreign Language Proficiency Bonus (FLPB) requires a 2/2 qualifying score (a 2 in listening and a 2 in reading) for receiving a minimum monthly pay of $200 and 3/3/3 (plus a 3 for speaking) or 4/4 for the highest pay of $500 every month\(^4\). Interest in the ILR scale continued to develop outside of government in the late 1970s with a number of oral proficiency testing workshops held to familiarize academics with its main assessment instrument, the oral proficiency interview (OPI). In 1981, in order to allow more students who have spent many hours studying foreign languages to register their progress by using a standardized test, the ACTFL and the Educational Testing Service (ETS) initiated the effort to develop a scholastic counterpart of the language proficiency assessment.

In the meantime, the spread of the ILR scale has not been validated in any research program examining its fundamental assumptions about language use and development. The only research to study the quality of the test and its rating scale that had attracted most academic investment at that time, was the investigation of inter-rater

\(^3\) “How did the language proficiency scale get started?” Interagency Language Roundtable. http://www.govtirl.org/skills/

reliability, which was consistently reported to be high (Chalhoub-Deville & Fulcher, 2003). However, it is precisely those positive reports of the high inter-rater reliability that quickly lead to the ACTFL’s uncritically adopting the FSI interview procedure and their belief that the interviews and rating scales in themselves possessed psychological reality. As a result, the ACTFL Provisional Proficiency Guidelines appeared in 1982 (ACTFL, 1982). The complete Guidelines, published in 1986, has experienced several rounds of revision, embracing the revision of Speaking in 1999 and 2012 and the revision of writing in 2001.

The latest edition of the ACTFL Proficiency Guidelines identifies five major levels of proficiency for each skill: Distinguished, Superior, Advanced, Intermediate, and Novice. Different levels of proficiency are described in a hierarchical sequence of performance ranges. However, although goals stated as to “update descriptions to reflect communication in the 21st century” and to “ensure usability for lay population” (Swender 2012: 13) are highlighted for the latest version of the Guidelines, it remains subject to some previous criticisms. We will review these in Section 1.1.2.

1.1.2 The Development of the Language Competence Models

The notion of competence can be traced back to Ferdinand de Saussure’s (1959) study of structural linguistics, in which he differentiates between \textit{la langue}, a system of signs, and \textit{la parole}, the realization of the system in a particular situation. In 1965, Chomsky published the definition of linguistic competence, in which he distinguished it from linguistic performance. In theory, the two concepts parallel the Saussurian’s \textit{la langue} and \textit{la parole}, in that Chomsky’s linguistic competence is realized in an ideal
speaker’s tacit knowledge of grammatical structures, which was introduced in the commonly quoted paragraph below:

Linguistic theory is primarily concerned with an ideal speaker-listener, in a completely homogenous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky 1965, 3).

Chomsky views the linguistic performance as the actual use of language in practical situations in a reductionist way by confining all cases of variation to “performance” which are “fairly degenerate in quality” (Chomsky 1965, 31) as they are filled with errors. Although Chomsky is arguably the first linguist to define the term “competence”, he is not considered the originator of the term as it is commonly used in everyday language (Spolsky 1989). The Chomskyan idea of competence and performance was soon criticized by scholars, such as Hymes (1972), Canale and Swain (1980) and Bachman (1990), who took a more situational or sociolinguistic approach. This debate can be regarded as the starting point of the development of multiple models of communicative competence.

Reviews of various models of communicative competence, which address the construct of language ability and language assessment (Llurda 2000, Johnson 2001, Chen 2011), can be found in numerous previous studies. Therefore, instead of reviewing these models chronologically, I only address the strengths and weaknesses reflected in the development of various models at different stages. Figure 1.1 is a graphical representation of the similarities and differences between three major models of
communicative competence, developed by Hymes (1972), Canale (1983) and Bachman (1990).

Figure 1. Similarities and Differences between Models of Communicative Competence

First, these three models are antithetical to Chomsky’s view of “an ideal speaker-listener” seeing it as a concept only existing in the theory. Hymes was the first scholar to make a distinction between performance as “the actual use of language in concrete situations” (Chomsky 1965, 4) and the underlying modes and rules of performance, which he calls “ability for use” (Hymes 1972, 282). Later, followers of Hymes attempted to relate the complexity underlying Hymes’ broad and ambiguous concept of “ability for use,” to real world situations, either implicitly or explicitly. For example, although the
Canale and Swain model (1980) and a revised model later proposed by Canale (1983) intentionally eschewed the term “ability for use” in their new models, the newly added “strategic competence” (Canale and Swain 1980, Canale 1983) and “discourse competence” (Canale 1983) are actually included as implicit ways to refine the elements of the original “ability for use”. Similarly, in Bachman’s Communicative Language Ability (CLA) model, the strategic competence refers to cognitive skills that mediate language use, such as assessing, planning, and executing, which are essential to achieving communication goals. Bachman’s inclusion of strategic competence as a critical component of communicative competence can be viewed as an explicit attempt to acknowledge the importance of Hyme’s “ability for use” concept.

Second, each model presented in Figure 1.1 can be seen as an improved and expanded version of its predecessor. However, the more detailed the division into subcomponents of a model is, the more abstract and technical it becomes in terms of practical implementation. For example, Johnson (2001) criticized Bachman’s model for separating cohesion from coherence by placing cohesion under organizational context and including coherence under pragmatic competence. At this time, although scholars began to focus on the multidimensional aspects of the language largely as a result of the expansion of subcomponents in the improved models, they did not offer useful insights into how these components interact with each other and how they could be assessed separately in foreign language education.

5 In Figure 1.1, a dashed arrow is used to indicate that although Bachman’s strategic competence derives from Canale and Swain’s model, it actually drastically differs from it. While Canale and Swain define “strategic competence” as a coping mechanism for the breakdown in communication, Bachman actually views “strategic competence” as a mechanism responsible for interaction.
Third, neither Canale nor Bachman were able to sufficiently address the notion of social context in their models, in which expressions are used as prompts “to construct meanings by working with processes we already know” (Turner 1991, 206). All these models view communication as a form of information-processing and information-transmission, in which context remains separated from other components. For example, Bachman’s definition of social context, which was previously defined by stable features of context including register, dialect, and variety, presents a static view of interaction.

It was not until the 21st century that Sandra Savignon (2002) attempted to address some of the problems discussed above in her Communicative Language Teaching (CLT) classroom model, shown in Figure 3. Savignon adapted this classroom model from her “inverted pyramid” model (1983) as shown in Figure 2, trying to illustrate the interactions among components of communicative competence in an increasingly wide range of communication contexts. Savignon’s improved model (2002) brings “context” to the forefront in the sense that the development of any kind of competence under this model requires an understanding of the social context in which language is used.

Figure 2. The “Inverted Pyramid”: Components of communicative competence (Savignon, 1983: 46)
However, similar to the three models presented in Figure 1, which have never been validated by empirical research, Savignon was not able to provide any empirical studies, either for her model proposed in 1983 or the one proposed in 2002 to support her hypothesis. The lack of an empirical basis, unfortunately, diminishes the model’s heuristic value. More importantly, it is the lack of empirical studies to support her models’ complexity that prevents her from further distinguishing between proficiency and competence. According to Savignon, “Language proficiency is communicative competence and should be defined and evaluated as such” (1983, 246).

1.1.3 Conclusion: Problems Associated with Terminology

To sum up, both competence and proficiency have been subject to various, sometimes limited, and occasionally conflicting interpretations, and are therefore used interchangeably. However, the present study distinguishes between the theoretical

Figure 3. Components of Communicative Competence (Savignon, 2002: 8)
development and the academic contexts of proficiency and competence. Here we not only recognize their respective differences, but also summarize the conceptual issues and definitional problems associated with the two terms revealed in the previous scholarship.

Originated in its Chomskyan formulation, the definition of “competence” has experienced countless revisions or, more accurately, expansions. However, despite these efforts, it is still open to debate whether these changes have contributed positively or negatively to a more precise understanding of this complex concept. This study regards those later attempts to incorporate aspects such as “skill”, “ability” or “performance” as an unsuccessful endeavor, which has confused the meaning from then on. As Taylor (1988) suggests, the concept of competence is an “investigatory and descriptive device, valid in a certain domain, but causing many difficulties outside it” (165), mainly because these variations and changes were beyond the scope of Chomsky’s analysis. Therefore, a scholar who attempts to apply “competence” to any given reality should expect some potential confusion. Moreover, even if we begin to view “competence” outside the Chomskyan tradition as a multidimensional feature of language ability, an increasing number of its subcomponents may have just made the implementation of this concept more abstract and technical, and therefore more difficult.

Secondly, having been completely dependent on the ACTFL Proficiency Guidelines, the notion of “proficiency”, as Lantolf and Frawley (1985) criticize, has become “what proficiency tests measure” (185). Although the concept of proficiency does entail some aspects of communicative competence, which are displayed and rated in oral proficiency interviews, the Proficiency Guidelines and the manner in which they are applied fail to address the underlying notion of communicative competence in a
meaningful way. The global rating adopted by the proficiency test makes it impossible to identify and measure the abilities that are subsumed under the nebulous heading of “oral proficiency.” Moreover, the very restricted form of “the interview,” although it allows for a thorough assessment of grammatical competence, can never fully measure other subcomponents of language competence, such as discourse because those being interviewed are always deprived of chances to initiate or guide a conversation or to react to the interviewer’s responses.

Thirdly, there are no empirical studies to support the underlying principles of proficiency and competence and their implied relations to foreign language syllabi and curricula. As mentioned above, the notion of proficiency is completely dependent on the ACTFL Proficiency Guidelines, which have no empirical foundation. Although many scholars have stated that “the levels of proficiency are ranked in a progressive order that points to a sequencing of abilities in which first things must come first,” (Heilenman and Kaplan 1985, 61) in a proficiency-based curriculum, a validated “progressive order” has in fact never been identified and the validity of a more or less universal progression that applies to all languages and cultures has not been explained. More importantly, the notion of “competence,” which was originally formulated through the Chomskyan tradition aimed at defining an innate system that could only exist in an ideal monolingual situation, which has been recognized as problematic in many ways (Paikeday 1985; Bruner 1990; Tomasello 1999; Sinha 2010). Applying such a monolingual concept to foreign language education while providing no empirical validations wakens the concept into a lesser heuristic for all types of interpretation, many of which are conflicting. Simply, it is the
lack of empirical studies to support the complexity of the competence models and justify the proficiency test’s validity that have resulted in the confusion of these two concepts.

1.2 A Critical Appraisal of Foreign Language Proficiency Assessment

While the proficiency approach retains currency in the field of foreign language teaching and testing, its inherent shortcoming as either a specific or generic term are illustrated extensively in studies that detail the general skepticism surrounding proficiency assessments. Along with the criticism associated with this concept, scholars have also voiced their condemnation over the institutionalization of the OPI (oral proficiency interview) on which the theory of proficiency is completely dependent, as it lacks the support of any empirical study. According to Freed:

A major issue that now faces the foreign language teaching profession is how to build on and preserve the positive contributions of the proficiency movement while at the same time recognizing the need to rethink, revise, or redefine certain acknowledged limitations of the guidelines and testing procedures as they are currently conceived (Freed 1989, 52).

In agreement with Freed, Bachman (1990) points out that the primary problems in measuring speaking ability through an oral interview procedure are not questions of efficiency or reliability but rather are associated with examining the validity of the interview ratings as measures of competence in speaking and the application of such ratings. The questionable validity of proficiency assessment of foreign language, i.e., the OPI, is examined in the following sections.
1.2.1 The Undefined “Real-World”: Questioning the Content Validity of the OPI

According to ACTFL, the content of the OPI test is sampled from a “real-world situation” for a “real-life purpose” (ACTFL, 2012). However, in practice these seemingly intuitive terms invite great controversy regarding the test’s content validity. The domain from which the content is sampled is too vast to be realistic. For example, the “real world” of business differs from the “real world” of academia in numerous ways. A case discovered in this study illustrates that persuasive strategies which enable a company in the gaming industry to sign numerous contracts fail to garner the same success in the field of art brokerage. As Johnson argues, “The claim that the OPI maintains its content validity by including in each interview questions that test real-life situations is undermined by the fact that the basic requirement for establishing content validity is not satisfied” (2003, 30).

Recalling my experience of participating in a Japanese language OPI as an interview subject, I can still remember how the ambiguity of the test context confused me and unduly influenced my performance. During the latter part of the interview, the interviewer asked me to discuss a news story that I had recently watched. It was several weeks after the 2013 Nobel Prize for Literature was announced and awarded. I, therefore, told the interviewer about the news that the Japanese novelist Haruki Murakami, who topped everyone’s list of favorites that year, again was snubbed after being nominated several times. After I briefly shared this piece of news, the interviewer asked me what I thought of the Nobel Prize itself. Although I was a little bit surprised about this sudden switch of conversational tone, I still responded to his question positively by explaining how I liked the idea that the prize can recognize exceptional people all over world from a
variety of fields. Immediately after answering his question, the interviewer expressed his strong objection to the Nobel Prize, saying that it puts too much emphasis on winning or losing and, therefore, tends to discourage those who are pursuing excellence in their respective fields. He asked me if I wanted to defend my opinion. Due to fact that he was an authority figure (a middle-aged Japanese male interviewer who had the power to decide whether I passed or failed), my first instinct was to avoid engaging in any serious discussion, let alone debate, with him. According to my understanding of the Japanese culture, I knew I should display my respect and even obedience by minimizing my disagreement. However, my knowledge about the ACTFL Proficiency Guidelines at that time indicated that the instructor was trying to elicit a persuasive discourse from me. So from that perspective, I should not only respond, but also attempt to eloquently defend my point of view. After the interview, the interviewer acknowledged my hesitancy to debate that topic with him, but he just simply suggested that I practice more to enhance my persuasive techniques. When I asked him what role I was expected to play when answering that question, he seemed quite confused and responded, “You are just yourself.”

One may still defend the OPI by saying that my experience with this interviewer may not be indicative of others’ experiences and the fact that the interviewer was still in training to improve his elicitation skills. However, the interviewer’s confusion regarding my question about our respective roles throughout the interview revealed that the context of the test was very loosely defined. Obviously, it is not something an OPI tester should be concerned about during the process. Such experiences, however, call into question the kind of “reality” that is represented and measured in the OPI. The truth can be found in
Lantolf and Frawley’s (1985) comment, “If the Guidelines measure reality by definition, they have construed a reality and therefore are the prescriptions of a theorist deciding what speakers ought to do” (p. 342).

1.2.2 The Non-conversation: Questioning the Construct Validity of the OPI

The experience I recounted also reveals another shortcoming of the current OPI concerning the construct validity rooted in its self-claimed “natural conversation.” Van Lier was the first scholar to underscore the discourse structure of the OPI by pointing out, “The interviewer has a plan and conducts and controls the interview largely according to that plan” (van Lier 1989, 496). Following van Lier and other scholars’ hypotheses (Young and Milanovic 1992; Young 1995) that OPI does not accurately measure speaking ability, Johnson uses discourse analysis to evaluate 35 OPI telephone interview samples at a variety of language levels. She concludes that OPI scenarios are not realistic conversations as the ACTFL claims and that each stage of the OPI is a distinct genre. For example, she argues that the warm-up stage resembles a sociological interview, while the probing stage is a controlled interview. Moreover, the new speech event created within the OPI does not correspond to any authentic speech event in daily life.

1.2.3 The Problem with Comparisons to the “Native Speaker”: A Norm-Referenced Test?

Popham (1981) distinguishes between a criterion-referenced test and a norm-referenced test in the sense that the former “is used to ascertain an individual’s status with respect to a well-defined behavioral domain,” while the latter “is used to ascertain
an individual’s status with respect to the performance of other individuals on that test” (p. 28). According to this definition, the OPI is a norm-referenced test because it creates an artificial yardstick, “a well-educated native speaker,” to whom the test candidate’s performance is constantly compared. Although the idea of “native-like proficiency” is essentially ignored by Chomskyan scholars, who usually claim that the native speaker is the authoritative source of grammaticality judgments, the same question does trouble many foreign language pedagogues.

The term “native speaker” in English seems to be a buzzword among linguists to the extent that it appears in the dictionary. For example, the Collins Dictionary (1979) defines “native speaker” as “a person having a specified native language”. According to Longman, a native speaker is “someone who has learned a particular language as their first language, rather than as a foreign language”6. However, is every single person who grew up speaking English equally competent in its usage? Indeed, we all know “native speakers” who communicate poorly in social settings, struggle to speak fluently, and whose writing is filled with errors that prevent comprehension. However, many of us have savored the extraordinary literary achievements of non-native speakers such as Conrad or Nabokov.

As Bachman (1990) claims, native speakers simply do not exist- only types of native speakers exist. It is useless for the OPI to use the term “native-like proficiency” as an artificial yardstick as our intuition still clouds a clear understand as to what constitutes a “native speaker.”

1.2.4 The Unrealistic “Global Rating”: Question the Rating Procedure of OPI

Numerous previous studies reinforce the hypothesis that a language rating scale be multidimensional and foreign language communication be evaluated from multiple dimensions (Palmer 1973; Hadden 1983; Chen 2011). Although the OPI includes in its design and rating procedures a functional trisection composed of function, content, and accuracy, these three components are not rated separately as the OPI claims, to provide a global rating. Pedagogically speaking, this global rating cannot provide helpful information to instructors, as the same score does not necessarily represent the same quality of performance. For example, a candidate who is capable of being functionally persuasive but has relatively poor listening comprehension skills may receive the same score as someone who has the opposite situation. However, to an employer who is looking for a bilingual salesperson, this failure to demonstrate different dimensions of language competence can render the test results unhelpful as the single global score does not indicate which dimension plays a more important role than the other in determining that global score.

1.3 Shifting the Focus from Competence to Performance and from Proficiency to Expertise

1.3.1 Rampton’s Proposal: The Emergence of the Term “Expertise” in Foreign Language Education

From our discussion in Section 1.1, it is obvious that the meaning of terms such as “proficiency” and “competence” is fuzzy, which complicates their implementation as revealed in 1.2. Therefore, it is time to present more useful and realistic terms to clarify
the goal of our endeavor. Since the late 1980s, many scholars in the field of language
teaching have begun to explore alternative terms, such as “proficient user of the
language,” suggested by Paikeday (1985); “multi-competent speaker,” put forward by
Cook (1991), and “competent language user,” proposed by Lee (2005). Adopting a
similarly critical position, Rampton (1990) argues for replacing the term “native-like
competence” with that of “expertise” by shifting the emphasis from “who you are,” as
reflected in native competence, to “what you know”. Rampton further elucidates this
difference in the quote below:

If native-speaker competence is used to set targets and define proficiency, the
learner is left playing a game in which the goal-posts are being perpetually moved
by people they cannot often challenge. But if you talk about expertise, then you
commit yourself to specifying much more closely the body of knowledge that
students have to aim at (1990, 90).

By proposing the new term, Rampton rejects the innateness underlying the native-
like competence and recognizes the domain-specific feature of expertise as both
“relative” and “partial” (1990, 99).

In fact, Rampton does not propose “expertise” as the only single term to take over
the role of “native-like competence” as he believes that the term “expertise” does not
adequately encompass the role that language plays as a symbol of social group
identification. He therefore adds the term “language loyalty,” which consists of language
affiliation and inheritance as well as expertise to address the symbolic value of language.
Rampton perceives that these terms as draw attention to language learning as a social
activity and prevent speakers from becoming trapped in language categories. In this
sense, learning a new language requires the skills to to “manage continuity, change, and
the relationship between social groups” (Rampton 1990, 100).
Although Rampton wisely perceives the necessity of initiating and maintaining the social relationships involved in learning a foreign language, I will argue that the term “expertise” alone also informs us about social group identification in the sense that an expert is a recognized member of a certain domain in which group identification must be constantly negotiated and that domain is contained within an identifiable component of a culture. Foreign language learners who attempt to establish themselves in a certain group within the target culture not only require extensive experience and competence with the language itself, but also a sophisticated level of cultural negotiation to initiate and to maintain the identification. Therefore, following Rampton in this dissertation, the previously widely held concept of foreign language assessment “native-like proficiency” and “native-like competence” will be replaced with the term “expertise;” however we will also assume that the concept of “language loyalty” is a large component.

Rampton proposed his argument for replacing “native speaker of a language” with “expert user of language” as early as 1990. However, unfortunately, this did not significantly impact the field of foreign language pedagogy as the ill-defined term “native-like proficiency” still prevails within foreign language assessment criteria and job advertisements recruiting foreign language teachers. Although expertise studies in the field of cognitive psychology were well established by 1990s, Rampton’s article does not sufficiently address the underlying characteristics of expertise, which may also explain why his proposal has not been fully accepted in the field of applied linguistics and foreign language education afterwards. Moreover, Rampton also failed to relate the significance of “expertise” to bringing the need to recommend foreign language learners for employment. This need requires communicating more than a global rating that is not
understood by people outside the foreign language teaching community, and demands that pedagogues focus beyond the formal language program.

Rampton listed five advantages of the term “expertise” over “nativeness”\(^7\) to explain why such a shift is necessary:

1) Although they often do, experts do not have to feel close to what they know a lot about. Expertise is different from identification.
2) Expertise is learned, not fixed or innate.
3) Expertise is relative. One person’s expert is another person’s fool.
4) Expertise is partial. People can be expert in several fields, but they are never omniscient.
5) To achieve expertise, one goes through processes of certification, in which one is judged by other people. Their standards of assessment can be reviewed and disputed. There is also a healthy tradition of challenging ‘experts’.

(Rampton 1990, 98-99)

Advantages three and four particularly address the central tenet of the expertise theory that the superior performance of experts is often domain specific, and the ability to transfer these performances outside their narrow area of expertise is surprisingly limited. However, what does this feature of expertise say about foreign language learning? Why does it make the term expertise more desirable than “proficiency” or “competence”? In addition, Rampton does not include the fact that expertise is acquired through learning and is maintained by processes of assessment and certification (advantage 2 and 5).

However, what concerns foreign language pedagogues most are the features that are able to inform us about foreign language teaching and assessment. Unfortunately, these are not well elaborated upon in Rampton’s article.

---

\(^7\) As reviewed in 1.2, “native” is always used as the modifier prior to “competence” or “proficiency”, which can be seen in many common combinations, such as “native-like proficiency”, “native-like competence”, “near-native proficiency” or “near-native competence”.
Coincidentally, in 1995, the term expertise reappeared in the field of teaching Chinese as a foreign language (TCFL). Two TCFL experts, Galal Walker and Scott McGinnis (1995) proposed that “The only reasonable goal for learners and teachers of LCTLs is to develop expertise in their respective career fields”. By proposing such a goal, it becomes clear that they define an expert learner as someone who has demonstrated success interacting with members of a particular society. However, it was until the 21st century that these predecessors’ proposals were fully recognized and the field began to borrow more theoretical insights from the expertise research tradition to analyze learning and teaching expertise in foreign language pedagogy. These ideas will be expanded upon in Section 3.1.1 during a discussion of Keith Johnson’s book *Expertise in Second Language Learning and Teaching*.

1.3.2 Advantages of Adopting the Term Expertise in Foreign Language Education

To remedy the oversight mentioned above in Rampton’s five advantages of the term “expertise” over “nativeness,” this study summarizes three advantages of using expertise as a preferred term to describe the ability of using language for communication.

First, the concept of domain as it relates to the theory of expertise reflects a realistic view of what a professional who is a native speaker is able to achieve in his or her native culture, which therefore, also sets a more attainable and realistic goal for foreign language learners. Studies on expertise in non-linguistic domains point out that the superior performances of experts are very domain specific, and the transfer of performances between even two similar areas of expertise is still limited. Chase and Smion’s (1973) theoretical account of chess expertise illustrates how the grand masters
rapidly calculate the best possible moves from long-term memory by chunking recognition of configurations of chess pieces as cues, while failing to demonstrate the same superior performance in basic memory tests. In *The Sports Gene*, David Epstein (2013) recalls a story about a female underhand softball pitcher who consistently strikes out the world’s best professional baseball (hardball) hitters to once again illustrate how the transfer of expertise between domains--specifically softball and baseball, two intuitively similar domains--is surprisingly limited.

By shifting our focus from developing proficiency to developing expertise in foreign language education, we can avoid many of the distractions associated with a traditional holistic approach to learning Chinese culture and instead focus our limited resources on the process of guiding learners into their respective domains within the target culture. By realizing that it is impossible to master the immensity of an entire language and learn every nuance of its culture, we are led to develop levels of curriculum designed to provide learners with a venue in which to begin to pursue expertise during the limited instructional hours within a program. More importantly, we can equip students with useful strategies to sustain lifelong learning outside formal education environments.

Second, obtaining the level of expertise must produce concrete, observable and demonstrable results through performance, which is the observable and measurable manifestations of expertise. Although superior performance in some nonverbal domains can be established more easily through games in the domain of sports or world-class concerts in the domain of music, in some linguistic domains, such as using foreign language at work, expert performances cannot be assessed as easily as counting matches
won, home runs hit, or the number of one’s well-attended solo concerts. Shifting the goal of the foreign language program from proficiency to expertise, therefore, requires courses to carefully select performances from the target culture as the basic pedagogical unit in teaching and assessment. Just as the Elo rating system\(^8\) is a recognized metric to identify chess expertise, the field of foreign language education also must agree upon some kind of objective measure of performance to identify expertise. Moreover, determining the process of identifying expertise in various domains can inform foreign language skill assessment in terms of who should be included as authorities to assess expert performance and what criteria should be relied on to identify expert use of language.

Third, as an inescapable dimension of all human activity, time and the use of time have become central to the discussion on the acquisition and retention of expertise. This is also a crucial element in foreign language learning. As the earlier genetic explanation of expert performance is giving way to the pervasive belief in the major role of practice in attaining eminence (Ericsson 1996; Glaser and Chi 1988), scholars have found that the investment of time, which applies to many different domains, leads to consistent patterns of performance levels over time. It is reported that all performers, even the most talented ones need approximately ten years of intense involvement before they reach an exceptional level in established sports, science, and arts (Ericsson 2008).

In 1973, the Foreign Service Institute (FSI) categorized foreign languages into four groups and estimated the number of hours of training required for reaching each expected level in each category of foreign language. According to FSI’s calculation,

\(^8\) Named after its creator Arpad Elo, in the domain of chess, Elo rating is a metric used to calculate the relative skill levels of chess players. Expert in chess is defined as a range of chess ratings approximately two to three standard deviations (200 rating points) above the mean (1600 rating points) and five to six standard deviations above the mean of chess players starting to play in chess tournaments.
presented in Table 1.1, approximately 1,320 contact hours are needed for learners with average aptitude for Chinese, a Category IV language, to reach the second proficiency level on the Interagency Language Roundtable Scale, which is equivalent to the ACTFL [2012] “Advanced” category. Unfortunately, no language program can dedicate this amount of time to its students. Also, this time challenge is further intensified for American learners of East Asian languages because these languages differ significantly from their base language and culture in terms of orthography, linguistic structure, vocabulary, and expected behavior.

Of course, although the picture appears clearer in the field of foreign language education, these figures are woefully out of date. In other words, the 1,320 hours cited here probably reflect very unusual learning conditions in the 1960’s and 1970’s. However, the crucial role that “time” plays in foreign language education should suffice to bolster that the concept of “expertise” is a more realistic goal for foreign language pedagogues. The number of years it takes to attain expert performance within a particular domain and the characteristics of deliberate practice underlying the domain experiences provide us with a solid framework for understanding the time challenge imposed on foreign language learning. Establishing a realistic goal for college-level foreign language education should help teachers to develop effective approaches for maximizing the limited amount of time inherent in any given language program.
### Table 1. Categories of Language Difficulty

<table>
<thead>
<tr>
<th>Group</th>
<th>Languages Offered</th>
<th>Length of Training</th>
<th>Minimum</th>
<th>Average</th>
<th>Superior</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP I: Afrikaans, Danish, Dutch, French, German, Haitian, Creole, Italian, Norwegian, Portuguese, Romanian, Spanish, Swahili, Swedish</td>
<td>8 weeks (240 hours)</td>
<td>1</td>
<td>1/1+</td>
<td>1+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 weeks (480 hours)</td>
<td>1+</td>
<td>2</td>
<td>2+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 weeks (720 hours)</td>
<td>2</td>
<td>2+</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>GROUP II: Bulgarian, Dari, Farsi, Greek, Hindi, Indonesian, Malay, Urdu</td>
<td>16 weeks (480 hours)</td>
<td>1</td>
<td>1/1+</td>
<td>1+2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 weeks (720 hours)</td>
<td>1+</td>
<td>2</td>
<td>2+3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 weeks (1320 hours)</td>
<td>2/2+</td>
<td>2+/3</td>
<td>3/3+</td>
<td></td>
</tr>
<tr>
<td>GROUP III: Amharic, Bengali, Burmese, Czech, Finnish, Hebrew, Hungarian, Khmer (Cambodian), Lao, Nepal, Filipino, Polish, Russian, Serbo-Croatian, Sinhala, Thai, Tamil, Turkish, Vietnamese</td>
<td>16 weeks (480 hours)</td>
<td>0</td>
<td>1</td>
<td>1/1+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 weeks (720 hours)</td>
<td>1+</td>
<td>2</td>
<td>2/2+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 weeks (1320 hours)</td>
<td>2+</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROUP IV: Arabic, Chinese, Japanese, Korean</td>
<td>16 weeks (480 hours)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 weeks (720 hours)</td>
<td>1</td>
<td>1+</td>
<td>1+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>44 weeks (1320 hours)</td>
<td>1+</td>
<td>2</td>
<td>2+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80-92 weeks (2400-2760 hours)</td>
<td>2+</td>
<td>3</td>
<td>3+</td>
<td></td>
</tr>
</tbody>
</table>

* The number of hours is the theoretical maximum at 30 hours a week

(Foreign Service Institute, 1973)

However, scholars studying expertise have pointed out that continued improvements in achievement are not automatic results extended experience. As Ericsson (1996, 3) argues, “a long period of preparatory education followed by an apprenticeship is required” to attain the expert level of performance in most professional domains of expertise. By focusing our attention on developing expertise, we are reminded that the ultimate responsibility for becoming an expert always lies with the individual learner. Becoming an expert in a language very different from one’s own, such as Chinese, is

26
always a lifelong endeavor, requiring many years of deliberate practice. According to their extensive studies of expert musicians, deliberate practice includes training activities involving a well-defined task with an appropriate level of difficulty for the particular individual, informative feedback, and opportunities for repetition and corrections of errors. The type of deliberate practice may vary from domain to domain. It can be seen in the expert violinists’ solitary practice (Ericsson et al. 1993), the National Spelling Bee contestant’s dedication to studying spelling and word origins (Duckworth et al. 2011), and wrestlers’ mat work with teammates and coaches (Starkes et al. 1996). Therefore, studying the notion of expertise in foreign language education will motivate us to further explore what features correspond to the “deliberate practice” of qualifying drills and exercises in our field.

The field of foreign language education has a strong tradition of seeking parallels with first language acquisition (Johnson 2005), which means that the suggestion to analyze findings in non-linguistic domains and consider relating them to language learning and teaching not well-received until the 21st century. However, this set of views has begun to be supplanted by ones that permit and indeed invite cross-domain comparisons. Considering this background, it is important for foreign language pedagogues to propose an expertise-oriented foreign language assessment program by utilizing concepts that spring from general cognitive and learning theories.

1.3.3 Defining Expertise

In this dissertation, expertise is defined as characteristics, skills, and knowledge that distinguish experts from novices and less experienced professionals. Expert
performance therefore, refers to types of superior reproducible performances of representative tasks that can capture the essence of the respective domains. With the growing impetus to study expertise in order to create machines that possess artificial intelligence, our understanding of the nature of expertise has increased as well. While popular domains for early study of expertise centered on chess and general problem solving, increased attention has been given to fields such as the arts including music, dance and acting (Sloboda 1991; Solso and Dallop 1995; Noice and Noice 1997), literacy education (Scardamalia and Bereiter 1991; Wagner and Stanovich 1996) and sports (Hodges, Starkes and MacMahon 2006; Epstein 2013). With an expanded understanding of expertise, in the late 80s and early 90s, a number of studies were done which attempted to identify characteristics of experts that are generalizable across various domains. For example, Glaser and Chi identify seven characteristics of experts that can relate to any field:

1. Experts excel mainly in their own domain
2. Experts perceive large meaningful patterns in their domain
3. Experts are fast; they are faster than novices at performing the skills of their domain, and they quickly solve problems with little error
4. Experts have superior short- and long-term memories
5. Experts see and represent a problem in their domain at a deeper level than novices; novices tend to represent a problem at a superficial level
6. Experts spend a great deal of time analyzing a problem qualitatively
7. Experts have strong monitoring skills

(Glaser and Chi 1988, xvii-xx)

What each of these characteristics refers to across a wide range of domains is reviewed in Chapter Two.
1.3.4 Defining Domain as it Relates to Learning Chinese as a Foreign Language

The notion of domain is an essential concept in cognitive psychology’s approach to studying expertise. As previously stated in Section 1.3, a central tenet of expertise theory is that the superior performance of experts is often domain specific, and the capability to extend their performance beyond a narrow area of expertise is surprisingly limited. However, the definition of domain in current expertise studies has not received the scrutiny it deserves in the sense that it has simply been equated with an occupation, such as baseball player or pianist. In other words, the definition of domain seems too intuitive to be appreciated. Although this might be true and has not posed a problem to scholars studying expertise in non-linguistic fields, the lack of a definition of domain may bring great challenges to foreign language pedagogues when we attempt to incorporate the idea of “being domain-specific” into curriculum and assessment.

Therefore, this study proposes to define domain as a restricted range of social interactions within a culture, bound by commonly shared knowledge and motivated by identifiable and shareable goals. Each domain includes a series of domain-specific activities. The structural properties of each activity serve as the source of inferences, predisposing members of a domain to see utterances as fulfilling certain functions. As Wittgenstein suggests, having a grasp of the meaning of utterances involves knowing the nature of the activity in which the utterances play a role (Wittgenstein, 1958: 23).

Under this framework, every student moving from elementary to advanced professional proficiency in the target language proceeds through a process of narrowing one’s domain and developing expertise within it. For example, in the US, undergraduate students’ choice of academic major may be their first step to identifying a domain they
wish to study in the target culture. Having such default domains as back-ups can help
students establish themselves in groups related to their domain in the new culture.
However, this alone is not sufficient for achieving successful group membership in the
target culture. The process in moving from peripheral participation to becoming a
recognized member of the domain group requires identification and initiation. Graduate
students and professionals going to the target country have at least one clearly
identifiable academic or professional domain. However, the process of establishing
oneself in a new group requires extensive experience and a sophisticated level of cultural
negotiation. Spending time in the country where the target language is spoken provides
an excellent opportunity to experience this kind of cultural negotiation, especially once
specific domains are identified.

Identifying a domain restricts the scope of social interaction within a culture, in
which meanings are generated and shared among domain members through the use of
deeply contextualized language. A domain-specific advanced level Chinese program,
such as the Chinese Flagship MA program at The Ohio State University (OSU), connects
American learners with their Chinese counterparts, which provides them with a shareable
frame of discussion, presentation and negotiation in the target culture. In fact, beginning
early in their careers as learners of Chinese, those who aim to achieve expert performance
in the target language have already assumed the responsibility of managing their own
learning. Therefore, even if some students decide to change their domain many years
later after leaving the program, the previous training within their original domain will
give them a sense of how expertise can be achieved through deliberate practice and how
it is evaluated through a variety of assessment tools. Content and strategies included in
such a program can be broad enough to be applicable to a variety of professional domains or to a more general domain, such as “working in a Chinese organization.” A more detailed discussion, based on empirical findings, on how to incorporate the role of domain into the assessment of foreign language abilities is covered in Chapter Five.

However, it is also important to note that when we analyze how experts function in a foreign professional environment, their demonstrated expertise may be viewed from two different aspects – how they communicate in the foreign language in a professional setting and how they handle domain-related work. Although these two levels are always integrated and can hardly be evaluated independently, the current study focuses more on the first type of expertise, which can be referred to as communicative expertise, rather than one’s domain expertise. In other words, this study looks more into how one negotiates with business partners, rather than whether or not the business is a financial success. Of course, unlike the traditional concept of “proficiency,” which often ignores the existence of domain, the idea of domain still plays a significant role in identifying and evaluating this type of communicative expertise as it defines the nature of every communicative activity taking place within it.

1.3.5 The Path from Competence to Performance: Defining Performance

As mentioned above, under the proficiency-oriented approach, both teachers and researchers focus their attention on developing students’ knowledge of various linguistic subcomponents of the target language, as described in a proficiency scale or a communicative syllabus. The goal of the expertise-oriented approach sees language learning as an undertaking of developing strategic mapping of the language onto domain
needs and strategic control so as to be able to use it. Therefore, pedagogy should focus on what is known about the variety of skills that encompass expertise and how these skills are integrated into entire pragmatically driven performances.

Expertise is observable and demonstrable through performances. This study adopts Walker’s definition of performances as “situated events” (Walker 2010, 8), the enactment of scripts or behaviors situated at a specified time and place with roles and audiences specified. Performance is therefore a useful tool for conceptualizing and organizing our daily encounters in life. Effective language learning precisely requires a prolonged experience of performing the language in meaningful contexts.

According to this definition, one may wish to argue that taking a standard proficiency exam is also a “performance” and an OPI in this sense is equally performance-based. Indeed, it is. However, performances involved in most current proficiency tests do not really reflect the type of performances that a CFL learner is likely to encounter when working in his or her respective domains with Chinese counterparts. Therefore, the performances related to proficiency tests, in most cases, are only accurate predictors of one’s ability to perform on a test, not of one’s success in a future career.
Chapter 2 Expertise in Nonverbal Domains

This chapter explores how expertise has been defined, described, framed and investigated in fields ranging from chess to physics, from medicine to music. By considering general expertise across a variety of domains, this chapter suggests avenues worthy of pursuing in the study of expertise in using a foreign language.

Based on the Glaser and Chi (1988: xvii-xx), seven general characteristics of experts were introduced in Chapter One. In the following sections this study delves into how each of the characteristics is reflected across various domains and to what extent they may relate to the acquisition of skills for communicating profession-related issues in a foreign language.

The seven characteristics are repeated below:

1. Experts excel mainly in their own domain;
2. Experts perceive large meaningful patterns in their domain;
3. Experts are fast; they are faster than novices at performing the skills of their domain, and they quickly solve problems with little error;
4. Experts have superior short- and long-term memories;
5. Experts see and represent a problem in their domain at a deeper level than novices; novices tend to represent a problem at a superficial level;
6. Experts spend a great deal of time analyzing a problem qualitatively; and
7. Experts have strong monitoring skills;

(Glaser and Chi 1988, xvii-xx)

2.1 Experts Excel Mainly in Their Own Domains

There was a short-lived belief that expertise is a kind of general heuristic searching skill requiring minimal knowledge about the specific content of any particular
field (Holyoak 1991). In reality, little evidence exists to suggest that an expert in one domain can easily achieve mastery in another area. In fact, numerous studies across different domains indicate that the opposite is true, highlighting the significance of detailed domain knowledge, specialized memory skill, and inference patterns.

Among the research conducted to date, Chase and Simon’s (1973) theoretical account of chess expertise, which illustrates that memory is domain-dependent, is considered a pioneering and highly influential study. The researchers asked chess players from beginners to international masters, to recall the location of all the chess pieces after being shown for a short period of time (approximately five seconds) a position from an actual chess game. Their results indicate that the ability to recall the correct location of chess pieces increases with a player’s skill. In fact, an expert could identify the locations of sixteen pieces more than could a novice player. However, when they asked all the players to recall chessboards with randomly placed pieces, all the participants displayed the same level of poor recall performance. This result invalidates the long-held theory that the superior performance of chess experts is simply a reflection of a generally superior visual memory. Instead of relying on a photographic memory, a superior chess players encode arrangements of chess pieces by drawing upon associations with their extensive knowledge of chess. In a real game, what enables a chess grand master to rapidly retrieve the best possible moves from long-term memory is his or her ability to recognize configurations of chess pieces as cues (de Groot 1965). This skill has been termed “chunking.”

Following Chase and Simon’s classic study, scholars from a wide range of domains have provided evidence that the superiority of an expert’s performance is
mediated by increased domain knowledge and domain-specific skills. However, this superiority generally does not carry over to domains in which the subject has little or no experience. For example, Voss and Post’s (1988) study of problem solving in political science shows that non-domain experts, chemists in this scenario, go about solving political science problems much like novices. They focus only on concrete issues when discussing political science rather than detailed abstract issues that they typically encounter in their everyday professional lives. As Glaser and Chi (1988) suggest, the reason for this lack of transfer from one area to another is because expertise in a particular domain is the result of years of study and exposure to information in that field, which subsequently yields a high degree of specialized knowledge.

One may argue that a lack of transfer in expertise between chemistry and political science is too obvious due to the very different nature of the two fields. However, in, *The Sports Gene*, David Epstein (2013) recalls a story about how Jennie Finch, a softball pitcher, beat the world’s best baseball hitter Albert Pujols, revealing the surprising discovery that even between such two intuitively similar domains as softball and baseball, the transfer of expertise is very limited. The obvious reason for baseball hitters’ superhuman reaction speeds is that they have acquired a perceptual skill, which allows them to pick up on cues from the pitcher’s body language before he or she throws the ball. Therefore, the hitters are actually focusing on the motion of the pitcher’s shoulder, torso and hand. For these reasons, when an expert baseball hitter is faced with a softball pitcher, who throws underhand, he or she can no longer rely on familiar cues and is reduced to being a novice, as far as hitting a thrown ball is concerned. The extreme anger and humiliation felt by Albert Pujols as described by Epstein shows that, even as an
expert, he was tricked into thinking that he could hit the ball because of his ability to see the ball instead of the ability to pick up anticipatory cues from another player. Just as an expert chess player may fail at a basic memory skill test, an expert baseball hitter may not necessarily have better eyesight than ordinary people.

This crucial role of domain in expertise studies reflects a realistic view of what a professionally engaged native speaker is able to achieve in his or her native culture. This presents the possibility to set a more attainable and desirable goal for foreign language learners. However, as mentioned in chapter one, the traditional concept of “proficiency” seems to eschew the importance of domain, in preference for restricting the ability of a foreign language learner in terms of four skills of speaking, listening, reading and writing independent of any particular domain. Therefore, findings from studies of expertise in nonlinguistic domains should encourage foreign language pedagogues to rethink the traditional designation of advanced skills and to question whether the transfer of superior communicative skills from one area of expertise to another isn’t similarly restricted.

2.2 Experts Perceive Large Meaningful Patterns within Their Domain

How do experts go about solving a problem in their domain? Does an expert “perceive” the situation differently than a non-expert? As discussed in the previous section, the superior recall ability of experts, illustrated in the example of chess grand masters, has been explained in terms of the way they “chunk” various related elements of a configuration that are related (de Groot 1965). Chase and Simon’s follow-up study on perception in chess found that superior players have the ability to encode positions into larger perceptual chunks. When researchers asked experts to reproduce from memory
board positions that had just been shown to them, or to copy a position from one board to another, the subjects’ use of the grouping technique became clear from their periodic glances at the board.

Further analysis of perception’s role in chess indicates that an expert’s ability to perceive large meaningful patterns does not necessarily reflect his or her exceptional perceptual skills, but rather a well-organized knowledge base. For example, chess pieces within a single chunk perceived by an expert chess player “are always bound by relations of mutual defense, proximity, attack over small distances, and common color and type” (Chase and Simon 1973, 80). The key conclusion we can draw is that this kind of perceptual superiority based on inventories of mental representations can be found among experts in a variety of other domains.

For example, expert radiologists are able to quickly envision mental representations of patients’ anatomy and can, therefore, evoke more effective and pertinent schemata (Lesgold et al. 1988). Unlike inexperienced radiologists who ignore abnormal features that could signal pathology because they do not fit into their schemata of normal anatomical structure, expert radiologists are able to adjust their prior expectations of normal anatomy to the needs of a particular patient. Their skillful use of localization cues allow them to detect abnormalities subsumed under normal variations that are usually neglected by non-experts. Moreover, their perceptual superiority is also facilitated by their “opportunistic” problem-solving capability. Whenever new information arises, such as a lab report or a new features on an x-ray, expert radiologists tend to use them to greater advantage.
In the field of electronics, Egan and Schwartz (1979) found that skilled technicians were able to reconstruct detailed drawings of circuit diagrams according to the functional nature of the elements in the circuit such as amplifiers, rectifiers, and filters. Novice technicians, however, produced chunks based more on the spatial proximity of the elements.

Findings from various domains, as reviewed above, reinforce Chase and Simon (1973)’s proposal that expert performance in any “skilled task” can be attributed to vast amounts of knowledge and pattern-based retrieval acquired over many years of experience in the relevant domain. According to skill acquisition theories, knowledge is first acquired and is then organized into procedures for responding when encountering situations characteristic of the relevant domain (Ericsson 1996, 15).

Foreign language learning also incorporates such processes of acquiring, organizing and storing knowledge, which Walker and Noda (2000) term “second-culture worldview compilation.” From the first day that foreign language learners experience performances in the target culture, they retain memories of those experiences in the form of stories, or discrete items of personal knowledge of the culture, and begin to accumulate applicable knowledge of the culture in stories by organizing their acquired language and cultural knowledge into useful categories, such as cases, sagas and themes. These categories are meaningful patterns that expert users of a foreign language are able to

---

9 Walker and Noda (2000) define a case as a collection of stories about doing something in a culture. For example, a case of shopping may consist of a series of stories including responding to greetings from a shop assistant, comparing different commodities, and negotiating a purchase when bargaining is expected. Sagas are series of stories about a specific set of people or a specific location, representing what a learner knows about behaving around particular people or at particular places. Themes are overarching cultural norms that manifest themselves in various associated behaviors. For example, the cultural theme of explicitly displaying modesty can be seen manifested in actions like declining a compliment, yielding to other people and making self-criticism.
perceive and draw upon when working in their domains. The greater the number of performances one has previously experienced and compiled in the target culture, the more skillfully one is able to recognize the cases, sagas and themes underlying a newly encountered situation and the more quickly one can retrieve the most appropriate “script” and adapt it to the current situation.

2.3 Experts Solve Problems Quickly and Accurately

There is a major difference between the amount of time a novice spends in attempting to solve a problem versus that spent by an expert (Chi, Galser and Rees 1982; Simon and Simon 1978; Larkin 1981). The difference is obviously quantifiable: Simon and Simon (1978) noticed a 4:1 time difference between expert physicists and novices in solving physics problems. Glaser and Chi (1988) point out two reasons for experts’ superior speed in solving problems in their domains. It is not difficult to see that faster speed in solving problems is still attributed to the superior pattern-based retrieval developed within one’s domain, as discussed in the previous section. As experts are able to perceive meaningful patterns in their domains, there is no need for them to search as extensively for a solution. The previously discussed chess expertise reveals that experts are able to perceive the problem in a way that simplifies it. What differentiates an expert chess player from a non-expert is not the use of different heuristics as all chess players consider the same number of moves (de Groot, 1965), but an ability to choose the best path to a win without having to consider all the others. This ability has been found in many other areas of endeavor. For example, in medical diagnosis research, both experts and novices use the same kind of “generate-and-test” heuristic. The expert, however,
begins the process with a more accurate hypothesis. Janet Starkes (2003) conducted extensive experiments on sports expertise in a variety of domains including volleyball, tennis and baseball and concluded that elite athletes tend to sift through visual information to extract critical data from the arrangement of players or from subtle changes in an opponent’s body language or posture so they can make predictions about the direction the ball is likely to travel. The high speed at which the experts perform the task can be seen as a result of the superior way they have learned to perceive the task, not simply the instinctual ability to react quickly.

Moreover, many practices allow an expert to automatically perform some basic skills in his or her domain, which frees up memory capacity for processing other aspects of the task. For example, a study revealed that a professional secretary typically types 25 million words in 10 years, during which time he or she would have typed the word “the” two million times and typed a common word like “system” 10,000 times (Gentner 1988). In such cases, when an expert’s skill becomes more automatic, it allows him/her to free up memory capacity for processing other aspects of the task.

This type of automaticity underlying expert performances is also a desired outcome in foreign language learning. Since rehearsals focusing on particular parts of an overall performance are important in attaining an excellent performance in any domain, an expert user of a foreign language must develop automaticity through vast amounts of practice, such as performing countless repetitions to build a repertoire of authentic expressions that sub tend automaticity.
2.4 Experts Have Superior Short- and Long-term Memory

Wilding and Valentine (2006) propose two methods for measuring superiority in memory. First, people with superior memory are able to retrieve a substantial amount of material without error. Second, an example of superior memory is one’s fairly long-term retention of an unusually large proportion of the original material.

The contrast between novice and expert memory performance is also striking. In one of the most influential studies in this area, Chase and Simon (1973) contradict the notion that the superior memory of an expert reflects a generally exceptional ability to store information. Instead, their study indicates that the superior memory of an expert is always linked to his or her extensive domain-knowledge and how that knowledge has been previously stored in his or her memory. Chase and Ericsson (1981) termed this phenomenon “skilled memory,” referring to the rapid and efficient utilization of recall within a specific knowledge domain to perform a task at an expert level. Even when they asked subjects to remember meaningless information, such as lists of numbers, given sufficient practice, subjects were able to acquire meaningful encoding methods, allowing them to recall information based on associations with pre-existing knowledge and patterns in long-term memory. For example, some former collegiate runners tend to remember the number 3493 as “3 minutes 49.3 seconds,” which is considered a world-class completion time for the mile. They also found that once this information was stored in long-term memory, the experts could retrieve it easily by activating specific retrieval cues.

Chase and Ericsson’s findings about how subjects can be trained to extend their long-term memory to store and retrieve information is consistent with other memory
studies which focus on the difference between novice and expert skill performances. All these studies shed light on the cognitive substrate governing performance at high levels of skill execution across diverse skill domains (de Groot 1965; Starkes 2003; Richman et al. 1996). Experiments of this type usually ask experts and less experienced professionals to recollect exact steps in performing certain tasks so as to make comparisons of their abilities to recall the processes and properties of skill execution as the performance actually unfolds. These experiments show that the unpracticed performances of novices are controlled by declarative knowledge that is held in working memory and can be attended step-by-step. In contrast, highly practiced or over learned performances become automatic, and are supported by procedural knowledge that operates without the need for explicit or attended monitoring.

Such expert-like skilled memory, as described above, is also developed in learning and using a foreign language. L2 procedural memory is established through interacting in the target culture and is strengthened by the extensive use of well-designed contextualized drills and exercises available in pedagogical materials. Once they have accumulated a sufficient number of performance experiences, expert learners trained under this method are able to recognize and learn from new situations by effectively accessing previously established cultural memories.

2.5 Experts Perceive and Represent a Problem on a Deeper Level

The vast experiences of experts enable them to visualize the complexity of a problem and to understand the context in which the problem must be solved. According to Galser and Chi (1988), both experts and novices possess conceptual categories for
organizing domain knowledge and sorting domain problems. It must be noted that, the experts’ categories are “semantically or principle-based,” while the categories of the novices are “syntactically or surface-feature oriented” (p.xix). Such differences of problem representation can be found in many domains, especially in physics and mathematics, the two domains that have received the most attention in expert research in the recent years.

Studies on problem solving in the field of physics show that the knowledge that experts use in problem solving is characteristically much more abstract than that of novices (Voss 1989). For example, Larkin (1981) found that before developing any equations, expert physicists often draw a diagram to help them analyze a problem. Then, they place the problem in a particular category, such as a “force” problem, and proceed to apply appropriate equations that lead to the solution of the problem. Moreover, due to the fact that the sets of equations stored in experts’ memories are always based on particular principles or laws, they are able to form functional units or useful “chunks,” as discussed in Section 2.2.

Nonetheless, novices tend to tackle the problem in a “piecemeal manner” (Voss 1989, 263). Unlike experts who tend to work forward in solving problems, novices tend to employ a means-ends analysis strategy to work out the solution. When novices attempt to solve a problem in physics, they usually build their problem categories around literal objects mentioned in the problem description, such as a variable or set of variables in the problem. Then, they simply come up with an equation that contains the parameters. Similar methods have been used to solve algebra equations (Silver 1979), in which
experts categorize problems on the basis of the underlying equation or principle while novices tend to categorize problems based on the content.

In sum, research indicates experts and novices organize knowledge differently. Experts organize their domain-related knowledge in relation to underlying principles, while novices tend to organize knowledge more superficially. Such different tactics for organizing knowledge influence the ways experts and novices perceive and categorize problems.

Similarly, expert users of a foreign language do not simply memorize an utterance based on its phonological, syntactic and semantic features, but also incorporate the stories constructed around the utterance, which allows them to form larger knowledge domains (Walker and Noda, 2000). When communicating in the foreign language, experts are able to associate a newly encountered situation with previously stored stories, which they have organized into the categories of case, saga and theme. Equipped with the ability to recall the most pertinent story at the right time, an expert user of a foreign language interacts with the target culture by extracting appropriate scripts from memory that are associated with the stories instead of thinking up a response in an impromptu way.

2.6 Experts Spend Much Time Analyzing a Problem Qualitatively

As discussed in Section 2.3, experts’ decisions are based on intuitive judgment, which allows them to act more quickly than nonexperts. However, studies also show that experts tend to approach a problem more qualitatively so as to understand the overall purpose of a task, while novices may plunge immediately into attempting to solve for an unknown.
One such example can be seen in the study of literary expertise (Scardamalia and Bereiter, 1991), in which expert writers spend more time in the planning stages and the writing of the first sentence of a simple narrative, while the novices simply plunge in and writes as fast as they can. Moreover, compared with non-experts, expert writers always work harder at the same assigned tasks, “engaging in more planning and problem solving, more revision of goals and methods, and in general more agonizing over the task” (Scardamalia and Bereiter 1991, 172).

However, such observations do not contradict findings in Section 2.3. Glaser and Chi (1988) characterize this phenomenon as a way for experts to construct a model of a problem, during the process of which they are able to identify and add constraints to that problem. By investing more effort in adding constraints to a problem, experts are in effect reducing the search space that they need to find the solution. For instance, Voss and Post (1988) explore how experts and non-experts come up with solutions to increase crop production for the Minister of Agriculture in the Soviet Union. Through an analysis of their respective protocols, they found that more experts (approximately 24%) tend to elaborate on the problem by associating it with possible constraints such as Soviet ideology and the amount of arable land, which helped them eliminate some impossible solutions, such as the fostering of private competition or increasing planting. However, only very few (approximately 1%) of non-experts were able to perceive the problem in this way.
2.7 Experts Possess Strong Self-monitoring Skills

How do the best get better? Acquiring knowledge or a skill in any discipline involves constantly seeking out the most reliable and comprehensive instructional sources available. Experts’ self-monitoring skills are manifested in experts’ superior ability to notice errors they have made in the past and to accurately predict the level of difficulty of tasks in their domains. Self-monitoring skill allows experts to accurately estimate the amount of time and resources to allocate to the tasks ahead. As Glaser and Chi (1988) point out, the superior monitoring skills and self-knowledge of experts reflect their greater domain knowledge as well as a different representation of that knowledge.

There has been extensive research on the significant role that self-monitoring plays in the development of exceptional physicians (Epstein, Siegel and Silberman 2008), actors (Lan and Morgan 2003), athletes (Zimmerman and Kitsantas 1997), and athletic coaches (Schempp et al. 2006). Experts in all these domains demonstrate an ability to reflect on their own experience, knowledge, hypotheses, and inferences and use that information to modify their practices accordingly.

Scholars have also pointed out that what enables an expert to hone self-monitoring skill is the achieved automaticity that reduces the demand for cognitive resources to process the tasks themselves (Winne 1995; Zimmerman 1995). Namely, limited cognitive capacity may be a reason for self-monitoring deficiencies in novices, who tend to panic when they become overwhelmed by a task. This certainly explains why experts with a relatively high level of routinization demonstrate better self-monitoring skills. Based on this assumption, many scholars devised experiments to help subjects improve their self-monitoring skills. For example, Lan and Morgan (2003)’s study
determined that videotaping is an effective strategy for theater students to improve their self-monitoring skills when they cannot focus on learning and self-monitoring simultaneously. When the two cognitive processes (learning process and self-monitoring process) are separated, this reduces the demand on one’s cognitive capacity.

Superior self-monitoring skill does not only include the ability to assess one’s own performance, but also the ability to gauge the level of difficulty of a task. According to Glaser & Chi’s study of expertise in physics, experts are more accurate than novices in judging the difficulty of physics problem. Their data show that one-third of expert physicists’ assessments of the level of difficulty of problems were based on knowledge of the underlying physics principles governing the solutions, while only 9% of the novices’ decisions were based on the underlying principles. Instead of relying on physics principles, novices tend to access the problem based on non-problem characteristics, such as “I’ve never done problems like this before” (p.xx). Thus, Glaser and Chi’s findings reflect that superior monitoring skills are based on experts’ superior domain knowledge and a high level of representation of that knowledge. This finding echoes Ericsson’s (1996) comments about the two goals of experts when constructing mental representations – “to constantly improve the given level of performance” and “to exhibit the best possible performance that is attainable given the current level of skill” (p.3).

According to studies of good language learners in the field of ESL (Rubin, 2005), expert learners of a foreign language usually demonstrate strong self-monitoring skills. Expert learners are skilled at setting observable and useful criteria to recognize accomplishments and to evaluate their learning. This finding will be discussed in Section 3.1 when reviewing previous literature on expert foreign language learners.
2.8 Conclusion

Including a review of the generic characteristics of expertise serves this dissertation in two ways. First, as these characteristics are prevalent across a variety of domains, they can be used as a reference in the current study to identify qualified subjects - expert users of Chinese as a foreign language. The two subjects in this study manifest many of these characteristics in both learning and using Chinese as a foreign language, which will be discussed in detail in Chapter 3. Second, findings from these non-linguistic domains provide a starting point for the current study. Keeping these characteristics from the literature in mind, this dissertation will explore how these generic characteristics are specifically operationalized in communication in a foreign language.
Chapter 3 Identifying expertise in using Chinese as a foreign language

3.1 Literature Review: Describing Experts in Learning/Using a Foreign Language

The impetus to study expertise has gained momentum over the decades in a large number of domains. Along with its growth, our understanding of the nature of expertise and the armory of research methods used for the study of expertise developed as well. Yet, it seems difficult to translate these findings into educational practice or language pedagogical practice (Johnson 2005; Rubin 2005). Two reasons may account for this difficulty of applying the fruitful findings in expertise since late 1970s to foreign language arena.

First, there was little motivation for cultivating expert users of foreign language during the 1980s and well into the 1990s. During that period, foreign language education still focused on how to get students to the Intermediate-level, which according to the ACTFL proficiency guideline is the realm of the linguistic survivor (Martin 2015, xiii). Unlike other disciplines such as sports, arts and sciences, in which the highest levels of performance and achievement have always been an object of fascination, the field of foreign language pedagogy in the US, especially the field of teaching Chinese as a foreign language, has rarely entertained goals focused on learners attaining expert performances in a professional settings. Consequently, even though there has been a growing desire for students to attain upper levels of performance in L2 since the
beginning of the 21st century, there still exists a kind of “disjuncture” (Byrnes 2002, 34) among foreign language programs in American higher education. Namely, the first few semesters aim at enabling students to develop basic interactional abilities in L2. These courses are always followed by subsequent content courses in the rest of the semesters, rarely incorporating explicit pedagogical interventions that target advanced-level L2 learning.

As mentioned in the first chapter, the second reason is the lack of efforts to look at findings in non-linguistic domains and to relate them to language learning and teaching. According to Johnson (2005), this is related to the strong tradition of seeking parallels with first language acquisition in foreign language pedagogy.

However, in the 21st century, both of the two situations described above have been improved, driven by the precipitous increase in demand for language professionals. The lack of motivation and the existing mismatch that may have been tolerable in the past decades are no longer acceptable. The following sections address the progress taking place over the past few decades from three different perspectives: the tradition of Good Language Learner (GLL) studies in second language acquisition (SLA), studies on expertise in different skill areas in using foreign languages, and some newly emerging studies on characteristics of foreign language users with advanced proficiency and beyond.

3.1.1 Expertise in Foreign Language Learning

In the field of SLA, the book *Expertise in second language learning and teaching*, edited by Keith Johnson (2005) can be considered the first and so far the only scholarship
that directly borrows theoretical insights from the expertise research tradition to discuss learning and teaching expertise in foreign language pedagogy. Following Johnson’s review of general expertise in non-linguistic domains from psychology studies of expertise, Joan Rubin (2005) traces the history of GLL studies to identify learning strategies used by expert learners by proposing a model of GLL. Although scholars at that time did not think of themselves as studying expertise as the term “expertise” was not the language they used in their discussion, the enormous growth in understanding of the cognitive and affective interactions between knowledge and processes contributes to a general understanding of expertise in communicating in foreign language.

Rubin’s model of expert language learner is called Learner Self-Management (LSM), which is defined as “the ability to deploy procedures and to access knowledge and beliefs in order to accomplish learning goals in a dynamically changing environment” (Rubin 2005, 37). Based on Rubin’s model, I summarized the characteristics of expert learners in Table 2, categorizing them into three areas: expert learner strategies, expert learner knowledge, and expert learner beliefs.

Table 2. Characteristics of Expert Learner

<table>
<thead>
<tr>
<th>Strategies</th>
<th>Knowledge</th>
<th>Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Planning strategy. Expert learners set realistic goals within a realistic time frame.</td>
<td>1. Metacognitive knowledge. Expert learners know well about task requirements needed to select appropriate strategies.</td>
<td>1. Facilitating beliefs about language learning.</td>
</tr>
</tbody>
</table>

(Cont’d)
Table 2 (Cont’d)

<table>
<thead>
<tr>
<th>2. Implementing strategy. Expert learners work on a task based on a thorough analysis of task purpose, task classification and task demands.</th>
<th>2. Self-knowledge. Expert learners know well about their learning styles and multiple intelligences, and motivation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Self-monitoring strategy. Expert learners are skilled at setting observable useful criteria to recognize accomplishment and to evaluate their learning.</td>
<td>3. Contextualized knowledge. Expert knowledge represents contexts of applicability in which the learner knows when and how to use their knowledge.</td>
</tr>
<tr>
<td>4. Problem-solving strategy. Experts are flexible and adaptive. They are not disturbed by mistakes.</td>
<td>(Cont’d) Expert learners possess a strong sense of self-efficacy. They are comfortable with uncertainty and are willing to live with a certain amount of vagueness. They always actively try out guesses and are willing to make mistakes.</td>
</tr>
<tr>
<td>5. Affective strategy. Expert learners are able to laugh at their own mistakes. They usually have a sense of humor (Naiman et al. 1975, 54)</td>
<td>2. Responsibility. Expert learners are self-managed learners who believe responsibility for learning lies with themselves.</td>
</tr>
<tr>
<td>6. Social strategy. Expert learners use a few well chosen formulas to continue to participate in activities which provided contexts for new learning (Wong-Fillmore 1976)</td>
<td>(Summarized from Rubin 2005)</td>
</tr>
</tbody>
</table>

As presented above, the LSM model is characterized by interactions within each of the three categories, as well as between each other. Expert learners possess a well-developed repertoire of cognitive, social and affective strategies. They are also equipped
with adequate knowledge about the task, the context and themselves. When they carry out the strategies, they are constantly modifying their knowledge repertoire as well. All of these learning strategies seem not only apply to learning a foreign language, but also to learning in general. At the same time, they have also developed a productive belief about language learning and about themselves as language learners, which allow them to be more tolerant with ambiguity and mistakes. As Rubin notes, an expert learner “is willing to appear foolish in order to communicate and get his message across” (p.49).

Research instruments used to study the GLL and learning strategies include thinking aloud techniques, questionnaires, observation, diaries and dialogue journals, case studies, learner histories, interviews, and computer tracking (Cohen 1987, Cohen and Scott 1996). Although none of these instruments is perfect and has many limitations, researchers endeavor to minimize the disadvantages by combining a few of them together in one study.

3.1.2 Expertise in Foreign Language Use

Scholars in SLA have not only examined strategies that expert learners tend to use, but also introspected how an expert learner manipulates the interactions among process, knowledge and belief in the traditional division of four skills, i.e. speaking, listening, reading and writing. These researchers find that expert learners possess a well-developed repertoire of knowledge and strategies, accumulated through past experience and training. This repertoire specially includes knowledge about the nature of a specific task and the procedures needed for completing these tasks effectively. Such possession of
rich domain and field knowledge applies to experts in all different kinds of skill areas as discussed in the following sections.

**Oral Expertise**

Bygate (2005) sees that oral expertise lies in the possession of an oral repertoire, and the mastered strategies to exploit the repertoire skillfully. Although there exist different views of what constitute the main oral language repertoires, numerous previous studies (Chafe 1982; Chafe 1985; McCarthy and O’Keefe 2004) agree that the spoken language has some prototypical, or in some cases, unique features, distinguishing it from the written language. Distilled from a survey of literatures on different conceptualizations of oral language repertoires, Bygate (2005, 109) summarizes that the oral repertoires that expert speakers are expected to be familiar with include a range of discourse patterns, a range of convergence strategies, linearization, ellipsis, as well as key ways of marking personal involvement and deictic options.

Certainly, it is impossible to discuss the possession of these repertoires that contribute to oral expertise without referring to the application of the repertoire. Based on Levelt (1978, 1989)’s model of oral language production, Bygate (2005) further emphasizes two complimentary aspects that highlight the oral expertise: the ability to hold plans in working memory while acting on them as well as the achieved automaticity and flexibility at all levels of processing. The development of expertise can be seen as a growth in one’s ability to integrate low and high level skills in communication. Equipped with relatively rich lexico-grammatical knowledge, an expert speaker of a foreign

54
language not only attends to the lower levels of processing, but also is able to manage a great level of automaticity with higher levels of processing in on-going discourse.

**Listening Expertise**

In the same sense, it is the combination of knowledge and strategies that enables expert L2 listeners to process what they hear more effectively than non-expert listeners. L2 listening expertise is developed through an accumulation of systemic knowledge of the target language, including the phonological knowledge, semantic knowledge, grammatical knowledge, pragmatic knowledge and discourse knowledge. For example, due to their profound phonological knowledge, expert L2 listeners are found to be fast at converting sounds to words and can effectively hold aural input in their echoic memory without much interference (Greenberg and Roscoe, 1988). Moreover, expert listeners also know how to apply appropriate strategies to process stored linguistic and metacognitive knowledge, which is reflected in their ability to exercise maximum control of variables during listening. Goh (2002) adopts the retrospective protocol to examine expert listeners’ concurrent thinking while accomplishing a listening task, in which she finds that expert listeners always monitors their unfolding comprehension and at the same time, retains the sound for an unfamiliar word in working memory. It is also found that during the listening process, experts continue to relate what they have heard to their prior knowledge as well as the evolving context, through which they are able to process and interpret the part that eluded them the first time. This ability to identify important aspects of input and to use the most relevant linguistic and contextual clues for arriving at a reasonable interpretation corresponds to the finding about experts’ ability in judging which avenues
are likely to bring about the desired result (Bereiter and Scardamalia 1993, cited in Goh 2005).

**Reading and Writing Expertise**

As expertise usually refers to some concrete outcome that is observable, there seems little discussion around expertise in reading, a relative private and silent activity whose process is not so visible.

The general observation of experts’ reacting fast to solve a problem seems not applicable to expertise in reading and writing as in both fields, experts tend to interpret the task in ever more demanding ways. Although skilled readers may demonstrate a certain level of automaticity at the decoding level of text processing, that fluency in text processing alone does not equate expertise in reading. Adopting a think-aloud technique to examine readers’ interaction with the text while reading, Wallace (2005) proposed that reading expertise lies in a high level of metacognitive awareness about the reading process and the textual ambiguity as reading “is a thinking process evidenced and supported less by automaticity than by enhanced reflectiveness” (p.101). Expert readers’ think-aloud protocol reveals that they were able to engage in a more productive way of interpreting the text by speculating on the wider context of situation to include writer purpose and readership, which goes far beyond the stage of simply decoding and comprehending the text ‘s linguistic constituents.

Similarly, composition is also viewed as a contemplative activity so that experts may not necessarily spend less time and effort than non-experts on certain tasks. Weigle (2005) summarizes the essential areas of knowledge that are involved in writing expertise,
including language knowledge, topic knowledge, genre knowledge and audience knowledge. In addition to these domains of declarative knowledge, learners have also built up task schemas through experience, which is defined by Hayes (1996) as “packets of information stored in long-term memory that specify how to carry out a particular task”. Expert writers make good use of metacognitive skills to assess the demands of a given writing task and their available resources for completing a writing task. Think-aloud technique is also widely used in examining writing expertise. Through an analysis of experts’ think-aloud protocols, experts are observed work recursively, shifting back and forth between planning, composing, retrieving content from their memory, reorganizing or revising their writing, attending to their audience, and coming up with appropriate language to convey their message (Ransdell, Levy and Kellogg 2002, Olive, Kellogg, and Piolat 2002). Moreover, just as an expert listener, an expert writer is also skilled at attending to a wider variety of considerations simultaneously, including “the overall message, the organization, and the appropriateness of specific linguistic choices” (Weigel 2005, 130).

3.1.3 Understanding Advanced L2 Learners and Professional-level Chinese Use

Due to the lack of any tradition in bringing experts together to study the characteristics qualifying their expertise, very few empirical studies have been undertaken to date in order to explore the characteristic of higher levels of foreign language proficiency. Among these few, Leaver and Atwell (2002) report their close examination of the nature of ILR level 4 proficiency based on an extensive interview with a number of ILR level 4 learners in multiple languages and Patrick McAloon (2008)’s
case study of five advanced-level Chinese learners and their profession-related advanced performances. Although neither of them uses the term “expertise” to organize their theses, both studies share some similarities with the expertise study tradition in their approach to bring advanced learners together to look at what constitute their exceptional performances by using research techniques such as retrospective interviews and ethnographic observation. They remain the only precedents in this field to date that do much to inform learning and teaching at an advanced level.

Based on the in-depth retrospective interviews with ILR level 4 learners who were still using the language in their personal and professional life, Leaver and Atwell reported a list of attributes defining ILR level 4 proficiency including polyglotism, instrumental motivation, use of authentic texts in preparation, checking presentations with native speaker, wanting feedback, unified bicultural personality, load of short-term memory, and focus on sociolinguistics in speaking (2002, 267). However, with a limited pool of level 4 learners, Leaver and Atwell are very cautious about suggesting what in fact contributes to developing Level 4 proficiency by only proposing four aspects that play an important role in achieving level 4 proficiency: literacy, formal instruction, authenticity and motivation.

In addition, Leaver and Atwell’s study also proposes the ability to express one’s own personality and emotional states in culturally appropriate way as a newly suggested component of communicative competence. It was at level 4 which they observe that the elision of the learner’ personality in C1 and that in C2 occurred. Beyond negotiation of meaning, advanced learners are able to use words, gestures, and behaviors in telic ways.
Unlike Leaver and Atwell’s study that exclusively relied on interviews, McAloon shadowed four Americans and one Englishman who speak Chinese at work and interviewed their Chinese colleagues to evaluate their performances. McAloon’s study reveals what non-native speakers of Chinese are able to do with their advanced level Chinese within a professional setting. Moreover, the quantitative evaluations obtained from the native speakers of Chinese demonstrate how foreigners’ performances are perceived and evaluated by their native Chinese counterparts. Although McAloon’s study does not directly answer the question what characteristics constitute the advancedness of their performances, his ethnographic observations along with the qualitative interview data did identify a few behavioral aspects highlighting the advanced performances such as demonstrating domain knowledge and an understanding of Chinese culture, and avoiding conflict in a culturally appropriate way.

3.1.4 Conclusion: Inadequacies in the Previous Literature

Although “expertise” is not the discourse many studies use for the various reasons, there is a perceived interest and trend in exploring what defines the term “advanced” in learning and using foreign language. Previous scholarship as reviewed above, although limited in amount, provides us with a valuable starting point from which we may be able to approach a comprehensive concept of expertise in using Chinese as a foreign language.

While speaking fluent Chinese at work appears to be very different from hitting home runs or performing solos in concert, it is still valuable to realize that much of what we know about expertise in the nonverbal fields such as sports and music will anticipate today’s profession-related foreign language training movement. If we are able to
recognize that what we are after in our field has already been discussed in the field of
cognitive psychology for over 30 years, we may find that in our parallel searches for the
nature of expertise, we are actually discovering different parts of the same elephant.
Nevertheless, a review of the available literature actually discloses its major limitation as
the lack of connection with the rich tradition of expertise scholarship in the field of
cognitive psychology. This limitation can be further demonstrated in two specific
problems in the current scholarship – the absence of the concept of domain in their
discussions and the paucity of empirical studies looking at concrete performances in a
natural setting.

First, although all the eleven chapters in Johnson (2005)’s Expertise in Second
Language Learning and Teaching are organized and developed around the topic of
“expertise”, none of them can really fit into the paradigm of expertise research formed in
the past decades. For example, the four chapters looking at the expertise of the
traditionally divided four skills though provide interesting observations of advanced level
skills, their discussions are still centered on domain-free cognitive skills, which runs
contrary to the expertise theory.

Second, research methods adopted in the existing studies on expertise are relative
limited, lacking ethnographic data to support their findings. The majority of studies I
have reviewed in this chapter are exclusively dependent on retrospective interviews (Goh
2002; Leaver and Atwell 2002; Weigel 2005 ), the limitation of which will be discussed
in 3.3.2. The rest, such as Bygate’s discussion of oral expertise, however, is simply based
on a summary of linguistic theories or an extensive literature review of relevant topics.
Such modest and haphazard way of dealing with the concept will make it difficult for
scholars to understand its nature, as expertise should always connote an outcome or product which is visible and can be judged or evaluated.

By recognizing these existing limitations, we can address these limitations and criticisms without sacrificing the undisputed achievements of those pioneering studies in the past decade. Among all these efforts in exploring advanced foreign language use and understanding advanced foreign language learners, McAloon’s (2008) study is the closest to the current endeavor for its inclusion of domain in his discussion and its ethnographic approach. The current study considers McAloon’s study as a precedent, based on which, more investigations were carried out using a combination of research methods to address some of its limitations.

First, while McAloon’s study identifies what tasks in Chinese are engaged by the foreigners at work on a regular basis, the current study looks at how non-native speakers of Chinese demonstrate their expertise in handling these tasks. In this sense, McAloon’s finding provides the present study with a direction in terms of locating qualified subjects and identifying events for ethnographic observation.

Second, although McAloon points out that truly advanced foreign language skill is domain-based, the five cases presented in his study fail to demonstrate any empirical evidence to show how “domain” plays a role in advanced language proficiency. Some events that McAloon observed and video recorded are not domain-specific tasks that can identify expert performances. Moreover, due to the different research questions McAloon asked, subjects he recruited as well as the data he collected did not provide sufficient samples to demonstrate what “expertise” is, which is the focus of the current study.
Last but not least, unlike McAloon’s analysis of the observed performances that is completely based on the interview-based evaluation, the current study directly looks at and closely analyzes the verbal and nonverbal behaviors demonstrated in the recorded performances in order to identify characteristics of expertise. Interview data will only be used as a reference and will only be presented when it corresponds to the researcher’s observation in the natural working setting.

3.2 Research Questions

Given the growth of Chinese language education in the US and the pressing domestic need for more non-native speakers of Chinese equipped with demonstrable language and culture expertise, it would be helpful to establish a deeper understanding of what the goal of this endeavor looks like. Through observing, documenting and analyzing expert performances in Chinese given by non-native speakers of Chinese, this study aims at identifying a list of salient characteristics of expert performances. The main research question that guides such an attempt is what specific abilities does an expert user of Chinese possess and demonstrate in his/her interaction in Chinese in a professional setting? Previous literature on expertise in non-verbal domains has worked out a list of generalizable features defining expert performance across a wide variety of disciplines. This study will use these identified general characteristics as a starting point to further examine their respective domain-specific realization in using Chinese at a working setting. Based on this general inquiry, I am going to address the following research questions in this chapter.
First, what array of linguistic and cultural repertoire do expert users of Chinese possess? Previous scholarship (Bygate 2005) points out that experts are thought to have a highly developed repertoire, structured at both macro and micro levels. Just as an expert chess player is able to remember to a high extent of accuracy the layout of the pieces of a game and an expert tennis player has mastered a large number of strategic combinations of moves, what meaningful linguistic and cultural patterns are stored in the repertoire of an expert user of Chinese as a foreign language?

Second, how do experts employ their repertoire at different levels when communicating in Chinese? Repertoire is not stored as blocks of abstract knowledge but orientations to actions, as Walker (2000) comments “you cannot learn a foreign language, you can only learn to do things in a foreign language” (p.15). Therefore, this study also aims to identify procedures underlying experts’ performance in achieving their intention.

Third, how do experts manage their knowledge and experiences in order to solve domain-specific problems? Specifically, how do experts handle routine tasks by retrieving knowledge and experiences from their memory? Does the same kind of automaticity identified in expert performances in a wide range of non-verbal domains (Chase and Simon, 1973) also manifest in experts’ use of Chinese? At the same time, how do experts improvise their performances appropriately when they recognize novel situations?

Fourth, as reviewed before, Glaser and Chi (1988) listed exceptional monitoring skill as the last characteristic defining expertise. This finding is used as a starting point in this study to examine to what extent do experts take the responsibility to monitor and evaluate their communication in Chinese. Particularly, I am interested in looking at how
experts attend to their errors in communication? What compensation strategy do experts develop during the process of monitoring and evaluating their own performances?

In addressing these questions by examining the expert users of Chinese, the ultimate inquiry is among the list of characteristics identified from the data. Which of them are domain-determined and which of them can be generalized to domain-independent strategies that can be incorporated in different levels of foreign language instruction in a program?

3.3 Research Methods

The review of expertise studies in non-linguistic domains in Chapter Two reveals three types of fundamental inquiries. The first type of inquiry targets the identification of expertise, which is to find people consistently demonstrating expertise in their domain and to describe the characteristics of their exceptional performances. The second type of question concerns the explication of expertise, which focuses on providing explanation for the presence or absence of the observed expertise. The third type of question asks how one can achieve a certain level of expertise, in another word, it focuses on the process required for developing of expertise.

Each type of questions can be answered through two characteristic approaches: the absolute approach and the relative approach. The absolute approach collects together truly exceptional people to study how they perform in their specialization, which also requires the researcher to resort to some measure of performance. The relative approach studies experts in comparison to novices, under the assumption that novices can eventually achieve expertise through engaging in a large amount of deliberate practice.
Ericsson (2004) has reminded researchers that it is important to avoid the temptation to study differences in performance between experts and novices unless they are doing the “exactly same tasks.” As this dissertation aims to identify characteristics of expertise in using Chinese as a foreign language, it requires the study to identify those naturally occurring activities that correspond to the essence of expertise within a certain domain. However, within a certain domain, it is very difficult to compare different individuals’ levels of naturally occurring performance since their tasks will differ in difficulty and many other aspects. Take for example, for a business person who has come close to negotiating a partnership with a Fortune 500 company but has not yet achieved a signed contract. Is that person’s performance poorer than that of a business person who has successfully had a contract signed with a very small company? Unless all businessmen situations involve clients with nearly identical statuses, it will be nearly impossible to compare the quality of their performances.

Therefore, this study is based on the absolute approach by recruiting expert users of Chinese as a foreign language from Chinese workplaces. The following sections will explain the recruitment of subjects, methods used to collect and analyze data in details.

3.3.1 The Subjects

Subjects participating in this study are two Americans using advanced Chinese for work on a regular basis either in a business or an academic domain. Chapter One demonstrated the limitations of the currently available standardized assessment tools such as the OPI and HSK in evaluating domain-specific expertise; therefore, the current study used a combined approach by basing the recruitment of potential subjects on both formal
assessment results and self-identification. I first identified a pool of non-native speakers of Chinese who were tested at superior level according to the ACTFL Proficiency Guidelines within the last five years. This initial stage of identification began with contacting former students of advanced-level programs and coworkers who knew of potential subjects. Recruitment postings were also made on a number of expatriate Bulletin Board System (BBS) websites and a few social networking websites. Eleven potential subjects were identified at this point. Nine of them were male. This also reinforced “the predominance of males among foreign Chinese speakers” indicated in the previous scholarship. (McAloon 2008, 218) I then sent a recruitment letter\textsuperscript{10} to everyone in the pool, requesting those who considered themselves qualified and able to participate in my study to contact me for a telephone and/or Skype interview.

Eventually seven potential subjects who at the time of recruitment were working in Chinese organizations\textsuperscript{11} and believed he or she still retained the superior-level skills in Chinese had Skype interviews with me. During the Skype interviews, the researcher usually spent the first 10 minutes to ask about the potential subjects’ Chinese-related learning working experiences, from which, the researcher is able to make a good prediction about their language level. None of the seven candidates were dropped at this stage. Following this Q&A mode, the researcher gave a general explanation about her research procedures in Chinese, focusing on the responsibilities that a subject needs to fulfill. Six out of seven\textsuperscript{12} candidates expressed their concerns after they learned that being

\textsuperscript{10} Recruitment materials are available in the appendices.
\textsuperscript{11} Among the seven potential subjects, one is working in New York City, and the other six are all working in big cities in China, including Shanghai, Beijing, Shenzhen, and Wuhan.
\textsuperscript{12} Alan was the only one who seemed fine with the research procedures. His identity as a graduate student and an artist gives him a lot more freedom than other candidates.
shadowed at work is a part of the study. Four of them gave the researcher the contact information of their respective designated Chinese supervisors, requesting the researcher to contact them so as to get permission for observing them at work. The other two of them who had the same concern preferred to ask for permission from their supervisors or colleagues by themselves. At the end of the interview, the researcher also asked each interviewee to recount their daily routines at work, especially to recall the interactions conducted in Chinese.

After approximately two weeks of the interview, three of them withdrew as they failed to get permission from their supervisors to be observed at work by an outside researcher. Reasons for not allowing such a study mainly involve the concern about confidentiality of sensitive business information and interference with work productivity. For example, one of the potential subjects who later had to quit the study is working at a non-government organization that helps Chinese workers to defend their rights. Although he is an ideal candidate for this study, due to concerns about confidentiality, after several months’ negotiation with the candidate’s supervisor, a Chinese labor activist, the researcher was only able to provide two published videos in which this candidate engaged in interviews on TV, talking about various issues related to labor rights in China. To a certain extent, data like this also reveals information about how one uses profession-related Chinese to demonstrate expertise within a domain. However, the limited number of samples as well as the lack of spontaneity in those interactions eventually disqualified the candidate from being a subject for this study.

In addition, there were another two candidates who were eventually disqualified from the study as subjects by the researcher. During the summer of 2014, the researcher
visited three out of four candidates in Shanghai and Beijing to get further information about their working environment. During this pre-study visit, the researcher found that two of the candidates actually did not engage in as many verbal interactions in Chinese as expected. One is a scholar studying late Qing history at Beijing University, who spent the majority of time working alone in the library. His major domain-related interactions take place in academic conferences, involving giving academic presentations, attending seminars, and interacting with other scholars in his field. However, unfortunately, none of the three conferences he was going to attend during that year overlapped with the researcher’s data collection schedule. The other candidate was found to be a leader of a team of professional interpreters and translators, the majority of whom are native English speakers. Therefore, interactions observed in his office involving management and translation-related business rarely involved Chinese. Although he also actively engaged in communications with Chinese business partners outside his office, most of them cannot be observed due to issues of confidentiality.

Eventually, approximately three weeks after these Skype interviews, two of them, both males, were confirmed to be the subjects of this study.

One of the subjects is Jack Mayer\(^\text{13}\), who is a CEO of an International online gaming company with approximately four hundred employees. Jack mainly works in his main office in Shenzhen, which has about two hundred employees. He also visits his subsidiaries in Shanghai, Guangzhou, Beijing, Hong Kong and Los Angeles on a regular basis. Jack estimates that approximately 80% of all his daily business communication takes place in Chinese, including communications with the upper management, outside

---

\(^{13}\) The names of the subjects, as well as their Chinese colleagues are all pseudonyms.
partners, and vendors to build and grow cooperative ventures. The other subject is Alan Bahar, who is a third-year graduate student studying Chinese painting at the Central Academy of Fine Arts (CAFA). Alan reports that over 98% of his daily work-related and scholarly interactions are in Chinese. Routine activities involve discussion and explanation of his art as well as professionally commissioned translation work. During the data collection period of this study, Alan had just accepted a one-year position teaching a Chinese art and culture class at the Central Academy of Fine Arts High School. Therefore, interactions with students and fellow teachers, preparations of pedagogical materials and lecturing on Chinese art and culture are also added to the repertoire of his domain activities.

The recruitment of subjects for this study reveals the same difficulty that McAloon (2008) discussed in his dissertation. In his study, he eventually identified 5 research subjects after contacting more than 80 potential candidates and concluded that finding qualified candidates is “extremely difficult” (p. 217). According to his explanation, the reasons for non-participation varied from no longer using Chinese for work to not being allowed to have an outside researcher recording in the work environment. Similarly, Leaver and Atwell’s study on ILR level 4 learners also indicates that the pool of level-4 language learners in general is very small and locating level-4 speakers is difficult as “no lists of such learners are kept on file anywhere” (2002, 279).

3.3.2 Data Collection

As the stated goal of this study is to identify characteristics of expert performance in a natural working setting, it adopted a combined approach of ethnographic observation
at Chinese workplaces, and Stimulated Recall (SR) tasks designed and elicited by the researcher to explore the experts’ concurrent thinking. Both methods are commonly and widely used in the expertise scholarship, with both their strengths and weakness fully recognized and discussed. Moreover, the researcher also conducted open-ended interviews with both the two American subjects and the Chinese counterparts closely working with them to contextualize the captured performances and to gain a holistic understanding of what the subjects are able to achieve beyond a finite number of collected samples.

By using such data and method triangulation\textsuperscript{14}, this study aimed to enhance the credibility and validity of the findings and to fully explain the richness and complexity of expert behavior. The following sections will discuss in details about how each type of the data was collected and analyzed.

3.3.2.1 Ethnographic Observation of Working Practices at Chinese Workplaces

Expertise is about being a practitioner at work, in situations that are experienced as problematic and formulated as defined tasks. The current dissertation seeks to document and analyze work-related performances at workplace, the place where people are attempting to get some work done, for which ethnographic observation is well suited. Unlike the studies of experts purposely arranged by a researcher in a laboratory using simulated tasks, this study observes and records spontaneous practices occurring at the workplace.

\textsuperscript{14} Data triangulation refers to using more than one data source (e.g., the American subjects, Chinese counterparts), while method triangulation refers to using more than one method of data collection (e.g., ethnographic observation, interviews, and SR tasks) and analysis (e.g., iteration, coding to theme).
workplace. Within the workplace, an expert must have accumulated sufficient experience
to be acknowledged as experts by other people with whom they interact.

Ethnographic observation was chosen over other widely used methods of
gathering data on exceptional performance, such as retrospective interviews or the diary method. These methods were not used due to problems associated with their accuracy and, therefore, validity. A few studies that attempted verification of the reported strategies by observing experts’ performances found that great discrepancies existed between the reported strategies and the observation (Watson 1913; Verplanck 1962; Nisbett and Wilson 1977). This type of discrepancy is mainly caused by misleading summaries and after-the-fact reconstructions of what participants think they must have done. Therefore, the current study attempts to avoid such potential unreliability in the experts’ retrospectively reporting a problem-solving process by watching and documenting people at work in everyday settings.

In the current study, the researcher shadowed each subject for one week. The shadowing consisted of formal observation involving video or audio recording for the majority of time as well as participant observation in a few cases where a non-participant physically present in the event is considered culturally inappropriate. Several months before the official data collection period, the researcher paid a visit to both subjects for one day to pre-investigate their respective working environment. The researcher also contacted both subjects one month prior to her traveling to China, asking them to identify

---

15 98% of the data are collected through video recording. Only two interactions were audio recorded upon the subjects’ requests. In one case, Jack expected that he would probably rebuke some employees for their poor work at a meeting. He therefore requested the researcher to audio record the meeting only to make his employees less embarrassed. In the other case, Alan was discussing a cultural issue with his classmate over WeChat, a voice-messaging app, making video recording unnecessary.
specific work/academic-related routine activities that are expected to take place during the month when the researcher would be able to travel in China as needed. Modifications were made to data collection schedule based on the answers provided by the subjects so as to maximize the usefulness and efficiency of data collection.

During the week when the researcher shadowed the subject, she showed up twenty minutes before the subject’s work day to set up the camera. The recording began whenever the subject entered a situation using Chinese to communicate. When the conversation ended or switched permanently to English, recording was stopped. At the end of every shadowing day, the performances recorded on mini-DV tapes were downloaded to the researcher’s laptop, encoded, and transferred to both a hard drive and a DVD disc to store the data. The original data on the mini-DV tapes were deleted and the tapes were then reused for the next day of shadowing.

3.3.2.2 Stimulated Recall (SR)

In this study, the method of SR is chosen over the think aloud technique to elicit verbal accounts of subjects’ thought processes at the recorded performances. Despite the popularity of think aloud techniques used in studies on expertise in problem solving, many scholars have verified their concerns about the action of verbalizing one’s “inner speech” during the concurrent performances, such as the claim that verbalizations were subject to judgment biases (Nisbett and Wilson 1977). Even though Ericsson (2006) later has claimed that the think aloud technique will not change the underlying structure of the thought processes and thus avoids the problem of reactivity, i.e., the act of generating the reports may change the cognitive processes that mediate the observed performance, it is
obvious that think aloud techniques are at least not usable in situations in which significant verbal interactions are part of the individual’s behavior during the observed performance.

The other reason for choosing SR for this study is that it can largely benefit education/training as the recorded introspection portfolio can be used as effective training mechanisms to help novices learn through strategies used and commentated on by experts. As Lyle has suggested, SR provides a valuable research/educational tool to identify critical cues in some “difficult cases” used in simulation exercises with which to “train perception for pattern recognition” (2003, P.874).

Therefore, in this study, following each recorded event, the video was played back to the subject. While watching the video, the subject then was expected to recall what he had been thinking during the interaction. Although ideally this process needs to be done immediately after a certain event is video recorded, in some cases, due to the tight working schedule of the subject, the SR tasks had to be completed several hours after the recorded event. Among the 13 performance samples that are used as stimulus in SR tasks for Jack, the majority of them (i.e. 9 tasks) were completed at the end of the day of observation, usually within the 8 hours after the occurrence of the recorded event. The other four tasks were completed the following day. As for the other subject Alan who has greater flexibility in his schedule, almost all of the SR tasks (i.e. 11 tasks) were completed right after the end of the event. This study recognizes this existence of delay in completing the SR tasks and considers it as an important limitation with respect to the process of eliciting expert thoughts underlying their behavior. However, in terms of the data obtained, there were no discernible difference between the two subjects’ SR protocol.
as they both were able to recall in great detail and demonstrated a high level of self-awareness about what they have done and what they intended to achieve. This in itself suggests one characteristic of expertise in using foreign language for problem solving that will be discussed later in Section 3.5.

It is worth noting that the SR data in this study is complementary to the recorded interactions from the ethnographic observation and in the development of the paper there has been an interplay between the two data sources. Subjective reports in the SR process have suggested things to look for in the recorded interactions. Phenomena initially identified from the conversational transcripts have often been illuminated by experts’ comments in the SR protocol.

### 3.3.2.3 Interviews

Interview, or more specifically speaking, retrospective interview, has also been used as an imperfect but necessary method of investigation for expertise studies. Interview studies allow the researcher to discover the development of expertise from a long-term perspective, calling attention to researchable opportunities for other investigations. In some cases, subjective retrospection given by the subjects may also challenge the directions headed by researchers under the situation of limited data or constraining timeframes. The limitation of interview studies have also been well recognized as Freeman (2000) points out, we have to recognize that “we can never identify and measure the full context of anyone’s life, even in the present, and interpretation of data can only be as well informed as possible” (p.236).
The current study therefore uses interview as the third type of data source, being complementary to the recorded interactions and the SR protocol. There are two types of interview questions, respectively designed for the two American subjects who are shadowed in this study and their respective five Chinese colleagues or supervisors. Interview questions for both types were formed long before the fieldwork so as to meet the Institutional Review Board requirements. Questions for both structured interviews are in the Appendices E, F and G.

Questions for the two American subjects mainly focus on their personal background, Chinese learning history and their current work in China, concerning the nature of their Chinese language usage. Specific questions were also asked to elicit a few retrospective statements, such as the most successful profession-related experience that marks their expertise in using Chinese, encountered misunderstandings or, in a general sense, failure that indicates the inadequacy of their ability and cases of engaging negotiations with native Chinese in either professional or non-professional situations. Interviews with both subjects were conducted in Chinese on the last day of their respective shadowing period and audio recorded.

Questions for the Chinese subjects include personal background information, their working relationship with the American subject, and their evaluation of the American subject’s Chinese use at work. Similarly, Chinese subjects were also asked specific questions to elicit some retrospective statements, such as the American subject’s behavior that marked his exceptional skills in using Chinese at work and the American subject’s behavior that caused misunderstanding or made Chinese people feel uncomfortable.
3.3.3 Method of Analysis

In order to identify the features of expertise demonstrated by foreigners at work and to examine how these identified characteristics are perceived by native Chinese working with them, this research sought to provide “thick description and grounded interpretation” (Prior 1995, 321) of experts’ performances and their corresponding expert thoughts underlying everyday work. For data analysis, the commonly used iterative and progressive data analysis method was employed. Reflexive iteration allowed the researcher to visit and revisit the data and to connect them with emerging insights, progressively leading to refined focus and understanding about what is going on in experts’ work-related performances. Recorded videos from the natural ethnographic observation, SR protocols and audio-recorded interviews were all transcribed, and data was coded and categorized relative to the focus of the research questions presented earlier. The researcher compared and contrasted patterns generated through the analysis of different data sources, aiming at unveiling overlaps as well as inconsistencies.

In this study, recorded performances collected from the ethnographic observation served as the main set of data. Characteristics underlying an expert performance were first identified from the observed and recorded performances. Following such a process, comments on what constitutes the expertise of the subject collected from both interviews and SR protocols were included if they corresponded to the identified characteristics. In another word, characteristics of an expert performance recalled by an interviewee but not reflected in the observed performances were not raised as a final finding. However, it would suggest a direction for further study when more relevant data from real-life performances are available. Moreover, by exploring the dialectical relationship between
what the data are telling me and what I want to know according to the research questions guiding the current study, this study especially examined how the experts’ accounts of their thinking process revealed in the SR tasks correlated with the actual effects of their performances and the native Chinese’s general evaluation of their Chinese skills. The interactions among these different types of data will further address the self-monitoring skill of an expert Chinese learner. Moreover, this study is also interested in how those features of expertise identified from experts’ recorded performances echo the seven general characteristics identified by Glaser & Chi as discussed in chapter one.

3.4 Profiles of Two Expert Users of Chinese

3.4.1 Jack Mayer

3.4.1.1 Chinese Learning Background

Fascinated by some Chinese acrobats’ performance in a circuit show that he watched in his childhood, Jack started his journey learning Chinese in 2004 at the Ohio State University (OSU). Having been obsessed with Japanese and Korean video games, Jack was also motivated to learn Japanese and Korean at the same time. After four years’ undergraduate study at OSU, Jack continued to pursue his master degree in Chinese Flagship program at OSU. During his six years’ Chinese learning career in college, he attended intensive in-China study programs for four consecutive summers, three 8-week programs in Qingdao and one 6-week program in Beijing. During his graduate study, Jack also studied abroad in Nanjing University for five months from 2007 to 2008, where
he took several graduate-level classes in his domain with local Chinese graduate students, studying various subjects such as Social Psychology, Issues in Internet Media, and Chinese Social Behavior Analysis. During the same year, he had a chance to work as an intern in a factory of an American shoe corporation, located in the city of Dongguan in Guangdong, China. Although the company was just run by a friend of his family, Jack played an important role in the company, facilitating communications between US company offices and Chinese factories and performing quality inspection on shoe shipments coming out of multiple factories across southern China.

Jack was tested “Superior” on the ACTFL OPI test after three years’ learning Chinese in October 2007. In the same month, he also obtained a level 3 ranking on the US government’s ILR test. In 2008, Jack graduated from OSU with a BA in Chinese, a BA in Japanese and a MA in Advanced Chinese Language and Culture — all concurrent degrees Jack is currently pursuing and EMBA through the Cheung Kong Graduate School of Business (长江商学院), where he is the only Westerner participant.

3.4.1.2 Domain-related Experience

Jack’s career originated in his long history of engaging in the virtual worlds (online games), during the period of which his original dream was to become a professional online game player. However, it was during his internship in the shoe company in Dongguan in Guangdong, China, that he started buying and trading virtual goods and services throughout various online games to kill the boredom of being trapped in a non-air conditioned factory dorm and to make some pocket money in the meanwhile. Realizing that he could make a decent amount of money from the real money trading
(RMT) profession, Jack quit his internship in Dongguan and went to Nanjing to start his first Chinese gaming studio. This became the embryo of his current business.

In 2010, Jack started a Web and mobile game publisher operation, specializing in free-to-play games as its co-founder and CEO in Changsha in Hunan, China. After several years’ expansion and moving his firm to Shenzhen, Jack currently has approximately 400 employees, 200 in the main company while the other 200 in a few subsidiaries, spread in Shanghai, Beijing, Guangzhou in China and Los Angeles in the United States. The main business of his company is to bring Chinese games to Western players, operating Chinese games worldwide including English, French, German, Spanish, Turkish, Portuguese, Arabic and Russian markets. Jack was included on the “2014 Forbes China 30 under 30” list, for which he delivered a 30-minute speech in Chinese at the Forbes forum about how he launched his career at the Forbes forum.

3.4.1.3 Motivation for Learning Chinese and Working in China

Jack developed his motivation to learn East Asian languages from his rich experience of playing East Asian video games since he was a teenager. However, he did not develop his particular interest in learning Chinese until he had some first-hand experience in China and witnessed the great changes it had gone through within the past decade. Jack has been considering China as a place of huge growth that offered plentiful potential opportunities. This, along with the growing business of virtual trade industry in China, made him decide to “stay in China forever”. Jack also values the time and effort he has committed to learning Chinese, as he commented in the interview, “I have spent seven years in college, full time. If you have committed that much of time and energy to
learning Chinese, then I don’t understand the point of working in US. If you don’t have the environment, you will forget it, or you cannot sustain advance (sic).”

Without being asked, Jack especially shared his observation of some Chinese-speaking Westerners’ leaving China after staying there several years as they think the place is too “uncultured”\(^{16}\) to withstand. Jack admitted that he also went through a similar stage, such as complaining about the air pollution all the time or having problems doing business in Chinese culture. However, he was eventually able to get over all this after being in China for three or four years. He attributed this to the fact that he went to China at a relatively early age so that his brain is unformed enough to be open to a new life and new ideas.

Jack also attributed his motivation for learning Chinese to his personality. He described himself as shy and introverted, especially in his earlier ages as a student. He has been feeling more comfortable around Asian people, thinking he actually has much more Chinese friends than American friends. Jack identified several Chinese he thinks he trusts with his life and his money, including the other two co-founders of his current company. He also started to make friends with people from other industries after starting his EMBA in Beijing.

---

16 In the interview, Jack said: “I think a lot of people. They have been here for 5 years or 10 years when they see people spit on the ground they think China is a terrible place. This does not necessarily take place in US, but so what? You think it’s the right way to do things. Something you just have to not care too much. Screw China, I wanna go back home. I think it’s a waste of opportunity. Different people have different mindsets. It takes a very deep cultural understanding to be able to adapt to it.”
3.4.2 Alan Bahar

3.4.2.1 Chinese Learning Background

Alan’s Chinese learning experience started with a two-month self-study, in an immersed target culture environment in the summer of 2009. During that summer, upon an invitation of his martial arts teacher in San Francisco, Alan spent two months with a local official’s family in a small town in Shanxi, China, teaching English and martial arts to the family and some other local students. Before arriving in Shanxi, he purchased a Chinese 101 textbook from the university bookstore, taking it with him to China. The book, he recalled, became the tool he heavily relied on in order to communicate with the locals around him who could barely communicate in English. Alan returned to his college, San Francisco State University (SFSU), skipping the first two quarters’ Chinese classes and directly enrolling in Chinese 103. In addition to the Chinese 103, he also took a Flagship prep class, getting ready for the SFSU Chinese Flagship Cohort program that was still embryonic at that time. By becoming a Flagship student, he was able to have a one-on-one conversation tutor, which he recalled was very helpful at that time. In the summer of 2010, Alan went to a study abroad program at the Ocean University in Qingdao for two months. Although he was misplaced into a lower level at first, Alan was able to convince the academic director to be switched to the highest level in the program, taking an integrated Chinese class with a group of pre-graduate Flagship students and many heritage senior undergraduate students. He went back to SFSU after two months’ study in Qingdao as a senior, finishing his last quarter’s study at SFSU. During his last quarter in college, he received an exceptional treatment to be exempted from all the
regular Chinese courses that he was supposed to take, but to take a one-on-one tutoring session instead with a Chinese professor to read a few classical and modern Chinese authors’ works, such Lu Xun (鲁迅) and Yi Zhongtian (易中天). In the winter of 2011, he was sent to the Flagship center at Nanjing University, where he completed his Flagship program, graduating with a BA in French and a BA in Chinese. Alan was tested “Superior” on the OPI test in May, 2012, three years after being thrown into Xiaoyi, Shanxi with a basic Chinese textbook.

3.4.2.2 Domain-related Experience

Unlike Jack who developed his interest in doing business in the game industry at a very early age in his life, Alan was not pushed into visual art until he graduated from college and arrived at China. In Alan’s own words, his obsession with visual art was coming out of “a kind of fatigue with verbalization, fatigue with narrative, fatigue with writing and fatigue with language”, which eventually took over his original plan to become a novelist. After finishing his intensive Chinese study and internship in Nanjing University and graduating from the Flagship program at SFSU, he moved to Beijing and started a short period of preparation for the graduate school entrance examination in order to apply for the MA program at the Central Academy of Fine Arts (CAFA). During those months, he primarily engaged in tutoring sessions with local art teachers to acquire basic painting skills, such as sketching. He also self-studied through reading the mainstream theories of art that prevails the Chinese academic world. Admitted into the program in 2013, Alan has been studying Chinese painting under the supervision of a well-known Chinese artist and is expected to graduate with his master degree within three years.
During his study at CAFA, Alan was also actively engaged in domain-related practices including exhibitions, writing art critiques for the mass media and even teaching. He was hired as a part-time art teacher at the High School of Fine Arts in Beijing, which is affiliated to CAFA, teaching “The Contemporary Application of Chinese Traditional Culture”. Through all these various types of domain activities, Alan gradually became “so CAFA” in his own words, building a strong interpersonal network with Chinese artists and even authorities in his field.

3.4.2.3 Motivation for Learning Chinese and Staying in China

Pursuing Chinese language and culture while becoming an artist in China was “an intellectual, emotional and philosophical decision” for Alan. Alan’s interaction with Chinese culture started with his first semester at college, when he began learning martial arts to fulfill the physical education requirement. It was the martial arts learning experience that sparked his interest in looking at Chinese culture as a comprehensive system of everything from food to physical culture, from medicine to visual arts, motivating him to continue to take numerous elective classes in Chinese culture at college. Alan attributed his fascination with Chinese culture to his strong desire to explore an alternative to the Western culture that he was born into. He raised two factors that intensified his desire: the identity crisis he had experienced and the culture community in Los Angeles that he was born into. These two factors had been interplaying in his life, as he commented below:

So growing up in that environment especially with the influence of Hollywood, and consumption and luxury goods and image, and, you know, basically everything that is talked about postmodern philosophy is manifested in the sense
of Los Angeles as an urban landscape. I’ve always been suspicious of it and have some critical perspective on American culture. Especially my father is from Iran, so that I have the constant tension between the American culture and the Eastern culture, which is “The Other”. This tension is what made me. My mother is Italian American. So I am always curious is there an alternative to Western culture. I mean Western culture is obviously the most powerful, economically and politically force in the world, especially with America as the leading nation, as the leading super power. At the same time, I found this culture very sick. It’s a very disease culture. There’s so much unhappiness in it. You grew up with it. You wondered is this “the culture” that the whole world views itself as number one. Is it really the best thing? My interest in Chinese culture started as an exploration of an alternative to this worldview.

Driven by this strong desire to understand Chinese culture, Alan sees the mastery of Chinese language as a necessity for him to avoid viewing the Eastern culture superficially as Eastern Spirituality, which he thinks is very common in California with the hippie movements in the 1960s. Alan specifically mentioned that being able to read Chinese classics such the I Ching, and the classics of Confucianism and Taoism without relying on translation allowed him to be very rigorous with Chinese culture. Alan views the Chinese literati culture as a path for him to pursue as a young artist and scholar.

Apart from his intellectual reason as a young scholar, Alan also mentioned his philosophical, or, in a narrow sense, religious reason for delving into the Chinese culture. Influenced by his parents’ very different and even contrasting religious background — a very developed Muslim and a very devout Christian — Alan has been seeking religion for himself. He finds China in its religion of life, teaching him everything from what one should eat in wintertime to how one should behave as a moral person to how one can deal with people, society and the universe. Alan said he found Chinese culture had very interesting perspective on these issues, which seemed to indicate a perfect way for him to live a balance life.
3.5 Identifying Expertise: An Analysis of Professional Use of Chinese at Work

3.5.1 Experts Possess A Highly Developed Communicative Repertoire

Observations of the two American subjects’ interactions in their workplaces and interviews with the Chinese counterparts working with them both reveal that expert users of Chinese possess a highly developed communicative repertoire consisting of a wide range of informal uses of lexis and phrasing including regional and dialectical usage and very formal expressions such as occupational jargons, literary allusions, and formulaic polite speech (客套话). Beyond the lexical level, experts are familiar with a variety of discourse genres and a wide range of conversational strategies. They also have a well-developed awareness of potential registers, allowing them to decode and encode parts of the discourse effortlessly. Observed from the two subjects’ interactions at work, they seem to have few difficulties in accessing appropriate vocabulary, and linking it up appropriately to their communicative intentions. At the same time, experts are also able to manipulate these discourse features skillfully in different contexts to signify personal involvement including individual identity, group membership as well as attitudes and emotion. Their skillful appropriation of the discourse features are motivated by their personal preference to some extent, but in most cases, are motivated by the dynamics of an unscripted conversation. A detailed analysis of an expert-like communicative repertoire is presented in the following sections.
3.5.1.1 An Untextbookish Colloquial Style

Video data collected from both subjects’ workplace settings shows that both subjects’ oral interactions resemble numerous features of language that are typical of, and in some cases unique to, spoken language. Some usages of language may even be considered ungrammatical, and, therefore, are usages not easily found in pedagogical material. Such usages are very likely to be acquired through rich post-program immersion experiences. Adopting such a colloquial repertoire is considered by native speakers as “not easy (不容易)” and “different from the majority of foreigners in China” (和其他都不太一样吧).” As one of Jack’s Chinese colleague explained:

“I think other foreigners’ speech are very methodical. I wonder if it is because that is what their teachers taught them to be. Or, it may resemble the textbook recording. Just like when we learned English, a lot of things that we learned from the textbook will never be heard in US. But Jack is different. He talks in the same way as we do. You don’t need to accommodate him. It just never sounds weird. No difference. It’s not an easy thing to accomplish.”

A close look at the two subjects’ oral interactions finds that their “non-textbookish style” speech consists of a large number of ellipses, hedges, inversions and discourse fillers. The two subjects are capable of manipulating these colloquial features to achieve various communicative intentions at a level below the conscious.

When complaining about an unwise decision made by the school officials on arranging exhibits, Alan disclosed his anger to an art museum staff by using inversions at

17 The original comment was made in Chinese: 就是其他老外吧，他们说话，反正，反正就是怎么说呢，挺一板一眼的。是不是我想就是他们老师教的就是那样的，或者课本录音可能就是。就像我们学英语的时候，课本上学的很多东西，你在美国听不到别人说嘛。但 Jack 不一样嘛，他就跟我们说的完全一样，你不用去适应，就是从来没有什么别扭的感觉，没区别。不容易。
several places. Although those inversions he used may read ungrammatical, it was the frequent use of these inversions that made the overall speech sound very natural and effective in exerting an appropriate level of emotion in Alan’s talk. The following excerpt is taken from their conversation:

<table>
<thead>
<tr>
<th>Time</th>
<th>Role</th>
<th>Chinese Transcription</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:00</td>
<td>Art Museum</td>
<td>为什么要放照片呢?</td>
<td>Why do you want to display photos?</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:02</td>
<td>Alan</td>
<td>我们也不懂真的，我们领导说的</td>
<td>We have no idea, really. The official of our school said that.</td>
</tr>
<tr>
<td>5:03</td>
<td>Art Museum</td>
<td>我们做了那么多展览，没有放照片的</td>
<td>We have done so many exhibitions but we have never displayed photos.</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:04</td>
<td>Alan</td>
<td>我也觉得很业余搞的我告诉你</td>
<td>I also think it very unprofessional, they made it, let me tell you.</td>
</tr>
<tr>
<td>5:06</td>
<td>Art Museum</td>
<td>对啊，所以你们就说服你们的领导，别让他放照片。就说美术馆那么多展览没放过照片的。</td>
<td>Right, so you just persuade the official. Don’t allow him to display the photo there. Just tell him among all the exhibitions held in art museums; no one has ever displayed photos there.</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this conversation, Alan strongly expressed his objection against the school official’s decision about including photos of the artists in the exhibition tags. Alan thought the inclusion of photos would not only take up too much space, but also made the exhibition look unprofessional as it is uncommon to display the artists’ photos next to their exhibits. The inversion of “真的 really” here emphasizes Alan’s strong disagreement with the official’s decision in his response to the art museum staff’s question “why do you want to display the photos?” When the art museum staff implied the unprofessional aspect of displaying photos, Alan immediately recognized her implied
intention by using another predicate inversion “我也觉得很业余搞的” to signify his strong dissatisfaction. The inverted verb “搞的” representing a low level of formality not only naturally fitted into Alan’s colloquial style but also perfectly signified the unprofessionalism of the school officials. The third inversion of “我告诉你 I tell you” used in Alan’s speech is a common way to express strong emotion such as anger, disappointment, dissatisfaction and warning in Chinese. However, Alan’s strong expression of his dissatisfaction seemed to be completely recognized and accepted by the art museum staff who first showed her agreement with Alan and then encouraged Alan to negotiate with the school official.

It is reasonable to predict that none of the three cases of using inverted structures in speech will be introduced in pedagogical material. McCarthy and O’Keefe (2004) suggest that “dialogues produced for classroom use are for the most part scripted”, which results in materials that “lack core spoken language features such as discourse markers, vague language, ellipses and hedges” (p.29). More importantly, these types of untaught usages of language are markers of the subjects’ advances in their communicative repertoire. This also corresponds to what Kubler said about pedagogical material. According to Kubler, advanced level instruction should include very colloquial and very formal terms known by any educated native speaker while the beginning and the intermediate level of Chinese instruction is “neither particularly colloquial nor particularly formal.” These he calls “unmarked” style (Kubler 2004, 127). From the interactions observed, we can see that as a result of years of post-program immersion in the target culture, both subjects have gone beyond the stage of using “unmarked style.”
3.5.1.2 Dialectal Expressions and Accents

The two subjects’ colloquial style is also highlighted by their flexible use of dialectal expressions and their skillful imitations of regional accents. Alan is observed switching adeptly among Mandarin, Beijing accented Mandarin and “vulgar” Beijing dialect, which Alan referred it to 胡同串子 (hutong chuanzi, alley argot) Beijing dialect. Jack can speak dialect of Shangsha, capital of Hunan province, to his colleagues and business partners from Hunan province. He even serves as the Mandarin-Changsha dialect interpreter for them as many of his Chinese counterparts cannot speak standard Mandarin. Such ability to incorporate dialectal variations in one’s speech is considered difficult to obtain even for native speakers of Chinese. This ability makes the two subjects sound “smart (机灵),” “amiable (亲切)” and “humorous (幽默)” according to their Chinese counterparts. Moreover, both subjects are very aware of their ability to understand and to speak Chinese dialects and can always exploit it to their own advantages.

Commenting on one of his performances teaching in the classroom setting during the SR interview, Alan also recalled how he was able to achieve different communicative goals by manipulating different accents in his speech:

[In this video] my tones are pretty crisp. I think I sound very Beijing. Being a teacher is more like being an actor, you know. That’s what I am saying. I have worked very hard to get a tone like the Beijing teacher. Teaching is very theatrical. And having that tone of voice, I think, is important to establish a certain level of authority. So I mean even it’s a little, even you think my comment is superficial, there’s something achieved by crisp tone and Beijing accent. Especially I sound like, when I talk in front of 60 people, my pronunciation is modeled on CCTV, like a loudspeaker…Beijinghua (Beijing dialect) has many different versions. People went to university beijinghua, people who sitting in hutong (alley) all day playing cards beijinghua. I am very sensitive to different kinds of dialect. There’ rustic beijinghua there’s vulgar beijinghua. The video in which I made an
announcement in class, I was speaking Beijing-centered Putonghua (Mandarin) in the classroom environment.

In this review, Alan was fully aware how the CCTV-like Mandarin or what he referred as “Beijing-centered putonghua” helped him establish authority in classroom. Besides, he also gave examples of how he naturally went back and forth between rustic Beijing dialect in small talk with the ticket people at Jingshan Park and standard Mandarin when chatting with the classmate going there with him.

In Jack’s case, he realized that it is advantageous for him to speak Chinese with a southern Chinese accent as Chinese southerners are always prototypically considered as more capable in doing business. A close look at Jack’s speech finds that his speech is filled with the word “ma 嘛”, a typical discourse filler used in Hunan dialects. Compared with Alan, Jack seems to be less capable of controlling his accent, which in fact became something he has been arduously working on after he found out he said “ma 嘛” too many times in a relatively formal talk he gave. During the time when the researcher shadowed and interviewed Jack, he said he made progress in gaining control of using “ma 嘛” in speaking and started to attend to other minor aspects of his accent in speaking. Although Jack is not satisfied with the fact that his Hunan accent is sometimes overwhelming to the listener or even to himself, he still views it as a powerful weapon for him to signify either an identity of an astute businessman or an image of a good buddy to his southerner colleagues and friends.

Even in some relatively formal situations, Jack knows very well about how using a certain dialectal expression can potentially bring a sense of humor to foster a lighter
atmosphere and to make himself appear intelligent. In a Chinese Competition TV show he attended as a young successful businessman, Jack played that card in his response to a tough question deliberately asked by one of the judges, an excerpt of which is presented below:

<table>
<thead>
<tr>
<th>Time</th>
<th>Role</th>
<th>Chinese Transcription</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>03:21</td>
<td>Judge Gu</td>
<td>对，你现在主要还是做代理商吧？</td>
<td>So now what you are doing is still like an agent, right?</td>
</tr>
<tr>
<td>03:24</td>
<td>Jack</td>
<td>我们做代理商，但我们不是提篮子的。我们是有自己的平台。</td>
<td>Well, we are agents, but we are not doing tilanzi. We have our own online platform.</td>
</tr>
<tr>
<td>03:27</td>
<td>Judge Jin</td>
<td>不是提篮子的</td>
<td>Not doing tilanzi</td>
</tr>
<tr>
<td>03:28</td>
<td>Judge Gu</td>
<td>哈哈</td>
<td>I see, HAHA (Laugh)</td>
</tr>
<tr>
<td>03:29</td>
<td>Jack</td>
<td>所以游戏是我们自己运营的。我不会签你这个游戏。我二十万签你的游戏，我们是不做这种生意的。</td>
<td>So we operate the games ourselves. I won’t buy a game from your company at 100,000 and then sell it to another company at 200,000. We are not doing that kind of business.</td>
</tr>
</tbody>
</table>

During this session when judges are asked to raise challenging questions to the contestants, one of the judges, judge Gu, questioned Jack’s business mode, asking if what he did is just simply an agent’s work. Although the judge used the relatively neutral term “代理商 agent”, Jack quickly sensed the judge’s negative implication and responded with an emphasis that his business is not doing tilanzi 提篮子 as those judges might have suspected. The Hunan dialectal expression tilanzi that literally means “carrying a basket” carries a defiant tone about what a middleman does. By emphasizing they are not doing tilanzi, Jack not only clearly conveyed the message that he recognized the hidden meaning of that question, but also signified a humorous and relaxed image that a TV
show greatly values. Moreover, Jack’s witty image was also recognized by both the judge and the MC marveled at the fact that this foreigner even knows dialectal expressions as they immediately repeated the particular expression *tilanzi* he used.

3.5.1.3 *Domain-specific Terminology*

Both subjects are able to demonstrate a sophisticated knowledge of the terminology that is current within their respective domains, and are thereby able to establish their epistemic authority. In Alan’s case, his skillful use of the jargon of the art world along with his ability to explain those expressions in a readily understood way to his students impressed the Chinese people working with him. Many examples of this can be seen in his communication with students in the classroom setting. In one of the video scenes in which he was giving instruction to students’ artwork, Alan clearly explained to students about the differences between two types of paper specifically used for Chinese painting “元书纸 Yuan-shu paper” and “毛边纸 Mao-bian paper” and thereby pointed out how the different textures of different types of paper will affect the quality of painting.

In Jack’s case, his daily communication is filled with jargon used in the gaming industry as well as a rich store of the cyber slang that is frequently used by young professionals in that field. All the five Chinese colleagues I interviewed particularly commented on Jack’s abundant knowledge of some games’ Chinese names, which many Chinese professionals in that field may have no idea about. From the collected data, Jack’s exceptional familiarity with domain terminologies was observed in several situations, all contributing to his authority as a CEO of a game company and his professional image as an astute businessman.
Below is a conversation at one of the senior staff meetings, in which Jack was reporting back from a business trip experience to his colleagues. In this conversation, Jack was discussing with some senior staff about the potential of buying the intellectual property of a Chinese game and making some adaptations based on it to make money. When his staff Manager Tang talked about how profitable this Chinese game is as their monthly income in November was 160 million, Jack immediately corrected him that it was not the income, but the *liushui* that was 160 million. The term *liushui* that Jack preferred over “income”, literally meaning “running water”, but is actually a jargon in his field, indicating an exaggerated turnover self-reported by a company. Jack’s cautious choice of words implied his reserved opinion about that company, but also created a positive image of a rigorous professional in the gaming industry. Another game-field jargon he used later was *huanpi*, literally meaning “changing the skin”, which is a common process in which game or app developers do not have to create a game from scratch by taking a premade source code of a game and changing its design completely to give it a brand new look. Both “*liushui* (an exaggerated turnover)” and “*huanpi* (changing the skin of a pre-existing game)” were particularly identified and commented on by Jack’s Chinese counterparts in the interview, who think they both contribute to Jack’s image as a rigorous boss and a sharp and intelligent professional in the gaming industry.

<table>
<thead>
<tr>
<th>Time</th>
<th>Role</th>
<th>Chinese Transcription</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:09</td>
<td>Jack</td>
<td>他们那个三国之刃这个产品的一个收益利润，他们好像……</td>
<td>The profit rate of their product</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Sanguozhiren</em> (the name of a game) seems not…</td>
<td></td>
</tr>
<tr>
<td>1:14</td>
<td>Manager Tang</td>
<td>还没公布</td>
<td>Was not made public yet</td>
</tr>
<tr>
<td>1:14</td>
<td>Jack</td>
<td>对，还没公布</td>
<td>Yes, was not made public yet (Cont’d)</td>
</tr>
</tbody>
</table>
1:15  Manager Tang  实际上 11 月份他们收了 1.6 个亿  Actually their income was 160 million in November

1:21  Jack  对，11 月那个产品……  Yes, in November, that product…

1:22  Manager Qi  赶快去开户啊，哈哈  We should hurry to open a (stock trading) account, haha (laugh)

1:24  Jack  产品，产品的流水是 1.6 个亿，不是他们收入是 1.6 个亿  The liushui (turnover) was 160 million, not their income

2:05  Jack  现在这个产品，虽然这个横版格斗这个模式不是特别看好，但是呢我觉得有一个机会，就是第一，这个游戏在国内是排名第一的，那我们现在正好是在跟 FOX 还有华纳这两个 IP 商谈一些合作。那我觉得正好如果我们去跟这个 IP 商去说，诶，这个国内排名第一的产品我们可以去换一个海外 ip 的皮，我们去搞一个欧美版本，他们应该是会特别感兴趣的。  The current product, though its model, the horizontal-version of a fighting game, is not that promising, I think there’s a chance. First, this game is ranked first in China now. We happen to be negotiating cooperation with Fox and Warner, this two IP (Intellectual Property) companies. We can tell them that we can huanpi (change the skin) this first prize game, making it into an American version. They should be particularly interested.

3.5.1.4 A Rich Store of Cultural Knowledge

In McAloon’s study, which we previously reviewed, he found that Chinese learners would professionally benefit from the ability to make cultural references, which seemed lacking in his subjects’ performances. In the current study, both subjects were able to demonstrate their deep understanding about Chinese culture, which not only can be observed from the recorded performances they gave, but are also very often mentioned by the Chinese people working with them.
The current study found that what caught Chinese people’s attention and impressed them most was a foreigner’s abundant knowledge of the geography or the history of China. A foreigner who is able to deal with a sizable amount of knowledge in these two fields are considered as someone who has a high appreciation of Chinese culture and has enthusiasm to delve into the diversity of Chinese culture that is a source of pride for most Chinese. Both subjects demonstrated their knowledge about Chinese geography in dealing with people from different areas of China. An assistant in Jack’s company recalled the first time she introduced herself to Jack, telling him her hometown was in Jiangxi province, Jack quickly responded that he knew that her hometown had very fancy china pottery as the town of Jingde (景德鎮) in Jiangxi is known as the Porcelain Capital of China. Jack’s comment impressed his assistant not only for the cultural facts he accurately knew, but also for the appropriate way in which he made connection with her in the conversation. In other words, what may impress the Chinese interlocutors is not only a foreigner’s knowledge about achievement culture, but more importantly, their knowledge of the behavioral aspect of achievement culture. “Even some native Chinese may not know that much about what Jiangxi has”, said the assistant in the interview. Alan’s professor and classmates also made similar comments. One of the professors mentioned that he was particularly impressed by Alan’s ability to predict someone’s hometown when he met Chinese for the first time and how he could always make relevant and sophisticated comments on the cultural features associated with that place. This type of small talk happened a lot in both subjects’ interactions, shortening the social distance between them and their Chinese counterparts. For example, Alan’s Chinese classmate expressed the same kind of surprise when Alan was trying to figure
out the exact character used in someone’s name during the first-time meeting. Alan’s friend was impressed by the fact that a foreigner could correctly say the name of an ancient state in Chinese history, although it was still basic to a native Chinese.

Once at my classmate’s exhibition, he was talking to someone there, and I was shocked by his Chinese. So he asked me ‘what’s your name’, and I said ‘Feng Shan’, then he asked me if my ‘Shan’ is the same ‘shan’ in shanshanlaichi (a Chinese idiom meaning ‘coming late and making people wait’). Then, I was thinking, wow, he even knows Chinese idiom. Also, I have a friend whose name is ‘Bu Qi’. Alan asked my friend if his ‘Qi’ is the same ‘Qi’ [as] in ‘Qi-guo’ (Qi State, an ancient state during the Zhou Dynasty). I was so amazed at that time. An American even knows the Qi State!18

Moreover, from the above comment, we can also find what impressed the Chinese audience is not only a foreigner’s knowledge about history, but also their recognition of those achievements of that culture due to China’s long history such as idioms, ancient poems and classical texts. One of Alan’s classmates evaluated Alan’s Chinese as “better than ordinary Chinese” as he has large vocabulary as well as a deep commitment to classical Chinese19. She further mentioned that Alan always quoted some poems to make a point or to describe his feelings, some of which may even be unfamiliar to Chinese people. Besides reciting a poem, Alan is also good at interpreting the poems for his Chinese friends using modern Chinese language, which sometimes may even have embarrassed some Chinese people who are not able to do it in the same way.

18The original comment was made in Chinese: “就是有一次同学的一个展览上边，他在那儿讲话，他把我惊到了，就是他问我叫什么名字，然后我说“冯珊”，然后他说“是姗姗来迟的姗吗？”然后，然后我当时就想，哇噻，他那个回答得，就是他那个成语什么的都知道。然后我有一个朋友，他叫那布齐嘛，然后他就问，说“你是那个齐国的齐吗？”，所以我当时顿时就惊呆了，就是一个美国人竟然他知道齐国！”
19 Her original comment was made in Chinese: “(他的中文)有可能比一般的中国人要好。指的是他的词汇量，包括他对中国古代文化的涉猎。”
While impressing their Chinese counterparts with their deep knowledge of Chinese civilization, the ability such as reciting a Chinese poem or using sophisticated discourse filled with Chinese idioms in fact also has a domain-dependent value. For people like Alan who work in an academic field of Chinese humanities, such behavior will definitely contributes to the image of a scholar who is highly literate. But for professionals like Jack who work in a relatively young, relaxed and energetic field such as the gaming industry, there seems no high expectation for him talking in a highly rhetorical way. Going through Jack’s interactions with Chinese people at work reveals that although he always sounded intelligent and professional as analyzed above, he rarely used Chinese idioms or quoted some Chinese classical texts to make his point. The only example found in his recorded interactions is his quoting one entry from The Analects when he attended a TV show to display his modesty. He explained to the MC at the beginning of the show that why he attended this show was because he wanted to learn from others as Confucius say “when a group of three people walk together, you can always find your teachers present” (三人行必有我师焉). Although Jack expressed his wish to improve himself so as to sound more literary in Chinese, he also admitted that he would only keep that as a personal ambition. It will not change his way of talking during work, as he should attend to his Chinese audience, the majority of whom do not talk in a highly rhetorical way.
3.5.2 Experts Process the Communicative Repertoire in An Automated and Creative Manner

Findings revealed in 3.5.1 partially correspond to Byrnes’s 2002 definition of the Advanced L2 (AL2) user as “someone who is able to draw on a sizable repertoire of fixed or chunked language forms that will ensure the sociocultural appropriateness and acceptability of utterances while, simultaneously, enhancing their fluency” (p.51-52). However, the possession of a rich communicative repertoire alone cannot guarantee expertise. Byrnes further points out that an AL2 user is “someone who also has a good command, including a significant level of metalinguistic awareness, of the meaning-form relationships that are encoded with various degrees of fixity and fluidity at the lexicogrammatical level” (p.52). The so-called “metalinguistic awareness of the meaning-form relationships” in Byrnes’s definition can be observed in both subjects’ on-going discourses, in which they were able to integrate low and high level skills of communication. In the following sections, I focus on two complimentary aspects that highlight the expert way of communication: Automaticity and Creativity.

3.5.2.1 Automaticity

In observing the two subjects’ verbal interactions at work, one characteristic that appeared time and again was their ability to hold new information (either new linguistic information or content information) in short-term memory while continuing the on-going interactions. More importantly, they were not only able to continue the communication without the interference of the new information, but could also acquire the information and ultimately use it on the spot. This trait seems to correspond to what the expertise
studies say about experts’ memory. Extensive practice allows an expert to automatically perform basic skills in their domains, which frees up memory capacity for processing other aspects of the task.

In the case of working in a second language environment, this form of automaticity was demonstrated in the experts’ way of dealing with unfamiliar linguistic expressions. While both subjects in fact had very few difficulties accessing appropriate vocabulary, once they encountered something they did not know, they appeared relatively calm and in many cases, did not even disclose the fact that they had encountered something unfamiliar on spot. At the same time, experts exploited all kinds of available resources to figure out the meaning first, and then try to memorize it through repeating, paraphrasing, or reusing it in the next appropriate context they can identify.

One example shown in the following transcription illustrates how Alan dealt with an unfamiliar expression in an on-going conversation by faking his way through the situation.

<table>
<thead>
<tr>
<th>Roles</th>
<th>Chinese Transcription</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:48 Alan</td>
<td>能不能说服这个黄老师？</td>
<td>Can you persuade teacher Huang?</td>
</tr>
<tr>
<td>5:51 Student Yu</td>
<td>怎么了？</td>
<td>About what?</td>
</tr>
<tr>
<td>5:52 Alan</td>
<td>就是不要照片，因为今日它有这个非常固定两个，只能要么这样，要么这样。他说这个，我们既然地方不够的话，真的是照片是无法使的。</td>
<td>Just take off the photo as Today’s Museum has two very fixed (size), either this one (showing the sample in his left hand) or this one (showing the sample in his right hand). So since we don’t have enough space, these photos cannot be used.</td>
</tr>
<tr>
<td>6:06 Art Museum Staff</td>
<td>对啊，你照片既然不够……</td>
<td>Yes, since your photos do not have enough …(space to display)</td>
</tr>
</tbody>
</table>

(Cont’d)
In the SR interview, Alan confessed that he did not know clearly when his classmate Lin said that their teacher Ms. Huang was just saying something thoughtlessly, implying that what she said was not convincing at all. Although Alan had never heard of the idiom yikouyishuo (一口一说 speaking something thoughtlessly) that Lin used, based on the complaining tone of her speech and her frowning facial expression, Alan inferred that it should be a criticism about what Ms. Huang said. Moreover, Alan was not only able to infer the right implication of that unfamiliar idiom based on other contextualized cues he recognized, but also to chunk this four-character idiom out and to repeat it in both an accurate and an appropriate way, showing his agreement with his classmate. His accurate chunking and repeating the idiom in the on-going conversation reminds us about what Goh (2002) finds about listening expertise that expert listeners are fast at converting
sounds to words and can effectively hold aural input in their echoic memory without much interference. This explains what Alan was able to do when he encountered the idiom from his classmate’s speech.

While experts are found to be able to consciously hold new expressions in their working memory, the current study did not find any evidence to show that they have superior memories in other aspects. In fact, in most cases, when the researcher confirmed with the subjects about a particular expression they used in an event just ended a few minutes previously, both subjects had a hard time recalling what they said until they were given the video prompt during the SR interview. This finding seems to agree with what other expertise studies have found about the difference between experts and non-experts in terms of their ways of recollecting memories of the exact steps in performing certain tasks. Previous studies find that when experts and non-experts are asked to recall the processes and properties of skill execution as the performance actually unfolds, non-experts are better at recalling what was going on than experts. This can be explained by the fact that the less practiced performances of non-experts are controlled by declarative knowledge that is held in working memory and attended to step-by-step. In contrast, experts’ highly practiced performances are automated, supported by procedural knowledge that operates without the need for explicit or attended monitoring.

3.5.2.2 Creative Indexing

In addition to the automaticity, the other aspect that highlights the expert’s way of exploiting their communicative repertoire is their flexibility in using linguistic forms to construct heterogeneous situations in which all kinds of socio-cultural information are
indexed through language practices. As recognized experts in their particular fields, both subjects have the authority of assigning situational meanings to the particular forms they choose to use, through which their membership within their respective fields is accrued. More importantly, both of the subjects are very conscious about this process, being able to repurpose the linguistic forms for their own communicative intentions.

In one of the examples touched upon in 3.6.1.1 when Alan explains to a museum staff about why they had to adopt a particular design for the exhibition tag, the size of which does not meet the museum’s requirement. Within the 90-second conversation, Alan used the Chinese word lingdao (leader, 领导) three times, signifying both his affective stance and his social identity within the group that is engaged in that event. Among the variety of Chinese expressions that can be used in reference to the head of institution, the particular expression lingdao is more often used in a political context, signifying the Chinese Communist ideology. This particular lingdao that Alan referred to, as confirmed in the SR interview later, is the Party Committee Secretary of CAFA. As Alan recalled, although it is normal to refer everyone working in the context of a university as a laoshi (teacher), he intentionally used the expression lingdao instead, so as to emphasize the fact that he was forced to choose that particular design as a compromise to an undesirable decision made by the school official. The third time when Alan emphasized the fact that including a photo in the exhibition label was an order taken from the school official, he further expressed his strong disagreement with the official’s decision by saying that “he made the exhibition appear unprofessional”. However, one Chinese artist interviewed later commented that there existed a mismatch between Alan’s
use of *lingdao* indicating an in-group member within the school official system\(^{20}\) and his bold comment about the school official’s unprofessionalness. However, in the SR interview, when the interview confirmed with Alan about this perceived mismatch, Alan justified his using this seemingly reverent expression *lingdao* by offering a very different reason as shown in the comment below.

[Interviewer: Do you think it is inappropriate for you to say your *lingdao* is unprofessional?] No, because we students all dislike *lingdao*. The *lingdao* refers to the Party Committee Secretary of CAFA. We students always think he knows nothing so students can always complain about him. When I mentioned *lingdao*, everyone understood what I was implying.\(^{21}\)

Alan’s comment revealed that he intentionally used the expression *lingdao* to reinforce his social identity as a student instead of an in-group member of the official system as the other Chinese artist had assumed. In the SR interview, Alan’s use of a newly coined Internet expression “吐槽” (*tucao*, to complain) to indicate that every student likes complaining about their *lingdao*, again reinforced his identity as a student. Although *lingdao* may index an inferior social stance with a strongly reverent tone in many other contexts, in the context of this student exhibition, Alan repurposed the word *lingdao* to distance himself from the administration and show his closeness to the students. From the transcription of his interaction with the art museum staff, we can tell that his intention was successfully conveyed and picked up by the young staff who

\(^{20}\) The original comment was made in Chinese “好像是官场上，那种体制内的人。”

\(^{21}\) This is the only SR interview comment that Alan made in Chinese as he naturally chose during the very first interview he did. After this interview, he switched to English upon the researcher’s request. The original Chinese comment is “没有，因为我们学生都讨厌领导嘛。领导就是党委书记，我们学生都觉得他什么都不懂的，学生都可以吐槽吐槽他嘛。我一说领导，他们就懂了.”
encouraged Alan to insist on his own decision and to persuade the school official to change his mind.

<table>
<thead>
<tr>
<th>Role</th>
<th>Chinese Transcription</th>
<th>English Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:30 Art Museum Staff</td>
<td>你做成这个样子的吗？这样的一个尺寸吗？</td>
<td>You want to make it like this? This size?</td>
</tr>
<tr>
<td>3:33 Alan</td>
<td>对啊，就是长条嘛，不就不符合你们的要求了嘛</td>
<td>Yes, just a long and narrow one. It doesn’t meet your requirement, does it?</td>
</tr>
<tr>
<td>3:36 Art Museum Staff</td>
<td>对啊，这样不好看啊，没见过这样长条的展签呀</td>
<td>Yes, it does not look nice. We have never seen an exhibition tag like this.</td>
</tr>
<tr>
<td>3:40 Alan</td>
<td>那咱们创一个，创新嘛</td>
<td>Then let’s just create one, innovate!</td>
</tr>
<tr>
<td>3:45 Art Museum Staff</td>
<td>嗯，不行，这可能不行</td>
<td>Hmm…no, that might not work.</td>
</tr>
<tr>
<td>3:47 Alan</td>
<td>你觉得哪方面不行？显得奇怪吗？</td>
<td>Which aspect did you refer to? Seems strange to you?</td>
</tr>
<tr>
<td>3:52 Art Museum Staff</td>
<td>就是我会觉得这样……另外就是那个美术馆是有这个规定</td>
<td>Yes, well, I think, …Also, our museum has a rule</td>
</tr>
<tr>
<td>4:05 Alan</td>
<td>这个规定特别没有余地是么？(和工作人员一同走近墙上的展品)因为我们主要是个什么考虑呢，就是说，要是太宽的话，太宽的话，就是这种宽度的话，作品之间显得特别挤，但是我们这个学校的领导又要求我们一定要照片。所以我们本来就想干脆就这么大一个标签，但是呢，那个展签呢，就领导说非得有照片了，所以呢，我们只能就是说折中一下，就是这么一个</td>
<td>So this rule does not have any room for negotiation, does it? (Walking closer to the artwork on the wall to make a further explanation.) Well, you know, what we mainly concern about is whether it is too wide. If keeping the current width, it will make the artworks look too close to each other. But the lingdao (school official) required us to keep the photo there. So actually our original thought was we just had a tag like this, but, that tag, our lingdao (school official) said, must have photo on it. So, we have to compromise, like this.</td>
</tr>
</tbody>
</table>

(Cont’d)
4:38  Student Lin  相当于两个展签了  It equals two labels
4:39  Alan  对，就这样  Yes, like this
4:43  Art Museum Staff  放得下吗，那么多字  Do you have enough space for so many words?
4:45  Alan  不，这个说明不要了  No, we don’t want this introduction any more.
4:47  Art Museum Staff  说明不要了？？  You don’t want?
4:48  Alan  不要了  No I don’t.
4:49  Art Museum Staff  这样很怪的。你，你这上面放照片，下面放，这样很怪的。要不然就把照片去掉，为什么非要放照片呢？
This looks very weird. You put a photo on the top, and below it you put…this is really weird. Otherwise, you can just get rid of the photo. Why do you have to display a photo?
4:56  Alan  我也不知道为什么要放照片  I have no idea.
5:00  Art Museum Staff  为什么要放照片呢？  Why do you want to display photos?
5:02  Alan  我们也不懂真的，我们领导说的  We have no idea, really. The lingdao said that.
5:03  Art Museum Staff  我们做了那么多展览，没有放照片的  We have done so many exhibitions but we have never displayed photos.
5:04  Alan  我也觉得很业余搞的我告诉你  I also think it appears unprofessional …they made it … let me tell you.

Indexical knowledge revealed in Alan’s case is found to be at the core of linguistic and cultural competence of both subjects, when they take an active part in the on-going process of meaning-construction within their respective domain activities. It is through this active process that their communicative goals get linguistically realized moment by moment over interactional time. Jack’s assistant Ai also recalled a business event in which Jack’s flexible manipulation of certain expressions helped the company avoid some sensitive questions asked by the business partner. According to Ai, even though Jack is always the person who played the Chinese game during the business
negotiation, some of their business partners may however asked some blunt questions, assuming that a westerner like Jack would appreciate or at least be very tolerant with the direct way of dealing with business. In one of such situations as Ai recalled, Jack was asked about what specific business strategies his company used to promote one of their game products, which involved some business secrets that Jack felt uncomfortable to disclosing. Jack then told his business partner to manmanlai (慢慢来, to slow down), which in Ai’s view, was used in a playful way so as not to take that direct question seriously. As the phrase manmanlai is always used to encourage or comfort someone, implying that the speaker is in a superior position, Ai recognized Jack’s use of manmanlai as a powerful way to signify his authority as the boss of the company, who chose not to disclose the information readily. Moreover, even though Jack fully recognized his business partner’s intention, his response clearly indicated his unwillingness to take the question. Instead of unconditionally meeting the expectation of members in the target culture, his use of manmanlai more or less challenged the boundary between how he is expected to behave in the target culture and what he is willing to commit in the present situation.

22 Coincidentally, a very similar case of using the phrase manmanlai was also found in another Chinese learner’s oral introspective memoir in which he recalled an experience of negotiating a deadline with his boss at work, as recorded in the researcher’s previous study. In that memoir, a young American intern, Daniel, talked about how he was able to recycle the word manmanlai that has been used by his Chinese teachers and friends to comfort and encourage him when he was struggling with his Chinese learning to negotiate with his boss about a project deadline. While the utterance manmanlai used by his Chinese teachers and friends carries the authority that a native speaker imposes on non-native learners, in Daniel’s case, he was able to reaccentuate the word for his own purpose, to transgress the boundary of obedience and to contest the authority of his boss in a light-hearted way.
To theorize Alan’s and Jack’s experiences of using a conventional expression to creatively index their own intentions, I appeal to Bakhtin’s proposal that every utterance is a repurposing or re-signification of other people’s words:

As a living, socio-ideological concrete thing, as heteroglot opinion, language, for the individual consciousness, lies on the borderline between oneself and the other. The word in language is half someone else’s. It becomes ‘one’s own’ only when the speaker populates it with his own intention, his own accent, when he appropriates the word, adapting it to his own semantic and expressive intention. (Bakhtin, 1981, 293-4)

Both subjects have mastered the strategy to appropriate “the other’s” to re-signify their own intentions. Observations of their adept but also creative use of some conventional expressions reveal that being able to appropriate the target culture’s expression to their own intentions while retaining their “otherness” always requires skill, experience, and in most cases, a certain amount of culturally appropriate irony or humor that accentuates their expertise in communicating in Chinese.

3.5.3 Experts Assume the Responsibility of Making Improvements

Advancement in any performed capacity means accepting the fact that you will never fully get it down, and that you must repeatedly resume the perspective of the novice. Like any expert in any domain, foreign language learners need to engage in a deliberate and demanding process of moving from novice to expert. Even when they achieved a certain level of expertise, they are still responsible for managing significant portions of their own learning process as they maintain expertise. In the current study, both subjects recall such situations in which they attentively pay attention to their mistakes, exploit all the available resources around them to improve their performance in
the target culture and deliberately engage in some systematically planned learning activities to further improve their Chinese.

Experts are able to hold new information in memory when they encountered new items in an on-going conversation. On the one hand, it is their achieved automaticity that frees up their memory capacity for processing the new item. But on the other hand, it is also their strong motivation to keep improving themselves that make them pause and ask questions.

The transcription below illustrates a general process of a CFL expert skillfully coping with an unfamiliar expression in the conversational mode. During the SR interview when Alan was discussing with the interviewer about one of the recorded performances, he interrupted the interviewer, asked her to repeat an idiom she used and confirmed the characters used in the idiom with her. The whole process lasted less than one minute, in which Alan efficiently focused on the idiom in the on-going discourse, made an educated guess about a specific character xiao 晓 based on context and repeated the whole idiom for himself to memorize. Even though Alan did not ask for an explicit explanation, the way he decomposed the idiom already demonstrated his understanding of the idiom’s structure. Moreover, this process also shows Alan’s expertise in chunking and grouping, the two of which together constitute the tenet of an expert’s perceptual skill in any domain.
28:55  Interview -er
Chinese Transcription 而且我觉得你也是动之以情，晓之以理的。就是说，因为好几个地方，你那个 transition 都做得特别好；然后你还说“我也理解，你怎么怎么样”。
English Gloss Also, I think you “moved him with emotion and also made him understand the reasons” (dongzhiyiqing, xiaozhiyili). You made good transition at several places; and then you also said “I do understand.”

29:08  Alan
动?

29:08  Interview -er
就是可能你教书教的特别长，就是你有一种特别老师的那种 discourse

29:18  Alan
你刚才说什么词，我没听懂，什么晓理什么什么的?

29:23  Interview -er
哦，动之以情，晓之以理。

29:29  Alan
呃……动之以情，晓是……那个是晓得的晓?

29:33  Interview -er
对对对，知晓的晓，晓得的晓得。

29:37  Alan
晓得- 什么?

29:39  Interview -er
晓之以理。

29:41  Alan
哦哦，明白了。

29:42  Interview -er
嗯，动之以情晓之以理。

29:43  Alan
知道了知道了，我理解了。动之以情，晓之以理。

Also, I think you “moved him with emotion and also made him understand the reasons” (dongzhiyiqing, xiaozhiyili). You made good transition at several places; and then you also said “I do understand.”

What expression did you use just now? I could not understand that part. What xiaoli?

Ah, dongzhiyiqing, xiaozhiyili.

Emm…dongzhiyiqing, “xiao” is the “xiao” in xiaode (which means, to know, to understand)?

Right. The “xiao” in zhixiao (which also means to know), the “xiao” in xiaode.

xiaode what?

I see. Got it.

Yeah, dongzhiyiqing, xiaozhiyili.

I see, I see. I understand. Dongzhiyiqing, xiaozhiyili.
According to Alan’s classmates at CAFA, such conversations as recorded above take place very often in their communications with Alan. He was not only actively asking questions about things he did not know, but more importantly, he would repeat the new item many times to himself so as to memorize it. Alan’s interview data further reveals what he does for improving his Chinese daily and more importantly, his attitude towards this kind of practice. Alan emphasized several times that learning a foreign language is not simply a skill for him, but a way of learning to be a human in that culture which has to take a lifetime whether you are a non-native speaker or a native speaker:

I read all the time. I take notes. It is mostly through reading, you know, and writing. Whenever I have a doubt, I asked my friend if it is right or not. Not only the choice of words, but also maybe if this phrase is used appropriately. I encourage my friends to correct me. I feel like on the daily basis, verbal Chinese is not really, especially because I am working in a domain with everything that I am so familiar. There’s not really a lot of new vocabulary on daily basis. … Anyway, the study of Chinese is the study of being a person. It’s not a skill. For me it’s not a skill. Chinese is not a skill that I can technically, that can be quantified. It’s the way of being a human. And studying this language implies the study of 3000 years of written philosophy and literature that I have just touched the tip of the iceberg. So I mean that practice and cultivation is the work of a lifetime. I mean for me it’s a very moral thing. But technically you have to refine your language all the time. You have to look things up. You have to make sentences. Whenever I learn new words, I always make sentences in my notebook, zaoju, sentence construction. I do all these things.

In Jack’s case, he mentioned some similar strategies he used to refresh his Chinese, such as reading and carrying a notebook to record new expressions. He was especially attentive to his use of Chinese, one example of which has been discussed.

---

23 Alan’s classmate Shan commented in the interview, “比如说遇到一些成语什么的，他可能懂的不是那么太多的话，这些东西就是他还是会去问，还是会反反复复地说一说，他还是在学。就比如他今天学到了一个什么成语，我今天跟他说了一个什么东西，但基本上他还是在学，偶尔会有一两个词汇他不知道的，他自己会念叨念叨，会反反复复地说一说。这个样子。可能也是一个学习的过程嘛。就跟他看书标注词、生字是一样的。他看的书上，就是生僻的字，哪怕是中国人都不认识的，很多我也不认识，他都去注音，都要知道。阅读习惯特别好.”

110
previously as he worked very hard on getting rid of the sentence-final particle *ma* in his spoken discourse to weaken his Hunan accent. He also organized a reading group consisting of Chinese senior staff in his company, reading books either written in or translated into Chinese and meeting for discussion on a monthly basis.

Moreover, although both subjects were tested as “Superior” according to OPI at the moment when they stepped out of their respective language programs, they both admitted in their interviews that their performance in Chinese is completely different than it was at that time—in Alan’s words, “the difference between the heaven and the earth” (天和地的差别). However, they both value the previous training in their home institutions, considering these programs as a training ground for them to engage in steps of self-managed learning, which therefore allows them to keep improving beyond the confines of a program. This also corresponds to what Ericsson (1996) points out, “a long period of preparatory education followed by an apprenticeship” (p.3), is required to attain the expert level of performance in most professional domains of expertise.

### 3.6 A High Expectation for Multilingual Experts

Observations and discussions presented in section 3.5 verify the two subjects’ expertise in communicating in Chinese as a foreign language. Moreover, from the interview data reviewed above, it is apparent that both of the subjects are admired as experts in something of value to their Chinese counterparts. Nevertheless, some of the interview comments made by the Chinese colleagues who work very closely with the subjects also disclosed a critical attitude towards what foreign experts are expected to achieve within their areas of expertise. As mentioned in 1.4.2, the current study solely
examines the two subjects’ expertise in communicating in Chinese within their respective academic and professional domains and not their domain expertise. However, through interviewing Chinese counterparts working with the two subjects, the researcher came to realize that the Chinese counterparts usually expected something beyond Chinese-language expertise when communicating profession-related issues. Moreover, their expectations of a foreign expert is somewhat different from their expectations of a Chinese expert in the same domain. Even though such differences might not be explicitly expressed in their interview, they can still be detected in some of their critical comments on the foreign subjects’ performances. The added value that those Chinese counterparts were looking for was a demonstration of characteristics that cannot be easily found in their interactions with native Chinese experts. I would call this “global competence.”

An interview with Jack’s business partner Tang illustrated such expectations for the multilingual competence. When asked about his opinion about a Chinese Ted-talk-like speech Jack gave recently on the topic of the Chinese game industry, Tang’s comments did not sound as enthusiastic about the talk as other Chinese people who watched and evaluated Jack’s performance. Although Tang admitted that the talk was a success in terms of Jack’s expressing his idea in Chinese, he said the talk is not as much on the “high-end” as he had expected. When further asked what he meant by “high-end”, he explained that Jack did not demonstrate a kind of international entrepreneurship, in Tang’s own words, Jack’s lacking “guojifan (国际范)”. Although it is reasonable to wonder if Tang was playing the game of modesty as he was talking to an out-group

---

24 In the interview, Tang used the popular internet buzzword “高大上” to express his expectation for Jack’s performance. Literally meaning “high, big and up”, this phrase is always used to describe a person or a thing of good taste and high level in the Chinese media.
member (the researcher) about an in-group member (his business partner) who is very close to him, Tang’s detailed criticism of Jack’s performance in the talk seemed to indicate that his criticisms were genuine.

Moreover, coincidentally, the comment on Jack’s lack of an international feeling corresponded to another comment given later by a Chinese evaluator of Jack’s performance. The rater said Jack looked like a businessman from Wenzhou. Although as China’s biggest manufacturer of small scale goods, Wenzhou is renowned for its entrepreneurs, the comment of “like a Wenzhou businessman” sometimes may also contain a negative connotation, implying that smart and astute entrepreneurs who only know how to make money out of small scale business. In other words, a Wenzhou businessman-like American entrepreneur may rarely demonstrate the international feeling that is expected by his Chinese counterparts. Although here the Chinese rater did not explicitly mention if this “Wenzhou-like” characteristic shown in Jack’s performance is something positive or not, his following comment that the performance is “nothing special” seemed to have disclosed his disappointment over the performance.

Tang even expressed his suspicion over Jack’s ability in dealing with business in American culture. Recalling the first time when he went to US with Jack on a business trip, Tang said “his performance in communicating with their American business partners seemed not a lot better than mine (他的表现好像也没比我好多少)”. Obviously, the “performance” mentioned by Tang here should not be Jack’s linguistic performance since Tang can barely communicate in English. What Tang meant here is Jack’s knowledge

25 Some native Chinese are recruited as raters of the two subjects’ performances. This is a study that is presented in Chapter 4.
26 That Chinese rater’s original comment in Chinese was “就是温州老板在做生意，没啥特别的” (Just a boss from Wenzhou who is doing business, nothing special).
about how business should be done in American culture with game publishers in the United States does not surpass his own. Similarly, the researcher also noticed that although Jack is able to perform bilingually, in many cases it seems very difficult for him to switch from Chinese to English so that he prefers to use Chinese in most cases. The most obvious example is that Jack felt very uncomfortable to use English to comment on his performances in Chinese during the SR interview even though he was required to do so by the researcher. After trying several times, Jack permanently switched to Chinese and refused to continue the SR interview in English.

From all these examples, it seems clear that one can be socialized into Chinese culture to an extent that he may be perceived by the very Chinese counterparts as having lost certain abilities to perform equally well in the native culture or using the native language is not a value that is appreciated by some Chinese counterparts. Sometimes, foreign experts may not have recognized the existence of such high expectations from the Chinese people they work with as they usually take pride in their being able to internalize the target culture so quickly and completely, such as Jack’s insistence on using Chinese to complete the interview. But in some cases, very self-reflective foreign experts may also ask the same question about what value they can bring to the domain in the target culture they have worked in, as shown in the following example involving Alan.

During the SR interview, after watching his own performance of giving instructions to several students on their sketches at the CAFA high school, Alan became very upset about what he did in the video and expressed his disappointment in a very emotional way. He expressed that his issue with the video is not about language, but the fact that he “acted so Chinese”. He thought the class was unsuccessful and blamed it to
the fact that he was completely using a Chinese pedagogical paradigm to teach the class. In his description, the Chinese pedagogical paradigm was represented by a master-disciple relationship between him and his students, in which the teacher imparts the standard to the students and the students take the responsibility of achieving the fixed standard in their responses. When asked about why he chose to follow the Chinese paradigm instead of an American one, Alan gave two reasons. First, he thought the Chinese pedagogical paradigm is what those Chinese high school students are very used to as that is what Chinese education is based on. Although he thought the choice of pedagogical paradigm took place unconsciously, he may have just chosen something that is the easiest to handle and can avoid conflicts. The second reason for teaching in a Chinese way is his lack of experience in receiving art education in US. His major at college has nothing to do with art and all his knowledge and skill about art were gained in China during his four year of study at CAFA. Therefore, even if he wants to bring some American pedagogical elements to his class, he may still have failed to do that as he himself has a very vague idea about what art education looks like in American K-12 programs. However, Alan’s strongly emotional comment on his own performance still reflected his dissatisfaction over the pedagogical paradigm he chose. A transcribed interview from the SR data can sufficiently illustrate Alan’s anger:

I sound just like a CAFA professor. It’s so CAFA. It’s so ridiculous. That’s so CAFA. Oh my god. 往下研究 (to further research), I sound like 徐冰 (Xu Bing). So CAFA, oh my god. The entire language is like everything is in this school. 你要往下研究 (you should do further research on this), 这个线要压得匀 (this line should be painted evenly). It sounds like 国画系 (the department of Chinese Painting). [Researcher: When you said “it was so CAFA,” did you refer to something positive or something negative?] It’s obviously negative. It’s a passive paradigm. It creates very skilled craftsman. It just not creates artists at all. It is absolutely the language I have heard for four years. I realized that’s almost
comical how perfect it is. It’s so funny. The reason that I am so disgusted by myself right now is because I suffered under the system for four years in my life, I worked like a devout monk disciple of this system and the result is I basically, while agreeing on some level that traditional materials have a way to them, then can be discussed, like you know, like if you make a line like this that, it has certain effects, it does certain things. But the entire narrative of being authoritative, wise master and deferent disciple is something I found very depressing. And I really don’t think it’s very stable. I think it was very immature, sort of very uniform and powerful paradigm. I do not think it is effective for participating in western global society.

In Alan’s case, the researcher did not receive any negative comments from the Chinese interviewees about his teaching performance at CAFA high school. However, similarly mocking comments like what Jack had received, e.g., “like a Wenzhou businessman”, were also found among the Chinese raters evaluating Alan’s performance samples of teaching. One Chinese rater said Alan was acting just like “some Beijing middle school class-teacher (北京某中学的班主任)” and the other related Alan’s behavior to “a typical image of a Chinese academic director (让我想起了中学教导主任).” Both of the two terms, “class teacher” and “middle school academic director” may connoted a conventional negative image of some authoritative figures at school (usually middle-aged females) who are very demanding and always worry too much about trivial matters concerning students. Obviously, these images run counter to what is expected for a young American art teacher who has received higher education in both US and China.

One of the researcher’s students once concluded after his two years’ stay in Beijing, “the foreigners who are treated well in China are ones who can take their expertise learned in the US and apply it to China. The best they can get without any other expertise is to be trotted out on a TV show.” However, from the examples observed in the current study, taking one’s expertise in the US and applying it to China seems a lot more
complicated than we had expected. Simply translating one’s expertise into Chinese, such as talking about the game industry in fluent Chinese or articulating the rules of painting an even line in concise Chinese, may not be enough to satisfy those Chinese experts with a higher expectations. To distinguish oneself from the foreigners trotted out on a TV show, a foreign expert may need to demonstrate a certain level of global professionalism, going beyond the state of either staying in one’s native culture or completely abiding by the conventions in the target culture. Without sufficient research and data, the current study will not further discuss what may constitute such multilingual competence. However, this multilingual competence must be added to the list of characteristics of expertise and is definitely a promising research area in the future.
Chapter 4 Assessing Expertise in Communicating in Chinese as a Foreign Language

4.1 Literature Review: Alternatives in Assessment

The past few decades have witnessed a paradigm shift in the concept of assessment, from a narrow view of testing and examination to one of a broader culture of assessment (Lynch, 2001), indicating not only a semantic change but also a conceptual shift. The increasing popularity in self-assessment and portfolio assessment conveys a message that assessment is no longer merely perceived as a way to observe and record learning, but a means to promote learning. Foreign language evaluation in the context of assessment culture can be perceived as a socially constructed activity embedded in the local context with teachers, students and other community members, all of whom are recognized as meaningful assessment partners.

This dissertation adopts the definition of assessment from Lynch (2001) as “the systematic gathering of information for the purposes of making decisions or judgments about individuals” (p.358). In addition, there also exist two intuitively similar concepts: testing and measurement, which are sometimes used interchangeably in the context of “assessment.” However, many scholars have pointed out distinctive differences among the three concepts and reminded us to avoid a superficial commingling of these terms (Bachman 1990; Lynch 2001). In Bachman’s words, “Not all measures are tests, not all...
tests are evaluative, and not all evaluation involves either measurement or tests” (Bachman 1990, 11). The relationships among the three concepts: assessment, testing and measurement (which are associated with the narrow perception of assessment) are illustrated by Bachman (1990)²⁷ in Figure 3.

Figure 3 is Bachman (1990)’s illustration of the relationships among measurement, testing and assessment. The five areas can be respectively described as follows:

Area 1: Evaluation not involving either tests or measures
Area 2: A non-test measure for evaluation
Area 3: A test used for purpose of evaluation
Area 4: Non-evaluative use of tests and measure for research purposes
Area 5: Non-evaluative and non-test measurement

²⁷ Lynch (2001) later proposed a slightly different graph in which test and measurement completely contained in the biggest circle of assessment. Lynch’s interpretation however excluded the possibility of 4 and 5 in Bachman’s graph. Although the discussion in the current study mainly focuses on assessment of area 1 and 2, illustrated in both Lynch’s and Bachman’s presentations, it still adopts Bachman’s graph in order to render a more open-minded perception of the relationship among assessment, measurement and test.
Based on the graph above, the term “alternative assessment,” usually seen as “an alternative to standardized testing” (Huerta-Macias 1995, 8), can, therefore, be understood as a type of evaluation involving neither tests nor measures (area 1), or a non-test measure for evaluation (area 2). Assessment procedures of this kind can include diaries, portfolios, self-assessment and peer-assessment. The following sections will review major alternative assessment instruments used for foreign language testing and professional performance assessment used for Human Resource, including their respective strengths and weaknesses, and their actual implementation.

4.1.1 Portfolio Assessment in Foreign Language Education

The recent shift away from teacher-centered instruction in foreign language education has forced these scholars to come up with assessment approaches that are compatible with their new roles and responsibilities. Many have embraced alternative assessment tools that encourage learner participation, cultivate a sense of ownership and provide a multidimensional and dynamic portrait of what students have achieved over time. A most notable form of alternative assessment, the portfolio, has become popular with its virtues well recognized by both educators and students.

Portfolio assessment first emerged as an alternative technique in writing instruction when it was proposed by Elbow and Belanoff as a substitute for the traditional proficiency examination in their freshman writing course at Stony Brook in the early 1980s (Elbow and Belanoff 1986). Today, portfolios are most commonly used for assessing writing proficiency, and focus heavily on student self-evaluation and teacher autonomy (Graves 1983; Tierney, Carter, & Desai 1991). However, as a purposive
process of collecting evidence to illustrate an individual’s accomplishments, the portfolio assessment has been utilized in various contexts from proficiency testing to program evaluation.

Barton and Collins (1997) list seven characteristics underlying the development of educational portfolios, stating that they should be multi-sourced, multi-purposed, authentic, dynamic, explicit of purpose, an integration of a learner’s academic course work and life experiences, and based on student ownership.

However, a review of the studies on portfolio assessment reveals two limitations regarding their application for foreign language assessment. First, although it has been more than three decades since they first appeared in writing curricula, their application in fields other than writing is still very limited. There are two possible explanations for this phenomenon. A traditional view of assessment usually considers the four skills (speaking, listening, reading and writing) separately. However, it is almost impossible for such a performance assessment to isolate the course of action so as to evaluate those receptive skills that rely more on inference, such as listening and reading. Moreover, compared to building a portfolio of writing performances, it is far more difficult to capture and document spoken performance samples, especially those improvised in authentic contexts.

Second, among the very limited number of portfolio assessment tools, the rating method is exclusively limited to self-assessment, which supports its claimed benefit of

---

28 Such speculation was inspired by Professor Mari Noda's presentation given at her Seminar on "Assessment in Material", in the Spring of 2015 at OSU, in which she discussed operationalizing the traditional versus the alternative assessment, leaving assessment of receptive skills to skill assessment while having performance assessment do their job to evaluate performances entailing multi-skills in authentic contexts.
fostering a learner’s sense of ownership. However, as seen in previous studies, self-assessment has been controversial in the scholarly world (Ross 2006, Brown, Dewey and Cox 2014). The only two widely recognized portfolio assessment techniques for capturing spoken performance in foreign language that the researcher was able to find, were the European Language Portfolio (ELP), based on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001, 2004) and the LingualFolio, both of which aim to facilitate learner reflection and autonomy.

Scholars who embrace the use of self-assessment in ELP defend against the general concern that learners do not have the knowledge required for accurate self-assessment. Self-assessment, however, refers to behavior, which allows learners to judge what they can do communicatively and with what general level of proficiency they can do it, without attending to linguistic accuracy (Little 2009). However, this explanation is farfetched in two ways. First, it contradicts the CEFR’s claim that it not only focuses on the behavioral dimension of L2 proficiency, but that it also offers a scaled summary of what it calls “qualitative aspects of spoken language use,” including range, accuracy, fluency, interaction, and coherence (Council of Europe 2001, 28-29). In other words, it is impossible for learners to only attend to the behavioral dimension of L2 proficiency without evaluating how accurately, fluently and coherently they are able to communicate. Second, even if learners could be exempted from the responsibility of assessing their own linguistic aptitude, the evaluation, which is based on behavioral criteria consisting of a list of can-do statements remains problematic. For example, the rating criterion for Spoken Production at the highest level (C2) of the ELP states “I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with
an effective logical structure which helps the recipient to notice and remember significant points.” Without sufficient knowledge of the target culture, a learner can only judge the appropriateness based on his or her default culture. The learner must also be aware of differing cultural expectations within the target culture. For example, an eloquently argued thesis on promoting the civil rights of LGBT persons may not impress, or might even offend, a Chinese listener. Therefore, this complex gap between an American learner’s self-perceived ability to “present a clear, smoothly-flowing argument” and the Chinese recipients’ reaction to a controversial topic and style of conversing can never be revealed in a simple checked can-do statement.

In fact, in terms of the strengths and weaknesses of relying on self-assessment in portfolio assessment, McAloon (2015, 156) reminds us that our perceptions of language and communication should guide the process of validating self-assessment:

If a researcher believes that the purpose of language is self-expression, then the final arbiter of successful communication is the nonnative speaker. If the researcher believes that language represents a part of a culturally situated act performed to elicit from listeners a sympathetic response and generate feelings of “being on the same team,” then the native-speaker listener’s assessment may be more valuable.

In this study, as mentioned previously, it is believed that the success of a communication is always in the eyes of its beholder. Although we assert that the portfolio assessment based on self-assessment ratings can increase learners’ self-awareness and foster learning autonomy, which may be good additions to a foreign language curriculum, it cannot be used effectively as a formal assessment for recommending foreign language learners for employment, for example. Therefore, in addition to Barton and Collins’ list of the seven characteristics of portfolio assessment, this study proposes to add “multi-
rater” as the eighth characteristic, the rationale of which will be discussed in the study presented in this chapter.

4.1.2 360-degree Feedback: Evaluating Professional Performance in Human Resource Management (HRM)

Well-designed assessment procedures are essential for helping supervisors to monitor performances of their employees and to assess their productivity in the workplace. During the mid-1960s and early 1970s, with the goal of establishing a broader and more accurate evaluation of employees’ performances, HRM practitioners shifted from supervisor-centered downward feedback to the concept of upward, multisource feedback (Lepsinger and Lucia 1997). The fundamental premise underlying the design of the 360-degree feedback is that data collected from multiple perspectives is more comprehensive and objective than data gathered from a single supervisor.

Studies describing the potential advantages and pitfalls of using this multi-source and multi-rater have appeared in numerous journals recently (Little 2002, 2009; Scharer 2008). The biggest reward of using 360-degree feedback, as reviewed in many previous studies, is its function in conveying and reinforcing the organizational values to its employees through the process of giving and receiving feedback (London and Beatty 1993, Lepsinger and Lucia 1997). This tool is able to reveal to an organization’s members how the organization defines expertise and what constitutes its expectations. Similarly, when the 360-degree feedback is applied to foreign language assessment, it can reveal the behavioral expectations of Chinese counterparts in the target-culture society and contribute to a refinement of pedagogical goals and practices.
Another advantage of the 360-degree feedback is that it collects data from multiple raters who are equipped with domain knowledge and domain conventions within the candidate’s domain, including the input of individuals who are familiar with the test taker’s academic or professional domain can lead to an effective assessment of professional-level foreign language use. Introducing the concept of domain within the assessment, as discussed in chapter one, will prevent the assessment from defaulting to the context of a tester interviewing a test-taker, as does an OPI. Therefore, 360-degree assessments will be more reliable for recommending foreign language learners for employment and placement in subsequent programs and assignments. (The review of OPI in Chapter One focused on the limitations of decontextualized OPI and the ILR scale on their applicability to assess professional-level foreign language use.)

One shortcoming of the 360-degree feedback is a relatively limited method of collecting data. Questionnaires and interviews are the two most common methods for collecting feedback from multiple sources (Lepsinger and Lucia 1997). The unreliability of the retrospective interview has been discussed in Section 3.3.2 and will not be relied upon here. Moreover, the use of these interviews to elicit evaluations of previously observed performances also limits the source of raters in that only those who have worked with the individuals in question can be considered. Even though this may still be relevant in the HR Management context, it will not be helpful when we apply the 360-degree feedback to the field of foreign education, as those learners have not yet started their professional lives in the target-culture, a scenario for which that assessment was made. Similarly, based on McAloon (2015), questionnaire-based evaluations are also of limited use to hiring organizations. “None of its own trusted employees will have had an
opportunity to observe and therefore have a recollection of an applicant’s previous performance” (p.162).

Another concern about the 360-degree feedback is its high cost, including both the investment of time in designing the assessment items and the investment of human resources in the process of evaluation and post-evaluation coaching sessions. One solution to this disadvantage is to limit the use of this method to managerial personnel at the more senior levels (Lepsinger and Lucia 1997, 17). This solution adopted by the human resource managers can be used to inform us about who should be the target candidate of a similar assessment in the field of foreign language learning. Like those executives and upper-level managers whose skills represent a significant investment of salary and benefits, advanced-level foreign language learners have invested huge amounts of personal time and money and have been the recipients of significant institutional resources in foreign language learning. Potential employers or advanced training programs focused on exploiting foreign language skills should be the target users of a multi-rater portfolio assessment.

Inspired by the 360-degree feedback, McAloon’s study (2008) was the first and the only empirical study to date to adopt the multi-rater portfolio performance assessment to evaluate professional uses of Chinese as a foreign language. By using the Advanced Language Performance Portfolio System (ALPPS) created at the Ohio State University (OSU), five foreign subjects’ performance portfolios were assessed by different groups of raters—the foreign subjects themselves, their colleagues, and randomly selected Chinese teachers. He found that portfolio evaluations conducted by native speakers is an accurate reflection of native speakers’ reactions to non-native speakers’ foreign language
performances. In an article published in 2015, McAloon elaborates on the benefits of using multi-rater feedback tools like ALPPS for employers of foreign language speakers, language instructors, language program administrators and language learners themselves. McAloon’s arguments (2008, 2015) lay a solid foundation for the current study. Following his suggestion, the current study further examines the validity of utilizing the multi-rater technique in conjunction with the performance portfolio assessment by exploring both the quantitative and qualitative feedback given by the different types of raters engaged in the assessment process.

The current study has made three notable improvements based on McAloon’s study. First, instead of simply using the subjects’ colleagues, this study invited “true domain insiders” who have no prior background knowledge of the test subjects\(^{29}\). Such a change attempted to reflect the reality in performance assessment as an applicants’ performances are always assessed by people who would have had no chance to observe their previous performances. Second, this study fixed the problems in articulation and translation of the evaluating scales and categories that caused misunderstandings in McAloon’s study (McAloon 2008, 229)\(^{30}\). Detailed improvements on this aspect will be introduced below in Section 4.3.3. Third, in this study, a new type of rater-- non-teacher domain outsiders--were recruited in order to triangulate the data. Although feedback from

\(^{29}\) Compared with Alan, Jack is relatively famous in his field. Due to his actively engaging with TV shows, he could be more or less considered as a public figure in his field. However, none of the recruited raters had known Jack before except for the domain insider Zhong, who vaguely recalled that he might had listened to Jack’s talk once at a conference a few years ago. However, all the three domain insiders expressed their strong interests in knowing Jack and his business personally after they finished the evaluation, which again justified Jack’s expertise in his field.

\(^{30}\) In the middle of his study, McAloon noticed that the level 0 representing “no performance” in English was mistakenly translated into “poor performance (表现差)” in Chinese, which caused confusions among raters. Unfortunately, his study failed to correct this mistake timely, which might potentially misled him in interpreting the quantitative data.
Chinese instructors and domain-insiders will still be the focus of this study as both of them are the potential evaluators of the performance assessment in foreign language education, the addition of non-teacher domain-outsiders’ evaluations can always be referenced to increase the richness of the data. The next section presents research questions to guide this study.

4.2 Research Questions

Jian and Shepherd (2010) recall an unfortunate story of an American intern working for a company operating a cruise line along the Yangtze River (p.97). Despite the intern’s nearly perfect pronunciation and knowledge of Chinese grammar, his limited understanding of the cultural mores and expectations affecting how people interact with each other within his domain resulted in unintended insults when his Chinese colleagues and customers misinterpreted his sense of humor. Observations of the student working in Chinese workplaces revealed the shortcomings of the current American assessment system of foreign language proficiency. That is, the assessment instruments in place exclusively rely on the opinions of English-speaking Chinese instructors working in the U.S., even though many students trained in Chinese later go to work for Chinese companies where the majority of their colleagues are non-English speaking. Although that intern’s fluency and sense of humor won the hearts of the company executives in its New York headquarters, his jokes fell flat with his new Chinese audience who had little or no exposure to American humor in their life. If the interviewers who assessed that intern’s language skills had been native Chinese who speak little English and understand
little about American culture, he might have received negative feedback about his humor much earlier and could have avoided upsetting his Chinese colleagues and customers.

A foreign language learner’s success in the target culture hinges greatly on what native speakers in that culture will permit him or her to be. When we look at professional performance in a foreign language, it is only a specific group of native speakers in that target culture who share the same experiences with the language learners that will have a significant effect on their success. In other words, it is those Chinese counterparts working with the learners who can inform us about what they see as positive and negative aspects about learners’ communication since the success of a communication always lies in the eye of the beholder. Given this perspective as foreign language educators, we cannot avoid challenging ourselves to consult native speakers living in the country for their evaluations of a foreign language performance.

Therefore, in the current study, there are two research questions that guide my attempt to measure the value of including evaluators other than Chinese teachers to assess advanced Chinese use. First, how will different types of evaluators including domain insiders, domain outsiders and Chinese instructors evaluate the same professional performance by a non-native speaker of Chinese in a Chinese workplace? Second, if evaluators with different areas of expertise perceive these performances differently, what factors may contribute to this trend? By understanding these factors underlying the differences between the evaluations of Chinese instructors and domain insiders, we may take a further step to suggest the possibility of recruiting evaluators with various backgrounds apart from Chinese teachers so as to incorporate different evaluative perceptions with regard to a professional performance. However, if neither quantitative
nor qualitative differences can be found within the three different types of evaluators, it might suggest little need for recruiting evaluators other than Chinese teachers.

### 4.3 Research Design

#### 4.3.1 Constructing A Professional Performance Portfolio in Chinese

A professional performance portfolio in Chinese was created online for each of the two American research subjects, Alan and Jack, who attended the study discussed in Chapter Three. Alan’s portfolio contains 11 performance samples, including ten videos and one text. The ten videos are all spontaneous performance samples collected during the ethnographic observation. The textual sample was a review of a local art exhibition, which he wrote for the exhibition’s catalogue. When the researcher asked him for performance samples as part of his portfolio, he specifically suggested including the piece of writing. Jack’s portfolio contains 11 video samples, eight of which were collected during ethnographic observation. The other three video samples were taken from two TV shows that Jack attended the same year as the ethnographic study, which contain spontaneous performances including a Q&A session and rehearsed performances such as a speech given at the 2014 Forbes Forum.

Details about how the subjects were shadowed and filmed during ethnographic observation were described in Section 3.3.2 and will not be repeated here. All the performance samples were uploaded to a YouTube-like website in China, called Youku (优酷), with a password set for each video to protect the privacy of the subjects and the
confidentiality of their businesses. Each video was also assigned a number, a brief title (e.g., report on a business trip), as well as a short paragraph of description introducing the context of the video. Moreover, there were another 10 randomly selected videos of foreigners doing things in Chinese uploaded to the same website, mixed with the 22 target performance samples. Raters were told to evaluate 32 samples without being told which of the foreigners were the target subjects of the study. This is an attempt to prevent the raters from fostering fixed impression towards a certain subject when the number of subjects is very limited.

4.3.2 Recruiting Native Chinese Evaluators

For each American research subject, three types of native Chinese evaluators were recruited.

The first group consisted of Chinese domain-insiders. In Jack’s case, three Chinese upper management level professionals (Zhong, Zhu and Zhang\(^\text{31}\)), who were actively engaged in the gaming industry, were recruited from Beijing and Shanghai. The researcher contacted six different game publishers in Shanghai and Beijing one month prior to the data collection period and eventually identified one CEO of a small-scale game publishers, one senior marketing manager and one director of research and development, all of whom were from various well-known game companies in Shanghai. Like Jack, Alan also functions in multiple domains, playing different roles, which include graduate student of Chinese art, high school art teacher and an artist in the field of

\(^{31}\) Like the names of the two subjects, all the names of the raters presented in this dissertation are pseudonym. Moreover, names of the raters from the same rater-type group are coded into Chinese last names sharing the same pinyin initial. For example, here, Jack's domain insiders’ names are all coded into last names with Zh.
Chinese painting. The researcher narrowed the scope of her search to focus on artists in the field of Chinese painting who also had extensive experience teaching high school students in order to match Alan’s skill set. Although it might have been much easier to find college art teachers who also regularly attend art events as independent artists, the two very different teaching venues may disqualify a college art professor from evaluating Alan’s teaching performance as a domain-insider. That is because even an experienced college art professor may not be aware of the expectations for a high school art teacher, which may lead to misunderstandings about of Alan’s communicative intentions, or even lead to interpretative loss. Therefore, in Alan’s case, the researcher eventually recruited two separate groups of evaluators in his two professional domains. Two Chinese artists (Yu and Yi) actively involved in exhibitions with over 10 years’ experience teaching Chinese painting to children from first grade through high school were recruited to evaluate Alan’s performance samples. Moreover, one Chinese artist who had little teaching experience (Yang) and a high school Chinese painting instructor (Yin) whose art-related experience was mostly restricted to his course work were recruited to evaluate Alan’s performances within their respective fields. Such a compromise was made to ensure that all the domain-insider evaluators had not only abundant knowledge in the domain fields of the subject but also rich domain-related working experience.

The second and the third types of evaluators were both domain outsiders. However, the second type included instructors teaching Chinese as a foreign language at U.S. colleges (Xu, Xiao and Xin), while the other type were Chinese who had no expertise in pedagogy (Hu, Huang and Hong).
Three female instructors were recruited from different parts of China (one from the north and two from the south), representing a wide range of ages, educational backgrounds and working experiences. However, they also shared some common characteristics, making them qualified Chinese-teacher evaluators. First, all three of them had over three years’ experience teaching Chinese in U.S. colleges, including advanced level. They were familiar with standardized foreign language assessment instruments used the U.S. and China, such as OPI and HSK. Two of them who received training from the OPI workshop respectively in 2011 and 2012 (Xu and Xin) were very familiar with its testing procedures and rating scales. The other evaluator (Xiao), although without any formal training in OPI testing, had also self-studied the test and adopted its testing procedures and rating scales to give informal assessment to her students regularly. Moreover, with previous experiences teaching foreign students in China, this instructor was also experienced in teaching HSK preparation class.

In conclusion, the two subjects had access to the same groups of domain-outsider evaluators, including three Chinese teachers, and three not involved in education. Also, the two subjects respectively had their own domain-insider evaluators, three for Jack and four for Alan, with two working cooperatively to be chosen as a domain-insider evaluator.

4.3.3 Obtaining Quantitative Data from Portfolio Assessment Ratings

This study used a scale from 0 to 5 to evaluate various aspects of all the 22 performance samples collected from the two American subjects. The scale was adapted from the scoring rubric used in McAloon (2008)’s research that was reviewed in detail in Section 3.1.3. According to McAloon, the 0-5 scale was developed by the Chinese
Flagship program at the Ohio State University and has been adopted for their advanced language portfolio performance assessment (p.228). The scoring rubric is shown in Table 3 below:

Table 3. Scoring Rubric Used for Portfolio Assessment in the Current Study

<table>
<thead>
<tr>
<th>评分标准</th>
<th>Scoring Rubrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>5- 表现非常好</td>
<td>5- Superior Performance</td>
</tr>
<tr>
<td>4- 表现挺好</td>
<td>4- Strong Performance</td>
</tr>
<tr>
<td>3- 表现不错，合格了</td>
<td>3- Competent/Passable Performance</td>
</tr>
<tr>
<td>2- 表现一般</td>
<td>2- Some Performance</td>
</tr>
<tr>
<td>1- 表现不太好</td>
<td>1- Minimal Performance</td>
</tr>
<tr>
<td>0- 看不出表现</td>
<td>0 – No Performance Observed</td>
</tr>
</tbody>
</table>

(Adapted from McAloon 2008)

In this study, although each evaluator was given a bilingual version as shown above, the majority only relied on the Chinese version to understand the rubric since they had a limited understanding of English. Based on some points established from McAloon’s study, the researcher made several changes to the Chinese version to avoid potential confusion caused by translation problems.

First, the researcher replaced the original “表现差” (poor performance) with “看不出表现” (no performance observed), representing the 0 score, which more accurately corresponds to the English version. Secondly, to native Chinese speakers, unlike the significant difference between the two English phrases “passable performance” and “strong performance”, respectively, which correspond to score 3 and 4, the Chinese phrases “表现不错” and “表现挺好” are not distinguishable, as both of them can be properly translated as “quite good.” Therefore, the researcher added the term “合格了”

32 Although McAloon noticed this problem in the middle of his study and notified all the participants through email to make clarification, very few of them read that email and were thus unable to make the correction (p.229).
(passed) to “表现不错”， making it more equivalent to “passable performance” in the original English version. Thirdly, the original Chinese translation of “minimal performance” in McAloon’s version is “表现有进步”, literally meaning “the performance reflects progress.” Although it is a more positive and encouraging way to express the meaning of “minimal competence” to say, “You are making progress,” such a euphemism will cause confusion in the professional performance assessment context for those who are unaware of this nuance. Therefore, in the current version, the researcher adopted a more direct expression “表现不太好” (not very good performance) as an equivalent of “minimal performance” in the English version. In addition to the three changes made to the Chinese translation, the researcher also orally emphasized the straight-line increase in the current scoring rubric to all evaluators beforehand. According to McAloon, unlike the 0-5 ILR scale in which advancing from a 4 to 5 score is much more difficult than from a 3 to 4, the difference between a 4 and a 5 is equal to the difference between a 3 and a 4 (p.229).

Each evaluator was given evaluation forms, corresponding to the number of performance samples he or she agreed to evaluate. A sample of the evaluation form can be found in Appendix I. Each evaluation form number corresponds to a video uploaded online. The evaluator was also given a brief description of the performance sample, which is also available online.

During the texting process, the evaluator was asked to use the 0-5 scales to rate the performance based on eight categories of evaluation also adopted from McAloon’s study and comment on the three open-ended questions added by the researcher. Table 4 shows a bilingual version of the eight categories and the three open-ended questions.
Table 4. Categories of Evaluation Used for Portfolio Assessment in the Current Study

<table>
<thead>
<tr>
<th>Category</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 表达能力 Delivery</td>
<td>2. 语言组织能力 Organization of Ideas</td>
</tr>
<tr>
<td>3. 文化得体度（言谈举止）Cultural Appropriateness</td>
<td>4. 听众/对象适应调整 Adjustment to Audience</td>
</tr>
<tr>
<td>5. 修辞效果 Rhetorical Effectiveness</td>
<td>6. 文化常识相关度 Cultural Reference</td>
</tr>
<tr>
<td>7. 专业知识相关度 Domain reference</td>
<td>8. 回答问题圆满度 Question and answer</td>
</tr>
</tbody>
</table>

问题一：这个人给你什么样的印象？
Question 1: What were your impressions of this person?

问题二：根据这个人的这段表现，你愿意与他继续交流吗？为什么？
Question 2: Based on this person’s performance, would you like to further communicate with him or her? Why/Why not?

问题三：除了以上八个项目之外，您对这段表现还有何评价？
Question 3: In addition to your evaluations of the eight aspects listed above, do you have any other comments on this performance?

(Adapted from McAlloon 2008)

Although the researcher noticed some problems\(^{33}\) associated with the eight categories and recognized their potential for leading to confusion among the evaluators, the current study was still willing to adopt the same categories so that the findings can be easily compared with McAlloon’s conclusions. However, in the current study, an interview-based evaluation was incorporated following the portfolio assessment in an attempt to clarify any existing misunderstandings about the categories. Moreover, the three newly added questions were also designed to reveal some subjective aspects about the evaluation and to go beyond the eight categories.

\(^{33}\) McAlloon also recognized some problems about the eight criteria although he did not have a chance to improve the method in that study. For example, as the lowest-scoring criterion, the criterion “rhetorical effectiveness” seemed to be too vague to count as McAlloon’s study found Chinese instructors often abstained from scoring this criterion and there existed widely varying interpretations of this criterion among evaluators.
4.3.4 Obtaining Qualitative Data from Interviews of the Evaluators

Following each evaluator’s submission of his or her assessment results, the researcher scheduled an in-depth interview with each of them individually within one week. This interview allowed the researcher to qualitatively look into the underlying reasons for each evaluator’s rating and what he or she focused on in the assessment. A few general, open-ended questions were added to help the evaluator recall his or her impressions of each performance sample. In addition, the interview provided a good chance for the researcher to clarify any potential confusion or misunderstandings as a result of any unclear evaluation category. By interviewing the evaluators about their experience in using the portfolio assessment, the researcher was able to collect some beta-testing feedback that could be valuable for future implementation of this type of assessment instrument.

As mentioned above, every interview, which was approximately two hours long, was conducted within one week after the evaluator’s completion of the assessment. The three Chinese-teacher-domain-outsider evaluators’ interviews took place via Skype as the researcher was not able to travel to the cities where they lived. However, all the other interviews were face-to-face in Shanghai. All the interviews were conducted in Mandarin Chinese. The majority of the interviews were audio-recorded. Only two had to be recorded on paper, as the act of audio recording severely influenced the interviewees’ performances to the point that the interview had to stop. Both interviewees were from the domain-insider evaluator group.
4.4 Data analysis: Comparing Portfolio Evaluations by Evaluator and Criterion

4.4.1 The Two Subjects’ Portfolio Evaluations

The two subjects’ respective portfolios were established by including those performances observed and recorded from the ethnographic study presented in Chapter 3.

The following two sections will respectively introduce the two subjects’ portfolios including length of the performance, format and content, and more importantly, the specific domain or domains involved in the performance.

4.4.1.1 Alan Bahar’s Portfolio Evaluations

To recap, a total of 10 individuals evaluated Alan’s 11 performance samples from his portfolio. These samples included nine video clips, one audio clip and a text that is a review article of an article reviewing an art exhibition. The average length of the oral performances was approximately five minutes 24 seconds. All nine video clips contained verbal communications between Alan and one or several Chinese interlocutors although some identities were hidden upon request. The majority of the interactions were conversations; however, a few of them were closer to the presentation mode. The audio clip was a message recorded on WeChat, an audio-messaging app on Alan’s cellphone\textsuperscript{34}. This was the only sample in which the interlocutor’s responses were not available as it was not an on-going conversation but more like an offline voice message.

\textsuperscript{34} The WeChat app only allows for a 60 second audio message. Therefore this audio sample consists of four consecutive short WeChat audio messages. Since all four messages were on the same topic and were not interrupted, the researcher collected them all from Alan and edited them into one 4-minute long message.
Raw evaluation data for each performance sample in the portfolio can be found in Appendix J. Table 5 presents brief descriptions of the content of each performance sample, the role Alan played in the activity, and the major domain involved.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Format</th>
<th>Length</th>
<th>Description</th>
<th>Domain/Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Text</td>
<td>/</td>
<td>An art exhibition review written for the exhibition catalog.</td>
<td>Art/Artist</td>
</tr>
<tr>
<td>2</td>
<td>Video</td>
<td>8’32”</td>
<td>A conversation with exhibition staff about the design of the exhibition tag during a pre-exhibition set up.</td>
<td>Art/Artist</td>
</tr>
<tr>
<td>3</td>
<td>Video</td>
<td>2’11”</td>
<td>A conversation with exhibition staff and the other artist about the display of a piece of artwork during a pre-exhibition set up.</td>
<td>Art/Artist</td>
</tr>
<tr>
<td>4</td>
<td>Video</td>
<td>4’2”</td>
<td>Giving instructions at the beginning of a class on some “housekeeping” issues.</td>
<td>Pedagogy/Teacher</td>
</tr>
<tr>
<td>5</td>
<td>Video</td>
<td>5’54”</td>
<td>Giving instructions to three students regarding their respective pieces of art.</td>
<td>Art Pedagogy/Teacher</td>
</tr>
<tr>
<td>6</td>
<td>Video</td>
<td>1’33”</td>
<td>A conversation with a student who refused to participate in class activities.</td>
<td>Art Pedagogy/Teacher</td>
</tr>
<tr>
<td>7</td>
<td>Video</td>
<td>4’18”</td>
<td>Demonstrating calligraphy to a group of students.</td>
<td>Art pedagogy/Teacher</td>
</tr>
<tr>
<td>8</td>
<td>Audio</td>
<td>3’53”</td>
<td>A WeChat voice message on a few art/culture-related topics including idioms, Chinese and Western painting.</td>
<td>Academic/Graduate student</td>
</tr>
<tr>
<td>9</td>
<td>Video</td>
<td>8’28”</td>
<td>Discussing a graduate project to a friend in the dorm.</td>
<td>Academic/Graduate student</td>
</tr>
<tr>
<td>10</td>
<td>Video</td>
<td>4’16”</td>
<td>A conversation with a student committee leader to recommend someone’s artwork.</td>
<td>Academic/ Graduate student</td>
</tr>
<tr>
<td>11</td>
<td>Video</td>
<td>10’54”</td>
<td>Discussing a scholarship on Chinese history at a reading group meeting.</td>
<td>Academic/ Graduate student</td>
</tr>
</tbody>
</table>

4.4.1.2 Jack Mayer’s Portfolio Evaluations

Similarly, Jack’s portfolio also contained 11 performance samples, all of which were oral performances, including 10 video clips and one audio clip. The average length
of a performances was approximately five minutes 15 seconds. Three of the ten video clips were obtained from public resources with Jack’s permission, as they were clips of TV shows or public speeches Jack delivered in the past. The rest of the video clips contained verbal communications taking place at work, collected from various meetings with colleagues inside and outside Jack’s company. In most cases, the interlocutors requested that their faces be hidden in the videos. Moreover, some video clips of upper management meetings had to be edited so as to reduce the possibility of disclosing business secrets.

Three of the ten videos were presentations, including a brief speech on a TV show, a long public speech given at a business forum and a pre-recorded short speech to Jack’s employees about the Chinese New Year. The rest of the interactions were all in the conversation mode. The audio clip was from a similar company conference setting. It was audio-recorded instead of video-recorded mainly because Jack felt that the audio recording would provide more privacy than video during a meeting in which he addressed a few sensitive issues and criticized an employee.

Raw evaluation data for each performance sample in the portfolio can be found in the Appendix K. Table 6 presents brief descriptions of the content of each performance, the role Jack played in the activity, and the major domain involved.

Table 6. Descriptions of Performance Samples in Jack’s Portfolio

<table>
<thead>
<tr>
<th>Samples</th>
<th>Format</th>
<th>Length</th>
<th>Description</th>
<th>Domain/Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Video</td>
<td>1’15”</td>
<td>Talking about issues of cleanliness in the office space at a weekly meeting</td>
<td>Management/CEO</td>
</tr>
<tr>
<td>2</td>
<td>Video</td>
<td>3’59”</td>
<td>Talking about communication the various departments at a weekly meeting</td>
<td>Management/CEO</td>
</tr>
</tbody>
</table>

(Cont’d)
Table 6. Descriptions of Performance Samples in Jack’s Portfolio (Cont’d)

<table>
<thead>
<tr>
<th></th>
<th>Type</th>
<th>Duration</th>
<th>Description</th>
<th>Evaluator</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Video</td>
<td>53”</td>
<td>Discussing the importance of an internal audit of the company at a weekly meeting</td>
<td>Management/CEO</td>
</tr>
<tr>
<td>4</td>
<td>Video</td>
<td>1’52”</td>
<td>Giving a brief speech at the company’s Chinese New Year party</td>
<td>Management/CEO</td>
</tr>
<tr>
<td>5</td>
<td>Audio</td>
<td>7’24”</td>
<td>Criticizing a group of staff members who made a substantial error in their work</td>
<td>Game and management/CEO</td>
</tr>
<tr>
<td>6</td>
<td>Video</td>
<td>4’57”</td>
<td>Telling coworkers about new cell phone games</td>
<td>Game/CEO</td>
</tr>
<tr>
<td>7</td>
<td>Video</td>
<td>4’28”</td>
<td>Interview during a TV show to discuss the Chinese gaming industry</td>
<td>Game/CEO</td>
</tr>
<tr>
<td>8</td>
<td>Video</td>
<td>1’</td>
<td>Presenting during a TV show to introduce the business philosophy of his company</td>
<td>Game/CEO</td>
</tr>
<tr>
<td>9</td>
<td>Video</td>
<td>24’10”</td>
<td>Delivering a speech on the topic of video games</td>
<td>Game/CEO</td>
</tr>
<tr>
<td>10</td>
<td>Video</td>
<td>3’45”</td>
<td>Reporting on the situation of their business partners learned from a recent business trip</td>
<td>Game/CEO</td>
</tr>
<tr>
<td>11</td>
<td>Video</td>
<td>4’7”</td>
<td>Recommending a video game to a friend</td>
<td>Game/Friend</td>
</tr>
</tbody>
</table>

4.4.2 Analysis of Quantitative Data: Aggregate Portfolio Ratings by Evaluator

4.4.2.1 Alan Bahar’s Aggregate Portfolio Scores

Table 7 shows that, in general, the average scores Alan received from all three types of raters were above 4.0, which according to the 0-5 scales, corresponds to a “Strong Performance.” On average, the Chinese instructors awarded slightly higher scores than the domain insiders: the average of the evaluation scores from the Chinese instructors was 4.66 points and the average domain insider evaluation was 4.13 points, a difference of 0.53. However, there seems no discernible difference between the two domain outsider groups: the average non-teacher domain outsider evaluation was 4.69, only 0.03 points above the average Chinese instructor evaluation.
Table 7. Aggregate Portfolio Scores for Alan Bahar

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider</th>
<th>Domain Outsider</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Score for All Performance Samples</td>
<td>4.13</td>
<td>4.69</td>
<td>4.66</td>
</tr>
</tbody>
</table>

Upon closer analysis of average scores Alan received in each domain, it is clear that Alan’s performance was a little better in the domain of art, both in professional and academic contexts. While in the domain of pedagogy, Alan received lower ratings from the domain-insider raters, the average score of which was below 4.0. This difference however, is quite understandable since Alan is relatively new to this field. Discrepancies in the ratings between the domain insider and the domain outsider consistently appeared across the three domains, as shown in Table 8.

Table 8. Aggregate Portfolio Scores in Various Domains, Alan Bahar

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider</th>
<th>Domain Outsider</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average, <strong>Domain of Art</strong></td>
<td>4.23</td>
<td>4.61</td>
<td>4.78</td>
</tr>
<tr>
<td>Average, <strong>Domain of Pedagogy</strong></td>
<td>3.99</td>
<td>4.70</td>
<td>4.52</td>
</tr>
<tr>
<td>Average, <strong>Domain of Academic</strong></td>
<td>4.19</td>
<td>4.72</td>
<td>4.72</td>
</tr>
</tbody>
</table>

Table 9 shows the average scores respectively given by the three groups of raters on Alan’s portfolio. Based on the average scores of every type of rater on each individual performance sample, the current study identified the core performance samples as two, five and nine, respectively from Alan’s three domains. As the DI-CI differences of these three samples were greater than the other samples in each domain group, they are chosen
to be further studied so as to determine why the difference\textsuperscript{35} between domain-insider and Chinese-instructor ratings were significant (close to 1.0). Moreover, sample six, the rater type difference of which is less than 0.1, is also worth further exploration so as to understand on what aspects domain insiders and Chinese instructors tend to agree with each other. A thorough investigation about the controversy between different types of raters will be presented in Section 4.4.3, in which feedback given on each criterion in the interview will be examined in detail.

Table 9. Average Portfolio Scores by Rater Types, Alan Bahar

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher</th>
<th>Chinese Language Instructor (CI)</th>
<th>Differences (DI-CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>4.61</td>
<td>5</td>
<td>5</td>
<td>0.39</td>
</tr>
<tr>
<td>S2</td>
<td>3.83</td>
<td>4.3</td>
<td>4.71</td>
<td>0.88</td>
</tr>
<tr>
<td>S3</td>
<td>4.25</td>
<td>4.54</td>
<td>4.63</td>
<td>0.38</td>
</tr>
<tr>
<td>S4</td>
<td>3.63</td>
<td>4.63</td>
<td>4.17</td>
<td>0.54</td>
</tr>
<tr>
<td>S5</td>
<td>4</td>
<td>4.84</td>
<td>4.96</td>
<td>0.96</td>
</tr>
<tr>
<td>S6</td>
<td>4.43</td>
<td>4.67</td>
<td>4.48</td>
<td>0.05</td>
</tr>
<tr>
<td>S7</td>
<td>3.88</td>
<td>4.67</td>
<td>4.46</td>
<td>0.58</td>
</tr>
<tr>
<td>S8</td>
<td>4.14</td>
<td>4.91</td>
<td>4.62</td>
<td>0.48</td>
</tr>
<tr>
<td>S9</td>
<td>3.96</td>
<td>4.75</td>
<td>4.84</td>
<td>0.88</td>
</tr>
<tr>
<td>S10</td>
<td>4.38</td>
<td>4.63</td>
<td>4.5</td>
<td>0.12</td>
</tr>
<tr>
<td>S11</td>
<td>4.29</td>
<td>4.62</td>
<td>4.91</td>
<td>0.62</td>
</tr>
</tbody>
</table>

4.4.2.2 Jack Mayer’s Aggregate Portfolio Scores

As shown in Table 10 below, Jack’s aggregate portfolio ratings seem very similar to Alan’s. First, in general, the average scores Jack received from all three types of raters were above 4.0, which as previously mentioned, indicates “Strong Performance.” Second, the discrepancy between different rater types can be found in Jack’s portfolio ratings as

---

\textsuperscript{35} The “Differences (CI-DI)” is calculated by subtracting the average domain-insider rater score from the average Chinese-instructor rater score for all samples in the subject’s portfolio.
well. On average, the Chinese instructors awarded higher scores than the domain insiders; however, the difference of 0.47 was quite small. The two groups of domain outsiders disagreed more with each other on Jack’s rather than Alan’s performances, although the difference was smaller than that between domain-insider and Chinese-instructor raters.

Table 10. Jack Mayer’s Aggregate Portfolio Scores

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider</th>
<th>Domain Outsider Non-teacher</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average score across all performance samples</td>
<td>4.23</td>
<td>4.31</td>
<td>4.70</td>
</tr>
</tbody>
</table>

As introduced previously, there were also two relatively distinct domains in which Jack interacted every day at work: the domain of management, in which he is the CEO of the company, and domain of gaming, in which he is a professional who is knowledgeable about the gaming business in China and probably all over the world. When the researcher analyzed Jack’s performances in the two different domains separately, as displayed in Table 11, the previous finding based on the aggregate data that domain-insider group tend to be the harshest raters was found to be no longer valid. This indicates that the evaluations Jack received were different depending on the domain.

First, according to domain insiders, Jack’s performance in management was much weaker than his performance in the domain of gaming, a difference of 1.64. Moreover, the average score 3.14 points that Jack received from domain insiders for his performance in management indicates that domain insiders found Jack’s performance as a CEO to be only “passable performance”. This huge difference between the two different domains are further illustrated by an anecdote in which one of the raters thought the subject in
Samples One, Two and Three was not the same person as the one in the later samples. Second, there was a substantive discrepancy between the Chinese-instructor ratings and domain-insider ratings of Jack’s performances in management, a difference of 1.66. Third, there was very little difference between the evaluation of domain insiders and Chinese instructors of Jack’s performance in the field of gaming. One reason for this reduced discrepancy may be that performance samples in the field of gaming include three performances (Samples Seven, Eight and Nine) in two TV programs, consisting of both rehearsed and improvised parts, which made the performances more impressive. Moreover, the appearance on TV itself may have given the evaluator a preconceived positive impression as it is usually considered an honor to be recognized by the mass media. From the qualitative interview data on these three samples, comments were relatively short but very positive, such as “完美 (perfect)”,”“没什么好说的，一个字，棒 (Have nothing to say, just one word, ‘wonderful!’)” and “非常非常了不起 (very, very admirable).” Also, from Table 12, the quantitative data show that all three types of raters evaluated Samples Seven, Eight and Nine very high, close to the full score of 5.0.

However, more importantly, the tendency of Chinese-instructors to give higher scores completely changed in Jack’s case, in the domain of gaming, in which, they gave a a slightly lower evaluation than the domain insiders.

36 During the evaluation, the researcher found that raters tend to evaluate those performance samples that did not include the subject’s face (such as the audio sample or the textual sample) higher than those video samples in which they can see a foreigner’s face. Therefore, in some cases, raters would mistakenly think the same subject’s two samples were done by different people when they could not see the subject’s face in one of the samples. However, the situation mentioned here is the only case in which that rater failed to recognize Jack in some of his video performance samples due to his relatively weak performances.
Table 11. Jack Mayer’s Aggregate Portfolio Scores in Different Domains

<table>
<thead>
<tr>
<th>Average, Domain of Management</th>
<th>Domain Insider</th>
<th>Domain Outsider Non-teacher</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average, Domain of Management</td>
<td>3.14</td>
<td>4.26</td>
<td>4.80</td>
</tr>
<tr>
<td>Average, Domain of Game</td>
<td>4.86</td>
<td>4.35</td>
<td>4.64</td>
</tr>
</tbody>
</table>

Based on the average scores from every type of rater on each individual performance sample, as listed in Table 12, the current study identified performance samples one, two, three, and four as core performances for further analysis in Section 4.3.3 as the difference between domain-insider and Chinese-instructor ratings were large, all above 1.0. Moreover, it also worth noting sample six in the field of gaming, the only improvised performance on which the two types of raters tend to agree with each other, as the difference was less than 0.1.

Table 12. Average Portfolio Scores by Rater types, Jack Mayer

<table>
<thead>
<tr>
<th>Domain Insider</th>
<th>Domain Outsider Non-teacher</th>
<th>Chinese Language Instructor</th>
<th>Differences (DI vs. CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>3.34</td>
<td>4.3</td>
<td>4.88</td>
</tr>
<tr>
<td>S2</td>
<td>2.96</td>
<td>3.71</td>
<td>4.88</td>
</tr>
<tr>
<td>S3</td>
<td>3.13</td>
<td>4.54</td>
<td>5</td>
</tr>
<tr>
<td>S4</td>
<td>3.14</td>
<td>4.48</td>
<td>4.43</td>
</tr>
<tr>
<td>S5</td>
<td>5</td>
<td>3.66</td>
<td>4.63</td>
</tr>
<tr>
<td>S6</td>
<td>4.46</td>
<td>4.25</td>
<td>4.42</td>
</tr>
<tr>
<td>S7</td>
<td>5</td>
<td>4.96</td>
<td>4.96</td>
</tr>
<tr>
<td>S8</td>
<td>5</td>
<td>4.96</td>
<td>4.96</td>
</tr>
<tr>
<td>S9</td>
<td>5</td>
<td>4.84</td>
<td>5</td>
</tr>
<tr>
<td>S10</td>
<td>4.62</td>
<td>3.48</td>
<td>3.86</td>
</tr>
<tr>
<td>S11</td>
<td>4.96</td>
<td>4.3</td>
<td>4.71</td>
</tr>
</tbody>
</table>
4.4.2.3 Conclusion

From the aggregate portfolio scores, firstly, we can conclude that at work, both subjects demonstrated strong performances, which again verified their qualifications as experts in their respective domains. However, as their daily work also consists of multiple domains in which they have varied levels of experience, both subjects also demonstrated different levels of expertise in different domains. This phenomenon further illustrates that the capability of communicating in a foreign language is domain-dependent. The domain-free language proficiency (as both subjects were tested OPI superior) cannot guarantee the same level of performance across all fields. Jack’s case especially illustrates the fact that one’s domain expertise in one field cannot be easily transferred to another.

Secondly, based on the average scores of rater types, a discrepancy was found between domain-insider ratings and domain-outsider ratings. In general, domain-insider raters tended to be harsher in evaluating their domain counterparts’ performances, while there was relatively no difference among domain-outsider raters no matter whether one is a Chinese instructor or not. Chinese instructors tended to rate higher than domain insiders although with some exceptions that warrant further investigation in the next section. The same tendency was also revealed in McAloon’s (2008) study, in which three out of the five portfolios showed that Chinese instructors on average gave higher evaluations than did the domain insiders (p.374). Since the domain-insider group in McAloon’s study only consists of the subject’s colleagues, who may have tended to rate more generously than strangers, the finding that domain insider raters who were strangers in this study were even harsher than the Chinese instructor raters is not surprising.
Finally, differences in ratings between rater types can be minimized when some added value of the performance samples were recognized by the raters. For example, a performance shown on television indicating recognition by the mass media may have led certain types of raters to rate the performance high.

In the next section, interview data as well as the answers to the three open-ended questions\(^\text{37}\) on the evaluation sheet will be analyzed qualitatively. In order to answer the question what caused the large rater type differences in some performances, those categories that gave rise to most controversies will be analyzed in detail with a focus on the core performance samples identified in this section.

4.4.3 Analysis of Qualitative Data: Interviews with Chinese Raters

As previously stated, all 13 Chinese raters participated in one-on-one interviews with the researcher after they submitted their evaluation sheets. The average length of an interview was approximately two hours. A list of interview questions can be found in Appendix H.

Following a brief background investigation, the researcher first asked raters for their general impression about the foreign subject’s use of Chinese in a professional setting, which corresponded to the first open-ended question on the evaluation sheet. Responses to this question from all 13 raters were particularly positive despite a wide range of ratings, as reviewed in the previous section. A typical positive comment, “He sounded no different than native Chinese (听上去和中国人没什么差别)” was made by

\(^{37}\) As a recap, the three open-ended questions are: 1. In addition to your evaluations of the eight aspects listed above, do you have any other comments on this performance? 2. What were your impressions of this person? 3. Based on this person’s performance, would you like to further communicate with him or her? Why/Why not?
10 (out of 13) interviewees. Interestingly, three interviewees, including two domain insiders from Jack’s and Alan’s domains, respectively, and one non-teacher domain outsider, made the comment, “People would think he is Uighur from Xinjiang (人家还以为他是新疆人).” This comment, which shows a little sense of humor, was also mentioned in McAloon’s study when he recalled how a Chinese colleague positively commented on the performances given by one of his subjects (p.244).

Comments from all the three groups of raters focused on two aspects of the subjects: their personal attributes and their Chinese skill levels. Typical descriptions of the former type included humorous (幽默), cunning (油滑), astute (精明) and amiable (亲切随和) for Jack, and serious (认真), rigorous (严谨), responsible (负责) and knowledgeable (有见地) for Alan, all of which appeared in more than half of the raters’ interview comments. Typical comments of the latter type can be further divided into two types: measurements of linguistic accuracy, such as “some minor defects in tones (声调上还是有小瑕疵),” “clear pronunciation (吐字清晰)” and “a good command of various collocations (各种搭配掌握得不错),” and descriptive comments about the rhetorical aspects of the subject’s speech, such as “humorous (说话风趣),” “talking passionately (侃侃而谈),” and “natural and appropriate (自然得体).” Both types of comments can be found in the Chinese instructors’ interview responses. However, none of the non-teacher raters commented on any aspect of linguistic accuracy.

The second interview question focused on if the interviewees would be willing to have further communication with the subjects, as indicated in response to question two on the evaluation sheets. In Jack’s case, 68% of the responses were positive, indicating a
willingness to have further communication with him. Very similarly, 70% of the responses were positive in Alan’s case. However, only 13% of the responses indicated a reluctance to communicate with Jack, while 22% of the responses indicated an unwillingness to communicate with Alan. The rest of the responses were “N/A,” as shown in the Appendix. These responses usually included comments such as “It is hard to judge based on this video. I’d like to see more.” “I have no strong desire to talk to him, but I cannot say I am not willing to do so.” or no response. This decision, however, had very little to do with the subject’s performances, as indicated in the ratings and comments. In other words, a performance sample that received the most responses of “I do not wish to communicate with him” may not necessarily be the one that received the lowest average rating with regard to their performances.

Through an analysis of raters’ responses to this question, the researcher was able to identify several factors that influence the raters’ choice of whether or not they would like to further communicate with a foreigner in Chinese. First, if a performance fully demonstrated the subject’s personal characteristics, especially a pleasant personality or a positive state of mind, it usually motivated the rater to initiate further communication with the subject. This can be seen in raters’ positive responses to Jack’s public speeches and presentations on TV, in which he was able to showcase his humor. Comments such as “I am willing to talk to him as he seems to be an interesting person.”

38 It is also important to note that among the 13% negative responses Jack received, a certain number of them were given by domain outsiders who refused to communicate with Jack because of the cross-domain barrier. They either indicated that the topic was too unfamiliar to have a discussion with Jack or they were simply not interested in that topic. This situation, however, was not found in Alan’s case. In other words, it seems that Alan’s domain of Chinese art had a relatively lower threshold than Jack’s.
像很有意思)” were prevalent. For example, when raters saw Alan assisting a young artist in setting up his artwork at the exhibition (sample 3) and his patient conversations with his students (sample 5) raters were more willing to talk to him because they saw him as “a sincere and responsible person (他很真诚，而且认真负责).” On the contrary, Alan was criticized when he recommended a student artist to the graduate committee with comments such as, as “It is hard to tell his emotional state from his talk (看不出什么情绪)” and was relatively less well received. Second, raters were inspired for further conversation, when the subject’s performance demonstrated a deep understanding of Chinese culture, especially on the subjects of its history, geography and literary and cultural traditions. Alan’s beautiful writing filled with literary allusions and quotes from Chinese classical texts (sample 1), his conversation with a friend on Chinese idiom (sample 8) and his elaborate comments on a Chinese history book at the reading group meeting (sample 11) received 100% positive responses, generating a desire for further communication. On the contrary, performances that involved sensitive topics that are not acceptable in Chinese culture or speech and contained criticisms of Chinese society that would offend raters, would cause the raters to avoid a conversation at all costs. Although such cases were not found in the two subjects’ portfolios, it was reflected in Chinese raters’ attitudes towards some other foreigners’ performance samples that were mixed into the evaluation process. For example, although he spoke in impeccable Chinese, an American young man who talked about how some factories in China exploit child labor was turned down by most Chinese raters, who described his performance as “aggressive (咄咄逼人)” and “meddling in others' (or another country’s) business (多管闲事).” Third,
Chinese raters tended to describe a redundant and longwinded speech as “arduous (费劲)”, which made further communication with the subject unlikely. Alan’s performance in persuading the art museum staff (sample 2) was an example of this type.

Aside from these two questions, the rest of the interview focused on a qualitative investigation of the scores given by the raters. When the researcher asked the interviewees to elaborate on why they had given a certain score for a particular criterion, she was surprised to find greatly differing opinions among the rater types concerning the following three criteria: cultural appropriateness, rhetorical effectiveness and domain reference, which was also in accordance with the quantitative findings as reflected in their ratings. These differences were caused by either varying perspectives of the evaluations of the performances, or different interpretations of the criterion itself. The following three sections will explore in detail the differences between the evaluations of domain-insider raters and Chinese-instructor raters. Quantitative data will also be cited to support the findings reflected in the interview comments.

4.4.3.1 Evaluations of Cultural Appropriateness: Do Native Speakers Think Alike?

The difference between the domain insiders’ and Chinese instructors’ evaluations was 0.64, which contrasts with McAloon’s finding, in which he claimed that the difference between colleagues’ and Chinese instructors’ evaluations was very small since both types of raters are Chinese natives, who are equally qualified to assess cultural appropriateness. However, in this study, the identification as “native Chinese” alone did not guarantee the same standard for the evaluation of cultural appropriateness. This study
used Haohsiang Liao’s (2012) definition of cultural appropriateness, which refers to those identifiable elements of an encultured performance, usually co-occurring with particular scripts (p.70). When presented to the evaluators, the term “cultural appropriateness” was translated into “文化得体度” in Chinese, with a further explanation in parentheses: 言谈举止, which stated that cultural appropriateness should include both verbal (言谈) and nonverbal behavior (举止). The researcher explained this category to each evaluator through an example: If a student of Chinese greets his or her Chinese teacher with a perfectly pronounced and intoned Nihao 你好 (Hello) instead of laoshi hao 老师好 (Hello, teacher!), this should still be considered as culturally inappropriate despite his or her linguistic accuracy.

After closely reviewing the performance sample scores given for “cultural appropriateness,” the researcher found that all the evaluators tended to give higher scores when the performance manifested the cultural value of politeness and gave lower scores when the performance contradicted this cherished trait. Politeness can be manifested in such actions as declining a compliment of one’s ability by giving credit to the teacher, treating someone to a meal to show friendliness, helping someone save face by not mentioning his or her name when offering criticism. The three types of evaluators (domain insider, Chinese instructor, and non-teacher domain outsider) mainly agreed on the concept of politeness and had similar comments about it, including positive ones such as youlimao 有礼貌 (polite), qianxu 谦虚 (modest), qianxu zhongken 谦虚中肯 (modest and pertinent) liuyou yudi 留有余地 (leaving adequate leeway), bu qiangpo 不强迫 (not coercive), and youliyouju 有理有据 (reasonable), and negative ones including bureqing
不热情 (not enthusiastic), *bunaifan* 不耐烦 (impatient), and *quefa qinheli* 缺乏亲和力 (lacking friendliness, not easygoing).

However, although all types of evaluators were able to identify polite or impolite behaviors, domain insiders tended to go beyond this stage by associating the level of politeness with the role the subject played and the characteristics of that particular domain activity. An action consisting of a few seemingly polite behaviors may still be considered as impolite, or inappropriate when the action as a whole goes against expectations for that particular role or domain activity. One example comes from Alan’s performance, in which he was trying to persuade an art museum staff member to adopt his suggestion about the design of the exhibition information cards.

From Table 13, we can see that both Teachers Xiao and Xin gave Alan the highest score of five for cultural appropriateness. Xiao commented in the follow-up interview that Alan’s behavior was self-controlled, “moving forward or proceeding at a good pace” (进退有度). Xiao regarded Alan’s method as “persuasive, but not excessively so” (据理力争但又不过分), he was “polite and restrained” (有礼有节). Xin was also very satisfied with Alan’s polite persuasion. She specifically commented on Alan’s use of *lingdao* 领导 (leader)\(^{39}\) as an appropriate term of address to show respect to his supervisor, which she thought also contributed to his persuasiveness as it added a feeling of authority to his speech. Although, Teacher Xu’s ratings were slightly lower, she was also impressed by Alan’s persuasive technique. But she also pointed out that Alan did not smile, which she thought made the persuasion “not smooth enough” (不够圆滑).

\(^{39}\) The indexical meaning of *lingdao* has been discussed in details in 3.5.2.2.
Table 13. Scores on for Alan on cultural appropriateness, Performance Sample 2

<table>
<thead>
<tr>
<th>Domain Insider</th>
<th>Chinese Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang</td>
<td>Xu</td>
</tr>
<tr>
<td>Yu</td>
<td>Xiao</td>
</tr>
<tr>
<td>Yi</td>
<td>Xin</td>
</tr>
</tbody>
</table>

Sample 2 4 3 3 4 5 5

However, two out of the three domain-insider evaluators questioned the persuasive act itself, indicating that it did not match Alan’s role as an artist attending the exhibition. In the interview, Yu criticized that as an artist, it was not even acceptable to discuss the design of the exhibition label with the museum staff since it is not the artist’s responsibility, let alone to persuade them to adopt his own suggestion. Below is a quote from Yu’s interview:

我们出去书画展，不会这样，没人跟你啰嗦。人家有人家的工作，你把作品带去，怎么放人家是专业，对吧？你这样指手画脚算什么？资格再老，像张伟声老师这样的，也不会去指挥人家的，说你这个标签要怎么设计好看，我觉得这个……我要是工作人员，肯定是让人有点抵触的。我不知道和他们这个展览学生比较多，有没有关系。他们好像年纪相仿对吧？但我觉得你越位了总归不对。

When we attend an exhibition of calligraphy and Chinese painting, it is not usually like this (persuasive), no one will bother to talk to you (referring to the artist). They (referring to the museum staff) have their job, you bring your work there, but how to display your work, it’s their profession, right? No matter how respected you are, even the calligrapher Weisheng Zhang [Here, Yu mentioned a well-known calligrapher in Shanghai, known to the researcher as well], he will not give directions to others, saying that you should design the label in this way. I think this … If I were on that museum staff, I would definitely resent that. I am not sure if it is because this exhibition has more student artists. They (referring to Alan and that museum staff) are same age, right? But I think it’s always not right to step out of your role and overstep your boundaries.

Another domain insider, Yi, although did not directly question the validity of persuading the museum staff, commented that Alan had “a strong sense of originality, but an arrogant attitude (创新意识强，态度傲慢).”
The last domain insider, Yang, who rated the performance a little higher, also criticized Alan for not smiling. But according to Yang’s response in the interview, he thought the performance was still appropriate (得体). However, when the researcher shared the other two domain insiders’ criticism with Yang after the first round interview, Yang agreed that if he were the artist, personally, he would not adopt that persuader’s position when he attended an exhibition. However, he thought Alan’s assertive but gentle tone was still acceptable, especially since he was a laowai 老外 (literally meaning “an old-outsider,” which refers to a foreigner). He further commented that Westerners always gave him an impression of jiaozhen 较真 (being rigid, extremely serious) and pingdeng 平等 (being equal, without hierarchical difference), an attitude from which he thinks Chinese people should learn. Alan’s frowning face showed that he was very serious to the extent that he could ignore the existence of hierarchical differences (here, between a young artist and the staff member of the host museum), when challenging authority. Obviously, Yang’s opinion verified the fact that Alan’s method of persuasion was not appropriate behavior for a professional Chinese artist. However, indicating a persistent tendency to compliment Chinese speaking foreigners, Yang’s comment also further inspires ideas worthy of our attention. First, when a Chinese professional evaluates a professional performance within his or her domain, it is still likely that he or she would adopt a different standard for foreigner even though the performance was entirely conducted in a Chinese environment. Second, some features of “foreignism” may appear positive to Chinese evaluators even though they may not agree that the behavior would be considered culturally appropriate in Chinese culture. This, in fact, corresponds to
McAloon’s (2008) findings, in which he claimed that “giving compliments easily” and “using the Chinese court system” were the two skills that seem to impress Chinese evaluators even though they are also described them as “foreignisms” (McAloon 2008, 394). Similarly in Alan’s case, his serious demeanor and his frankness in expressing his disagreement to the authority were appreciated by some Chinese who speak highly of these two qualities in Americans.

To approach such a difference in a further step, it seems that the domain-insider evaluators possessed a more dynamic evaluation of the subjects’ performances. They paid closer attention to whether or not the subject’s behavior can fit into all the five elements that define a performance, including time, place, role, audience and the script. They tended to deduct points whenever a certain aspect of a subject’s performance appeared incongruent with some parameter of a perfect performance. For example, points were deducted when evaluators believed that Alan’s persuasiveness was not congruent with his role as an artist, and his interlocutor’s role as a museum staff member who is responsible for the exhibition.

Nonetheless, Chinese instructors tended to measure the performance with relatively objective standards, paying little attention to how appropriate those standards were for the situation. For example, using polite terms to persuade someone definitely met the Chinese instructor’s standard of politeness. Other standards they used included but were not limited to “being formal,” “wearing a smile,” “talking in a gentle tone,” “sitting still and straight” and “displaying modesty.”

---

40 The five elements are proposed by Galal Walker (2000), in his conceptualization of performance as “situated events.” According to Walker, a performance can be defined by five parameters including time, location, roles, audience and script.
Sample 4 also illustrates this difference, in which all the three domain insiders rated “cultural appropriateness” lower than the other three Chinese instructors. Sample 4 is the performance in which Alan was giving instructions to the whole class (in Sample 4) during his lecture on Traditional Chinese Culture at the CAFA High School. The specific ratings can be seen in Table 14.

Table 14. Scores for Alan on Cultural Appropriateness, Performance Sample 4

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider</th>
<th>Chinese Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 4</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

From the Chinese instructor, descriptions such as “carefully (细心)” and “patient (耐心)” prevailed in the evaluation sheets. However, although these characteristics were also recognized and mentioned by some domain insiders, they did not believe that these characteristics necessarily contribute to a positive image of a professional high school teacher. For example, although Yang commented that Alan was a patient teacher, he also pointed out that he lacked the authority required for a teacher when he smiled as some of his students ignored him and were slow about getting to their seats. Feng also mentioned “the lack of authority (没有威信),” which referred to Alan’s lack of concision and clarity in the instructions he gave in Sample 4. Feng not only deducted points in the area of “organization,” but he also took points off on “cultural appropriateness” because he said it was inappropriate for a teacher to act like that. On the other hand, Yu used another strong word popomama (婆婆妈妈, literally meaning “grandma/mom-like,” which indicates that someone’s speech is rambling and redundant, and behavior is slow and
hesitant) to criticize the same performance. However, none of these issues were identified or commented on by the Chinese instructors.

The researcher also noticed that domain insiders not only paid attention to whether the performances were consistent with the subject’s role in that particular communication context, but also whether the subject’s behavior was congruent with his or her identity as a foreign professional. Sometimes, such measurement seems extremely harsh and very personal. However, this discovery, in fact, corresponds to some other findings in previous studies, which, warrants further discussion below even though it is difficult to make any predictions at this point in the discussion.

In performance Sample 1, domain insider Yi gave a 3-point score to cultural appropriateness, which was two points lower than the other eight evaluators. Performance Sample 1 is an article Alan wrote reviewing an exhibition, published in the exhibition catalogue. The average rating on this sample was 4.87, the highest score of the 11 samples. Raters were impressed by the fact that a foreigner was able to write in Chinese at such a sophisticated level. Positive comments included “Beautiful writing style with a typical Chinese aesthetic perspective (文笔漂亮, 有中国式的审美),” “The writing contains many Chinese elements, which was not an easy job (中国元素很多, 不容易),” “Writing was concise and beautiful. He has a good grasp of the subject and can express himself so well. (文字简洁漂亮, 应该是很懂, 所以才能表达得那么好),” “Insightful and thought-provoking. Has a good understanding of Chinese culture. (有思想, 有见地, 懂中国文化),” “Perfect (完美)” and “I especially appreciate his opening and ending, which are very beautiful. (特别欣赏他的开头和结尾, 漂亮).”
However, domain insider Yi criticized that Alan’s writing was “piled high with too many Chinese elements” (中国元素太多了，堆砌了), which did not match his identity as a foreigner. Yi described such an attempt as zuozuo 做作 (unnatural, artificial), on which he further elaborated in the interview:

一个外国人说出“我年少不知礼”，这是很怪的，给人感觉为了用而用，比较虚，包括那个“明镜亦非台”。为什么用这种特别文绉绉的中国文人式的笔调？我不懂。……你看，他叫自己“研墨的人”，说不出来的味道总归，不太符合他的身份。但这只是我个人看法。

A foreigner saying something like “I am too young to understand li (rituals)” is very weird. It makes people think he used the expression just for the sake of using it, quite pretentious. This includes his use of that quotation “The mirror inherently is formless.” Why did he have to write in such a literary style that only belongs to Chinese literati? I do not understand… Look at this, he called himself yanmo de ren (a man grinding Chinese ink), I cannot tell exactly how I feel, but anyway, it does not match his identity. But it’s only my personal opinion.

Although Yi’s evaluation of this performance sample stood out as an outlier and he also emphasized that the comment might only be his personal opinion, he is definitely not be the first Chinese to question an American’s attempt to act “too Chinese.” An anecdote from Xiaobin Jian describes how an American who strictly followed Chinese banquet rituals when dining out with Chinese businessman actually wound up annoying them because they unfortunately interpreted his behavior as insincerity. The surface-level harmony and enthusiasm of the American was interpreted as formalism, (形式主义),

41 Here, the two Chinese expressions and one quotation Alan used indeed reflect the “Chinese literati-style” as Yi commented, not only because of its content such as the Confucian concept of li and the traditional literati’s life style of grinding ink, but more because of its discourse, which is very formal and literary conventional expressions. All these expressions may contribute to indexing an image of traditional Chinese literati who is very gentle and modest.

42 This anecdote was from one of my conversations with Professor Xiaobin Jian during our independent study sessions in autumn, 2013.
which although it guides and constricts aspects of banqueting in China, was perceived as insincere and even as a “parody,” in this context.

Both Yi’s critique and the example above raise the question of how far should one go to achieve acculturation? It is obvious that both foreigners in the two cases have a good knowledge about the inner workings of Chinese society. However, a Westner, acting completely in accordance with the conventions of Chinese society seems insincere or even offensive to the native Chinese. Cornelius Kubler (1997) provides an explanation for this phenomenon in an earlier publication:

Pretending to be a member of another culture, when one is obviously not, can be seen by members of that culture as presumptuous or even intrusive. This perception of “invasion” by members of a culture is not so much linguistically based as it is communicatively based. A foreigner’s excellent control of pronunciation, grammar, vocabulary, and the like is seen as a positive factor. However, if foreigners use communicative conventions that reflect an adherence to Chinese cultural values, attitudes, and traditions, there is a natural suspicion that they are pretending to be Chinese when they are not.

(NFLC Guide, 1997)

Going back to Alan’s case, even though he indeed is a person who has to grind the ink and use a brush to paint every day, the act of referring to himself as yanmo de ren or a “Chinese painting artist” to display his modesty was questioned and even criticized by some native Chinese. Similarly, when Alan wrote that he was too young to understand li, he was not perceived as a modest American young man striving to understand the Confucian culture, but someone who is “quite pretentious ( 比较虚 ),” according to Yi. His perfect writing, which demonstrated his excellent control of grammar, vocabulary and genre in Chinese composition, aroused suspicion from the native Chinese as it is
filled with communicative conventions that closely adhere to Chinese values and traditions.

4.4.3.2 Rhetorical Effectiveness: Focus on Rhetoric vs. Focus on Effectiveness

“Rhetorical Effectiveness” was the second area in which domain insiders and Chinese instructors tended to disagree. This is the only criterion ignored in McAloon’s study, even though according to his findings, it is the lowest-scored among domain insiders. McAloon abstains from exploring further as he claims it “may be too vague to count” (p.319). He notes widely varying interpretations of this criterion among evaluators and therefore, questions its validity.

In fact, the current study also noted the variations in the interpretation of the criterion among different types of raters. In general, we found that Chinese instructors seemed to have a more open-minded perspective about rhetorical strategy than domain insiders and non-teacher domain outsiders, who usually have little knowledge of linguistics. Chinese instructors seemed to be more willing to embrace the researcher’s explanation of the criterion -- any language that has a persuasive or impressive effect on its audience could be considered as a rhetorical strategy. However, the other two groups of raters tended to narrowly interpret rhetorical strategies as “比喻、拟人、夸张、排比 (figures of speech, personification, exaggeration, and parallelism),” which is based on what they were taught in primary and secondary Chinese education classes.

Based on this startling difference, one may possibly speculate that the domain insiders’ low rating of “Rhetorical Effectiveness” could be caused by their restricted view of the criterion. This prevented them from being aware of the diverse rhetorical strategies
used by the subjects. However, the current study dismiss this argument and contend that it is irresponsible to completely attribute the difference in ratings to the divergent interpretations of the criterion without further exploring those evaluators’ comments in detail. Particularly, such conjecture was unjustified as it cannot account for the significant difference that also existed between domain insider’s and non-teacher domain outsider’s ratings of this criterion since both groups had a comparatively narrow understanding of it.

Tables 15 and 16, respectively show the averages of Alan’s and Jack’s Rhetorical Effectiveness scores of their portfolios.

Table 15.Average Scores on Rhetorical Effectiveness by Rater Types, Alan Bahar

<table>
<thead>
<tr>
<th>Sample</th>
<th>Domain Insider (DI)</th>
<th>Chinese Language Instructor (CI)</th>
<th>Differences (CI-DI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>4.67</td>
<td>5</td>
<td>0.33</td>
</tr>
<tr>
<td>Sample 2</td>
<td>3.33</td>
<td>4.67</td>
<td>1.33</td>
</tr>
<tr>
<td>Sample 3</td>
<td>4</td>
<td>4.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 4</td>
<td>3</td>
<td>3.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 5</td>
<td>3.3</td>
<td>5</td>
<td>1.67</td>
</tr>
<tr>
<td>Sample 6</td>
<td>4.67</td>
<td>4.67</td>
<td>0</td>
</tr>
<tr>
<td>Sample 7</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Sample 8</td>
<td>4</td>
<td>4.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 9</td>
<td>4</td>
<td>4.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 10</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Sample 11</td>
<td>3.67</td>
<td>4.67</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 16. Average Scores on Rhetorical Effectiveness by Rater Types, Jack Mayer

<table>
<thead>
<tr>
<th>Sample</th>
<th>Domain Insider (DI)</th>
<th>Chinese Language Instructor (CI)</th>
<th>Differences (CI-DI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>3</td>
<td>4.67</td>
<td>1.67</td>
</tr>
<tr>
<td>Sample 2</td>
<td>2.33</td>
<td>4.33</td>
<td>2</td>
</tr>
<tr>
<td>Sample 3</td>
<td>2.33</td>
<td>4.67</td>
<td>2.34</td>
</tr>
<tr>
<td>Sample 4</td>
<td>3</td>
<td>4.33</td>
<td>1.33</td>
</tr>
<tr>
<td>Sample 5</td>
<td>5</td>
<td>4.67</td>
<td>-0.33</td>
</tr>
<tr>
<td>Sample 6</td>
<td>3.33</td>
<td>3</td>
<td>-0.33</td>
</tr>
<tr>
<td>Sample 7</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 8</td>
<td>5</td>
<td>4.67</td>
<td>-0.33</td>
</tr>
</tbody>
</table>

(Cont’d)
A close examination of the samples containing significant discrepancies between domain-insider and Chinese-instructor raters (as shaded in grey) along with the interview data associated with them revealed the impact of the different perspectives held by various rater types on the evaluation. Chinese instructors tended to focus more on the rhetorical devices used by the subject, and rated the performances accordingly by counting the number of rhetorical devices used. However, domain insiders in general, only recognized those rhetorical strategies that served the subject’s communicative intentions by consistently measuring their effectiveness.

One typical example that illustrates this point can be seen in Alan’s portfolio Performance Sample 11, in which he shared his experience of discussing a study on Chinese history at a reading group event. This approximately 10-minute sample was a plethora of sophisticated rhetorical flourishes, which were recognized by all evaluators, particularly the Chinese instructor Xin, who was able to enthusiastically recollect most of Alan’s impressive expressions. In the interview, Xin cited a series of expressions that Alan used to emphasize the suffering that the country had endured throughout history, which was eloquently described in the book. These expressions include *wanxi* 愧惜 (lament), *qiliang* 凄凉 (desolate), *beizhuang* 悲壮 (solemn), *huagmiu* 荒谬 (preposterous), *huangtang* 荒唐 (absurd), *yichongchong de kanke* 一重重的坎坷 (being burdened with ups and downs), *lingren luolei* 令人落泪 (close to tears), and
"nanyuanbeizhe" (literally meaning “going south by driving the chariot northward”, an idiom meaning that one’s actions are contrary to one’s purpose), all of which are very literary and elegant. Although other raters were also able to cite some of these expressions, Xin was the only one who included them on the evaluation sheet. From the instructor’s perspective, Xin specifically commented on Alan’s use of the measure word *yichongchong* (literally meaning layer upon layer, which is used here as a metaphorical use here to indicate being burdened with something), which she considered an excellent example of Alan’s superb knowledge of Chinese language and culture:

Knowing the word *kanke* (ups and downs) is already very impressive but he even used *yichongchong kanke* not just *henduo* (many). The expression *yichongchong* is not only used correctly, but also precisely. Also, it has a rhetorically aesthetic feeling.

Among those expressions Xin cited, four out of nine evaluators explicitly mentioned Alan’s use of *yichongchong kanke*, eight out of nine mentioned the Chinese idiom *nanyuanbeizhe*, and two Chinese-instructor raters positively commented on the use of the expressions *wanxi* and *qiliang*. However, not all the evaluators shared Xin’s appreciation of Alan’s attempts to use these elegant expressions, even though they were surprised to see a foreigner using so many sophisticated conventional expressions in a 10-minute talk. Domain insider Yang’s response was more typical of the raters:

思想的表达有些生涩，因为不知道要怎么表达，所以主要靠修辞。可堆砌这些修辞帮不了他，所以他还是一个外来人，只表达了感受，没有交流书的内容。
The way he expressed his thoughts was somewhat rough. Since he had no idea how to express himself, he had to heavily rely on rhetoric. However, over reliance on these rhetorical expressions cannot help either. So he still looked like an outsider (in Chinese culture), attempting to express his feelings, without actually conveying ideas about the content of the book.

It is particularly interesting to see how Domain Insider Yang explicitly commented on the uselessness of piling on rhetorical strategies that did nothing to convey the content of the book to his audience. Without being able to effectively introduce the book he read, which he assumed was the purpose of the event, Alan’s “precise” use of those rhetorical strategies was lost on Yang. Interestingly, the comment *duiqi* (piled with something in an excessive way) appeared again here, in Yang’s comment. As discussed in the previous section, the word *duiqi* was also used by domain insider Yi when he commented on Alan’s Performance Sample 1, criticizing that his writing was excessively filled with Chinese elements that he might not have the authority to use.

Similar criticism of Alan’s less pertinent comments on the book itself and his excessive elaboration on his personal feelings were expressed in the other two domain-insider raters’ comments. Yi appreciated Alan’s empathy for the life of Chinese intellectuals and the fact that he seemed to have developed a cultural awareness. However, he harshly criticized his speech as being repetitive and void of significance (“对知识分子生活很关心，对文化也很关注，但是讲的东西都是重复来重复去，很空洞”).

From these concrete comments, it is clear that while the Chinese instructor Xin was moved by Alan’s precise elaboration of his empathy for China’s suffering throughout history, domain insiders like Yang and Yi, however, considered those elaborations as ineffective compensation for his failure to express complex ideas with depth. More
importantly, to the domain insiders, sophisticated uses of rhetorical devices do not necessarily contribute to a high level of rhetorical effectiveness if it fails to serve a communicative purpose.

In addition to Alan’s Performance Sample 10, some contrasting comments from domain-insider raters and Chinese-instructors on other samples can further illustrate the findings presented above. To avoid redundancy, I will cite only a few comments from both types of raters in Tables 17-19 below, in which key points are bolded. Explanations will follow those comments when necessary.

Table 17. Selected Comments by Rater Types, Alan Bahar, Sample 2

<table>
<thead>
<tr>
<th>Rater Type</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DI</td>
<td>Yang: 很费力，摇摇摆摆，修辞不明显。 Yang: Very arduous. He is wavering. The rhetoric is not obvious. Yu: 语音准，语法棒，中文好，用词相当丰富，可就是说不出个道理。他最后说服成功了吗？ Yu: Impeccable pronunciation, good mastery of grammar, good Chinese and profound choice of words. But he just could not make his point. Did he successfully persuade her? Yi: &quot;干脆利索&quot;不能形容一个设计，要么&quot;干净利索&quot;还可以。其他好像没有用什么修辞。他在知道别人有规定的时候马上鼓励人家创新，蛮机灵的。 Yi: gancuilisuo干脆利索 (coming straight to the point without hesitation) cannot be used to describe a design. Rather, ganjingliluo干净利索 (clean and neat) is acceptable here. It seemed that he did not use rhetoric other than this idiom. When he heard from the staff that they had strict rules for the size of the design, he immediately encouraged them to be innovative. He’s very smart.</td>
</tr>
</tbody>
</table>

(Cont’d)
| CI  | Xin: 用词准确丰富。第一次使用“干脆利落”的时候，貌似应该是“干净利落”，不过不影响我给他的评分。“对不起那照片了”，“参差不齐”，“那咱们创一个”等词句使用得相当漂亮。中国人习惯称上司为“领导”，他使用上了，而且搬出领导是中国人的技巧，增加了说服力。Xin: His choice of words is precise and diverse. **He made a mistake when he first used the idiom gancuililuo (coming straight to the point without hesitation), which should be ganjingliluo (clean and neat). But it will not affect my rating.** He used some expressions beautifully, including “对不起那照片了 (wasted those photos)”, “参差不齐 (unevenly matched)”, “那咱们创一个, 创新 (Let us innovate one, be innovative).” Chinese people are used to addressing their supervisors as lingdao 领导, and he used that term. Also, frequently quoting your lingdao’s words is a strategy used by Chinese. **This made him more persuasive.**  
Xiao: “没有余地”用得漂亮。他还用了成语“干脆利索”和“参差不齐”。“创新”说得好。Xiao: meiyouyudi 没有余地(having no room for negotiation) was used beautifully. **He also used two idioms, gancuilisuo (coming straight to the point without hesitation) and cencibuqi (unevenly matched).**
Xu: 自己的道理说得很清楚。Xu: He made his point very clearly. |

There were two completely opposite evaluations of Alan’s Performance Sample 2: domain insiders viewed Alan’s performance as “arduous” and “wavering,” commenting, “He could not make his point clearly”; while Chinese instructors considered his performance “persuasive” and thought “He made his point very clearly.” Domain insider Yu’s comment indicates that he drew a clear distinction between speaking good Chinese and being an effective member of Chinese society. According to him, “Although Alan is undoubtedly qualified according to the former standard, his being not so successful in persuading his interlocutor made his excellent command of Chinese only an impressive but ineffective accessory.”

Table 17 (Cont’d)
There are three more evaluations worthy of attention. First, compared to domain insiders, Chinese instructors seemed to have a more acute sense of rhetorical strategies used in speech, as speculated previously. This can be seen from their elaborate comments and long lists of rhetorical effects. Nonetheless, both domain insiders Yang’s and Yi’s comments reveal their difficulty in identifying the rhetorical devices Alan used. Second, among the rhetorical elements, *chengyu*, the four-character Chinese idiom, usually stands out to catch the evaluators and is considered a reflection or a foreigner’s sophisticated mastery of the language. Moreover, the misuse of this idiom (*ganjinglisuo* vs. *gancuilisuo* in this example), although obvious to native speakers, seemed not to lead to a negative evaluation. As Xin claimed, “It will not affect my rating.” Third, *piaoliang* 漂亮 (beautiful) and *jiling* 机灵 (smart) seem to be another two characteristics of rhetorical effectiveness, other than the widely used comment “you shuifuli 有说服力 (persuasive).” The term *piaoliang* is not necessarily exclusively used to describe elegant or literary expressions, it is also used to express appreciation of originality, intelligence and humor. The two expressions instructor Xin mentioned (对不起那照片了 and 那咱们创一个，创新) are actually very simple and colloquial. However, they both evoked a somewhat playful tone that alleviated the anxiety of the situation.

---

43 Remember, this is the performance in which Alan was trying to persuade the art museum staff to remove the photo from the exhibition tag and to readjust the size of the label, which were against the museum’s regulations. In the first situation, Alan tried to explain to staff member who spent a lot of time taking those photos that the change of label design “would not make them feel sorry for those photos (对不起那照片了)”, indicating that the photos would be used elsewhere. In the second situation, he half-jokingly encouraged the museum staff to violate the museum regulation by “inventing a new size (咱们创一个，创新)”.

169
Table 18. Selected Comments by Rater Types, Alan Bahar, Sample 5

| DI  | Yang: 语气生硬 (修辞扣分), 老师不需要端着架子。  
Yang: A stiff tone. (Points, therefore, were deducted from “Rhetorical Effectiveness.”) Teachers do not need to put on airs.  
Yu: “概念了”一用得太抽象了，不适合指导学生,如果说“观察得不够仔细”就好了。“头疼治头，脚疼治脚”，大局上没有把握，不停地去重复，没什么效果。  
Yu: *gainianle (conceptualize) was used in a too abstract way, which is not appropriate for instructing his students.* It would be better if he had said, “You did not observe carefully enough.” The Chinese slang says “Cure your head if you have a headache; cure your feet if you have a pain in your feet,” which means you should hit to the point. He failed to take the big picture here, repeating himself a lot, which, therefore, was not effective.  
Yi: “小玩意儿”很北京的说法，“糙的纸”，说“讲究了”，也很地道。  
Yi: *xiaowanyir (a little gadget) is a typical expression in Beijing dialect. He also used *caodezhī 糙的纸 (coarse paper) and *jiangjiule 讲究了 (exquisite). Both expressions are very genuine uses of Chinese.*

| CI  | Xin: 修辞用了 “糙”、“润”，特别专业。  
Xin: As far as rhetoric was concerned, he used *cao 糙* (coarse), and *run 润* (smooth), *very professional.*  
Xiao: 润一点，用词到位。我想就算是个中国教国画的教授也不会讲的比他好了。  
Xiao: *runyidianr 润一点儿 (smoother), very accurate usage.* I think even a professor of Chinese painting could not have taught better than he did.  
Xu: 他说那个学生把手机“概念了”，这个说法比较高级，但不知道学生理解了没有（看不到学生的表情）。  
Xu: *He said that student gainianle 概念了 (to conceptualize) that cellphone, which is a very advanced usage.* But I am not sure if those students understood his meaning. (The students’ facial expressions cannot be seen in the video.)

In the above example, both Chinese instructors and domain insiders recognized Alan’s precise use of most domain-related terminologies. However, one of the terms, *gainianle 概念了 (to conceptualize)*, he used to criticize a student’s careless painting, was appreciated by Xu, a Chinese instructor as “a very advanced usage.” However, Yu, a domain insider, perceived it as ineffective instruction for the students. Yu pointed out that
telling a high school student something like “You need to be more observant.” would be more helpful than using a very abstract term that may be misinterpreted. As an experienced art teacher, Yu was able to recognize the difficulty that high schools students may have with abstract terminology in art. However, Chinese instructors tended to be impressed with complex and high level uses of language, and the term was indeed “advanced,” if not completely appropriate.

Table 19. Selected Comments by Rater Types, Jack Mayer, Sample 4

| DI  | Zhong: “比较洋气”用得挺巧，知道要凑谐音，不过知道的词不多，所以只能说洋气了。作为公司领导层的讲话，水平好像比较一般，没什么出彩的地方，就是平铺直叙的。
Zhong: bijiao yangqi 比较洋气 (very fashionable) was used intelligently. He knows well that he should use a homophone of yang 羊 (sheep). But he did not know many words, so that is why he could only say yangqi (fashionable). As a (supposedly sophisticated) talk from the upper management, this seems very ordinary and lacking in impressive elements. It’s very straightforward.
Zhao: 对活用“羊”字的文化有一定了解，但是用得牵强。“给力”用得熟练。
Zhao: He knows the cultural convention of using words containing the homophones of yang 羊 (sheep) (at the beginning of the Year of Sheep). However, he use of the word was farfetched. Geili 给力44 was used smoothly.
Zhao: 说得太随便、散漫。没讲出多少内容。新年应该讲得更鼓舞人心。或者讲得很实，或者讲得很幽默。他不太像个 CEO，就是什么“给力”啊，“洋气”啊，档次不高。
His speech was too casual, very careless. His speech lacked content. During the New Year, he was supposed to say something very motivating. He could have either talked in a very sincere way, or talked in a very humorous way. He did not talk like a CEO. Expressions such as geili or yangqi, are not very high-class expressions.

(Cont’d)

44 Geili 给力 was the most popular Chinese cyber buzzword of 2010, which describes something as "cool," "awesome" or "exciting."
Table 19 (Cont’d)

| CI | Xin: 内容不错，她说了“洋气”。  
Xin: The content was good. He used *yangqi* 洋气 (fashionable).  
Xu: 讲话得体，能鼓舞士气。作为 CEO, 有幽默感，又平易近人。  
Xu: His speech was appropriate, which can motivate his employees. As a CEO, he has sense of humor and is also approachable.  
Xiao: 很会 motivate 人啊，是个让人喜欢的领导，“洋气”说得应景。还有“给力”这个词当年多火啊。  
Xiao: He’s very good at motivating people and is likable. The term *yangqi* was appropriate for the occasion. Also, *geili* was so popular during that year (2010). |

Jack’s performance in Sample 4 was a videotaped speech for the company’s Spring Festival Gala, which is a very common event in Chinese organizations. In Jack’s approximately two-minute speech, he recalled the hardships and growth of the company during the past year, positively affirming the efforts of his employees. He also mentioned several upcoming challenges in the new year, including enlarging the global market, switching the focus to cell phone gaming and having the company go public. At the end, he presented his New Year’s wishes to everyone in the company.

In general, Jack’s talk was very colloquial in style and his speech was described as “straightforward” (平铺直叙). However, the two rater types interpreted this simple and unadorned speech style very differently. Domain insiders such as Zhong and Zhang expected the speech of a CEO to be more formal and considered Jack’s speech to be “not very high-class” (档次不高). However, Chinese instructors thought the speech created an “approachable” and “humorous” image of Jack, which was “very motivating.” These significantly contrasting comments might be caused by very different expectations for a New Year’s speech given by upper management. When evaluating the performance,
domain insiders usually picture themselves doing the same kind of task in their career, which leads to a more critical evaluation, as they clearly know what would be expected of them. Although Chinese instructors, however, may have had a clearer expectation for standard language use, they may have been less able to envision a costly gala thrown by a gaming company. Chinese instructors seldom questioned if the pairing of Jack’s colloquial and straightforward style with the image of an approachable leader was appropriate for such an occasion.

In this particular speech, Jack followed the Chinese custom of using a word containing the homophone Chinese character of the zodiac animal of the year, which was yang 羊 (sheep) in 2015, in his New Year’ wishes. Five out of six evaluators explicitly mentioned this in the interview, recognizing Jack’s familiarity with Chinese culture and his willingness to put his knowledge of cultural norms to work in his speech. However, unlike Chinese instructors who were 100% positive about Jack’s use of the term yangqi 洋气 (fashionable), domain insiders questioned the appropriateness of such a blessing word, as it is very uncommon to wish someone “a fashionable new year” in Chinese culture. Domain insiders Zhong and Zhang attributed this faux pas to Jack’s limited knowledge of Chinese idioms, so that the incorrect term, yangqi was the only word he was able to think of at that time.

From all four examples discussed above, it can be seen that both groups of raters were very appreciative of the subject’s sense of humor, eloquent persuasion and uses of conventional expressions (especially four-character idioms) and were able to recognize them rhetorical strategies. However, domain insiders usually evaluated these aspects based on their situational appropriateness, while Chinese instructors were easily
impressed by the level of sophistication of the expressions themselves. (A typical Chinese-instructor comment was “This is a very advanced expression” or “He even knows this idiom. It’s a HSK level 5 vocabulary term.”) In these circumstances, Chinese instructors always evaluated this criterion by adding up points based on the number of rhetorical strategies they recognized, while domain insiders tended to deduct points for ineffective or inappropriate uses of rhetoric.

Two examples from Jack’s portfolio can further illustrate domain insiders’ attentiveness to the situational appropriateness of the rhetorical strategies used. In Performance Sample 2, Jack emphasized the importance of communication among different departments at an upper management staff meeting. He recalled a story about how the lack of communication between the two departments in his company embarrassed him in front of his business partners as his staff had no idea what projects others were doing. Jack’s complaint that wo haomeimianzi 我好没面子啊 (I really lost my face!) was recognized by the domain outsiders including the Chinese instructors as “friendly” (亲切), “natural” (自然), and “not pushy” (不给人压迫感). They appreciated his ease in revealing his emotional state to his employees, which again contributed to his image of an approachable boss. However, two domain insiders criticized this comment by pointing out that “saying ‘I really lost my face’ just called attention to a problem without suggesting a solution for it.” (“很没面子“只是一个效果，没有发现问题本质，也没有找到答案。) In addition, domain insiders felt that Jack’s comment was neither concise nor clever, which was not appropriate for the highest leadership in the company. The
ambiguity of his attitude prevented him from effectively addressing this critical issue in this company.

Likewise, in a similar situation, when Jack addressed the importance of supporting the internal audit of his company so as to prepare for going public, he said “I also feel very annoyed” (我也感到很烦) to show his understanding of how his employees felt about the trivial paper work required for internal auditing. Although this speech was identified by the Chinese instructors Xin and Xu as an effective way to show empathy to his employees, it was again disregarded by the domain insider, Zhong. According to Zhong, such a speech would only diminish his authority as a CEO without sufficiently addressing the need for being tolerant of the situation, which was the message he should be sending to the company as they prepared to go public.

4.4.3.3 Domain References: Do Great Minds Think Alike?

Differing evaluations of domain references may be the most predictable finding in the data. It is reasonable to believe that the variations in domain familiarity among different rater types will lead to different evaluations. Tables 20 and 21 display the average domain reference scores given by domain insiders and Chinese instructors, respectively for Alan’s and Jack’s portfolios. It is obvious that Chinese instructors consistently gave much higher ratings than domain insiders. Particularly in Alan’s case, the Chinese-instructor rater group gave the highest possible scores (5.0) to eight out of 11 performance samples, while the domain insiders were not as generous, and gave no 5s. These significant differences (Difference ≥ 1.0) between the two groups, which will receive further investigation, are highlighted in green in the two tables. Also, there were a
few outliers, bolded in the two tables, in which Chinese instructors were even harsher than domain insiders. These exceptions indicate that people who are less familiar with the domains were even more critical than the domain experts, which seemed to contradict our hypothesis that a lack of domain familiarity would make one more lenient.

Table 20. Average Scores on Domain References by Rater Types, Alan Bahar

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider (DI)</th>
<th>Chinese Language Instructor (CI)</th>
<th>Differences (CI-DI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>4.67</td>
<td>5</td>
<td>0.33</td>
</tr>
<tr>
<td>Sample 2</td>
<td>4.33</td>
<td>5</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 3</td>
<td>4.33</td>
<td>5</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 4</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Sample 5</td>
<td>4.33</td>
<td>5</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 6</td>
<td><strong>4</strong></td>
<td><strong>2.67</strong></td>
<td><strong>-1.33</strong></td>
</tr>
<tr>
<td>Sample 7</td>
<td>3</td>
<td>3.67</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 8</td>
<td>3</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Sample 9</td>
<td>2.67</td>
<td>5</td>
<td>2.33</td>
</tr>
<tr>
<td>Sample 10</td>
<td>4.33</td>
<td>5</td>
<td>0.67</td>
</tr>
<tr>
<td>Sample 11</td>
<td>4.67</td>
<td>4.67</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 21. Average Scores on Domain References by Rater Types, Jack Mayer

<table>
<thead>
<tr>
<th></th>
<th>Domain Insider (DI)</th>
<th>Chinese Language Instructor (CI)</th>
<th>Differences (CI-DI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>3.67</td>
<td>5</td>
<td>1.33</td>
</tr>
<tr>
<td>Sample 2</td>
<td>4.33</td>
<td>5</td>
<td>1.67</td>
</tr>
<tr>
<td>Sample 3</td>
<td>4.33</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Sample 4</td>
<td>2.67</td>
<td>4</td>
<td>1.33</td>
</tr>
<tr>
<td>Sample 5</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 6</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 7</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 8</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 9</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Sample 10</td>
<td><strong>4.67</strong></td>
<td><strong>2.33</strong></td>
<td><strong>-2.34</strong></td>
</tr>
<tr>
<td>Sample 11</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>

A comprehensive investigation into both types of samples may provide two potential explanations for such seemingly inconceivable and contradictory findings. First,
generally speaking, a lack of knowledge about the domain and the nature of activities taking place within it would be very likely to inspire a rater to give a more generous score. When domain insiders observe someone accomplishing tasks that they themselves are very familiar with, they will likely become more critical, as their domain-related knowledge and experiences will lead them to maintain a high standard. This high standard reflects the pride they take in their area of expertise, as revealed in their responses during the interviews. For example, domain insiders tended to compare Jack's performance on samples one-four with those of some highly respected CEOs in their area of expertise. For example, Chen Tianqiao (陈天桥), the board president and co-founder of the most influential developer and operator of online games, was mentioned many times during the interviews with domain insiders. A typical criticism of performance sample 3, in which Jack discussed the importance of supporting auditing in his company to help prepare it for going public is stated below:

陈天桥让盛大上市的时候肯定不是这样鼓励员工的吧。还是可以学习优秀游戏行业企业家的发言，更上一层楼。

When Chen Tianqiao encouraged his employees before Shanda Games45 go public, he definitely would not say things like this. (Jack) can still learn from those excellent entrepreneur in the field of gaming industry, to reach a higher level. (Domain insider Zhang)

When domain-insider raters criticized Jack’s performances for its lack of charm and charisma, they also mentioned role models in the broader context of business. For example, two domain insiders suggested that Jack read Ma Yun's books on management

---

45 As the first Chinese operator of online games that went public on the Nasdaq, Shanda Games and its founder Chen Tianqiao were easily mentioned as the model in that area.
strategy (马云), a Chinese business magnate who is the founder of a highly successful Internet-based business.

From these examples, we can see that the empathy that a domain insider may have developed during the process of observing the performance of his or her counterpart may not necessarily result in sympathy or kindness in the actual evaluation. Instead, it may instill pride in maintaining high standards of excellence, which may cause the domain insider to score more harshly. One of the two outliers, performance sample 6 in Alan’s portfolio, also reflected this phenomenon. The ratings show that the Chinese instructors, as domain outsiders, evaluated this criterion much lower than domain insiders. However, a close examination of this performance sample revealed that although the performance took place in a Chinese art class, the content of the performance was actually exclusively pedagogical. In this sample, Alan was criticizing a student for his lack of participation in class activities. From their interview responses, it is apparent that the Chinese instructors perceived themselves as domain insiders when evaluating this particular sample. Their typical responses include, the following: “If I were that teacher (如果我是那个老师),” or “If my student … (encountered that situation).” Table 22 presents a selection of responses given by the three Chinese instructors regarding domain references.
Table 22. Chinese Instructors’ Comments on Domain References in Performance Sample 6, in Alan’s Portfolio

Xu: (It demonstrated) a lack of domain knowledge. In addition to flaunting his own authority, he repeated himself. I think if we were put in a similar situation, we would have just failed that student without embarrassing him in front of the class. If don’t think it is appropriate in America: you would not talk about this in front of other students. He repeatedly asked the student why he never participated in his class just because as a teacher, he had no idea about how to teach students.

没有什么专业知识，他除了展现了威严以外，一直都在重复自己的话。我觉得如果是我们中国的老师，直接就挂了他，不会愿意说那么多。如果是美国的教育方式，那也不会在大庭广众之下去说，我觉得挺不妥。他不停重复问他为什么上课不参与，就是因为他作为老师也不知道该怎么教育。

Xiao: He was not encouraging to his student. I myself would not have acted like that. I cannot say it’s professional. It does not involve any domain references, domain knowledge or techniques, except for his understanding of the high school educational system, as Alan repeatedly mentioned that his class was a “required course.”

他和学生很疏远，我自己不会这么做。谈不上专业，没有涉及到什么专业知识技巧。不过他对高中的教育系统还是有一点了解的，好像提了好几次“必修课”。

Xin: Although his performance was consistent with his role, it was excessive. If my student watched online videos in class (as Alan’s student did), I would not talk to him in this harsh manner, because you can tell that Alan’s student was not enthusiastic about the class at all. I think Alan lacks the basic techniques that a teacher should know.

他的表现虽然符合他的身份，但是太过了。如果我的学生上课看视频，我也不会用这种生硬的方式，因为你可以看到他的学生不为所动。我觉得他缺少做老师的基本技巧。

Second, the lack of domain familiarity not only means that a domain outsider will be unaware of the hidden weakness behind those seemingly confident and fluent speeches, but also it prevents him or her from recognizing meaningful domain references in specific jargon-heavy communications. This may explain why domain outsiders become even “harsher” in a few cases. As discussed previously in Section 3.5.1.1, in sample 10, when Jack was commenting on how to make money through “changing the skin (换皮)” of an old game by marketing it with a newly designed cover, his accurate use of domain-
specific terms such as *huanpi* (changing the skin) and *liushui* (an exaggerated turnover self-reported by a company) was immediately identified by the domain insiders as intelligent and professional. However, lack of knowledge of this occupational jargon prevented domain outsiders from assigning meanings to these lexical items. As one Chinese instructor commented, “There’s nothing special about this performance. He is just an employee working in a gaming company (没什么特别的，就是一个游戏玩家).” Moreover, domain outsiders could not appreciate that performance as much as the domain insiders, not only because they failed to understand the jargon, but also because they lacked familiarity with the nature of the domain activity itself. Another Chinese instructor indicated on the evaluation sheet, “The performance is not professional at all (一点也不专业)” and gave a 0 to this criterion. When asked about her harsh words, this instructor criticized the overall style of Jack’s speech as “too casual, he stopped whenever he wanted (太随便了，想到哪儿讲到哪儿).” According to this rater, since the sample description said it was a performance in which Jack was reporting back from a business trip, she was bothered that Jack’s performance was very disorganized, containing a few random topics such as encouraging someone to open a stock account. These disorganized topics covered in Jack’s report plus a few indecipherable terms completely destroyed that Chinese teacher’s appreciation for the performance. However, all the domain insiders, whose ratings of domain references in this performance sample were close to a full score, defended Jack and were able to explain why he had to mention those seemingly random elements in his report. As domain insider Zhong responded:

He was just introducing the company he had visited during a business trip. Isn’t that company Feiyu Technology? Well, the only theme of his report was that
Feiyu is a promising company with a high profit margin. But he did not exaggerate that part, as did his employee. There is a difference between liushui and income. So I think it’s a good catch, very professional. Then, he just said he wanted to purchase the intellectual property of a game that company had produced, “change the skin of it” and make money. He has a businessman's mind, and is very sensitive to the market. (Interviewer: It seemed that he made a comment about the stock market which made everyone laugh. Do you think that was too casual?) The stock he mentioned is that of Feiyu Technology, as it is very profitable. This seemed an inside message that he shared with his employees. But he mentioned it also to emphasize that this company has a promising future. Or maybe he just said that for fun. I think it’s good. We should all buy that stock. Haha.

Finally, the researcher also found that what underlies a critical evaluation from domain-insider raters is not always the higher standards they embrace, but simple disagreement about a certain domain subject in some cases. This kind of situation in which low ratings are given just because the rater disagrees with the subject is considered biased and must be prevented if a researcher plans to implement such an assessment in the future. Examples of this kind were mainly found in Alan’s portfolio (samples 8 and 9), which reflected the old Chinese idiom, “Intellectuals tend to disparage one another (文人相轻).” In performance sample eight, Alan and his classmate exchanged their views on why Chinese people tend to use idioms more often than Western people. Alan mentioned a famous Chinese contemporary artist Xu Bing (徐冰)’s view, that the preference for using idioms is rooted in Chinese people’s appreciation of formulaic expressions. He
expanded Bing's analogy by stating that the preference for formulaic expressions is just like the preference for following an exemplar over valuing originality in traditional Chinese art. Alan elaborated that this kind of preference can be reflected in the emphasis on painting from exemplars in Chinese art versus painting from nature in Western art. He also stated that both calligraphy and Chinese Kung-fu share the same spirit. In the four-minute performance, Alan’s talk was filled with domain-specific expressions, such as “师古人、师造化 (learning from old masters and nature),” domain knowledge of representative artists, exhibitions and artworks related to this topic, such as “天书 (Book from the Sky by Xu Bing),” “芥子园画谱 (Manual of the Mustard Seed Garden),” “八九前卫艺术 (1989 Avant-Garde Exhibition)” and “马蒂斯 (Matisse)” and a domain-specific rhetorical style, reflected in his highly literary and elegant speech (e.g. 把你的视角纳入一个文脉之中 How Your Perspective Fits into the Culture). Moreover, his delivery of these points was also clear, fluent and concise. However, during the interview, domain insiders seemed to be unsatisfied with Alan’s opinion, which actually was the artist, Xu Bing’s, view. According to domain insider Yu, who rated the criterion a 3.0, "Alan firmly believed that the preference for using Chinese idioms has nothing to do with formulaic expressions, but the historical stories associated with every idiom." Based on this disagreement, this rater deducted points from this criterion.

Very similarly, in sample 9, in which Alan explained a work that he was preparing for his M.A. graduation exhibition to a friend, his borrowing of the Chinese artist Qi Baishi (齐白石)’s philosophy “似与不似之间 (painting between similarity and non-similarity)” was again challenged by the two domain insiders. They thought Alan’s
view was “outdated 老生常谈的调子” and “shallow 比较浅,” and therefore gave this performance low scores.

4.4.3.4 Conclusion

A close examination of the qualitative interview data identifies three areas in which domain insiders and Chinese instructors disagree with each other most: cultural appropriateness, rhetorical strategies, and domain references. These disagreements between the two types of raters deeply reflect their different focuses of attention in evaluations.

First, in evaluating cultural appropriateness, both types of raters tend to identify performances that manifested culture values of politeness as being culturally appropriate. However, domain-insider evaluators possess a more dynamic evaluation of the subjects’ performances, going beyond the stage of simply identifying polite or impolite behaviors. They pay close attention to whether or not the subject’s behavior can fit into all the five parameters defining a performance, through consistently associating the level of politeness with the role the subject played and the characteristics of that particular domain activity. Based on this standard, an action consisting of a few seemingly polite behaviors may still be considered as impolite, or inappropriate when the action as a whole goes against expectations for that particular domain activity or a certain aspect of that subject’s performance appears incongruent with the role he or she played in that activity. Chinese instructors instead, tend to measure the performance with some relatively objective standards, such as using the polite term of address, wearing a smile,
and talking in a gentle tone, while paying little attention to how appropriate those standards were for the situation.

Second, in evaluating rhetorical strategies, both groups of raters were very appreciative of the subject’s sense of humor, eloquent persuasion and uses of conventional expressions, such as the four-character idioms and were able to recognize them rhetorical strategies. However, domain insiders usually evaluated these aspects based on their situational appropriateness, while Chinese instructors were easily impressed by the level of sophistication of the expressions themselves. In other words, domain-insider raters would only recognize those rhetorical strategies that effectively served the subject’s communicative intentions, while the Chinese instructors tend to focus more on the rhetorical devices used by the subject, and rate the performances accordingly by counting the number of rhetorical devices used.

Last but not least, the familiarity with a domain plays a critical role in determining the evaluation of domain references. The lack of domain familiarity prevents domain outsiders from noticing either ineffective demonstrations of professionalism or meaningful domain references that beyond their understandings, which therefore, lead to either implausibly positive comments or unreasonably harsh evaluations. Domain insiders, however, usually keep a critical eye towards their counterparts’ performances. The rise of empathetic consciousness in the process of evaluation encourage domain insiders to consistently maintain a high standard, reflecting the pride they take in their own field.

All these three characteristics differing between domain-insider and Chinese-instructor evaluations make domain insiders more critical and rigorous evaluators in most cases. However, there also exist two types of potential biases in domain insiders’
evaluations. First, in evaluating rhetorical effectiveness, domain insiders hold a relatively restricted view in understanding the term “rhetorical.” Some of them tend to narrowly interpret rhetorical strategies as “比喻、拟人、夸张、排比 (figures of speech, personification, exaggeration, and parallelism),” as a result of their primary and secondary education from the Chinese classes (语文课). This less open-minded understanding sometimes prevent domain insiders form paying attention to some other meaningful expressions that have a persuasive or impressive effect on its audience. Second, in evaluating domain references, domain insiders tend to confuse the demonstration of domain references with expressing one’s points of view in domain-related subjects. Therefore, when they disagree with the argument the subject presented, they tend to evaluate low of their ability in demonstrating domain references even though the “disliked argument” is very professional. Moreover, the tendency that “scholars always disparage one another” is reflected in some of the domain-insider evaluations, which obviously diminish the objectivism in their evaluations.
Chapter 5 Towards A Performance-based Portfolio CFL Assessment

Following the proposal of a paradigm shift in foreign language assessment from proficiency to expertise in Chapter One and a review of how the concept of “expertise” is defined in Chapter Two, I conducted two studies with the goal of identifying and assessing expert performances in Chinese as a foreign language. In this chapter, I discuss how the findings in the previous chapters might translate into a new set of assessment instruments for CFL – a performance-based portfolio assessment.

The first study, presented in Chapter Three, operationalizes the concept of expertise by identifying a series of characteristics observed in the two subjects, both of whom are foreigners regarded as experts in their respective fields. These traits provide lenses through which the content of foreign language assessment at an advanced level can be brought into focus. Based on this implication, in this chapter, I suggest examining certain skills in the assessment of the professional use of Chinese as a foreign language. All of which is discussed in detail.

The second study presents the significantly different ways in which various demographic groups approach this type of assessment, which provides a rationale for recruiting domain expert evaluators to assess advanced professional foreign language use. In addition, an examination of the ratings and interview responses in the portfolio assessment portion of this study further highlights the necessity for establishing transparent statement-based criteria within the assessment tool that reflect domain-
specific expectations. I then propose a set of rubrics to be used in a performance-based portfolio assessment tool, which is based on the feedback collected from the raters participating in this study. Rooted in the lessons learned from the experimental implementation of the portfolio assessment, I also discuss a series of proposed changes in the techniques for implementing and administering new assessment instruments.

5.1 A Content Selection Framework for Creating a Representative Portfolio

The 22 video samples featuring the two subjects, presented in Chapter 3, effectively showcase their expertise in communicating professionally in their respective domains. Identified characteristics underlying the two subjects’ expert performances in these videos contribute to establishing a content selection framework for building a dynamic and holistic portfolio to demonstrate expertise.

Although performance samples collected for this study were all from professional settings, characteristics identified from the expert performances are not necessarily restricted to a particular domain. As a matter of fact, these identified aspects can be further divided into two categories: those of the domain-general and the domain-specific, which collaboratively contribute to creating the traits of the ideal colleague and experienced professional. Based on the findings presented in Chapter Three, I suggest including in one’s portfolio performance tasks that reflect a list of features discussed in the following three sections. The purpose of such a portfolio is to present as complete a portrait as possible of the subject as a professional foreign colleague with whom Chinese people are willing to work.
5.1.1 The Domain-general Aspects of a Portfolio

Domain-general characteristics that are prized by native speakers of Chinese usually include politeness, amiability, intelligence, and a good sense of humor. Moreover, these characteristics of a foreign professional’s communication style indicate that there is no need for a native Chinese colleague to make accommodations in terms of pronunciation, pace, or style in any intercultural communication.

First, a foreign professional’s use of informal language that contains key features typical of spoken language usually conveys a message to native speakers that they are able to maintain effortless communication that requires less accommodation from their native colleagues. Foreign colleagues demonstrated their ability to understand colloquial speech, including ellipses, hedges, inversions and discourse fillers, even when they were used in an unconventional way. This knowledge is a likely guarantee of a smooth interaction. On the contrary, speaking in a scripted way in daily interactions may convey to the Chinese colleagues a lack of post-program immersion in the target culture, which would have helped them create a more natural and informal speech style more readily understood by their colleagues. Moreover, expert interactions with Chinese colleagues on a daily basis also highlight a high level of control of these spoken discourse features to achieve various communicative intentions. For example, Alan’s use of inversions of predicates in several places naturally exerted an appropriate level of emotion in his speech. His method of revealing strong dissatisfaction was easily detected and understood by his Chinese colleagues, making the communication effortless and clear.

Second, the ability to demonstrate one’s knowledge about Chinese dialects in dealing with people from different areas of China can easily impress a native Chinese and
reduce the distance between them. Due to the large number of varieties in Chinese language, Chinese people grow up accommodating a variety of dialectal accents. Moreover, using vernacular dialectal expressions is a key strategy in achieving humorous effect in one’s speech. Both experts in this study were observed switching adeptly between Mandarin and at least one another dialect so as to fit in different roles they played and to achieve different communicative intentions. They are not only equipped with the ability to understand the dialectal expressions used by others, but are also aware of exploiting their own abilities in speaking dialect to their own advantages. For example, Alan knows well how using a Beijing-centered Mandarin will help him establish his authority in teaching at a high school while some appropriate uses of rustic Beijing dialect can present him as an interesting person in front of his Chinese colleagues outside the professional setting. Similarly, Jack worked hard to keep a balance between adopting a Southern dialectal elements so as to win the trust from more business partners and getting avoiding his overwhelming Hunan accent (such as his effort in dropping the distinctive discourse marker “ma 嘛”) when delivering a formal speech in the public media venues. If a foreign subject is able to include some performance samples in which they accommodate the accent of the interlocutor they converse with or they use certain dialectal expression to bring a sense of humor to foster a lighter atmosphere, they will usually appear intelligent and make themselves look interesting to talk to.

Third, it is suggested that the portfolio users should include evidence of demonstrating their informational culture knowledge through behaviors, especially in the field of Chinese history and Chinese geography. A foreigner who is able to deal with a sizable amount of knowledge in these two fields are considered to be someone who has a
high appreciation of Chinese culture and has the motivation to delve into the diversity of Chinese culture that is a source of pride for most Chinese. The behavioral aspect of one’s knowledge about Chinese geography can be demonstrated in dealing with people from different areas of China by making relevant and sophisticated comments on the cultural features associated with their interlocutors’ hometown. This type of small talk frequently occurred in interactions inside and outside professional settings, shortening the social distance between the interlocutors. Moreover, knowledge about Chinese history can be demonstrated in one’s behavior of recognizing those achievements of the culture due to China’s long history such as idioms, ancient poems and classical texts. Although this ability to recite a Chinese poem or use sophisticated discourse filled with Chinese idioms has some domain-dependent value, they are generally appreciated by Chinese people, especially when used by a foreigner.

On a side note, these domain-general characteristics also suggest qualities that are important at all levels, which should be incorporated into the instruction even at an early stage of learning. For example, typical Chinese discourse fillers that highlight a natural and informal speech style can be introduced and practiced at the beginning level while the behavioral aspect of the informational cultural knowledge should always be emphasized whenever a piece of cultural information is introduced. Although curriculum and instruction are not the focuses of the current dissertation, these pedagogical implications from the current findings should not be neglected and worth further researches.
5.1.2 The Domain-specific Aspects of a Portfolio

The duality of expertise as discussed in the previous chapters determines that the domain-specific aspect of a portfolio is equally important. Expertise in language is a tool for developing and demonstrating expertise in a domain that uses language for expression, transaction, and socialization within the community of experts and non-experts in that domain. At the same time, a domain expertise is prerequisite to claiming expertise in a foreign language. To focus on this issue, this dissertation defines domain as a restricted range of social interactions within a culture, bound by commonly shared knowledge and motivated by identifiable and shareable goals.

In the context of CFL assessment, domain plays a significant role as it is the knowledge of the nature of the domain activity that allows domain members to grasp the meaning of utterances playing a role within that domain. A series of domain-specific activities constituting each domain should be identified and used as valid performance tasks in a domain-specific assessment. The structural properties of each activity serve as the source of inferences, predisposing members of a domain to respond to utterances as if they are fulfilling certain functions. On the one hand, it is an understanding of the social situation of a domain that allows the learner to assign meanings to lexical, grammatical, phonological, and discursive structures encountered in domain activities. On the other hand, it is also the competence in using the language that enables one to engage in all the social dimensions of a domain (including social interactions, social relationships, and social institutions), which should also be fully reflected in the performance in assessment.

Based on the findings from the two subjects’ portfolios, it is suggested that foreign subjects should include tasks that demonstrate their epistemic authority within the
domain. This is usually manifested through the subjects’ skillful use of the occupational jargon. By adopting domain terminologies in one’s interactions, a foreign professional is able to maintain a concise and highly efficient communication with their domain counterparts. Moreover, based on the domain insiders’ feedback revealed in Chapter Four, it is important to make sure that actions included in the portfolio should be congruent with the role that the foreign subject played in that particular domain activity, meeting the domain expectations in terms of the level of rhetoric, cultural appropriateness and domain references.

5.1.3 The Meta-cognitive Aspects of a Portfolio

As found in most portfolio assessment definitions, the student’s self-reflection on his or her work is a mandatory component of the proposed generic content selection framework (Graves 1983; Tierney, Carter, & Desai 1991). Commenting on one’s learning process and style is often considered a form of metacognition, which becomes essential for self-regulation.

Findings from the previous chapters already demonstrate that experts’ self-monitoring skills are manifested in their superior ability to notice errors they have made in the past and to accurately predict the level of difficulty of tasks in their domains. Self-monitoring skill also allows them to accurately estimate the amount of time and resources to allocate to the task ahead. Both subjects observed in this study are skilled at setting observable and useful criteria to recognize accomplishments and to evaluate their performances. Performances analyzed in this study further reveal a general process of a CFL expert skillfully coping with an unfamiliar expression in the conversational mode.
and assuming the responsibility for improving and refreshing their Chinese from time to time.

However, had it not been the interaction between the recorded performances collected from the ethnographic observation and the two subjects’ thinking process revealed in the SR tasks, these experts’ superior self-monitoring skills would have never been fully presented. Therefore, it is crucial to suggest every user of such a portfolio assessment to gather metacognitive evidence on a regular basis as a separate type of entry that describes the development of the competency from the learner’s point of view. At the documentation stage, just as what the two subjects did in the SR interview, learners can be asked to justify their portfolio entries and comment on their relevance. It can take the form of tape-recorded verbal justifications accompanying each performance entry collected in the portfolio. It can also involve a personal overview of all the various entries, the links that emerge among these and their likely reflection of the learners’ development of the competency.

5.1.4 Conclusion: A Content Selection Framework

Although there should be fixed categories in the content selection framework as discussed in the previous three sections, the selection process within each (e.g., number of entries and format) remains flexible. The goal of such a portfolio is to showcase a non-native Chinese speaker’s ability to demonstrate one’s domain expertise through an expert use of the target language in a working environment of the target culture. Table 23 presents the content selection framework proposed above as a conclusion.
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Standard</th>
<th>Example of entries</th>
</tr>
</thead>
</table>
| Domain-general | Domain-general interactions that demonstrate one’s politeness, amiability, intelligence, and a good sense of humor as a good colleague that native Chinese are willing to work with, e.g., interactions in very colloquial speech, being able to deal with people from different parts of China, demonstrating a sizable amount of knowledge in Chinese history and Chinese geography. | 1. Evidence of accommodating different accents in conversation  
2. Small talks with Chinese by commenting on their hometown and the dialects they speak  
3. Evidence of making culture reference in a casual conversation |
| Domain-specific | Domain-specific tasks that demonstrate one’s epistemic authority within the domain, e.g., a skillful use of the occupational jargons to converse domain-related topics in an efficient and concise way. | 1. Evidence of an effortless performance on routine work.  
2. Evidence of three problems solved within the professional setting.  
3. A 5 to 10-minute presentation on a domain relevant topic to domain outsiders (e.g. Jack’s audience in front of TV)  
4. A 5 to 10-minute presentation on a domain relevant topic to colleagues |
| Metacognitive  | Self-reflection and self-regulation, e.g., awareness of one’s own problem-solving process and its effectiveness; coping with unfamiliar expressions in conversation; exploiting available resources to further improve one’s competency in the target language on a regular basis. | 1. Verb justifications accompanying a problem-solving performance  
2. Comparison of two problem-solving processes in two different contexts by analyzing different strategies used and their respective effectiveness  
3. Evidence of coping with an unfamiliar expression, by either requesting for repetition and explanation or making educated guess and chunking the expression from the flow of conversation  
4. A verbal documentation of one’s self-managed learning to maintain expertise |
5.2 Recruit Non-English Speaking Domain Experts as Evaluators

The average scores of rater types presented in Chapter Four clearly pointed out a discrepancy existing between domain-insider ratings and domain-outsider ratings, while there was relatively no difference among domain-outsider raters no matter whether one is a Chinese instructor or not. Specifically, the lack of domain familiarity prevents Chinese instructors, as domain outsiders, from noticing either ineffective demonstrations of professionalism or meaningful domain references that beyond their understandings, which therefore, lead to either implausibly positive comments or unreasonably harsh evaluations. Unlike the Chinese instructors who tend to count the number of rhetorical devices used by the foreign subjects, domain-insider raters only recognized those rhetorical strategies that effectively served the subject’s communicative intentions. Moreover, in most cases, domain insiders are more critical and rigorous evaluators, who always evaluate the expert performances based on their situational appropriateness, without simply being interfered by the high level of sophistication of the expressions used in the interactions.

However, in 4.4.3.4, the data also pointed out two existing types of potential biases in domain insiders’ evaluation, which requires the assessment to recruit at least more than one domain-insider evaluators as compensation. Moreover, the potential biases existing among domain-insiders make the Chinese instructors’ participation as evaluators equally meaningful. With more training on increasing one’s sensitivity towards the context of communication, Chinese teachers will be able to recognize meaningful expressions that contribute to the rhetorical effectiveness instead of narrowly restricting their attention on some literary devices and rhetorical strategies only. Moreover, without
an empathetic consciousness in evaluating the subjects’ performances, Chinese instructors should be relatively fairer in evaluating performances involving arguments they do not agree with.

Therefore, based on these findings, it seems that feedback from both domain-insiders and Chinese instructors should be included so as to foster an effective evaluation of the domain-related activity.

The vast variety of domains existing among advanced level learners calls for a significant amount of investment in identifying domain experts in China and working out a long-term collaboration between experts in China and organizations administering the assessment. This makes it crucial to invest on developing a network of potential “evaluators” from a variety of domains to form a repertoire that can serve long-term needs for CFL assessment. These evaluators should be relatively accessible so that a test-taker can schedule an assessment efficiently. Moreover, these evaluators should also be hired for providing consistent evaluations of the rating metrics so that they can be timely revised if they fail to reflect the changing expectations for foreign professionals within a certain domain.

5.3 Establish Transparent Categories and Criteria for Evaluating Expertise

Once we shift our focus from proficiency to expertise, we expect a performance-based portfolio assessment to provide the most transparent standards to help evaluators decide the extent to which a CFL learner is able to use his/her Chinese language skills to effectively communicate with Chinese counterparts in their respective domains. The evaluation categories and their criteria adopted in the current study were adapted from the
ALPPS, which were designed mainly for educators in the field of foreign language education and are therefore academic in nature. For this reason, many problems became apparent during the process of implementing the portfolio assessment, particularly revealed in the interviews with the non-teacher evaluators, which included both domain insiders and domain outsiders. For our purposes, the goal of such an assessment is for employers of international organizations to be able to use the portfolio assessment tool to choose the most qualified candidate. Therefore, the rating categories and their respective criteria need to be revised to reflect the most important questions a potential employer would ask any job candidate.

Interviews with the portfolio evaluator in the current study revealed a few problems of vagueness with the categories we used since the same rating category may mean different things to different raters. If such portfolio assessment is to become a commercial product to be used by domain insiders around the world, the training program must be flexible and easy to use so that evaluators will actually complete the review process. For example, as previously mentioned, the rhetorical effectiveness criterion posed a particular challenge to domain insiders and Chinese instructors in terms of determining a common definition.

Moreover, the translations of the categories from English to Chinese were also confusing on top of the already-very-academic terms. Since most evaluators' native language is Chinese, it would make more sense to draft these categories in Chinese while providing some general English explanations as reference.

In order to highlight the transparency and cultural accuracy of the rating categories during the assessment process, I would translate the previously used categories
and their criteria, along with the three open-ended questions asked in the interview and place them into a list of statements presented as a Likert Scale. Table 24 (the original version in Chinese) and 25 (the translated version in English) provide a sample of such statement-based rating categories, maintaining the 5.0 scale used in the current study.

Table 24. Sample Rating Categories and Criteria for CFL Portfolio Assessment

<table>
<thead>
<tr>
<th></th>
<th>1. 您将如何评价视频中这个外国人的语言表达？ *</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>表达非常流利：</td>
<td>1 十分不符</td>
<td>2 不太相符</td>
<td>3 勉强符合</td>
<td>4 比较符合</td>
</tr>
<tr>
<td></td>
<td>发音标准，无洋腔洋调：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>选词精准、地道：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2. 您将如何评价视频中这个外国人的语言组织能力？ *</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>表达简明扼要，不啰嗦：</td>
<td>1 十分不符</td>
<td>2 不太相符</td>
<td>3 勉强符合</td>
<td>4 比较符合</td>
</tr>
<tr>
<td></td>
<td>逻辑清晰，观点非常清晰：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>表达逻辑性很强，有条理：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>3. 您将如何评价视频中这个外国人的文化得体度？ *</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>言谈举止给人感觉舒服：</td>
<td>1 十分不符</td>
<td>2 不太相符</td>
<td>3 勉强符合</td>
<td>4 比较符合</td>
</tr>
<tr>
<td></td>
<td>懂得中国人的礼节规范，并能照之行事：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>4. 您将如何评价视频中的这个外国人在对话中的应对能力？ *</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>根据对方的话迅速调整自己的谈话内容</td>
<td>1 十分不符</td>
<td>2 不太相符</td>
<td>3 勉强符合</td>
<td>4 比较符合</td>
</tr>
<tr>
<td></td>
<td>是个好的倾听者（不只是自顾自地说）：</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Cont’d)
Table 24 (Cont’d)

5. 您将如何评价视频中这个外国人的说话风格？
   1 十分不符  2 不太相符  3 勉强符合  4 比较相符  5 非常相符
   谈话生动，有文采：
   很有说服力，让人印象深刻：
   能淋漓尽致地展现个人风格（如幽默）：

6. 您将如何评价视频中这个外国人的文化知识水平？
   1 十分不符  2 不太相符  3 勉强符合  4 比较相符  5 非常相符
   知识丰富，能引经据典：
   了解并热爱中国文化：

7. 您将如何评价视频中这个外国人的专业水平？
   1 十分不符  2 较不相符  3 勉强符合  4 比较相符  5 非常相符
   对自己的领域十分了解：
   在交谈中熟练运用大量行业术语：
   仪态专业到位：

8. 基于本段视频，您是否愿意与视频中的外国人继续交流？
   1 很不愿意  2 不太愿意  3 一般，没什么特别感觉  4 比较愿意  5 非常渴望
   继续交流的意愿：
   与之共事的意愿：

9. 基于本段视频，您对该外国人的印象是？
   1 十分不符  2 不太相符  3 勉强符合  4 比较相符  5 非常相符
   他／她是一个十分聪明的人
   他／她是一个友好、容易打交道的人
   他／她是一个十分负责、可靠的人
   他／她是一个学习能力很强的人
Table 25. Sample Rating Categories and Criteria for CFL Portfolio Assessment (Translated Version)

1. How will you evaluate this subject's delivery? *

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1 Strongly disagree</th>
<th>2 Disagree</th>
<th>3 Neutral</th>
<th>4 Agree</th>
<th>5 Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>His/her speech is fluent:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His/her pronunciation is accurate with little foreign intonation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The choice of words is always precise:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. How will you evaluate this subject's organization of speech? *

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1 Strongly disagree</th>
<th>2 Disagree</th>
<th>3 Neutral</th>
<th>4 Agree</th>
<th>5 Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>He/she speaks concisely and to the point:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He/she expresses his/her argument clearly:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His/her speech is well organized with strong logical connections:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. How will you evaluate the level of cultural appropriateness in this subject's speech? *

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1 Strongly disagree</th>
<th>2 Disagree</th>
<th>3 Neutral</th>
<th>4 Agree</th>
<th>5 Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>His/her speech makes me feel comfortable:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He/she is familiar with Chinese norms and acts accordingly:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. How will you evaluate this subject's ability to adjust his or her speech to the audience? *

<table>
<thead>
<tr>
<th>Criteria</th>
<th>1 Strongly disagree</th>
<th>2 Disagree</th>
<th>3 Neutral</th>
<th>4 Agree</th>
<th>5 Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>He/she can quickly adjust the speech to accommodate that of the interlocutor:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He/she is a good listener who seldom dominates the conversation:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Cont'd)
### Table 25 (Cont’d)

<table>
<thead>
<tr>
<th>Question</th>
<th>1 Strongly disagree</th>
<th>2 Disagree</th>
<th>3 Neutral</th>
<th>4 Agree</th>
<th>5 Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. How will you evaluate this subject's rhetorical style? *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His/her speech is beautiful and highly rhetorical:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>His/her speech is persuasive and leaves a good impression on the audience:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>His/her speech fully demonstrates his/her personal charm (e.g. being humorous):</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6. How will you evaluate this subject's ability to make cultural references? *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He/she demonstrates abundant cultural knowledge and is able to discuss it intelligently:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>He/she demonstrates an appreciation for Chinese culture:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7. How will you evaluate this subject's ability to make domain references? *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He/she demonstrates strong familiarity with his/her domain:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>He/she uses numerous domain terminology in conversation:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>His/her manner is professional:</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8. Based on this video, are you willing to further communicate with this subject? *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am willing to further communicate with him/her.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I am willing to work with him/her as a domain counterpart.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
It is important to note that, the above tables only provide a sample of how the rating categories should be worded to evaluators outside the academic world. However, the beauty of such a portfolio assessment instrument is that it allows organizations and their employers to identify and reinforce a specific set of behaviors that they believe are necessary for success in the target industry or organization. Ideally, a pool of rating-category statements should be designed and made available to the organizations so that any recruiter who plans to use the assessment can first select the most relevant categories for foreign employees and then add some other special categories that reflect the culture of their organizations.

5.4 Apply Modern Technology to Implement the Assessment

Over the past decades, we have benefited greatly from the development of technology in the field of education, which has allowed us to best utilize the resources available to us. The current section will discuss the technology required to implement a multi-rater portfolio assessment, as described above. This should allow us to effectively connect foreign professionals to native speaking domain insiders at almost any point on the globe and to have their performances accessed anywhere at any time.

5.4.1 Create an Online Platform

We are so fortunate that we are in the Internet age, when assessments can be implemented from almost anywhere in the world. Moreover, long-term databases can be maintained for people with specific needs to identify qualified personnel or to match up organizations through accessing stored performances with multi-faceted ratings. Through
such a platform, students, teachers, and eventually recruiters can view evaluation reports in English, identify the groups of evaluators, and specify the actual criteria on which the evaluations are based. An examination of commercial platforms designed to perform summative assessments revealed that this type of platform already exists, although at this time each of them just targets a narrow aspect of the assessment. For example, The Grady Profile, which was developed by Aurbach and Associates, allows teachers to create portfolios of student work and evaluate them using alternative assessment methods. A frame, designed by Salmat Learning is specifically used for corporate employee training and assessment. Fanping 凡评, a U.S. company partnered with The Ohio State University, which assists international students in applying to universities in the United States, shares a similar assessment philosophy to that applied in this study. By creating their own performance portfolios, international applicants can demonstrate their English proficiency and showcase their academic and interpersonal abilities to U.S. institutions before they submit an official application. The technology adopted by these web-based application assessments can inspire us to develop a multi-rater performance-based portfolio assessment, as previously discussed.

5.4.2 You Tube-type Technology

Language learning is a lifelong undertaking, which requires that an effective foreign language assessment that creates a longitudinal record of learners’ progress throughout their learning career. A You Tube-like technology will allow students or instructors to upload collected video performance samples to the Internet and easily share the link with qualified evaluators anywhere in the world. Videos are the most effective
and transparent demonstration of the learners’ progress, through which not only can we assess their linguistic skills, including paralinguistic features of their speech, but also be able to access their non-verbal behaviors that are equally important in evaluating a performance. At the same time, a stable video website with easy access, like You Tube, can also be used to store reliable databases for long-term use. Since a significant percentage of the evaluators are very likely to be located in China, it is important to test the video downloading speed in both the U.S. and China to make sure it meets most people’s video downloading expectations.

5.4.3 Video-conference Technology

As mentioned previously, the collected performance samples should target domain-specific activities, which in most cases involve interactions with domain counterparts in China. Therefore, the Internet may be the only technology to reach professionals around the world. Through widely available video-conferencing tools, such as Skype, learners from anywhere in the world are able to communicate about occupational and academic topics with people in China.

In addition, though we encourage learners to frequently videotape their performances in the target language in a naturalistic setting, it is unavoidable and actually necessary to include simulated domain-specific tasks to elicit performances if our goal is to gather a repertoire of in-depth, rich performances. In this case, the use of video-conferencing technology is especially helpful in presenting a realistic venue that captures the essential features of a particular work setting.
5.4.4 Digital Data Storage

Digital storage technology is essential for maintaining a longitudinal performance-based portfolio to store and retrieve data. When reviewing ALPPS, McAloon found that the full resolution videos would not be practical as all video file sizes must be less than 10mb for uploading and easy viewing by evaluators with varying Internet connection speeds. Therefore, a stable digital media storage website with large storage size would be ideal for such assessment.

5.5 Administer the Assessment thorough a Collaborative Participation of Various Institutions

As revealed previously, such multi-rater, multi-media portfolio assessment will always incur a high cost, which should be viewed as a reasonable investment for a certain group of users. Ideally, the institution associated with language teaching should not directly administer such assessment instruments. Instead, an independent organization, such as ACTFL, can be recruited to develop and administer the assessments. This will not only increase the creditability of the assessment results, but will also allow for an objective evaluation of various language-learning programs.

Moreover, similar to McAloon’s recent suggestion:

Organizations that have invested in producing high-value foreign language speakers, such as the Language Flagship and multinational corporations, can build on the experience of the 360-degree feedback practitioner community to perfect similar tools for assessing performance in foreign language environment” (2015, 168).

These assessment tools will be invaluable to governmental organizations, prestigious Chinese language programs and Chinese organizations, whose reviews can
inform others about its effectiveness. Additionally, institutions and individuals with pedagogical expertise and native speakers with domain expertise should also be recruited to evaluate and revise these metrics on a regular basis to ensure they reflect the most appropriate and up-to-date standards for assessing expertise.

5.6 A New Look at CFL Assessment: Should We Change? Why? What? And How?

With China emerging on the global stage and U.S. exports to China increasing (from $40.5 billion in 2005 to $120.8 billion in 2014\(^{46}\)), it is clear that the two countries’ economies are becoming more intertwined. Certainly, the Sino-American relationship extends well beyond trade in goods, reaching into all significant fields on an ever-increasing scale. The headlines over the last several years depict the perception among Americans that China will play a significant role in defining their future. The current increasingly competitive global environment therefore requires students to become truly advanced in their ability to speak Chinese and engage in extended interactions with counterparts in China as well as to interpret the intentions of Chinese individuals and organizations to achieve common goals. The more Chinese and Americans work together, the more of value it will become to develop assessment devices beyond the language program to measure learners' capacity to engage in future life activities using Chinese. In this concluding section, I will discuss why the proposal laid out above is possible and pressing in the 21\(^{st}\) century, what kind of resources and further research is required to carry out such a proposal and how it can be realized through a collaborative participation of various organizations.

\(^{46}\) The US-China Business Council (USCBC) State Export Report 2015
https://www.uschina.org/reports/us-exports/national
5.6.1 The Pressing Need

Discussions presented in the previous chapters have addressed from different perspectives the pressing need to create a new assessment procedure and instrument that can more efficiently evaluate Americans with Chinese language and culture expertise. The critical review of the concepts of proficiency and competence given in Chapter 1 revealed that the proficiency guidelines encompass the observed outcome mainly in learning of commonly taught European languages. This indicates that the proficiency test is less effective for measuring culturally appropriate performances in languages that significantly differ from American and European cultures such as Chinese and Japanese. As a result, unfortunately the interactions a student is faced with in proficiency tests frequently do not resemble the reality that a learner is likely to encounter in the target society, making the test results potentially poor indicators of a learner’s capability to engage in purposeful conversations with members of the target culture. The substantive discrepancies found in the evaluations of various demographic groups not only point out that the non-academic community is unfamiliar with the academic rating categories adopted in the current assessment, they also indicate that the qualities that the non-academic group perceives as beneficial and necessary are not reflected in the current academic procedures and standards.

The increase of enrollment in Chinese language courses at U.S. universities will result in a greater number of students achieving an advanced level of Chinese than ever before. According to a 2013 Modern Language Association (MLA) report47, these enrolments totaled 61,055. Particularly, the graduate language course enrollments

increased 175.9% between 2009 and 2013. These changes in numbers together with a high expectation for foreign experts, as observed from the cases in this study, once again indicate that the time when foreigners with insufficient language skills could succeed in business has past. Many American students of Chinese report the intention to pursue Chinese-language-related careers that require in-depth interactions with members of Chinese culture. Moreover, these careers will require a global competence and knowledge about their Chinese counterparts to gain the reputation of a foreign expert. Such a major shift in expectations for Chinese speaking foreigners in professional contexts makes it even more pressing to establish new assessment procedures and devices to reflect all the changes previously discussed in this chapter.

5.6.2 The Potentiality of A Change

If this call for a change is valid, it is time to evaluate the feasibility in carrying out a change based on what we have known so far about the new “expertise” paradigm of assessment.

First, the two studies conducted in Chapter 3 and four provide two models, respectively, describing the processes involved in identifying and evaluating expertise. Although it is extremely difficult to recruit qualified subjects for a study requiring ethnographic observation due to the restrictions imposed by various workplaces, there is a perceived interest\textsuperscript{48} among advanced Chinese language learners in knowing how their language skills are perceived by counterparts from the target culture. This may correspond to one of the features underlying the expertise as identified in chapter 3--

\textsuperscript{48} Both subjects in this study requested the researcher to share with them the ratings and evaluative comments from both people working with them and people rating their performances.
experts always assume the responsibility of their performances beyond the confines of a language program. Such curiosity and motivation in soliciting evaluative opinions from target culture natives to sustain a growing capacity to succeed in professional communications in Chinese should be fully addressed so as to recruit more experts as subjects for future studies. Even though these experts have already been exposed to various kinds of evaluations mostly involving daily life situations, they may still become potential users of a portfolio assessment, due to their motivation in establishing a longitudinal record of their progress. Moreover, Chinese raters participating in the current study show that native Chinese outside the academic community are comfortable with the idea of evaluating Chinese-speaking foreigners’ performances by using video technologies. Although the researcher identified some purely subjective comments mostly from the non-academic groups, they are part of the workaday life that both students and educators will need to deal with. After all, as long as our learners are expected to interact with people other than their Chinese teachers, these subjective evaluations, and even the biases, will become assets when taken into account.

Second, as reviewed at the beginning of Chapter 4, the wide application of 360-degree feedback in HR assessment and an emerging interest in alternative assessments in education have laid a foundation for both employers and educators to recognize the value of providing more qualitative assessments of their employees’ and students’ performances from multiple raters. Especially, the way in which the 360-degree feedback conveys and reinforces organizational values to its employees through the process of giving and receiving feedback sets a practical example for foreign language educators in terms of how to reveal the behavioral expectations of Chinese counterparts in the target-
culture society. The discrepancy found between the domain-insider and Chinese-instructor evaluations indicate that Chinese teachers are used to having a baseline against which to compare their students to each other (a representative comment of this type can be “My students at level X won’t be able to use that expression that subject used.”) without considering the interests of those students’ potential employers. However, as the number of advanced foreign Chinese speakers has risen in recent years, there is an accompanying and growing awareness among foreign language educators that Chinese learners need to be prepared for working in careers that involved sustained contact with Chinese counterparts. All these changes in awareness create the climate for implementing a performance-based portfolio assessment in foreign language education.

Third, thanks to technological advances, assessment can be implemented from any location. People of certain needs can identify qualified personnel or matched organizations through accessing stored performances and their multi-faceted ratings through a well-maintained longitudinal database. As outlined in 5.4, a stable video website with easy access, a widely available video-conferencing tool (such as Skype), a stable digital media storage website with large storage size plus the portable low-cost devices for videotaping that are becoming more obtainable operationalize performance-based assessment in an effective and efficient way. With the establishment of performance-based portfolios, everyone with a stake in foreign language learning at the advanced levels can have direct access to authentic and meaningful foreign language assessment.
5.6.3 Areas for Future Research

The two studies conducted for this dissertation have laid out models to be further built upon. At the same time, they have also identified a few areas that deserve more resources to carry out future studies and to refine the design of the studies.

First, the limited number of subjects and performance samples in this study point out the necessity for keeping track of our students after they leave our programs at advanced level. Considering how many resources are spent training these young people to be global professionals, it is reasonable to invest a more effort in maintaining a post-program network that provides contact with these valuable resources. Ultimately, surveys can be conducted across institutions to investigate what kinds of jobs and organizations attract individuals with expertise in Chinese as a foreign language. The result of such a survey may not only help us identify qualified research subjects, but more importantly, will be able to inform us about the current state of employing multilingual experts in China.

Second, as the majority of learner-users of such performance-based portfolio assessment will be graduates-to-be, it will be extremely useful to research the validity of using digitally recorded simulated performances to replace the spontaneous performances collected from actual working settings. The ethnographic data collected for this study certainly highlight the value of collecting improvised performances for portfolio assessment. However, in reality, it is almost impossible for the professionals to record their own performances naturally taking place at work without the involvement of a third party, let alone those who have not yet started a job. If simulated performances are found to be equally effective in predicting a learner’s capability to engage the domain activities
of the workplace, it will largely increase the efficiency of creating a performance portfolio.

Third, the high cost, including the investment of both time and human resources in the process of evaluation and pre-evaluation rater training, encourage us to try all means to increase the efficiency of instituting such an assessment tool. Although, as stated above, some commonly used and widely available Internet technology will suffice to realize the current assessment model, it will be more efficient if we can establish an independent online platform to connect raters, store performances and reflect evaluations. Multi-media coaching materials for first-time raters can be developed to walk them through the process.

Last but not the least, it will be useful to launch research projects to continue the investigation of what native speakers of Chinese in the workplace see as beneficial in the communication efforts by foreign professionals. The more data we are able to collect, the more detailed and transparent evaluation categories and criteria we can propose for the assessment device. More importantly, as foreign language assessment tools seem to eventually also frame the content and methodology of instruction, such data will help foreign language educators initiate a virtuous cycle. In another word, if performance-based portfolio assessment is used to measure students’ foreign language skills to communicate with their foreign counterparts, those types of performances will be emphasized in the instruction by teachers who are inclined to teach their students to conform to the nature of the accepted assessments. If the instruction at advanced levels is closely tied to those communicative activities that learners are going to encounter in professional settings in the target culture, their performances that end up in the portfolios
should be improved accordingly. This is a naturally occurring cycle that makes such a huge undertaking as instituting a new assessment paradigm well worth the costs in treasure and effort.
Appendix A Recruitment Letter for American Subject

Subject: A request for your help on a research  
Dear [potential American subject],

This is Zhini Zeng (曾稚妮), a PhD student at the Ohio State University, studying Chinese Pedagogy. XX recommended you to me as someone who is currently working in a Chinese environment, using advanced Chinese for work on a daily basis. Moreover, you are qualified potential subjects because you were tested at OPI Superior at college. Therefore, I would like to ask for your participation in a study I am doing for my doctoral dissertation, in which I explore how non-native speakers of Chinese demonstrate expertise at the workplace and how native Chinese from different domains assess the expertise differently.

This study has been approved by the Institutional Review Board (IRB). Details of the study can be found in the attached consent form. But I would like to briefly summarize what I am going to do with my subjects as below:

1) Shadow each subject at work for 5-7 days.
2) Videotape all the naturally occurring work-related interactions in Chinese (with a permission from the subject first)
3) Interview the subject for an hour about his learning background, motivation, current working status, and etc.
4) Stimulated recall tasks: Engage the subject in an introspection procedure in which videotaped interactions are replayed to the subject to stimulate recall of their concurrent thinking (about 30 minutes everyday, after work)
5) Interview 5 native Chinese who work closely with you at work. These interviewees will be identified by you.

If you are willing to participate in my research, I would like to schedule a time for a brief screening chat on Skype. If we can continue the study together, please review the attached consent form, sign it and email it back to me. Following receipt of your completed consent form, I will work out the dates with you for my visit. Please let me know if you have any questions or concerns about this study. I look forward to hearing from you.

Thank you very much,
Zhini Zeng  
PhD Candidate, Chinese Language Pedagogy  
The Ohio State University
Appendix B Subject Recruitment Posting for Expatriate BBS

Hello!

I am a PhD candidate at the Ohio State University, writing my dissertation on how non-native speakers of Chinese demonstrate expertise at the workplace and how native Chinese from different domains assess the expertise differently.

I am now looking for foreigners in Beijing, Shanghai or Shenzhen to participate in my research, which involves shadowing the subjects at work for one week and creating performance portfolios to be assessed by native Chinese. As a payoff, you can also make use of these portfolios to demonstrate your expertise to your current and future employer.

If you are a foreigner who use advanced Chinese ability for work and could participate in my research (or you know someone who meet these criteria), I would like to get in touch with you to discuss in detail. My email is zeng.51@osu.edu.

Thank you!

Zhini Zeng
PhD Candidate, Chinese Language Pedagogy
The Ohio State University

Here you can find my bio:  http://deall.osu.edu/people/zeng.51
Appendix C Recruitment Email for Chinese Interviewees (Identified by the American Subjects)

Original Email in Chinese:

主题：采访邀请

(potential Chinese interviewee)您好！

我是曾稚妮，美国俄亥俄州立大学、汉语教学法方向的博士生。您可能已经从 XX (the foreign subject) 那里听说了，我最近一直在追踪调查他在工作／学术场合的汉语使用情况，为我的博士论文收集数据（我在博士论文里探讨高级水平的汉语学习者如何在自己的专业领域中用汉语展现专业性。）

在我的研究中，我需要采访在工作／学习中与 XX 接触较多，对他比较熟悉的中国人，XX 向我推荐了您，认为您是我想要找的人。不知您是否愿意在下班后或者周末，在对您方便的地方，接受我 30 分钟左右的采访。如前所说，采访内容主要关于 XX 在中文工作环境中的表现。采访用中文进行，过程在得到您的允许后将会被录音。

关于参加研究的同意书，我已附在邮件中。请您仔细浏览同意书中的细节。若您能参加我的研究，我会在见面时给您打印版本的同意书，让您签字。

谢谢您，希望您能考虑我的请求！

曾稚妮
俄亥俄州立大学、汉语教学法博士生
联系电话：13916409716

Translation in English:

Subject: Request for your participation in an interview

Dear (potential Chinese interviewee),
This is Zhini Zeng, a PhD student from the Ohio State University, studying Chinese language pedagogy. I am currently working on my doctoral dissertation project, which explores how non-native speakers of Chinese demonstrate expertise at the workplace. As you may have already heard from XX (the foreign subject), I have been shadowing him/her for one week to record his work-related performances. I am now looking for native Chinese who have worked closely with XX, to take a 30-minute interview with me. In the interview, I will ask you several questions, mainly focusing on your comments on XX’s performance at work. The interview may be audio recorded with your permission.

I’ve attached the consent form to this email, which contains all the details about this study. If you think could help, please go through it carefully. If you agree to participate in my study, I will have you sign the form when I meet you in person.

Thank you so much for considering my request.

Sincerely,

Zhini Zeng
PhD Candidate, Chinese Language Pedagogy
The Ohio State University
Cell: 13916409716
Appendix D Recruitment Email for Chinese Raters

Original Email in Chinese:

主题：邀请您参与我的研究

(Potential Chinese rater)您好！

我是曾稚妮，美国俄亥俄州立大学、汉语教学法方向的博士生。您可能已经从 XX (a friend who recommended the person to me) 那里听说了，我在寻找一些汉语为母语的人，参与我的博士论文研究（在论文里我探讨的是中国人如何评估高级水平的汉语学习者在专业领域中用汉语交流的表现），帮助我评估一些外国人在工作领域用汉语交流的表现。简言之，我会邀请您上网，对一个视频网站上的数个视频中外国人的表现进行逐个评估。评估结束之后，我会就您提交的评估结果，与您进行一次采访，大约两个小时。

随信附上的是参与研究的同意书，里面详述了需要您参与的部分以及参与的流程。如果您愿意参与研究，请与我联系。我会跟您约时间见面（或在条件不允许的情况下通过网络视频或电话），将此份同意书交予您签字，并且再次向您当面解释、演示评估的操作过程。

谢谢您，希望您能考虑我的请求！

曾稚妮
俄亥俄州立大学、汉语教学法博士生
联系电话：13916409716

Translation in English:

Subject: Request for your participation in a research

Dear (potential Chinese rater),
This is Zhini Zeng, a PhD student from the Ohio State University, studying Chinese language pedagogy. As you may have already heard from XX (a friend who recommended the person to me), I am looking for some native Chinese to participate in my doctoral dissertation project, to evaluate some foreigners’ performances in Chinese at work. In brief, I will invite you to watch some videos online and to evaluate the foreigner’s performance shown in each video with the evaluation sheets I have prepared for you. After you finished the evaluation, I would like to schedule a two-hour interview with you to discuss the evaluation results you submitted.

I’ve attached the consent form to this email, which contains further details of this study and your potential participation in it. If you agree to participate in my study, please let me know. We will then set up a time to meet at a convenient location for you, when I will ask you to sign two copies of the attached consent form, one for you and one for me. I can also answer any further questions if you have.

Thank you so much for considering my request.

Sincerely,

Zhini Zeng
PhD Candidate, Chinese Language Pedagogy
The Ohio State University
Cell: 13916409716
Appendix E Interview Questions to the American Subjects

1. Please talk about your Chinese learning experiences, including language program(s) you attended, classes you took, study abroad experience, and your experiences in taking OPI test experiences and scores.
2. Why did you choose to learn Chinese at the beginning?
3. Do you consider yourself still studying Chinese now? If yes, what is the frequency? How much time do you spend? Can you identify all kinds of practices you usually engage to improve your Chinese? Is there any resources you used?
4. Please describe your Chinese-related working experience.
5. How did you make the choice to work in China?
6. Do you think you have close Chinese friends? Can you mention three people and comment your friendship with them?
7. How do Chinese people around you (your teachers, classmates, friends, colleagues, employers and business partners) evaluate your Chinese?
8. Recall a few situations described as follow, if they once happened in your life:
   a. Your misuse of Chinese caused misunderstanding or embarrassment that harmed your business.
   b. Your misuse of Chinese caused misunderstanding or embarrassment that harmed your relationship with Chinese colleagues or friends.
   c. Your successful use of Chinese that benefited your business.
   d. Your successful use of Chinese that enhanced your interpersonal relationship with Chinese people.
   e. Your demonstration of your identity as a foreigner or any foreignism in communication with Chinese people brought advantages to you.
9. In what aspect of your Chinese you think you need to improve most?
10. There is one observation of advanced-level foreign language learners — it seems that the higher proficiency level in the target language they reach and the more in-target culture experience they have, the less likely they are willing to perform to meet the prescribed “native expectations” and the more strongly they desire to assert their “self,” in communicating in the target language. Does this observation apply to you? Does this observation apply to any foreigners in China you know? What is your comment on this phenomenon?
Appendix F Stimulated Recall Interview Questions to the American Subjects

1. Please comment on the frequency of the recorded event – how often does it happen in your work?
2. (By pausing the video) What did you say here?
3. Why did you say that?
4. What were you thinking when you were doing that?
5. What do you think of that person’s response?
6. How will you evaluate your performance in this video? Was your communication effective? Why do you think so?
7. If there’s a chance for you to redo it, will you do it differently? How? Why?
8. Do you think the recorded event reflect your average level of performance?
9. In addition to the recorded events, is there any other work-related activity you used to do? What’s the frequency?
Appendix G Interview Questions to the Chinese Counterparts

1. Please tell me briefly about how did you come to know XX (name of the subject)?
2. How would you comment on your relationship with XX? (Colleagues? Friends? Very close friends?)
3. How would you evaluate XX’s Chinese level? Please give some concrete examples.
4. How would you evaluate XX’s performance at work? Please give some concrete examples.
5. Do you know any other foreigners who speak Chinese? Please mention a few and compare their communications in Chinese to XX’s.
6. Has XX ever consulted you any Chinese learning-related issues? Have you ever observed XX engaging any practices to improve his Chinese?
7. Please recall a few situations described as follow, if they apply to XX
a. His misuse of Chinese caused misunderstanding or embarrassment that harmed your business.
b. His misuse of Chinese caused misunderstanding or embarrassment that harmed your relationship with Chinese colleagues or friends.
c. His successful use of Chinese that benefited your business.
d. His successful use of Chinese that enhanced your interpersonal relationship with Chinese people.
e. His demonstration of his identity as a foreigner or any foreignism in his communication with Chinese people brought advantages to him
Appendix H Interview Questions to the Chinese Raters

1. 请简单地介绍您自己（年龄 / 职业 / 教育背景 / 工作经历 / 是否会方言？是否会外语？）
2. 在对这个视频的评价中，您为什么说不愿意和他交流？
3. 在对视频 X 的 (name of the criterion) 评分中，您给了 X 分，请问您的理由是？
4. 看了这个人的一些表现后，你对他有什么印象？
5. 您的生活、工作中，还接触过这样会说中文的外国人吗？他们的表现和视频里的外国人有何不同？

Translation in English:

1. Please briefly introduce yourself, including age, occupation, educational background, any foreign language or dialect you speak.
2. In your evaluation of this performance sample, why did you say that you would not be willing to communicate with this foreigner?
3. In your evaluation of the criterion … in video X, you gave a score of x points. Can you explain your reasons?
4. After watching a few videos of him, what is your general impression of him?
5. Have you ever communicated with foreigners speaking Chinese in your life? Can you compare the subject in the video with other Chinese-speaking foreigners you know?
Appendix I Evaluation Form Sample

视频 07 (Link to the online video was inserted in the real evaluation sheet I used)
视频背景描述：贾先生是一家游戏公司的 CEO。在每周例行的高管会议上，商务部的总监向贾总汇报了财务部员工的一些抱怨。贾总听后向高管们强调“财务部”和“商务部”之间要加强沟通。

| 5-Superior Performance 表现非常好; 4- Strong Performance 表现挺好; 3-Competence/Passable Performance 表现还不错（合格了）; 2-Some Performance 表现一般; 1-Minimum Performance 表现不太好; 0-No Performance Observed 没有表现 |
|---|---|---|---|---|---|---|
| 表达能力 Delivery | 打分 Scores | 评语 / 笔记 Comments/Notes |
| 语言组织能力 Organization of idea |
| 文化得体度（言谈举止） Cultural Appropriateness |
| 听众 / 对象适应调整 Adjustment to audience |
| 修辞效果 Rhetorical effectiveness |
| 文化常识相关度 Cultural reference |
| 专业知识相关度 Domain reference |
| 回答问题圆满度 Question and answer |

1. 这个人给你什么样的印象？ What were your impressions of this person?

2. 根据这个人的这段表现，你愿意与他继续交流吗？为什么？ Based on this person’s performance, would you like to further communicate with him or her? Why/Why not?

3. 除了以上八个项目之外，您对这段表现还有何评价？ In addition to your evaluations of the eight aspects listed above, do you have any other comments on this performance?
## Appendix J Alan Bahar’s Portfolio Evaluation Data

### Performance Sample 1 (Text)

**Description:** This is an art exhibition review that Alan was invited to write for the exhibition catalog.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters</td>
<td>Yang</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>5</td>
<td>5</td>
<td>3.83</td>
</tr>
</tbody>
</table>

DI Average: 4.61  DO Average: 5  CI Average: 5

| Q3 | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
Performance Sample 2 (Video, 8 minutes 32 seconds)

Description: Alan was invited to participate in a CAFA graduate-student art exhibition at “Today’s Museum” (今日美术馆) in Beijing. In the video, during the pre-exhibition setup, Alan was trying to persuade the museum staff to adopt his suggestion by removing the artist’s photo in the exhibition tag and revising the its size.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery</td>
<td>Yang 5</td>
<td>Yu 4</td>
<td>Yi 5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>4.5</td>
<td>3.25</td>
<td>3.75</td>
</tr>
</tbody>
</table>

DI Average: 3.83  DO Average: 4.30  CI Average: 4.71

Q3 | Y | N | N | Y | N | N | N | Y | N

227
**Performance Sample 3 (Video, 2 minutes 11 seconds)**

**Description:** Alan was invited to participate in a CAFA graduate-student art exhibition at “Today’s Museum” (今日美术馆) in Beijing. In this video, Alan offered his suggestion to another artist on how to arrange several sculpture exhibits of his. He joined the discussion with both that artist and a staff working in the museum, trying to figure out the best way to present that artist’s work.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yang, Yu, Yi</td>
<td>Hu, Huang, Hong</td>
<td>Xu, Xiao, Xin</td>
</tr>
<tr>
<td>Delivery</td>
<td>5       4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5       4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5       4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5       3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5       3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>0       4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5       3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5       4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>4.38    3.63</td>
<td>4.75</td>
<td>4.13</td>
</tr>
</tbody>
</table>

DI Average: 4.25  DO Average: 4.54  CI Average: 4.63

Q3: Y  N  Y  Y  Y  N  Y  Y
**Performance Sample 4 (Video, 4 minutes 2 seconds)**

**Description:** In this video, Alan was doing some classroom housekeeping at the beginning of his class “Classical Chinese Culture”, a course he has been teaching for one semester in a local art high school as a part-time art teacher. In this particular class, he was assigning tasks (to imitate the masterpieces in the *Manual of the Mustard Seed Garden*) to his students, reminding them about the upcoming end-of-semester evaluation.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yin</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>3.75</td>
<td>3.38</td>
<td>3.75</td>
</tr>
</tbody>
</table>

DI Average: 3.63  
DO Average: 4.63  
CI Average: 4.17

| Q3 | N   | N   | N   | N   | Y   | Y   | N   | Y   | N   |
### Performance Sample 5 (Video, 5 minutes 54 seconds)

**Description:** In this video, Alan was giving feedback to three different students on their respective drawing homework.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters</td>
<td>Yin</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>4.63</td>
<td>3.75</td>
<td>3.63</td>
</tr>
</tbody>
</table>

**DI Average:** 4  
**DO Average:** 4.84  
**CI Average:** 4.96

| Q3 | Y | N | Y | Y | Y | NA | Y | Y | Y |

230
**Performance Sample 6 (Video, 1 minute 33 seconds)**

**Description:** In this video, Alan was talking to a student who refused to participate in the class activities.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yin</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>4.29</td>
<td>4.57</td>
<td>4.43</td>
</tr>
</tbody>
</table>

**Q3**

| Q3 | Y | N | N | Y | N | Y | NA | NA | Y |

DI Average: 4.43
DO Average: 4.67
CI Average: 4.48
**Performance Sample 7 (Video, 4 minutes 18 seconds)**

**Description:** In this video, Alan was giving instruction to students on how to write a title for their respective artwork in *Li* calligraphy (隶书).

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yin</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>3.75</td>
<td>3.88</td>
<td>4</td>
</tr>
</tbody>
</table>

DI Average: 3.88  DO Average: 4.67  CI Average: 4.46

Q3:  N  N  Y  Y  Y  NA  NA  Y
Performance Sample 8 (Audio, 3 minutes 53 seconds)

**Description:** This near four-minute audio consists of four WeChat audio messages between Alan and his fellow graduate student. During the chat, Alan was sharing his understanding of the difference between Chinese painting and Western painting and why Chinese people like using idioms.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yang</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>4.14</td>
<td>4.43</td>
<td>3.86</td>
</tr>
</tbody>
</table>

DI Average: 4.14  DO Average: 4.91  CI Average: 4.62

Q3: Y Y Y Y Y Y Y Y Y
**Performance Sample 9 (Video, 8 minutes 28 seconds)**

**Description:** In this video, Alan was explaining to his friend about his MA graduation project he is now working on. He first introduced some general idea underlying his design and then pointed out several painting drafts hanging on the wall to talk about the specifics.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Raters</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yang</td>
<td>Yu</td>
<td>Yi</td>
<td>Hu</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>4.38</td>
<td>4</td>
<td>3.5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 3.96  DO Average: 4.75  CI Average: 4.84

Q3  Y  Y  Y  Y  Y  Y  Y  Y  Y
**Performance Sample 10 (Video, 4 minutes 16 seconds)**

**Description:** In this video, a graduate student committee leader in Alan’s school asked him to make a recommendation about potential exhibits for an upcoming graduate student artwork exhibition. Alan therefore recommended his friend’s work and explained to that student committee leader about why that work would worth such a recommendation.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Raters</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Yang</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td></td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td></td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td>4.75</td>
<td>4.25</td>
<td>4.13</td>
</tr>
<tr>
<td><strong>DI Average:</strong></td>
<td></td>
<td>4.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DO Average:</strong></td>
<td></td>
<td>4.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CI Average:</strong></td>
<td></td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q3

|        | Y | Y | N | NA | Y | Y | NA | Y | NA |
Performance Sample 11 (Video, 10 minutes 54 seconds)

**Description:** In this video, Alan was attending a reading club monthly meeting, during which he shared his experience reading a Chinese history book written by an American sinology.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yang</td>
<td>Yu</td>
<td>Yi</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>4</td>
<td>4.86</td>
<td>4</td>
</tr>
</tbody>
</table>

DI Average: 4.29  DO Average: 4.62  CI Average: 4.91

Q3: Y Y Y Y Y Y Y Y Y
## Appendix K Jack Mayer’s Portfolio Evaluation Data

**Performance Sample 1 (Video, 1 minutes 15 seconds)**

**Description:** In this video, Jack was addressing some issues such as the cleanliness of the office space and prohibition of smoking in company at a regular upper management meeting.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Raters</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
<td>Hu</td>
</tr>
<tr>
<td>Delivery</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>3.13</td>
<td>3.38</td>
<td>3.5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 3.34  
DO Average: 4.30  
CI Average: 4.88  

<table>
<thead>
<tr>
<th>Q3</th>
<th>N</th>
<th>N</th>
<th>NA</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
</tr>
</thead>
</table>
**Performance Sample 2 (Video, 3 minutes 59 seconds)**

**Description:** In this video, hearing some complaints from the director of the financial department, Jack was emphasizing the importance of communication among different departments at an upper management staff meeting.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Average</td>
<td>2.75</td>
<td>2.63</td>
<td>3.5</td>
</tr>
</tbody>
</table>

DI Average: 2.96
DO Average: 3.71
CI Average: 4.88

Q3   | N  | NA | N  | NA | Y  | Y  | Y  | Y  | Y  |

238
**Performance Sample 3 (Video, 53 seconds)**

**Description:** In this video, Jack was addressing the importance of supporting internal audit to some senior staff in his company.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong     Zhu      Zhang</td>
<td>Hu     Huang      Hong</td>
<td>Xu       Xiao     Xin</td>
</tr>
<tr>
<td>Delivery</td>
<td>3         3         3</td>
<td>5      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>3         3         4</td>
<td>5      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>3         3         4</td>
<td>5      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>3         3         3</td>
<td>4      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>2         2         3</td>
<td>5      5          3</td>
<td>5       5         4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>3         3         4</td>
<td>4      5          4</td>
<td>5       5         0</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3         3         3</td>
<td>5      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>3         4         4</td>
<td>5      5          4</td>
<td>5       5         5</td>
</tr>
<tr>
<td>Average</td>
<td>2.88      3         3.63</td>
<td>4.75   5          3.88</td>
<td>5       5         5</td>
</tr>
</tbody>
</table>

DI Average: 3.13  DO Average: 4.54  CI Average: 5

Q3: NA   NA   NA   Y   Y   N   NA   Y   Y
### Performance Sample 4 (Video, 1 minute 52 seconds)

**Description:** In this video, Jack was make a short speech to all the staff in his company, as a part of the video to be presented at the company’s 2015 Chinese New Year’s party.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Domain reference</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Average</td>
<td>3.29</td>
<td>3.14</td>
<td>3</td>
</tr>
<tr>
<td>DI Average: 3.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DO Average: 4.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI Average: 4.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3</td>
<td>N</td>
<td>N</td>
<td>NA</td>
</tr>
</tbody>
</table>
**Performance Sample 5 (Audio, 7 minute 24 seconds)**

**Description:** In this audio, Jack was criticizing a project team that made some severe mistakes in the maintenance of a newly published game. At this project meeting, Jack first listened to the report from the project coordinator and then addressed several issues related to their miss.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**Q3**

|       | Y | Y | Y | N | NA | N | N | Y | Y |

**DI Average:** 5  
**DO Average:** 3.66  
**CI Average:** 4.63
Performance Sample 6 (Video, 4 minute 57 seconds)
*Description:* In this video, Jack was introducing two games the company recently purchased to some senior staff.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>4.63</td>
<td>4.5</td>
<td>4.25</td>
</tr>
</tbody>
</table>

DI Average: 4.46
DO Average: 4.25
CI Average: 4.42

Q3 | Y | Y | Y | Y | N | Y | NA | Y | NA |
Performance Sample 7 (Video, 4 minute 28 seconds)

Description: In this video, Jack attended a TV show “Have lunch with whom? (和谁一起午餐)” engaging in the Q&A session in which he responded to many questions raised by some experts in the field of business and investment, concerning his business concept.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 5  DO Average: 4.96  CI Average: 4.96

Q3  Y  Y  Y  Y  Y  Y  Y  Y  Y
**Performance Sample 8 (Video, 1 minute)**

**Description:** This is one-minute-limit talk that Jack gave at a TV show “Have lunch with whom? (和谁一起午餐)”. In this talk, the introduced his experience in doing business in the field of game industry and the business mode of his company.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 5  DO Average: 4.96  CI Average: 4.96

Q3  Y  Y  Y  Y  Y  Y  Y  Y  Y
**Performance Sample 9 (Video, 24 minute 10 seconds)**

**Description:** This is a TED-talk like speech that Jack gave at the China Forbes Forum, in which he talked about his experience of coming to China and doing business in the field of game industry. He also shared his own opinion on the development of game industry in China.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 5  
DO Average: 4.84  
CI Average: 5

Q3 | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y
**Performance Sample 10 (Video, 3 minute 45 seconds)**

**Description:** In this video, Jack was reporting back from a business trip to some senior staff in his company. He talked about some outside vendors he visited during the trip and what potential business his company can develop in future.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters</td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Average</td>
<td>4.86</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

**DI Average:** 4.62  **DO Average:** 3.48  **CI Average:** 3.86

<table>
<thead>
<tr>
<th>Q3</th>
<th>Y</th>
<th>Y</th>
<th>Y</th>
<th>NA</th>
<th>N</th>
<th>NA</th>
<th>NA</th>
<th>N</th>
<th>NA</th>
</tr>
</thead>
</table>
**Performance Sample 11 (Video, 4 minute 7 seconds)**

**Description:** In this video, Jack was introducing a new game his company just purchased to his friend who had little experience playing games. Jack tried to persuade his friend to download the game on her cell phone and gave it a try.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Domain Insider (DI)</th>
<th>Domain Outsider Non-teacher (DO)</th>
<th>Chinese Language Instructor (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters</td>
<td>Zhong</td>
<td>Zhu</td>
<td>Zhang</td>
</tr>
<tr>
<td>Delivery</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Organization of idea</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural Appropriateness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Adjustment to audience</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Rhetorical effectiveness</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Cultural reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Domain reference</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Question &amp; Answer</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>4.88</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

DI Average: 4.96  DO Average: 4.30  CI Average: 4.71

Q3 | Y | Y | Y | Y | Y | Y | NA | Y | NA
Bibliography


Hadden, Betsy L. 1983. “A comparison of teacher and non-teacher perceptions of the second language communication of advanced ESL students who are native speakers of Chinese.” PhD diss., The Ohio State University.


Miyake, Naomi and Donald A. Norman. 1979. "To ask a question, one must know enough to know what is not known." *Journal of Verbal Learning and Verbal Behavior* 18:357-364.


Olive, Thierry, Ronald T. Kellogg, and Annie Piolat. 2002. “Studying text production with the triple task technique: Why and How?” In *Contemporary tools and


Paikeday, Thomas M. 1985. The Native Speaker is Dead. Toronto: Paikeday Publishing Inc.


http://www.govtilr.org/Publications/ACTFL%20Guidelines%202012%20ILR%20Presentation%201.6.pdf


http://www.languagetesting.com/understanding-proficiency


255