

**A two-phase mixed methods project on gender stereotypes  
targeting English learners in Chinese senior high schools**



**Jing LI**

Downing College

Faculty of Education

University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*

March 2020

## **Declaration**

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit (up to 80,000 words, excluding appendices, footnotes, and reference list) for the Degree Committee for the Faculty of Education.

Signature

## **Abstract**

Thesis title: A two-phase mixed methods project on gender stereotypes targeting English learners in Chinese senior high schools

Author: Jing LI

Student learning in language-related subjects has become a gendered field, featuring higher participation rate and better performance by girls and women. Among others, the stereotypes depicting females as better language learners are thought to have led to such observed gender differences. Yet, few studies have examined the exact contents and effects of these stereotypes. This two-phase project investigated the issue using mixed-methods designs. In Phase 1, a questionnaire survey (Study 1) and semi-structured interviews (Study 2) were conducted simultaneously. MANOVA and thematic analysis were applied to two data strands, respectively. It was found that students, their guardians, and their teachers of English stereotypically considered females as better language learners in three dimensions: aptitude, achievement and affect. Additionally, a stronger stereotypical additive connection between affect and achievement was discovered, compared to that between aptitude and achievement. Phase 2 employed a sequential design. First, Study 3, a field experiment, uncovered gender-divergent effects of stereotypes on learners: results from a  $2$  (gender)  $\times$   $2$  (stereotype activation) between-groups ANCOVA indicated that male participants suffered from performance decrements in an English test after having been exposed to negative stereotypes. Female participants' performance, instead, was marginally boosted. Subsequently, in Study 4, eight interviews with students in groups of three were carried out to explore students' encounters with gender stereotypes concerning language learning. A range of cognitive, emotive, and behavioural responses were identified, indicating that the female stereotyping of language fields have been exerting predominantly adverse influences on boys and girls alike. This project has furthered scholarly understanding of gender stereotypes and language education. It also promotes a gender-equitable and -inclusive environment for language learners by highlighting the practical need for awareness-raising programmes sensitising students, guardians, and teachers to their gender-stereotypical beliefs and a synergy of intervention strategies targeting at each of the three dimensions of gender-language stereotypes (aptitude, achievement, and affect).

Thesis word count: 79,007 words

## **Acknowledgement**

I would like to express my deepest and most sincere gratitude to my supervisor, Dr. Ros McLellan. Her guidance and support throughout my PhD journey has been essential to me. Her expertise in the field of psychology and gender, her critical and constructive comments on my work, and her encouraging and kind style have enlightened and inspired me, not just as a researcher, but also as a person. Ros also helped me advance my career, as we collaborated on two journal papers during the course of my doctoral study. I feel ever so lucky to have her as my supervisor, and I will always value her as my mentor.

I am also grateful to my advisor, Prof. Anna Vignoles CBE FBA. As an expert in gender and education, she has provided suggestions on improving my research design and data analysis skills. She also gave me insightful advice on how to build my academic career as a researcher. Similarly, Dr. Yongcan Liu and Dr. Karen Forbes from the Second Language Education Group also assisted me in developing my research ideas into feasible, rigorous plans and academic publications. My mentors from Tsinghua University, China, Prof. Xia Wu and Prof. Zhongshe Lv, have also showed substantial interest in my research and cared about my experience in Cambridge, offering me suggestions regarding academic publication, career choices, and personal life. I am truly indebted and thankful to all their help. In addition, I appreciate the opportunity to learn from Ruth Kershner and Dr. Sarah McGeown, who have kindly agreed to be my examiners.

I would also like to acknowledge the financial contribution from the China Scholarship Council. Without their scholarship, I would find it hard to focus on research while shouldering financial burdens.

Finally, I want to say thank you to my family and friends. My parents, Wanli Li and Jinting Zhang, have always backed me wholeheartedly; otherwise, I could not have pursued my academic and professional interests. My PhD colleagues, Danyang, Eman, Junlin, Esther, and Charlotte in particular, offered me valuable feedback concerning my research and career. I also want to say thank you to my dearest friends from Cambridge and the rest of the world, Zeyao, Kunjian, Wanying, Yuxiao, Yanran, Wenxiao, Yumiao, Lu'ou, and Yukun, because it has been your company that makes my life so colourful and fruitful. The past three-and-half years in Cambridge will always been my life's treasure. I will always cherish this experience.



## Table of Contents

List of Tables .....	i
List of Figures .....	iii
List of Abbreviations .....	v
List of Appendices .....	vi
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1 Significance of Current Research .....	3
1.2 Overview of Thesis Structure .....	4
<b>Chapter 2: Context .....</b>	<b>6</b>
2.1 Conceptual Issues on <i>Gender</i> .....	6
2.1.1 Terminological Disputes over <i>Sex</i> and <i>Gender</i> .....	6
2.1.2 Idea of <i>Gender</i> in Chinese Philosophy .....	7
2.2 Research into Psychological Aspects of Gender .....	8
2.3 English Language Education in China .....	10
<b>Chapter 3: Literature Review .....</b>	<b>12</b>
3.1 Gender Stereotypes in General: Processes, Dimensions and Features .....	12
3.1.1 Formation of Gender Stereotypes: Cognitive and Socio-Cultural Processes...	12
3.1.1.1 Cognitive Perspective: Social Categorisation .....	12
3.1.1.2 Socio-Cultural Perspective: Social Role Theory and Gender Essentialism .....	13
3.1.2 Children's Acquisition of Gender Stereotypes: Awareness, Endorsement, and Flexibility .....	15
3.1.3 Dimensions: Agency vs. Communality .....	16
3.1.4 Features of Gender Stereotypes .....	17
3.1.4.1 (In)accurate Gender Stereotypes .....	18
3.1.4.2 Durable Gender Stereotypes .....	19
3.1.4.3 Cross-Cultural Gender Stereotypes .....	20
3.1.4.4 Complementary Gender Stereotypes .....	21
3.2 Gender Stereotypes in Education: Contents and Effects .....	22
3.2.1 Quantitative-Males and Qualitative-Females Stereotypes .....	23
3.2.2 Conceptualising 'Female-Advantage-in-Languages Stereotype' .....	25
3.2.2.1 <i>Aptitude</i> Component .....	25
3.2.2.2 <i>Achievement</i> Component .....	28
3.2.2.3 <i>Affect</i> Component .....	29
3.2.2.4 Review of Methods Used in Literature on Female Stereotyping of Languages .....	30
3.2.2.5 Accuracy of Female-Advantage-in-Languages Stereotype .....	31
3.2.3 Differences in Stereotype Endorsement Due to Demographic Background ...	34
3.2.4 Stereotype Threat Effect on Negatively Stereotyped Learners .....	37
3.2.4.1 Mechanisms of Stereotype Threat Effect .....	38
3.2.4.2 Consequences of Stereotype Threat Effect .....	40
3.2.5 Stereotype Lift and Boost on Non-Negatively Stereotyped Learners .....	40

3.3	Limitations in Previous Work and Proposed Contributions of Current Project ....	41
3.4	Research Questions .....	43
<b>Chapter 4:</b>	<b>Methodology .....</b>	<b>44</b>
4.1	Research Design .....	44
4.1.1	Two-Phase Mixed Methods Research .....	44
4.1.2	Epistemological Underpinnings and Methodological Considerations .....	46
4.2	Phase 1: Capturing Existence of Female-Advantage-in-Languages Stereotype ...	50
4.2.1	Study 1: Questionnaire-Based Survey .....	51
4.2.1.1	Questionnaires as Data Collection Methods .....	51
4.2.1.2	Developing Questionnaires .....	52
4.2.1.3	Pilot Study 1 .....	54
4.2.1.4	Sampling .....	59
4.2.1.5	Distribution and Collection Procedures .....	60
4.2.1.6	Data Screening and Analysis Procedures .....	62
4.2.1.7	Ethical Considerations .....	62
4.2.2	Study 2: Interview-Based Survey .....	64
4.2.2.1	Interviews as Data Collection Methods .....	65
4.2.2.2	Developing Interview Schedule and Pilot Study 2 .....	65
4.2.2.3	Sampling and Interviewing .....	68
4.2.2.4	Data Preparation and Analysis Procedures .....	70
4.2.2.5	Ethical Considerations .....	74
4.2.3	Integrating Data from Studies 1 and 2 .....	74
4.3	Phase 2: Detecting Effects of Female-Advantage-in-Languages Stereotype ...	75
4.3.1	Study 3: Field Experiment .....	76
4.3.1.1	Pilot Study 3 .....	77
4.3.1.2	Sampling .....	79
4.3.1.3	Procedures .....	80
4.3.1.4	Data Preparation and Analysis Procedures .....	82
4.3.1.5	Ethical Considerations .....	83
4.3.2	Study 4: Group Interviews .....	84
4.3.2.1	Developing Interview Schedule .....	85
4.3.2.2	Sampling and Interviewing .....	86
4.3.2.3	Data Preparation and Analysis Procedures .....	87
4.3.2.4	Ethical Considerations .....	88
4.3.3	Integrating Data from Studies 3 and 4 .....	88
<b>Chapter 5:</b>	<b>Findings and Discussions I — Perceiving Girls as Better Learners of English by Guardians and Students .....</b>	<b>90</b>
5.1	Guardians' Perspectives from Study 1 .....	90
5.1.1	Data Screening .....	91
5.1.2	Demographic Information of Guardian Participants .....	91
5.1.3	Findings from Three-Way Mixed MANOVA .....	95

5.1.3.1	Hypothesis 1.1: Guardians Endorsed Female-Advantage-in-Languages Stereotype .....	95
5.1.3.2	Hypothesis 1.2: Guardians' Endorsement Differed Due to Gender .....	97
5.1.3.3	Hypothesis 1.3: Guardians' Endorsement Differed Due to Region .....	97
5.1.3.4	Hypothesis 1.4: Guardians' Endorsement Differed on the Basis of Gender and Region .....	99
5.1.4	Summary of Guardian's Perspectives .....	99
5.2	Students' Perspectives from Study 1 .....	99
5.2.1	Data Screening .....	100
5.2.2	Demographic Information of Student Participants .....	100
5.2.3	Findings from Four-Way Mixed MANOVA on Macro Level .....	101
5.2.3.1	Hypothesis 2.1.1: Students Endorsed Female-Advantage-in-Languages Stereotype on Macro Level .....	102
5.2.3.2	Hypothesis 2.1.2: Students' Endorsement on Macro Level Differed Due to Gender .....	103
5.2.3.3	Hypothesis 2.1.3: Students' Endorsement on Macro Level Differed Due to Region .....	103
5.2.3.4	Hypothesis 2.1.4: Students' Endorsement on Macro Level Differed Due to Grade .....	104
5.2.3.5	Hypothesis 2.1.5: Students' Endorsement on Macro Level Differed on the Basis of Gender and Region .....	106
5.2.3.6	Hypothesis 2.1.6: Students' Endorsement on Macro Level Differed on the Basis of Region and Grade .....	106
5.2.3.7	Hypothesis 2.1.7: Students' Endorsement on Macro Level Differed on the Basis of Gender and Grade .....	107
5.2.3.8	Hypothesis 2.1.8: Students' Endorsement on Macro Level Differed on the Basis of Gender, Region, and Grade .....	107
5.2.4	Findings from Four-Way Mixed MANOVA on Micro Level .....	107
5.2.4.1	Hypothesis 2.2.1: Students Endorsed Female-Advantage-in-Languages Stereotype on Micro Level .....	108
5.2.4.2	Hypothesis 2.2.2: Students' Endorsement on Micro Level Differed Due to Gender .....	109
5.2.4.3	Hypothesis 2.2.3: Students' Endorsement on Micro Level Differed Due to Region .....	111
5.2.4.4	Hypothesis 2.2.4: Students' Endorsement on Micro Level Differed Due to Grade .....	112
5.2.4.5	Hypothesis 2.2.5: Students' Endorsement on Micro Level Differed on the Basis of Gender and Region .....	114
5.2.4.6	Hypothesis 2.2.6: Students' Endorsement on Micro Level Differed on the Basis of Region and Grade .....	115
5.2.4.7	Hypothesis 2.2.7: Students' Endorsement on Micro Level Differed on the Basis of Gender and Grade .....	115

5.2.4.8	Hypothesis 2.2.8: Students' Endorsement on Micro Level Differed on the Basis of Gender, Region, and Grade .....	116
5.2.5	Summary of Students' Perspectives .....	116
5.3	Generation Gap in Endorsement of Female-Advantage-in-Languages Stereotype..	116
5.3.1	Hypothesis 3.1: Generation Gap Existed in Stereotype Endorsement .....	117
5.3.2	Hypothesis 3.2: Width of Generation Gap Differed Due to Gender .....	119
5.3.3	Hypothesis 3.3: Width of Generation Gap Differed Due to Region .....	119
5.3.4	Hypothesis 3.4: Width of Generation Gap Differed on the Basis of Gender and Region .....	119
5.3.5	Summary of Generation Gap in Endorsement .....	119
5.4	Summary of Guardians' and Students' Perspectives .....	120
<b>Chapter 6: Findings and Discussions II — Perceiving Girls as Better Learners of English by Teachers of English .....</b>		<b>121</b>
6.1	Demographic Information of Teacher Participants .....	122
6.2	Integrated Finding 1: Teachers Endorsed Female-Advantage-in-Languages Stereotype .....	126
6.2.1	Evidence from Questionnaire Dataset .....	126
6.2.2	Evidence from Interview Dataset .....	127
6.2.2.1	Teachers' Agreement and Disagreement on <i>Achievement</i> .....	130
6.2.2.2	Teachers' Congruous Understanding of <i>Affect</i> .....	134
6.2.2.3	Teachers' Diversified Interpretations of <i>Aptitude</i> .....	136
6.2.3	Differences in Magnitude and Scope of Three Components .....	138
6.2.4	Teachers' Awareness of Female-Advantage-in-Languages Stereotype .....	139
6.3	Integrated Finding 2: Causal Links among Three Components of Female-Advantage-in-Languages Stereotype as Perceived by Teachers .....	141
6.3.1	Evidence from Interview Dataset .....	141
6.3.2	Evidence from Questionnaire Dataset .....	145
6.4	Integrated Finding 3: Gender Differences in Endorsement of Female-Advantage-in-Languages Stereotype .....	147
6.4.1	Evidence from Questionnaire Dataset .....	147
6.4.2	Evidence from Interview Dataset .....	147
6.5	Integrated Finding 4: Regional Variations in Endorsement of Female-Advantage-in-Languages Stereotype .....	148
6.5.1	Evidence from Questionnaire Dataset .....	148
6.5.2	Evidence from Interview Dataset .....	148
6.6	Summary of Integrated Findings about Teachers' Endorsement .....	150
<b>Chapter 7: Findings and Discussions III — Perceiving Gender as a Differentiating Variable English Classrooms by Teachers of English .....</b>		<b>151</b>
7.1	Salience of Female-Advantage-in-Languages Stereotype among Teachers .....	151
7.2	Function of Female-Advantage-in-Languages Stereotype among Teachers .....	156
7.3	Additional Gender Stereotypes Accompanying Female-Advantage-in Languages Stereotype .....	159

7.3.1	Competent Boys Stereotype .....	160
7.3.1.1	Contents of Competent Boys Stereotype .....	160
7.3.1.2	Relationship between Competent Boys and Female-Advantage-in-Languages Stereotypes .....	163
7.3.2	Quantitative-Orientated Boys and Qualitative-Orientated Girls Stereotypes ..	165
7.3.3	Slack Boys and Diligent Girls Stereotypes .....	169
7.3.4	Career-Driven Males and Home-Bound Females Stereotypes .....	172
7.4	Critical Perspectives about Female-Advantage-in-Languages Stereotype .....	174
7.5	Summary of Teachers' Perspectives in Study 2 .....	175
<b>Chapter 8: Findings and Discussions IV — Effects of Female-Advantage-in-Languages Stereotype on Learners .....</b>		<b>177</b>
8.1	Findings from Study 3 .....	177
8.1.1	Participants and Design .....	177
8.1.2	Results from Two-Way ANCOVA .....	178
8.2	Findings from Study 4 .....	179
8.2.1	Descriptions of Experiences with English Learning .....	181
8.2.2	Students' Endorsement of Female-Advantage-in-Languages Stereotype ...	185
8.2.2.1	Students' Endorsement of Female-Advantage-in-Languages Stereotype: From Conformity to Critical Thinking .....	185
8.2.2.2	Causal Links among Three Components of Female-Advantage-in-Languages Stereotype as Perceived by Students .....	188
8.2.3	Students' Awareness of Female-Advantage-in-Languages Stereotype .....	191
8.2.4	Students' Responses to Female-Advantage-in-Languages Stereotype .....	192
8.2.5	Beyond Female-Advantage-in-Languages Stereotype .....	198
8.2.6	Students' Emergent Attitudes towards <i>Gender</i> : Conformity, Confusion, and Critical Thinking .....	200
8.3	Integrated Findings from Studies 3 and 4 .....	200
<b>Chapter 9: General Discussion .....</b>		<b>203</b>
9.1	Endorsement of Female-Advantage-in-Languages Stereotype .....	203
9.1.1	Contextualising Female-Advantage-in-Languages Stereotype .....	203
9.1.2	Accuracy of Female-Advantage-in-Languages Stereotype .....	207
9.1.2.1	Aptitude Component as Exaggeration .....	208
9.1.2.2	Achievement Component as Self-Fulfilling Prophecy .....	209
9.1.2.3	Affect Component as Over-Simplification .....	211
9.1.2.4	Summary: Female-Advantage-in-Languages Stereotype and Real Gender Differences .....	212
9.1.3	Three Convergences among Guardians, Students, and Teachers .....	212
9.1.3.1	Aptitude as Least Endorsed Component in All Three Groups .....	212
9.1.3.2	Variabilities within Each Group .....	214
9.1.3.3	Stereotypical Causal Connections between Components among Teachers and Students .....	215
9.1.4	Five Divergences among Guardians, Students, and Teachers .....	216

9.1.4.1	Disparities in Magnitudes of Female-Advantage-in-Languages Stereotype across Three Groups .....	216
9.1.4.2	Discrepancies in Gender Attitude between Teachers and Students .....	217
9.1.4.3	Disagreement over Strongest Component across Three Groups .....	217
9.1.4.4	Gender Difference or Similarity in Stereotype Endorsement .....	218
9.1.4.5	Regional Variation or Consistency in Stereotype Endorsement .....	220
9.1.5	Gender Stereotypes Accompanying Female-Advantage-in-Languages Stereotype .....	221
9.2	Effects of Female-Advantage-in-Languages Stereotype .....	222
9.2.1	Gender-Specific Effect on Test Performance .....	222
9.2.2	Effects on Self-Cognition, Emotions and Behaviours Concerning English Learning .....	223
9.3	Educational Implications .....	224
9.4	Limitations .....	225
<b>Chapter 10:</b>	<b>Conclusions .....</b>	<b>228</b>
10.1	Summary of Findings in Phase 1 .....	228
10.2	Summary of Findings in Phase 2 .....	229
10.3	Theoretical and Practical Contributions of Doctoral Project .....	230
10.4	Future Directions .....	231
10.5	Personal Reflections .....	232
<b>References</b>	<b>.....</b>	<b>231</b>

## List of Tables

Table 3.1	Summary of Research on Gender Differences in Verbal Abilities	31
Table 3.2	Summary of Research into Gender Differences in Language Performance	32
Table 3.3	Summary of Studies on Gender and Age Differences in Endorsement of Language-Related Gender Stereotypes	36
Table 3.4	Summary of Stereotype-Activating Cues in Research on Stereotype Threat Effect	39
Table 4.1	Data Screening Process in Pilot Study 1	56
Table 4.2	Internal Consistency of Items in Two Questionnaires in Pilot Study 1	58
Table 4.3	Demographic Information of Participating Schools in Study 1	60
Table 4.4	Transcription Notes in Studies 2 and 4	70
Table 4.5	Coding Methods and Example Codes in Study 2	71
Table 4.6	Results of Manipulation Check in Stages 1 and 2, Pilot Study 3	78
Table 4.7	Results of Manipulation Check in Stage 3, Pilot Study 3	79
Table 5.1	Univariate Main Effect of Learner Gender in Guardian Dataset	96
Table 5.2	Simple Main Effect of Region on Learner Gender in Guardian Dataset	98
Table 5.3	Univariate Main Effect of Learner Gender on Macro Level in Student Dataset	102
Table 5.4	Pairwise Comparison between Grade and Learner Gender on Macro Level in Student Dataset	105
Table 5.5	Univariate Main Effect of Learner Gender on Micro Level in Student Dataset	108
Table 5.6	Univariate Interaction Effect between Student Gender and Learner Gender on Micro Level in Student Dataset	110
Table 5.7	Simple Main Effect of Student Gender on Learner Gender on Micro Level in Student Dataset	110
Table 5.8	Univariate Interaction Effect between Grade and Learner Gender on Micro Level in Student Dataset	112
Table 5.9	Simple Main Effect of Grade on Each Level of Learner Gender on Micro Level in Student Dataset	112
Table 5.10	Pairwise Comparison between Grade and Learner Gender on Micro Level in Student Dataset	113
Table 5.11	Simple Main Effect of Generation on Learner Gender in Guardian and Student Datasets	118
Table 6.1	Professional Background of Interviewees in Study 2	125
Table 6.2	Univariate Main Effect of Learner Gender in Teacher Dataset in Study 1	127

Table 6.3	Coded Segments Describing Female Excellence in English Achievement in Study 2	130
Table 6.4	Teachers' Diversified Interpretations of <i>Aptitude</i> in Study 2	136
Table 6.5	Multiple Regression Analysis of Teacher Dataset in Study 1	146
Table 6.6	Cross Tabulations between Gender and Three Aspects of Female-Advantage-in-Languages Stereotype in Study 2	149
Table 6.7	Cross Tabulations between Region and Three Aspects of Female-Advantage-in-Languages Stereotype in Study 2	149
Table 8.1	Mean Performance Scores by Gender and Condition in Study 3	178
Table 8.2	Demographic Information of Interviewees and Details about Group Interviews in Study 4	180
Table 8.3	Multiple Regression Analysis of Student Dataset in Study 1	190
Table 8.4	Eight Type of Responses to Female-Advantage-in-Languages Stereotype among Students in Study 4	193
Table 9.1	Female-Advantage-in-Languages Stereotypes in Guardian, Student, and Teacher Datasets	204
Table 9.2	Summary of Research on Gender Differences in Verbal Abilities	205
Table 9.3	Practical Interpretations of Aptitude Component Measured in Study 1	208
Table 9.4	Summary of Research on Gender Differences in English in China	209
Table 9.5	Practical Interpretations of Achievement Component Measured in Study 1	210



## List of Figures

Figure 3.1	Three Theoretical Explanations of Gender Stereotypes' Formation	15
Figure 3.2	Conceptual Framework of Tripartite 'Female-Advantage-in-Languages Stereotype'	26
Figure 3.3	Situational Cues, Vigilance Process, and Stereotype Threat	38
Figure 4.1	Overview of Two-Phase Mixed-Methods Research Project	45
Figure 4.2	Analytical Framework of Study 1	63
Figure 4.3	Design of Study 2	66
Figure 4.4	First-Cycle Coding System in Study 2	73
Figure 4.5	Data Integration of Teacher Datasets in Phase 1	75
Figure 4.6	Sample Size Needed in Study 3	80
Figure 4.7	Procedures of Study 3	81
Figure 4.8	First-Cycle Coding System in Study 4	89
Figure 5.1	Age Make-Ups of Guardians from Study 1 and the China Education Panel Survey	92
Figure 5.2	Educational Backgrounds of Guardians from Study 1 and the China Education Panel Survey	93
Figure 5.3	Occupations of Guardian Dataset	94
Figure 5.4	Achievement, Aptitude, and Affect Ratings as a Function of Learner Gender in Guardian Dataset	96
Figure 5.5	Aptitude and Affect Ratings as a Function of Region and Learner Gender in Guardian Dataset	99
Figure 5.6	Achievement, Aptitude, and Affect Ratings as a Function of Learner Gender on Macro Level in Student Dataset	103
Figure 5.7	Achievement and Affect Ratings as a Function of Grade and Learner Gender on Macro Level in Student Dataset	106
Figure 5.8	Achievement, Aptitude, and Affect Ratings as a Function of Learner Gender on Micro Level in Student Dataset	109
Figure 5.9	Ratings as a Function of Student Gender and Learner Gender on Micro Level in Student Dataset	111
Figure 5.10	Vocabulary-Related Ratings as a Function of Grade and Learner Gender on Micro Level in Student Dataset	114
Figure 5.11	Grammar-Related Ratings as a Function of Grade and Learner Gender on Micro Level in Student Dataset	115
Figure 5.12	Aptitude Ratings as a Function of Generation and Learner Gender in Guardian and Student Dataset	118
Figure 6.1	Age Make-up of Teacher Datasets in Studies 1 and 2	123

Figure 6.2	Gender and Experience Make-Up of Teacher Datasets in Studies 1 and 2	123
Figure 6.3	Achievement, Aptitude, and Affect Ratings as a Function of Learner Gender in Teacher Dataset in Study 1	128
Figure 6.4	Venn Diagram of Believers of Female-Advantage-in-Languages Stereotype Components in Study 2	129
Figure 6.5	Perceived Causal Links between Aptitude and Achievement, and Affect and Achievement by Teachers in Study 2	142
Figure 6.6	Predictions of Means of Perceived Gender Difference in Achievement Based on Four Combinations of Perceived Gender Differences in Aptitude and Affect from Teacher Dataset in Study 1	147
Figure 7.1	Salience of Stereotypical Comments Made by Endorsers of Female-Advantage-in-Languages Stereotype in Study 2	152
Figure 7.2	Venn Diagram of Believers of Competent Boys Stereotype Components in Study 2	160
Figure 7.3	Summary of Themes Emerged from Thematic Analysis in Study 2	176
Figure 8.1	Mean Performance Scores as a Function of Gender and Condition in Study 3	179
Figure 8.2	The Venn Diagram of Believers of Female-Advantage-in-Languages Stereotype Components in Study 4	186
Figure 8.3	Predictions of Means of Perceived Gender Difference in Achievement Based on Four Combinations of Perceived Gender Difference in Aptitude and Affect from Student Dataset in Study 1	191
Figure 8.4	Summary of Themes Emerged from Thematic Analysis in Study 2	202
Figure 9.1	Comparison of FALS Magnitudes across Guardians, Students (Macro Level), and Teachers	213
Figure 9.2	Mean Ratings for Male and Female Learners by Guardians, Students, and Teachers	215
Figure 9.3	Influence of Teachers' Gender-Stereotypical Beliefs on Students' Gender Attitudes	219

### **List of Abbreviations**

CBS	Competent boys stereotype
EVM	The expectancy-value model
FALS	The female-advantage-in-languages stereotype
FL	Foreign language
GS	Gender stereotype
HEI	Higher education institution
IAT	The Implicit Association Test
L2	Second language
MM	Mixed-Methods
NCEE	National College Entrance Examination (‘Gaokao’)
OECD	Organization for Economic Cooperation and Development
PISA	The Program for International Student Assessment
RQ	Research question
SCM	Stereotype Content Model
STEM	Science, Technology, Engineering, and Mathematics
ST	Stereotype threat

## **List of Appendices**

Appendix A	Online Questionnaire for Guardians and Teachers in Pilot Study 1	258
Appendix B	Online Questionnaire for Students in Pilot Study 1	260
Appendix C	Paper-and-Pencil Questionnaire for Guardians and Teachers in Study 1	267
Appendix D	Paper-and-Pencil Questionnaire for Students in Study 1	270
Appendix E	Component Matrix of Two-Component Model of Guardian/Teacher Questionnaire in Pilot Study 1	277
Appendix F	Scree Plots of Questionnaires in Pilot Study 1	278
Appendix G	Component Matrix of Two-Component Model of Student Questionnaire in Pilot Study 1	279
Appendix H	Consent Form in Study 1	280
Appendix I	Sequences of Items for Male and Female Learners among Eight Versions of Questionnaire Brochures for Participating Families	282
Appendix J	Interview Schedule in Pilot Study 2	283
Appendix K	Interview Schedule in Study 2	287
Appendix L	Consent Form in Study 2	291
Appendix M	Gender Difference Test Instructions in Stage 1, Pilot Study 3	293
Appendix N	Gender-Fair Test Instructions in Stage 1, Pilot Study 3	294
Appendix O	Instrument for Manipulation Check in Study 3	295
Appendix P	Interview Schedule in Pilot Study 4	296
Appendix Q	Interview Schedule in Study 4	301
Appendix R	Data Screening Process of Guardian Dataset in Study 1	306
Appendix S	Data Screening Process of Student Dataset in Study 1	308
Appendix T	Sub-Codes Categorising Teachers' Endorsement of Female-Advantage-in-Languages Stereotype in Study 2	310

## Chapter 1: Introduction

Despite efforts to promote gender equality and diversity, men and women seem to take distinct paths in educational institutions on a global scale (e.g., Evans, Schweingruber, and Stevenson, 2002; Kollmayer, Schober, and Spiel, 2018; Schoon and Eccles, 2014; Skelton, Francis, and Smulyan, 2006). For decades, women's low participation rate in Science, Technology, Engineering, and Mathematics (STEM) domains has been witnessed across countries and regions, including Asia (e.g. E-Stat, 2012; Jon and Chung, 2013; Liu, 2018; Ma, You, Xiong, Dong, Wang, and Kou, 2016), North America (e.g., National Science Board, 2019; Wall and Statistics Canada, 2019), and Europe (OECD, 2017). Among others, sex differences in cognitive abilities, academic preferences, and learning strategies have been discerned as plausible mechanisms behind the gendered landscape (e.g. Liu and Wilson, 2009; Miller and Halpern, 2014; Priess and Hyde, 2010). Nevertheless, given the gradual closing of the gender performance gap favouring boys (in most of Europe, see Schleicher, 2019; worldwide, see Voyer and Voyer, 2014), researchers and educators are increasingly acknowledging social influences preventing women from entering and persisting in STEM fields, especially cross-cultural gender stereotypes (GSs) associating STEM with men and masculinity (e.g. Dasgupta and Stout, 2014; Herbert and Stipek, 2005; Kessels, Heyder, Latsch, and Hannover, 2014; Smeding, 2012).

In contrast to the considerable attention received by women's underrepresentation in STEM, an equally significant issue, if not more pressing, deserves more intensive and in-depth inquiry than it currently faces: men's underperformance and disengagement in literacy and language. Girls almost always receive higher scores than boys in literacy tasks and language exams in various cultures and across age groups (e.g., China Education Panel Survey, 2013; Goetz, Fenzel, Hall, and Pekrun, 2008; National Assessment of Educational Progress, 2011, 2015, and 2019; Voyer and Voyer, 2014). Moreover, the same female edge is also observed in the learning of second or foreign languages, such as Dutch (e.g., van der Slik, van Hout, and Schepens, 2015) and English (e.g., The International English Language Testing System, 2018; Educational Testing Service, 2017). A further worrying observation is the pronounced disengagement from language-related activities and subjects on the side of boys (e.g. Carr and Pauwels, 2006; Fisher, 2001; Pritchard, 1987; Schleicher, 2019; Taylor and Marsden, 2014; Williams, Burden, and Lanvers, 2002; Yao, 2011). The relationship between boys and languages, in fact, can be more troubling, because men and boys tend to face more psychological barriers deterring them from pursue their interests and potentials

fully and freely in traditionally feminine domains (Croft, Schmader, Block, 2015). Worse still, the negligence of the issue is unfortunate and unfair for boys, considering the personal cognitive and professional benefits of bilingual/ multilingual skills (Cambridge Public Policy SRI, 2015).

Noticing these problems, some educational researchers inspected the possibility that GSs feminising languages might have undermined boys' self-concepts, interest, and attainment in literacy and languages, as those male-STEM stereotypes are doing to girls (e.g., Schmenk, 2004; Tenenbaum, 2009). As generalised images about men and women, as well as boys and girls, GSs in education describe and prescribe STEM and language studies as masculine and feminine, and serve as criteria against which individual boys and girls are compared. But the studies looking into GSs targeting language learners, unlike those analysing the male stereotyping of STEM, are few in quantity and fraught with inconsistencies. An additional gap in scholarly literature is the absence of cross-cultural samples that reach beyond western, industrialised societies. Henceforth, the current two-phase mixed-methods (MM) project was conceptualised and conducted to understand gender stereotyping in English classrooms in senior secondary schools in mainland China from the perspective of social psychology.

My route to this exploration of GSs dates back to early school years. Though an avid reader and a straight-A student, I hardly ever enjoyed my achievement. 'You get good grades. So what? You'll end up marrying some guy and all that matters is *his* success.' 'You won't always stay where you are, you know. Boys will catch up with you eventually, because *they* are the clever sort. They simply haven't really made any effort yet.' The derogatory comments from teachers and relatives haunted me so incessantly that I had been treading as if on thin ice for twelve years throughout elementary and secondary schools.

Entering Tsinghua University in 2010, I was taken over by an unfamiliar sensation of confidence and security: I thought I had proved myself because Tsinghua is one of the top universities in mainland China. Besides, since I had chosen to study English, a female-dominated area, I reckoned that finally, I could have bid farewell to the disdain for my gender. Yet, such self-assurance was soon superseded by overwhelming commiseration for Yuan, Xu, and Tang, the only three boys in our cohort of forty-one freshers. English was not their ideal choice, and with men being a minority in the Department of Foreign Languages, they never felt that sense of belonging I gratefully embraced. All three selected a second major and transferred to the corresponding field eventually: Yuan did a master's degree in economics and is now working for a VC firm, Xu is practising law, and Tang has devoted

himself to journalism. Henceforth, I found myself stepping onto the ‘road to Damascus’ of a gendered nature: I inevitably saw gender everywhere. Are boys and girls, men and women supposed to be different? Why do we think of males and females as contrastive groups? What happens to individuals challenging these social conventions about gender?

Therefore, during undergraduate years, I joined two research projects: one investigating gender differences in classroom interactions, and another examining women’s use of self-deprecating humour. These academic experiences introduced me to approaches assessing gender differences, as well as similarities. They also deepened my understanding of the psychology of gender, including gender identities, gender roles, and gender stereotypes. Gradually, the gender-languages issue captured my attention: are girls better language learners than boys? Do we consider girls to be better? If so, in which dimensions? Is such thinking embedded in our Chinese culture, or is it more prevalent? Are we right in assuming such gender differences? How do these beliefs make individual learners feel? These questions have inspired me to study GSs concerning language learners since the MPhil programme. I aim to understand the nature, structure, and effects of these GSs so that a gender-fair and supportive learning environment can be established for future learners.

In the next section, I will lay out the theoretical and empirical rationales for this doctoral project. The chapter ends with a final note outlining the structure of this thesis.

## **1.1 Significance of Current Research**

Studying GSs targeting language learners in Chinese high schools contributes to both GSs and language education literature. Although psychological research on academic GSs is extensive and with intellectual rigour, it is unbalanced: the male stereotyping of STEM subjects receives much more attention than the female stereotyping of language and literacy domains. With regards to the field of language education, a longstanding tradition is to deem gender a differentiating variable regarding motivation, attainment, and etc. (e.g., Barton, 1997; Dewaele, 2007). This tradition, however, witnesses criticism due to its reductionist and stereotypical bent. More and more researchers argue that it not only overlooks the dynamic relationship between learner gender and language learning, but also inherently builds on the premise that men and women are diverging groups (e.g., Schmenk, 2004; Sunderland, 2000). Thus, this thesis, by providing a comprehensive and in-depth analysis of gender stereotyping in English classrooms, can lead to a more critical and nuanced understanding of gender’s role in language domains.

From a practical perspective, looking into GSs helps to promote a gender-equitable and -inclusive environment for language learners. GSs prescribing distinct abilities, interests and achievements for boys and girls are widespread across societies and populations (Tsui and Venator, 2008), and they are suspected to hinder boys from engaging in and committing to female-dominated fields, such as languages and the Humanities (e.g., Broaded and Liu, 1996; McGeown and Warhurst, 2019). In order to encourage individuals to explore and develop their potentials freely and fully, an integrated effort against academic GSs is needed at homes, in schools, and on a policy level. The first step to this synergy is the knowledge of GSs' content, dissemination, and effect—exactly what this project aims to offer. Hopefully, its findings will help reform the existing language curricula to sensitise parents, teachers, and teacher educators to their gender-related beliefs, and inform policy-makers to implement intervention programmes reversing plausible negative impacts of GSs.

## **1.2 Overview of Thesis Structure**

This thesis consists of ten chapters. Chapter 2 will set the intellectual and pragmatic *context* for the entire research project, including conceptual issues on *gender* in western and Chinese traditions, pertinent psychological studies into *gender*, and a description of China's English language education and its importance for students in senior secondary schools.

Chapter 3 will analyse and synthesise scientific exploration into GSs. From an initial consideration of how GSs generally form and circulate, it moves to GSs in education. The female-advantage-in-languages stereotype (FALS), a three-component construct, was conceptualised, which characterises girls as more gifted, enthusiastic, and higher-achieving learners of languages than boys. Then, Chapter 3 will evaluate research into stereotypes' impact on learners. Lastly, the remainder of the chapter will identify the theoretical and methodological gaps in previous works and propose the aims of the current project: to detect the existence and influence of FALS in English classrooms in Chinese senior secondary schools.

The fourth chapter shall expound how the two-phase MM project to achieve these goals. Following a general description of the project and rationales for selecting MM design, relevant details of the four studies in two phases will be clarified in sequence: Phase 1, a convergent parallel design, captured FALS, employing a questionnaire survey among students, their guardians, and teachers of English (Study 1) and twenty individual interviews with teachers (Study 2). The subsequent Phase 2, an explanatory sequential stage, examined effects of FALS on students, with a quasi-experiment (Study 3) and eight group interviews



with students (Study 4). Relationships between studies in the same phase, together with data integration endeavours within each phase, will also be described accordingly.

The findings will be reported in four consecutive chapters. Chapter 5 will focus on how FALS-subscribing students and guardians viewed girls as better English learners. Gender difference, regional specificity, and generation gap in FALS endorsement will also be investigated. In comparison, Chapter 6 will present teachers' FALS perspectives reflected by integrated findings from the teacher participants from Studies 1 and 2. Next, Chapter 7 will continue to address how teachers perceived gender as a differentiating variable in English classrooms, relying on FALS and other GSs emerged from the interviews in Study 2. Lastly, Chapter 8 will exhibit results from Phase 2, where influences of FALS on learners were inspected. It will illustrate that, in addition to performance decrements among male learners discovered in Study 3, FALS were also found to negatively impact boys *and* girls regarding self-concepts, learning behaviours, and affective relationships with the English subject.

Chapter 9 will discuss findings from Chapters 5-8. It will first contextualise the magnitude of FALS measured in Study 1 with reference to relevant sections in Chapter 3. Thenceforth, the accuracy of FALS will be evaluated by consulting literature and/or statistics concerning gender differences in language aptitude, affect, and achievement. Similarities and disparities in FALS endorsement among students, guardians, and teachers will also be discerned. Subsequently, the role of gender stereotyping in English education will be pondered. Then, gender-divergent effects of FALS will also be deliberated based on related studies outlined in Chapter 3. In these above-mentioned sections, the theoretical and practical implications of each relevant finding will also be assessed.

Finally, Chapter 10 will summarise the main conclusions of the project and consider its limitations and contributions. Opportunities for future research will also be proposed.

## Chapter 2: Context

This chapter provides the background of the current research. Starting with two fundamental conceptual issues on *gender*, it then introduces pertinent psychological studies into *gender*, featuring how individuals develop gendered behaviours and thinking. The chapter ends with a description of English language education in China, especially its importance to senior secondary school students. Combined with Section 1.1 in the previous chapter, this chapter further highlights the rationales for this doctoral research.

### 2.1 Conceptual Issues on *Gender*

This section entertains the concept of *gender* in two strands of thought: the western scholarship (from mainly linguistics, psychology, and sociology) and Chinese philosophy (based on Confucian and Daoist traditions). The former appraises the duality of *sex* and *gender*, and the latter explains the dichotomy of men and women in Chinese culture. Together, they pave the way for understanding the design and findings of the current project.

#### 2.1.1 Terminological Disputes over *Sex* and *Gender*

In dispute among academics for years are the connotations and denotations of *sex* and *gender* (Halpern, 2012). The conventional use of the former, *sex*, was challenged by feminist scholarship in the 1970s due to its allusion to the widespread and longstanding essentialist tradition, which deems men and women as biologically defined groups programmed to possess distinct characteristics and emanate divergent behaviours (Wood and Eagly, 2010). Furthermore, with the recognition that not all individuals can be neatly assigned to the male or female categories, researchers borrowed *gender* from Greek Sophists, a term classifying the names of objects into masculine, feminine, or neutral types, to describe the social meanings of being male or female (Archer and Lloyd, 2002). The adoption of *gender*, Torgrimson and Minson (2005) recommends, should suit discussions about the psychological attributes, societal standards and ascriptions, and culturally variable characteristics associated with men and women. This is why in the current thesis, the term ‘gender stereotypes’ (GSs), instead of ‘sex stereotypes’, is employed.

Of course, some researchers oppose to treating the two terms as corresponding yet different concepts (Archer and Lloyd, 2002). Instead, they are considered closely intertwined threads, because biological factors, as well as sociocultural ones, jointly shape our behaviours towards, interpretations of, and feelings about men and women (Wood and Eagly, 2012). Another standing issue with the parallel of *sex* and *gender*, as raised by Francis and Paechter (2015), is the automatic, stereotypical association between males and maleness/masculinity,

and between females and femaleness/femininity. Both arguments further accentuate the need to study GSs in educational settings prohibiting boys and girls from challenging the accepted gender norms.

It is worth noting that I personally do not perceive *sex*, or *gender*, as a dichotomy. A feminist promoting equal rights for all, I respect and value individuals positing themselves anywhere along the continuum of sexuality. I also acknowledge the insufficiency of the *sex/gender* language in characterising our lived experiences. Synthesising related studies in linguistics, psychology, sociology, I appreciate the dynamic and context-sensitive relationship between learners' *sex/gender* and their educational experiences, especially when taking other dimensions of identity into consideration (race, social class, etc.).

### **2.1.2 Idea of Gender in Chinese Philosophy**

Like many other cultures, Chinese also harbours the binary of men and women, or *yang* and *yin*, concepts connoting males and females respectively. (Zhang, 2002). A review of Confucianism and Daoism, the two most influential philosophies native to China, indicates two time-honoured traditions: gender hierarchy and gender differentiation.

The interaction between *yang* and *yin*, according to Liu (2008), is the core underlying mainstream Chinese philosophy: 'Yijing' ('Book of Changes'), one of the earliest classic texts, values the balance between *yang* and *yin* and assigns equal importance to them. However, its likening *yang* to heaven, the monarch, the superior, and males, and *yin* to earth, the subjects, the inferior, and females inevitably justifies men's dominance over women (Xiao, 2017). Since both Confucius and Laozi, founding fathers of Confucianism and Daoism, held 'Yijing' in high regard, an inbuilt tendency of gender hierarchy manifests in the Chinese culture. For example, the traditional character for female ('女') in Chinese language resembles a person kneeling and holding hands, a posture of submission (Fan, 1996). The Confucian teaching 'men are superior than women' ('男尊女卑') also shows the gender divide in social ranking. Therefore, many contemporary thinkers agree that the first and foremost construction of *gender* in Chinese philosophy identifies the hierarchical relationship between men's and women's social roles, not their private, essential natures (Wang, 2003).

Gender differentiation is the other predominant pattern in Chinese culture. The saying 'men and women operate in separation' ('男女有别'), has been construed on three levels (Zhang, 2015). First, men take public roles, but women are tied to domestic and maternal duties ('男主外, 女主内'). Second, women should be obedient to men, which refers to their

fathers before marriage, to their husbands during marriage, and to their sons after the death of their husbands (‘未嫁从父，既嫁从夫，夫死从子’). Third, women are considered essentially different from men. While men can become sages (‘君子’) via personal development, women are substantially similar to ‘contemptible wretches’ (‘小人’), because ‘get close to them and they take liberties; keep a distance from them and they grumble’ (‘近之则不逊，远之则怨’). In addition, men are supposed to be stronger, braver, and more successful than women (‘男强女弱’).

The first two interpretations, apparently, allude to the gender hierarchy discussed earlier, prescribing high social status and overwhelming social dominance to men. The third understanding, however, implies the essentialist perspective behind observable behavioural differences. Taken together, the three layers of construal reflect the primacy of socio-relational nature of gender in Chinese thinking: after all, the biological and innate differences between the sexes seem to take a marginal role in the traditional conceptualisation of *gender*. Notably, this last character contrasts sharply with the core stance of gender essentialism in western popular culture, regarding gender differences as natural and deep-seated (Shields and Dicicco, 2011). In fact, such cultural distinction has influenced the theoretical framework of this work (see Section 3.2.2.1) and interpretation of relevant findings (see Section 9.1.3.1).

Modern versions of these traditional beliefs persist in China, even though some evidence suggests a general trend towards gender egalitarianism between 1990 and 2010 (Yang, Li, and Zhu, 2014). For example, in some dialects, the term for wives is ‘the one at home’ (‘家里的’), neglecting working women’s professional identity (Zhao, 2011). In addition, Xu (2016) has discovered a revival of conventional gender ideology between 1990 and 2010: a higher percentage of participants agree with the marriage pattern of breadwinning men and homemaking women; another notion gaining popularity is that a woman’s worth is measured by her marriage, ‘a decent job values less than a good husband’ (‘干得好不如嫁得好’). Against this backdrop, the interplay between gender and academic/professional choices in China merits meticulous investigation.

## **2.2 Research into Psychological Aspects of Gender**

The study of gender has been of interest in many disciplines, ranging from anthropology to sociology (Liben and Bigler, 2002). Psychologists approach the topic, too, through one or a combination of biological, cognitive, and social-cultural lenses (Halpern, 2012). For example, biological psychologists might look into men and women’s brain morphology, cognitive psychologists may be curious about the development of gendered behaviours and thinking in

children, and social psychologists can examine gender stereotypes (GSs), an angle pursued in this project. Given the breadth of the field, this section will focus on research into the cognitive development of gender identity, a stream of scholarly enquiry intertwining tightly with GSs, the core construct under investigation in this doctoral thesis.

Gender identity refers to individuals' sense of self as a male or female (Archer and Lloyd, 2002). As a multifaceted concept, it encompasses a) the categorisation of the self (not necessarily) on the basis of birth-assigned sex category, b) the feelings about one's own gender, and c) the expression of gender through behaviour and speech (e.g., Egan and Perry, 2001; Kohlberg, 1966; Steensma, Kreukels, de Vries, and Cohen-Kettenis, 2013). Academic literature accounting for children's development of gender identity is extensive, among which the three influential explanations are social learning theory (e.g., Mischel, 1970), cognitive-developmental theory (e.g., Kohlberg, 1966), and gender schema theory (e.g., Bem, 1983). Although the latter two underscore human agency in the process—the capacity for children to construct gender-based meaning and behaviours from an early age, all three recognise the role of social environment in shaping individuals' cognitions about gender categories. The factors present in children's upbringing may include parent-child narratives, perceptual differences between gender categories in schools, and gender essentialist tendency inherent to many cultures (Halim and Ruble, 2010), the same sources conveying gender-stereotypical beliefs to children (for a more detailed review of the latter, see Sections 3.1.1-2). In fact, researchers have also shown that the early learning of gender categories is characterised by internalising socio-cultural beliefs about stereotypical differences between men and women, and boys and girls (Kite, Deaux, and Haines, 2008).

Yet, the study of gender identity development, regardless of their theoretical bent, seems fixated on how boys and girls become sex-typed; in a way, the development of gender identity is regarded as the process of gender differentiation (e.g., Kollmayer, Schober, and Spiel, 2018; Liben, Bigler, and Hilliard, 2014). This perspective, therefore, faces an unrelenting tide of criticism, as Fine and Gordon (1989) have put it, 'this ... construction of gender-as-difference functions inside psychology as a political and scientific diversion away from the questions of power, social context, meaning, and braided subjectivities'. This is where the study of GSs comes in, contributing to a more critical and nuanced understanding of *gender* in educational contexts: instead of inspecting how boys and girls, men and women differ, it directly targets how GSs portray gender categories divergently, and how they prescribe and affect individuals' choices, aspirations, and interests. Particularly, the current

project examines GSs circulating in senior high schools, whose teenagers are at the stage of striving to become ‘socially recognised and personally expressive’ (Waterman, 1982). According to Erikson (1995), while adolescents explore their personal values and goals, they integrate ‘aptitudes developed out of endowment’ with ‘the opportunities offered by social roles’, the latter under apparent influence of GSs.

### **2.3 English Language Education in China**

Since its initial entrance into Chinese schools in the 1970s, English language has been taught nationwide, with an estimation of over 300 million learners (Hu and McKay, 2012; Wei and Su, 2012). Although the provision and quality of English education vary as a function of available resources, a desire for grasping English prevails in this geographically vast and socio-economically diverse country (Butler, 2015). According to Feng (2012), such passions arise because proficient English skills can enhance individuals’ educational and professional opportunities. Additionally, English-speaking bilinguals are needed for the nation’s prestige and prosperity regarding international trade, media coverage, and etc. (Gil and Adamson, 2011).

The teaching of English has changed from a sole focus on accuracy to a dual emphasis on accuracy *and* competence (Hu and McKay, 2012). English lessons used to involve teachers explaining vocabulary and grammar, and students doing translation exercises. There was ‘a keen interest in the exact understanding of every word, a low tolerance for ambiguity and a focus on discrete points and specific syntactic constructions’ (Cheng, 2011). A marked problem with this approach was the crippling communicative skills it leaves students with due to limited input and insufficient practice (Butler, 2015). Thus, since the 2000s, innovative teaching materials and methodologies have been advocated in order to help learners develop proficiencies in language skills (listening, speaking, reading, and writing) in addition to the mastery of vocabulary and grammar knowledge (Feng, 2012).

English lessons are compulsory from primary to tertiary education in mainland China. For students in senior high schools, it is one of the three core subjects, the remaining two being Chinese and math. In Grade 10, the first year in senior high schools, students (typically 15-16 year olds) are required to take the core subjects, as well as six subsidiary ones—history, politics, geography, physics, chemistry, and biology. The former three are often collectively referred to as ‘the Humanities’ subjects, and the latter ‘the Sciences’ courses. The curricula of 11<sup>th</sup> and 12<sup>th</sup> graders, however, follow one out of two patterns. In 17 provinces, students continue with the core courses, but they have to choose between the Humanities and

the Sciences. In 16 other provinces, the core subjects are also necessary, but they can select any three from the six subsidiary ones. In both patterns, students learn six courses in total, but the second pattern is promoted by the State Council (2014), as it abolishes the rigid distinction between the Humanities and the Sciences branches. Evidently, English learning is mandatory in both patterns. Towards the end of the third year, students (usually 17-18 year olds) take 'Gaokao', or the National College Entrance Examination (NCEE), where the three core subjects each take up 150 marks. Students are also tested in 2-3 subsidiary subjects in Gaokao; the specific number of tests varies from place to place, but each subsidiary subject typically accounts for 100 marks. Given that the outcome of Gaokao determines which university a student can attend, English, with its heavy scoring weight, is thought to be one of the most important subjects during senior secondary education. Due to the significance students attach to English results, the study of GSs targeting language learners becomes imperative.

## **Chapter 3: Literature Review**

This chapter will analyse and synthesise scholarly inquiry into gender stereotypes (GSs). Section 3.1 will focus on GSs in general. Thereafter, the rest of the chapter will narrow its focus down to academic GSs since Section 3.2, the two overarching ones being quantitative-orientated males and qualitative-orientated females. Building upon them, the female-advantage-in-languages stereotype (FALS) is conceptualised. A three-component construct, FALS characterises girls as more gifted, more enthusiastic, and higher-achieving learners of languages than boys. This is followed by a review of studies on the impact of stereotypes: stereotype threat effect on adversely targeted individuals, and stereotype lift and boost on others. Section 3.3 will identify the theoretical and methodological gaps in previous works. Finally, Section 3.4 will propose the research questions this project aims to address.

### **3.1 Gender Stereotypes in General: Processes, Dimensions and Features**

GSs are general assumptions about men and women shared among members of a community, society, or culture (Haines, Deaux, and Lofaro, 2016). Endorsing GSs, people may assign stereotype-consistent dispositions and roles to men and women, decide certain behaviours or speech as appropriate or desirable for a gender category, and assess individuals through gender-stereotypical lenses (Kite, Deaux, and Haines, 2008). That is, GSs are descriptive, prescriptive, and evaluative in nature. This section will first delineate how GSs generally form and circulate in Sections 3.1.1-2. Then, it turns to the common dimensions that GSs encompass, agency and communion among others, in Section 3.1.3. Finally, the four features of GSs—accuracy, durability, cross-cultural differences and complementarity—will also be deciphered in Section 3.1.4.

#### **3.1.1 Formation of Gender Stereotypes: Cognitive and Socio-Cultural Processes**

Since the 1940s, social psychologists have been investigating why and how stereotypes emerge (Haslam, Turner, Oakes, Reynolds, and Doosje, 2002). Social categorisation, the cognitive process of classifying individuals into social groups, is regarded as the fundamentally underlying mechanism (McGarty, 2002). GSs, specifically, are considered to develop under the additional influence of socio-cultural factors, such as the gendered division of labour proposed by the social role theory (Eagly, 1987), and gender essentialism (Halim and Ruble, 2010).

##### **3.1.1.1 Cognitive Perspective: Social Categorisation**

The formation of stereotypes, in general, is rooted in our capacity and need to make sense of the social world (Ellemers, 2018). We humans are pre-wired to organise social



events and people into groups based on perceivable similarities and differences, a process referred to as social categorisation (Tajfel and Forgas, 2000). The same process is also necessary in order to make sense of the social environment with limited time and minimal effort. After all, details associated with individuals are diverse and complex, but seeing individuals as members of a group effectively reduces that cognitive load (McGarty, Yzerbyt, and Spears, 2002). This is when stereotypes arise: as people locate and activate previously stored information about a given social category, such schematic information can become prototypes, giving rise to stereotypical expectations in future (Stangor and Schaller, 2000). These stereotypes, in turn, assist social cognition by speeding up the identification, recall, and interpretation of newly encountered people. Gender, among other readily accessible and perceptually salient social categories, is one dimension along which stereotypes are constructed (Brown, 2010).

The social categorisation approach explains how stereotyping, as a sense-making device, stems from observations of the social world, but it does not adequately address the specific factors leading to the stereotyping of a particular social category, such as gender. Nor does it explain why stereotypes of a given group cover certain contents instead of others. From a socio-cultural perspective, social role theory (Eagly, 1987) and gender essentialism (Halim and Ruble, 2010) feature, respectively, the roles played by social reality and cultural ideology in the formation of GSs.

### **3.1.1.2 Socio-Cultural Perspective: Social Role Theory and Gender Essentialism**

Social role theory suggests that GSs shared in a society, or gender role beliefs, originate from societal division of labour by gender (e.g., Eagly, 1987; Eagly, Wood, and Diekmann, 2000; Eagly and Steffen, 2000). In typically industrialised, western societies, a greater proportion of men take paid jobs with power and prestige, while women are more visible in caretaking and/or home-making positions. Engaging with their respective occupations, men and women are witnessed to carry out role-consistent behaviours and believed to possess corresponding qualities. Thereupon, GSs associating men with independence, confidence, leadership and women with nurturance, care, and pro-social tendency develop. That is, the perceivable co-occurrence of a social role and its typical, required, and/or preferred characteristics is stereotypically construed as evidence that the predominating gender assuming that role is endowed with the pertinent attributes (Petersen and Hyde, 2014). Similar to the perspective of social categorisation, social role theory also acknowledges that GSs are grounded in individuals' experience with and knowledge about the social world.

Therefore, both approaches adhere to the reality principle, asserting an element of truth in GSs (see Section 3.1.4.1 for a more detailed account of whether GSs reflect reality). What distinguishes the two theoretical perspectives, though, is social role theory's sole focus on gender-based stereotypes, and its capacity in unfolding why contents of GSs comprise particular traits rather than others. In other words, aside from clarifying the generation of GSs, social role theory also clarifies the formation of GS contents.

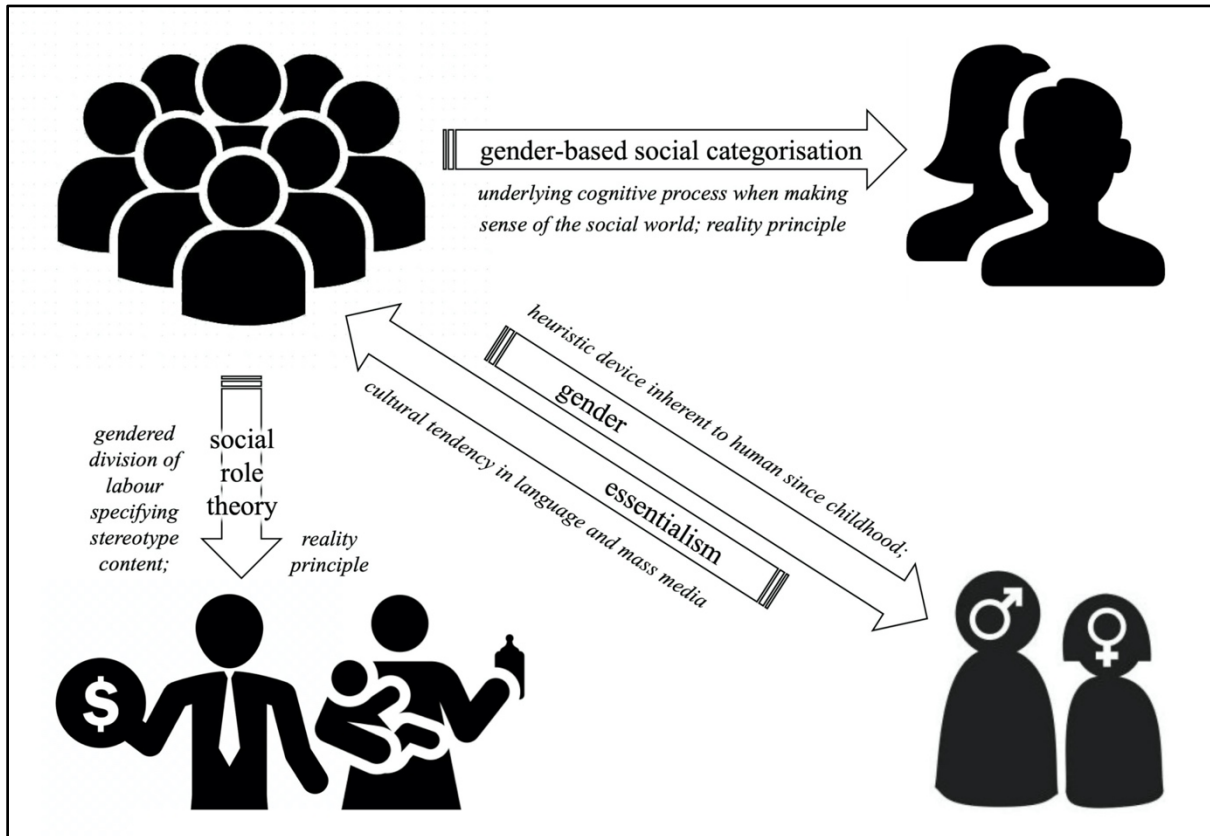
Social role theory raises that social reality, i.e. the gendered division of labour, paves the way for GSs, but a rival explanation is provided by gender essentialism, which argues for the interaction between human agency and culture in bringing about GSs (Halim and Ruble, 2010). Gender essentialism describes the tendency to attribute observable differences between men and women to distinct and innate dispositions between the sex categories (Gelman, Heyman, and Legare, 2007). Assuming an underlying, relatively stable 'essence' for men and a respective one for women, gender essentialism is a heuristic device emerging among pre-schoolers (Gelman and Taylor, 2000), *and* a shared gender ideology in many cultures (Bem, 1993). Researchers found ample evidence of individuals' automatic essentialising of gender: gender differences are viewed as natural and objective (Rhodes and Gelman, 2009); made-up gender differences are relied on to predict others' behavioural dispositions (Rhodes and Gelman, 2008, as cited in Halim and Ruble, 2010), to evaluate one's own perceptual ability (Prentice and Miller, 2006), and to affect one's commitment to improving certain skills (Rhodes and Brickman, 2008).

Cultural factors, such as language and mass media, also promote, popularise, and perpetuate gender essentialist reasoning (e.g., Sperber, 1996; Halpern, 2012). For example, in parent-child narratives, linguistic reference to generic forms of gender (such as 'girls') and contrasting boys against girls by parents implicitly present gender as a meaningful category, from which children learn to establish essentialist beliefs (Gelman, Taylor, and Nguyen, 2004). In addition, according to Halpern (2012), popular media are fraught with 'junk science' purporting that sex differences are caused by differences in men and women's brain morphology, or testosterone and oxytocin levels. These claims are not always scientifically reliable or valid (Ellemers, 2018), but they still prevail, ingraining essentialist claims in people's minds.

Figure 3.1 here illustrates the three approaches' respective contributions to accounting for GS formation. From the left to right at the top, individuals are shown to employ gender-based social categorisation to make sense of the world. From the top to the bottom on the left,

social role theory reveals that people observe gendered division of labour. Both processes lead to GSs, and both suggest that gender stereotyping is grounded in reality. The latter, unlike the former, further specifies the content of GSs.

**Figure 3.1: Three Theoretical Explanations of Gender Stereotypes' Formation**



From the top left to bottom right, gender essentialism thinking is exhibited as the third pathway that GSs emerge. It differs from the previous two because it highlights the interplay between individuals' inherent heuristics capacity and cultural factors in the social environment.

### 3.1.2 Children's Acquisition of Gender Stereotypes: Awareness, Endorsement, and Flexibility

Gender-based categorisation emerges among infants between 3- and 12-months-old, paving the way for the development of GSs and gender identities. In addition, as mentioned in Section 2.2, the two share common antecedents from children's social environment (Halim and Ruble, 2010). For example, parents can convey GSs to their children implicitly and explicitly, via encouragement for stereotype-consistent behaviours and discouragement or even punishment for counter-stereotypical ones (e.g., Bigler and Liben, 1999), or the gender-divergent choices they made regarding toys and clothes (Pomerleau, Bolduc, Malcuit, and

Cossette, 1990). Furthermore, agents in schooling systems also play a role: knowledge about GSs is learnt from teachers and classmates (Meece, 1987), cross-gender behaviours are noticed and criticised by teachers and playmates (Fagot, 1977), and GSs are presented in school textbooks (Shteivi, 2003). Meanwhile, media's part as sources of GSs is also widely recognised (e.g., Davies, Spencer, Quinn, and Gerhardstein, 2002; Ruble, Martin and Berenbaum, 2006).

Of course, as Liben and Bigler (2002) have demonstrated, children's awareness of GSs does not necessarily mean that they endorse those GSs. In addition, the strength of children's gender-stereotypical beliefs tends to vary from early to middle childhood: some scholars have found a declining tendency (e.g., Carter and Patterson, 1982; Miller, Trautner, and Ruble, 2006; Signorella, Bigler, and Liben, 1993), while others have detected a rising trend (e.g., Martin, 1989; Signorella et al., 1993). Age differences in stereotype endorsement also exist between children and adults, as Section 3.2.3 later will discuss.

### **3.1.3 Dimensions: Agency vs. Communality**

Two principal dimensions of GSs, *agency* and *communality* have been identified by researchers since the 1970s (e.g., Broverman, Vogel, Broverman, Clarkson, and Rosenkrantz, 1972; Fiske, Cuddy, Glick, and Xu, 2002; Wood and Eagly, 2010). Agentic traits are stereotypically associated with males—rationality, self-assertion, and competitiveness, to name a few (e.g., Haines et al., 2016; Kite et al., 2008). Communal qualities, on the other hand, are stereotypically feminine, examples of which include kindness, compassion, and trustworthiness (e.g., Becker and Sibley, 2016; Archer and Lloyd, 2002). In addition, these GSs seem remarkably consistent across countries. For example, undergraduates from 25 countries indicated whether a constellation of personality traits were used to characterise men or women in their cultures (Williams, Satterwhite, and Best, 1999). The results exhibit significant consensus among these cultures: adjectives demonstrating instrumentality and competence were deemed to portray men (for instance, 'ambitious' and 'logical'), and modifiers implying expressiveness or warmth were resorted to when depicting women (such as 'affectionate' and 'soft-hearted'). Another example comes from the analysis of recommendation letters for job applicants: researchers have uncovered that compared to female candidates, male ones are more likely to be labelled excellent and brilliant (Schmader, Whitehead, and Wysocki, 2007). Particularly, when spreading to educational and occupational settings, the agentic men and communal women stereotypes could manifest as

the stereotypical associations between males and STEM subjects, and that between females and qualitative domains (e.g., Bonnot and Jost, 2014; see Section 3.2.1 for more details).

Of course, agency and communality are not the only criteria men and women are compared against. Other identified dimensions include physical appearances, occupations, and role behaviours (e.g., Haines et al., 2016; Kite et al., 2008). Nevertheless, the two aspects are the fundamental aspects GSs tend to encompass. In fact, Fiske and colleagues found converging evidence that the same two dimensions are basic to human perceptions of various other social groups: in their proposed Stereotype Content Model (SCM), the two dimensions are named competence (i.e. agency) and warmth (that is, communality). (e.g., Cuddy, Fiske, and Glick, 2007; Fiske et al., 2002). Under this model, scholars unearthed cross-cultural GSs of warm and expressive women, and competent and independent men (e.g., Fiske et al, 2002; Operario and Fiske, 2004).

In addition to the descriptive function elucidated above, GSs also present certain dispositions, behaviours, or roles, as more appropriate or desirable for a given gender category. For example, men's interest and inclusion in professional and domestic roles involving caregiving are limited due to GSs assigning communality to women (Croft, Schmader, and Block, 2015). Another illustration of GSs' prescriptive nature is that not only are agentic traits seen as being typical of men, they also communicate what qualities people think men ought to possess; likewise, communal traits depict what women are *and* prescribe what women should be (e.g, Prentice and Carranza, 2002). The third manner through which people utilise GSs is to assess individuals: experimental studies have discovered that the quality of a piece of work, or an achievement, is rated higher when it is identified by a masculine name, instead of a feminine one (e.g., Banaji and Greenwald, 1995; MacNell, Driscoll, and Hunt, 2015). Such gender-based evaluative differences are prompted by GSs, and can further impact perceptions of men and women's potentials: men's success in education or at work is usually ascribed to innate ability; but women's accomplishment, especially in masculine areas, are attributed to external or temporary reasons, such as luck and effort (Swim and Sanna, 1996). In fact, when individuals display counter-stereotypical behaviours in the workplace, they risk being disliked and even devalued (e.g., Moss-Racusin, Phelan, and Rudman, 2010; Rudman and Phelan, 2008)

### **3.1.4 Features of Gender Stereotypes**

Academic inquiries into GSs probe into a diverse range. The previous sections account for the processes of GSs formation and acquisition, as well as the composition and functions

of GSs. This section now turns to four features of GSs: Section 3.1.4.1 discusses whether GSs accurately reflect gender differences in reality, and Section 3.1.4.2 analyses factors predicting stereotype change or stability. Then, Section 3.1.4.3 will explore cultural variations, or similarities, in the content of GSs. Finally, Section 3.1.4.4 pinpoints how GSs present men and women as complementary groups, each with their respective strengths and weaknesses. The four features are introduced here because they pertain to the academic GSs that will be expounded in the remainder of this chapter.

### **3.1.4.1 (In)accurate Gender Stereotypes**

As explicated in Section 3.1.1, GSs can contain an element of truth, because they emerge from people's observation of and direct experience with the social world (Becker and Sibley, 2016). Swim (1994) explored this fact-based nature of GSs by comparing the magnitude of perceived, stereotypical gender differences against the amplitude of genuine gender differences reported in several meta-analytic works. Out of the 17 GSs under inspection, nine were accurate (correctly reflecting the direction and strength of actual gender differences), six overinflated the extent of real differences, and only two were wrong in both direction and size. Similarly, Guimond (2008) synthesised differences in personality traits between men and women from North America, Western Europe, Africa and Asia; the identified differences were in line with GSs describing women as caring and men as agentic. Thus, overall, GSs seem capable of capturing differences between men and women. More specifically, academic GSs associating men with STEM and women with the Humanities seem to contain a kernel of truth, because there is evidence that male and female learners are academically inclined in stereotype-consistent ways (e.g., Priess and Hyde, 2010), although contradictory evidence of gender-atypical performance patterns do exist (e.g., girls outperform boys in math, as reported in Voyer and Voyer, 2014). Section 3.2.2.5 will show, in particular, that the GSs concerning language domains are likely to mirror genuine gender differences.

Four caveats have been raised regarding the seeming accuracy of GSs. The first argument is rooted in the definition of stereotypes, which are assumptions about social groups (Becker and Sibley, 2016). Thereupon, GSs are, by nature, broad generalisations about men and women, and apparently, not every man or woman necessarily fits such 'normalised', 'averaged' images. In this sense, GSs can hardly be accurate when they are applied to individuals. The second call of caution is voiced by researchers looking into actual gender differences: a review of hundreds of studies revealed a considerable similarity between men

and women's cognitive abilities, personality traits, and behavioural tendencies (Hyde, 2014). This similarity is defined statistically because a) the mean differences between men and women tend to be small, especially when compared to differences between individual men and individual women; and b) the overlap between men and women's distributions are substantially large. Therefore, in reality, men and women are similar despite their differences. GSs, however, only emphasise the differences. The essentialist tradition embedded in GSs, worse still, sometimes leads to far-reaching claims as if men and women were practically incommensurable (Ellemers, 2018). Therefore, although GSs may not be wrong, they cannot be acknowledged as reflecting the whole truth.

The third reason lies with findings that people's perception of the social reality, as well as their processing of counter-stereotypical information, can be filtered by gender-stereotypical lenses. For example, parents endorsing the academic GS that boys are more competent math learners consistently overestimate their boys' giftedness in math but underestimate their daughters' potentials (e.g., Eccles, Jacobs, and Harold, 1990; Frome and Eccles, 1998). Moreover, when encountering counter-stereotypical individuals, people sometimes disregard them as unrepresentative of their gender category (e.g., Fiske et al., 2002), or classify them as a specific subtype that diverges from the group norm (e.g., Kite et al., 2008). The final rationale is grounded in the line of research demonstrating how GSs lead people to behave in stereotype-consistent ways, which will be clarified in Sections 3.2.4-5. Apparently, these last two counterarguments cast similar doubts over GSs: if they influence social cognition and behaviours in stereotype-consistent ways, the 'reality' they claim to reflect becomes the end product of GSs themselves. Therefore, the accuracy of GSs, instead of winning them louder applause, actually makes them harmful and dangerous: as self-fulfilling prophecies, GSs can easily prevent individuals from perceiving the world more objectively and realistically, and they limit people's behaviours to gender-typical ranges.

#### **3.1.4.2 Durable Gender Stereotypes**

Considering that GSs develop from observations of gender-divergent patterns, as posited by social role theory (e.g., Eagly and Steffen, 2000), changes towards more gender-balanced participation and gender-equal performances in some societies should lead to parallel adjustment of GS contents. In addition, efforts to promote gender-egalitarian values and gender fairness in education and work might also give rise to more people questioning or even renouncing their stereotypical beliefs. Indeed, there is empirical evidence pointing out to stereotype change. Diekmann and Eagly (2000), for instance, has found that women in future

are believed to be more agentic than their peers in the present. Similarly, Plante and colleagues have also discovered that middle school students no longer think of girls as less talented or lower-achieving math learners (Plante, Theoret, and Favreau, 2009). Yet, in both studies, corresponding changes in stereotypes about men are absent: future men are not assumed more communal (Diekmann and Eagly, 2000), nor are boys regarded favourably in stereotypically female-dominating fields, such as language studies (Plante et al., 2009). Perhaps this is because men's participation in feminine fields are harder and rarer: typically, feminine qualities and roles are usually valued less in societies, and men can experience identity threat more often when they attempt to violate gender norms (Croft et al., 2015).

Actually, more studies have shown that GSs are more likely to endure. In Diekmann and Eagly's research (2000), for example, although the perceived agency increases in future women, they are still rated more communal yet less agentic than men. That is, the nature of GSs persists, even though the magnitude of the GSs is reduced. In another study, Haines et al. (2016) compared and contrasted GSs endorsed by people of this generation and those approximately three decades ago (1984). Their findings reveal that the amplitude of agentic men and communal women stereotypes remained practically the same. Furthermore, GSs appear resilient to change when their impact on social perception and behaviours are taken into account, as the last paragraph in the previous section has indicated. In academic settings, in particular, GSs can be self-fulfilling, channelling boys and girls to gender-disparaging paths (Kollmayer et al., 2018). For instance, a group of studies show that parents stereotypically believing in girls' lack of math ability would discourage their daughters from continuing with selective math courses (Tenenbaum, 2009), deliver more intrusive support while girls are doing their math homework (Bhanot and Jovanovic, 2005), and consequently, reduce the daughters' self-efficacy and demotivate them to pursue math in future (Tiedemann, 2000). Furthermore, similar effects are detected for teachers' GSs (for a review, see Gunderson, Ramirez, Levine, and Beilock, 2012). So long as these GSs continue to impact boys' and girls' educational engagement and pursuit, existing academic GSs will remain stable.

### **3.1.4.3 Cross-Cultural Gender Stereotypes**

Previous sections have pointed out that the agentic men and communal women stereotypes prevail in Western Europe, Asia, Africa, and North America (e.g., Fiske et al, 2002; Operario and Fiske, 2004; Williams et al., 1999). Nevertheless, variations in stereotype endorsement still exists. A common finding, for example, is the distinctions between



individualist and collectivistic societies. According to Hofstede (2011), the former values individuals' goals and rights higher than that of groups, and the reverse holds true for the latter. Compared to collectivistic cultures, individualistic societies generally tend to endorse less rigid GSs, because many collective countries are located in less wealthy regions in Africa and Asia, where people routinely adhere to more traditional gender roles than those wealthier, individualist parts of the world (Becker and Sibley, 2016). The same correspondence between gendered participation and related GSs also manifests when a particular field is under inspection: across 66 participating countries, those with higher proportions of female students and workers in the science field reported weaker male-science stereotypes (Miller, Eagly, and Linn, 2015).

Of course, individualism/collectivism is not the only dimension along which cultures vary; the relative weight attached to personal accomplishment or interpersonal relationship can also distinguish cultures (Hofstede, 1980). When a culture emphasises achievement over collaboration, it is task-orientated; but when cooperation is stressed more than performance, it is person-orientated. Transnational studies indicate that person-orientated cultures tend to hold less strong agency-related GSs: while men are stereotypically thought as more agentic than women in both Germany and Japan, Japanese women are not judged more communal than men (Steinmetz, Bosak, Sczesny and Eagly, 2014). The researchers reasoned that due to the priority of teamwork over personal attainment in a culture like Japanese, men and women are homogeneously expected to be co-operative, pro-social, and interdependent. Such expectations in turn bring out familiar communal behaviours in men and women, resulting in the typical Japanese men being perceived equally communal compared to Japanese women. In addition, the agentic men stereotype in Japan remains relatively weak, when being compared against another Asian culture—Malaysian (Lim, 2002). This study further shows that even cultures that are thought to be substantially similar can vary in their gender-stereotypical beliefs. In fact, some scholars argue for region specificity and sub-culture heterogeneity within national borders: Huo and Randall (1991) identified disparities in collectivism/individualism and person-/task-orientations among five Chinese cities, and stronger GSs are expressed by participants from more collectivist and person-orientated regions. Likewise, van de Vijver (2007) surveys four sub-cultural groups in the Netherlands, and discovered more gender-egalitarian beliefs in some groups than others.

#### **3.1.4.4 Complementary Gender Stereotypes**

The complementary feature of GSs is explored and established by SCM researchers, who have captured that almost universally, men are considered more competent but not necessarily nice, while women are thought as warmer but less effective (e.g., Archer and Lloyd, 2002; Fiske et al., 2002). Thus, GSs depict men and women as different yet complementary groups, each possessing their own strengths and weaknesses; in addition, one party's merit corresponds to the other's demerit, hence 'completing' each other. This is perhaps why existing GSs have been revealed as a means to justify current systems and maintain the status quo: complementary GSs can be interpreted as balanced and equal (Jost and Kay, 2005).

Academic GSs, as previous sections have implied, can also be complementary. Male students are stereotypically characterised more positively in STEM, besides being troublesome, less academically motivated and easily bored with schoolwork, while girls are perceived more favourably in the Humanities, as well as being compliant, averagely more devoted and enjoying school more (e.g., Bonnot and Jost, 2014; Carr and Pauwels, 2006; Jones and Myhill, 2004).

Although complementary GSs can be matched as pairs, the two GSs in the same pair might still differ in magnitude and result in diverging prescriptive powers. Some studies have detected that men occupying female-dominated positions (e.g., social workers) are more likely to be accepted (e.g., Siyanova-Chanturia, Warren, Pesciarelli, and Cacciari, 2015), but some others found that compared to communal men, agentic women are more favourably regarded (e.g., Miller, Bilimoria, and Pattni, 2000). In educational contexts, the complementary pairs seem to be endorsed to varying extent and affect learners with disparate forces, as the next section will delineate.

### **3.2 Gender Stereotypes in Education: Contents and Effects**

So far, the review of GS literature is not specified to any particular context. Yet, many recent works show that the comparative context where GSs arise and circulate can affect their magnitudes and effects on targeted individuals (e.g., Becker and Sibley, 2016; Ellemers, 2018; Halim and Ruble, 2010). GSs in schools, furthermore, merit more meticulous investigation, because of their profound influence on boys' and girls' enrolment, engagement, and career aspirations regarding a range of domains (Kollmayer et al., 2018).

This section analyses and synthesises academic inquiries into stereotypes about male and female students. Section 3.2.1 will first explain two overarching GSs—quantitative males

and qualitative females. Thereupon, the hypothetical female-advantage-in-languages stereotype (FALS) is proposed as a three-component construct describing girls as more gifted, more enthusiastic, and higher-achieving learners of languages than boys. Its theoretical underpinnings, including pertaining findings and major conclusions from previous literature will be expounded in Section 3.2.2. Then, Section 3.2.3 will move on to identify variations in stereotypical views due to gender, age, and region. The remaining two subsections will discuss effects of GSs (Section 3.2.4-5).

### **3.2.1 Quantitative-Males and Qualitative-Females Stereotypes**

As noted in Chapter 1, the majority of empirical studies concerning academic GSs examined, exclusively, the stereotypical association between males and STEM subjects (e.g., Miller et al., 2015; Tiedemann, 2000; Tomasetto, Mirisola, Galdi, and Cadinu, 2015; Yee and Eccles, 1988). The most widely endorsed male-STEM stereotypes, with an essentialist bent, deem that boys are naturally endowed with a stronger STEM aptitude than girls (e.g., Andre, Whigham, Hendrickson, and Chambers, 1999; Cejka and Eagly, 1999; Dasgupta and Stout, 2014). For instance, parents of first-graders in Italy rate boys as more competent math learners than girls, regardless of the gender of their own child (Tomasetto et al., 2015). Likewise, German adolescents and young adults assign stronger math aptitude to boys rather than girls (Steffens and Jelenec, 2011). Moreover, the male stereotyping of STEM also gives rise to stereotypical expectations of better attainment and higher performance on boys' side. Mothers from three parts of the world—Japan, Taiwan, and the United States—were asked to predict how well their children would do in a hypothetical math test (Lummis and Stevenson, 1990). Although the participating children showed no gender difference in actual math achievement, the mothers with sons appeared more optimistic about their children's performance than those with daughters. In their comprehensive review of gender-related math attitudes, Gunderson et al. (2012) also acknowledge that parents stereotypically overestimate their son's math grades, but the same cannot be said for daughters.

Some works disclose that the male-STEM stereotypes can encompass an affective dimension, maintaining that boys are more interested, or enjoy themselves more in STEM than girls. Indirect evidence comes from teacher nomination studies, for example: when math teachers are prompted to name and then describe two boys and two girls who either succeed or fail in math, they are convinced that successful boys have more fun in math than equally high-achieving girls (Fennema, Peterson, Carpenter, and Lubinski, 1990). Parents, too, share such stereotypically gender-specific appraisal of interest: parents of sons are more confident

that their children are attracted to science courses than those of girls, even if the children themselves report similar levels of interest (Tenenbaum and Leaper, 2003).

The three above-mentioned dimensions of male-STEM stereotypes, i.e. gender-distinctive aptitude, gender-differential achievement, and gender-divergent affect, are scrutinised in the expectancy-value model (EVM) of motivation (e.g., Eccles, 2014; Leaper, Farkas, and Brown, 2012; McGeown and Warhurst, 2019). The theory argues that domain-specific motivations develop out of a) expectancy beliefs, which include learners' competence self-beliefs and expectations of future success in the area, and b) perceived value, which represents the interest, enjoyment, and/or importance individuals attach to the field. When applied to understand gender differences in achievement-related choices, EVM studies reveal that cultural GSs, as well individuals' own gender identities, give rise to stereotype-consistent expectancy and value beliefs (e.g., Leaper et al., 2012; Li, 1999; Liu, 2018; McGeown and Warhurst, 2019; Schoon and Eccles, 2014). In these studies, boys tend to express more confidence in their math/science aptitude and future success, and they also appear more interested in math/science, corresponding to the aptitude, achievement and affect elements of academic GSs regarding STEM.

Contrary to the male stereotyping of STEM, the female stereotyping of qualitative areas receives much less empirical inquiry. Very few research on EVM probes into the Humanities subjects, the ones stereotypically associated with girls and women. McGeown and Warhurst's study is among the rarity (2019): compared to boys, primary schools girls are found to show higher expectations of success in reading and writing, and they also claim to find the two literacy activities more interesting and important. Another line of evidence arises from cross-cultural findings that girls express more positive self-efficacy beliefs and stronger interest in the Humanities courses (e.g., Andre et al., 1999; Francis, 2000; Liu, Hu, Jiannong, and Adey, 2010). Researchers, when contemplating on the sources of these gender differences, argue that cultural stereotypes assigning literary prowess to girls instead of boys must have pushed them to pursue the Humanities fields to some extent.

These existing evidence respecting females and qualitative courses, though not as extensive as the support for male-STEM stereotypes, seems to point out the same three stereotypical dimensions: aptitude, achievement, and affect. The emergent trio-component structure of the academic GSs, actually, resembles the ABC model of attitudes (Breckler, 1984). This classic model proposes that people form attitudes along three dimensions: A, affect, represents emotions and feelings embedded in the attitude; B, behaviour, stands for

behavioural intentions and action results; C, cognition, refers to understanding of the nature of the attitude object (Eagly and Chaiken, 1998). In terms of academic GSs, the gender essentialist notion that men and women naturally possess different cognitive and academic capacities corresponds to the C component in the ABC model, and the affect and achievement dimensions, respectively, correspond to the A and B elements. The affinity between the ABC model and academic GSs strengthens the theoretical underpinning of the latter, because academic GSs are fundamentally gender-stereotypical attitudes towards male and female learners.

### **3.2.2 Conceptualising ‘Female-Advantage-in-Languages Stereotype’**

Building upon the quantitative-males and qualitative-females stereotypes, the female-advantage-in-languages stereotype (FALS), is conceptualised by the researcher. It is hypothesised as a tripartite construct, characterising female language learners more positively in terms of a) language aptitude, b) achievement in language tasks, activities and exams, and c) affective relationships with language domains. Three major streams of research attest to its existence: 1) gender-related attitudes towards languages among learners, parents, and teachers; 2) gendered expectations and appraisals from parents and teachers; 3) gender stereotypes on cognitive abilities.

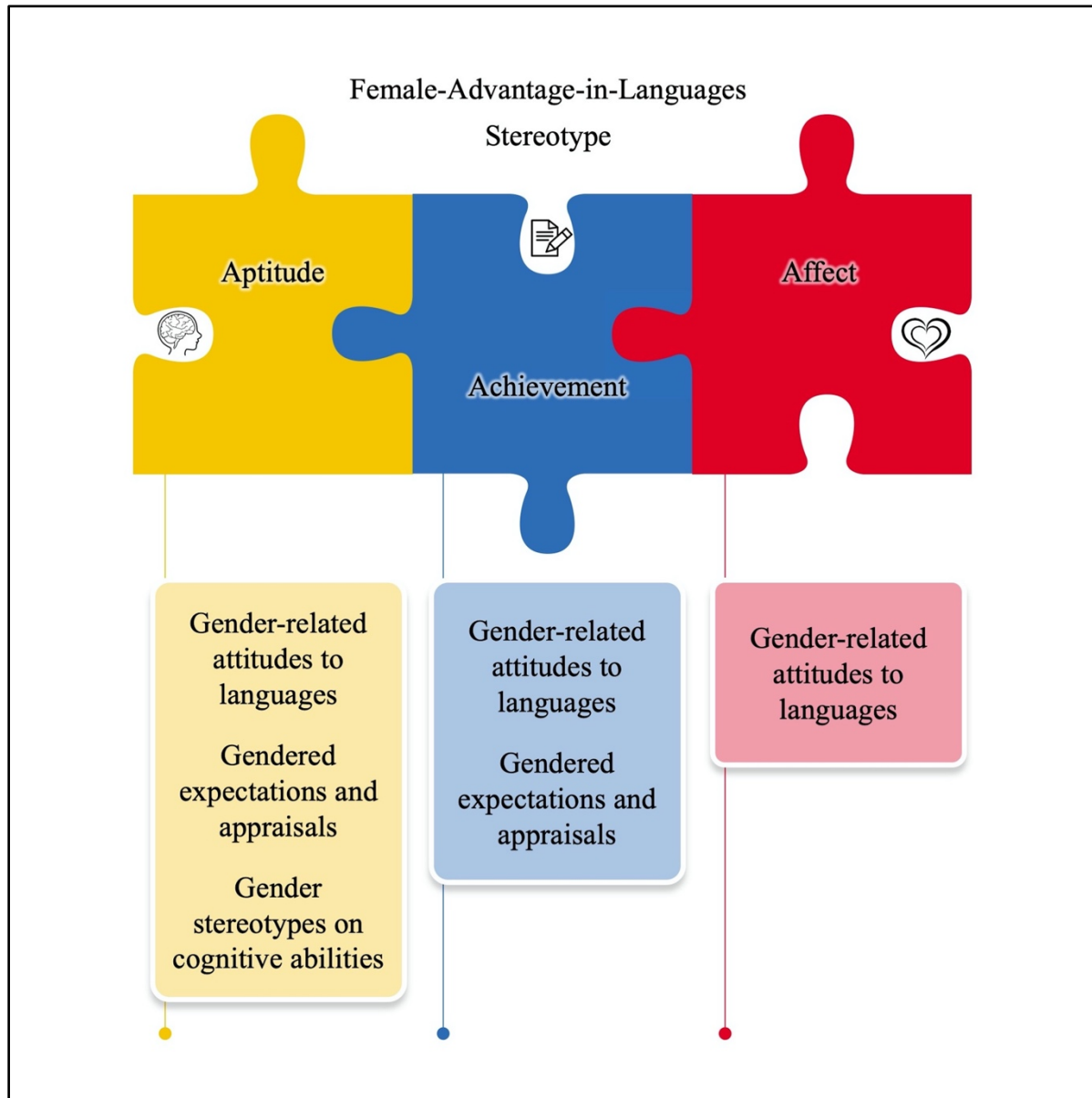
Figure 3.2 illustrates how each line of studies supports the existence of one or more of FALS components. In the remainder of this section, each component will be explained with references to academic sources. During the process, stereotype endorsement by learners themselves, their parents and/or guardians, and their teachers of languages will be entertained. A review of language-related GSs and an appraisal of their accuracy will also be introduced towards the end of this section.

#### **3.2.2.1 *Aptitude Component***

The *aptitude* component refers to stereotypical ideas that women and girls have a stronger *innate* linguistic capacity than men or boys. As Figure 3.2 here indicates, three lines of research have explored these ideas, although the results are somewhat conflicting. The initial evidence is provided by social psychological studies quantifying gender stereotypes concerning cognitive abilities (e.g., Haines et al., 2016; Kite et al., 2008; William et al., 1999). Explicit self-report measures are predominantly used, which typically contain a list of attributes for participants to rate regarding perceived masculinity and femininity. In these studies, language-related items are always considered feminine in various countries and

regions, including being verbal skilled, expressive (of emotion or tender feelings), talkative, and having literacy capacity (Archer and Lloyd, 2002; Cejka and Eagly, 1999; Liu, Huang, Jia, Gong, Huang and Li, 2011; Prentice and Carranza, 2002).

**Figure 3.2: Conceptual Framework of Tripartite  
‘Female-Advantage-in-Languages Stereotype’**



Research into gender-specific evaluations of linguistic abilities among parents and teachers also offer evidence for the aptitude component of FALS. Cross-culturally, parents tend to associate literacy tasks, foreign language courses, and social sciences in general, with females (e.g., Frome and Eccles, 1998; Guo, Tsang, and Ding, 2007; Schmenk, 2004; Muntoni and Retelsdorf, 2019). A few studies have further found that these parents of boys consistently rate male learners' language aptitude lower than those with daughters, even

when their own children's attainment in the relevant course is controlled for (Muntoni and Retelsdorf, 2019; Zhao, Li, and Yin, 2014). Some studies identified similar competence-based GSs among language teachers (e.g., Carr and Pauwels, 2006; Chen, 2007; Retelsdorf, Schwartz, and Asbrock, 2015; Zhao et al., 2014), although such studies are fewer compared to those investigating math and science teachers' GSs.

Still, these limited efforts are instrumental on theoretical and methodological grounds. First, when the same measurement scale is used on both parents and teachers, the latter seem to endorse the gender-language stereotype to a greater extent (Muntoni and Retelsdorf, 2019; Retelsdorf et al., 2015). A plausible explanation is that teachers are more actively engaged with language education: according to the reality principle of stereotype formation (Section 3.1.1), they have therefore established stronger stereotypical beliefs. Second, quantitative methods, featured by an abundance of questionnaire surveys, seem to predominate the area. Qualitative inquiries into how parents and teachers interpret 'language aptitude', or 'linguistic competence', and how they portray boys and girls distinctively is largely unexplored. The only exception is the work by Carr and Pauwels (2006), but the teachers' perspectives in this book is used to support findings from the learner data. Such insufficient attention directed at parents, even less for teachers, contrasts with the attention on learners, as the next paragraph will show.

The largest body of research concerning the aptitude component, actually, is academic literature on students' gender-related attitudes toward languages. For example, primary school children in USA are convinced that girls have a stronger gift in spelling than boys (Heyman and Legare, 2004). Familiar stereotypical perspectives prevail among older learners, including high school students (e.g. Plante et al., 2009) and undergraduates (e.g., Guimond and Roussel, 2001). In Carr and Pauwels's extensive book on boys' relationships with foreign languages (2006), Australian students express in interviews that girls are biologically better equipped for language courses than boys. Some other studies utilising adapted versions of the 'Beliefs About Language Learning Inventory' (BALLI), however, uncovered conflicting voices. BALLI is designed by Horwitz (1987); among over 30 items of foreign language learning, one targets gender difference in linguistic ability: 'Women are better than men at learning foreign languages'. Some researchers report that their participants agree with the statement (e.g., Altan, 2012; Peacock, 2001), while others find conflicting voices favouring men over women (e.g., Asassfeh, 2015; Siebert, 2003). The positive opinion regarding men, yet, is expected, because variations of the agentic men stereotype seem to

depict men as generally more intelligent and capable learners without specifying any domains (e.g., Bian, Leslie, and Cimpian, 2017). Notably, the nature of the interplay between the two academic GSs, FALS and male-agency stereotype, remains unclear, because no previous studies have approached both simultaneously. Another reason moderating the amplitude of the aptitude component of FALS might be the extent of gender essentialism in different cultures. For example, as Section 2.1 has illustrated, compared to western societies, Chinese people are less inclined to essentialise on the basis of gender. Therefore, the Chinese might hold a weaker stereotype regarding gender differences in language aptitude. Again, because there is little empirical research into GSs concerning language domains in China, this assumption awaits verification.

Therefore, to sum up, evidence from mostly quantitative studies reveal that learners, parents, and teachers stereotypically think that girls and women possess stronger language-related aptitude than boys or men. The inconsistencies from BALLI surveys and the explanation offered by male-agency stereotypes indicate that perhaps the aptitude component of FALS is somewhat contested, shared by fewer believers, or weaker compared to the other two components.

### **3.2.2.2 *Achievement Component***

The *achievement* component of FALS entails perceptions that girls usually outperform boys in language-related activities, tasks and assessments. It differs from the aptitude component in that it emphasises on learning outcomes, not learning potentials. Indirect sources can be found in review or commentary articles on gender and achievement in language and literacy. For instance, stereotypes prescribing language proficiency as a feminine pursuit, as argued by Watson, Kehler, and Martino (2010), have discouraged boys from fully committing to developing literacy skills. In the context of foreign language education, Schmenk (2004) also laments that despite many findings pointing out a lack of gender difference in certain English skills and achievement, some researchers still subscribe to stereotypical stances that females learn English as a foreign language (EFL) better than males.

Direct support for the existence of the achievement component comes from two streams of research: parents' and teachers' gendered expectations, and students' own gender-divergent efficacy beliefs concerning languages (e.g., Kurtz-Costes, Copping, Rowley and Kinlaw, 2014; Zhao et al., 2014). For example, Lummis and Stevenson (1990) have asked participating mothers how many marks they expect their children to get in two hypothetical



tests, one for math and another for reading. The results show that American and Chinese mothers anticipate their daughters to achieve higher in reading but sons in math. Teachers also endorse achievement-related GSs regarding literacy skills (Retelsdorf et al., 2015) and language subjects (Georgiou, 2008). Learners themselves are no exception: for instance, elementary and secondary school students from America, as well as China, are convinced that girls are more likely to become better writers and readers than boys (Kurtz-Costes et al., 2014; Zhao et al., 2014). McGeown and Warhurst (2019) uncovered similar gender-stereotypical expressions of confidence in reading and writing among 5<sup>th</sup> and 6<sup>th</sup> graders in England. A particularly intriguing finding from the study, additionally, is that the more students identify with feminine traits (e.g., ‘compassionate’ and ‘gentle’), the stronger their expectancy and value beliefs are, regardless of their biological sex. That is, literacy achievement is perceived as a feminine pursuit by both boys and girls.

Scholarly investigation into the achievement component, like the case with the aptitude component, also relies mostly on quantitative approaches. But the two types of research differ in quantity: less exploration of the achievement aspect has been undertaken, and the pertinent studies typically inspect language-related fields in addition to STEM ones (e.g., Lummis and Stevenson, 1990; McGeown and Warhurst, 2019). Thus, intensive, exclusive, and in-depth inquiry into the achievement component of FALS is in need.

### **3.2.2.3 *Affect Component***

The *affect* component of FALS asserts that girls tend to have more positive emotions with languages than boys, such as a keener interest and more enjoyment. It is the most under-researched component, with traces of evidence from gender-related attitudes towards language courses (e.g., Carr and Pauwels, 2006; Epting, Rand, and D’Antuono; 2014; Plante et al., 2009). Plante and colleagues (2009) have invited Canadian students to indicate the maleness and femaleness of two domains, math and languages. The 6<sup>th</sup>, 8<sup>th</sup> and 10<sup>th</sup> graders all believe that girls have more fun in language modules than boys. In addition, preschoolers and young adults also think that compared to boys, it is more typical of girls to prefer language-related activities (Francisca del Rio and Strasser, 2013; Koenig, 2018). Very few studies utilised qualitative interviews, which uncover the affect component in the peer group: Australian boys expressing interest or enjoyment in languages have recalled unpleasant experiences of being laughed at and even teased by classmates who find language-loving boys unmanly (Carr and Pauwels, 2006). In the follow-up interviews to a questionnaire survey, the 7<sup>th</sup> and 9<sup>th</sup> graders from England also claim that boys are routinely discouraged

from liking French or German lessons due to risks of peer teasing (Williams, Burden, and Lanvers, 2002).

#### **3.2.2.4 Review of Methods Used in Literature on Female Stereotyping of Languages**

Three streams of research has approached the female stereotyping of language domains, including gender-related attitudes towards languages, gendered expectations and appraisals from parents and teachers, and gender stereotypes about cognitive abilities. Together, they point out the existence of FALS, a tripartite stereotype describing girls as more gifted, enthusiastic, and higher-achieving language learners than boys.

In the existing GS literature, exclusive and intensive inspection of language domains is few, compared to that of STEM areas. With regards to data collection techniques, survey methods are most commonly used, including questionnaires with Likert scales as responses and structured interviews, the latter typically applied to young children (e.g., Francisca del Rio and Strasser, 2013; Heyman and Legare, 2004). Studies employing these self-report measurements usually invite respondents to indicate the degree to which they think a list of attributes or statements apply to a typical man or woman. Employing such direct and explicit measures, previous studies have identified dimensions where people would readily categorise men and women (as summarised by Wood and Eagly, 2010). In comparison, qualitative inquiries into idiosyncrasies of stereotypical, as well as counter-stereotypical, beliefs is largely unexplored. Another gap is the insufficient attention directed at parents and teachers, who have been recognised as sources of children's GSs. Furthermore, previous work hardly ever present perspectives from non-European or non-Anglo-Saxon societies, let alone those with a large population of foreign language learners like China.

Another method worthy of mentioning is the Implicit Association Test (IAT), which has gained popularity among experimental psychologists (e.g., Greenwald, McGhee, and Schwartz, 1998; Nosek, Banaji, and Greenwald, 2002). Using response latencies to represent the strength of the stereotypical associations between gender categories and particular traits or roles, IAT is instrumental in assessing the magnitude of GSs (Greenwald, Nosek, and Banaji, 2003). The longer the response time, the stronger the stereotype is. Yet, IAT was not introduced earlier because it does not qualitatively specify the nature of the stereotype under investigation. When IAT probes into academic GSs, the two concepts are gender categories (male or female) and domain (STEM or the Humanities). Whether the connection between the concepts pertains to aptitude, achievement, or affect remains elusive. Moreover, IAT studies usually compile male-STEM and male-qualitative stereotypes into one gender-domain

stereotype (e.g., Nosek et al., 2002). Without studying the two in isolation, implicit research can only inform the readers how strong the gender stereotyping of academic world is in general.

### 3.2.2.5 Accuracy of Female-Advantage-in-Languages Stereotype

A final note on FALS is its accuracy. Regarding aptitude and achievement components, they seem to contain a kernel of truth, because research into verbal capacities and linguistic achievement identifies, *in general*, gender differences favouring girls. Table 3.1 here summarises findings regarding language aptitude, where girls and women seem to surpass boys or men in various types of verbal tasks (expressed by the positive Cohen's *ds*). However, in three tasks, verbal items in SAT (Feingold, 1988), analogies (Hyde and Linn, 1988) and vocabulary (Hedges and Nowell, 1995), males display a marginal edge over females. These inconsistencies, according to Halpern (2012), might have been rooted in social and historical contexts, which profoundly impact the development of cognitive abilities. Thus, it would seem that when examining FALS and gender differences in linguistic aptitude, and contrasting the two to determine the accuracy of the former, contexts must be taken under consideration. Still, despite these three studies, it seems that the aptitude of FALS, overall, mirrors the real difference between males' and females' verbal abilities.

**Table 3.1: Summary of Research on Gender Differences in Verbal Abilities**

Study and verbal ability	Participants	Cohen's <i>d</i>
Feingold (1988)		
DAT language	9 <sup>th</sup> - 12 <sup>th</sup> graders	.40
DAT spelling	9 <sup>th</sup> - 12 <sup>th</sup> graders	.45
DAT verbal reasoning	9 <sup>th</sup> - 12 <sup>th</sup> graders	.02
PSAT (verbal)	11 <sup>th</sup> and 12 <sup>th</sup> graders	.04
SAT (verbal)	11 <sup>th</sup> and 12 <sup>th</sup> graders	– .01
Hedges and Nowell (1995)		
Reading comprehension	15- to 22-year-olds	.09
Vocabulary retention	15- to 22-year-olds	– .06
Hyde and Linn (1988)		
Anagrams	Adolescents and adults	.22
Analogies	Adolescents and adults	– .16
Essay writing	Adolescents and adults	.09
General verbal ability	Adolescents and adults	.20

Reading comprehension	Adolescents and adults	.03
Overall verbal ability	Adolescents and adults	.11
Speech production	Adolescents and adults	.33
Vocabulary knowledge	Adolescents and adults	.02
Johnson and Bouchard (2007)		
CAB verbal-proverbs	18- to 79-year old twins	.06
HB vocabulary	18- to 79-year old twins	.18
CAB verbal-vocabulary	18- to 79-year old twins	.05
Spelling	18- to 79-year old twins	.26
WAIS Vocabulary	18- to 79-year old twins	.05
Word Fluency	18- to 79-year old twins	.20
Word Beginnings/Endings	18- to 79-year old twins	.23
Maylor, Reimers, Choi, Collaer, Peters, and Silverman (2007)		
Category fluency	20- to 60-year-olds	.18
Reilly (2012)		
Reading comprehension	15-year-olds	.26
Steinmayr and Spinath (2008)		
Verbal intelligence	11 <sup>th</sup> and 12 <sup>th</sup> graders	– .32

*Notes.* DAT = Differential Aptitude Test. PSAT = Preliminary Scholastic Aptitude Test. SAT = Scholastic Aptitude Test. CAB = Comprehensive Ability Battery (developed by Hakstian and Cattell, 1973). HB = The Hawaii Battery (devised by DeFries, Vandenberg, McClearn, Kuse, Wilson, Ashton, and Johnson, 1974). WAIS = The Wechsler Adult Intelligence Scale (Wechsler, 1955). A negative value of Cohen's *d* here indicates that boys performed better than girls on a particular dimension.

Similarly, Table 3.2 here exhibits studies looking into gender differences in (first, second, or foreign) language performance in various countries. Again, except for two studies (Maqsud, 2006; Seginer and Vermulst, 2002), girls reliably and consistently outperform boys in language-related tests and assessments. Therefore, the achievement component might also hold true.

**Table 3.2: Summary of Research into Gender Differences in Language Performance**

Data source Subject, participants	Male students		Female students		Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
China Education Panel Survey, 2013 (representative of the population of middle schoolers)					
English, Grades 7 and 9	72.86	31.39	86.20	27.53	.45
Dewaele, 2007 <sup>a</sup> (Belgium)					

Dutch, Grade 12	/	/	/	/	.47
French, Grade 12	/	/	/	/	.34
English, Grade 12	/	/	/	/	.24
Goetz, Fenzel, Hall and Pekrun, 2008 (Germany)					
German, Grades 5 to 10	3.78	.85	4.16	.83	.32
Huang, 2011 (mainland China)					
English, undergraduates	365.22	41.52	387.23	42.99	.52
Maqsud, 2006 (South Africa)					
Setswana, Afrikaans, English, Grade 7	367.05	95.11	352.03	85.82	– .17
National Assessment of Educational Progress (the USA, national sample)					
Reading, Grade 4, 2019 cohort	217	39	224	37	.18
Reading, Grade 8, 2019 cohort	258	38	269	36	.30
Reading, Grade 12, 2015 cohort	282	41	292	39	.25
Writing, Grade 4, 2002 cohort	146	35	163	35	.49
Writing, Grade 8, 2011 cohort	140	34	160	33	.60
Writing, Grade 12, 2011 cohort	143	36	157	33	.41
Program for International Student Assessment (B-S-J-G <sup>b</sup> , China)					
Reading, 15-year-olds, 2015 cohort	486	108	503	109	.16
Program for International Student Assessment (OECD average)					
Reading, 15-year-olds, 2018 cohort	472	102	502	94	.31
Reading, 15-year-olds, 2015 cohort	477	98	504	92	.28
Reading, 15-year-olds, 2012 cohort	474	96	512	87	.41
Reading, 15-year-olds, 2009 cohort	471	94	510	86	.43
Reading, 15-year-olds, 2006 cohort	466	101	505	92	.40
Reading, 15-year-olds, 2003 cohort	477	98	512	90	.37
Reading, 15-year-olds, 2000 cohort	478	98	510	91	.34
Seginer and Vermulst, 2002 <sup>c</sup> (Arab and Jewish)					
English, Arabs in Grade 8	/	/	/	/	.13
English, Jews in Grade 8	/	/	/	/	– .18
Steinmayr and Spinath, 2008 (Germany)					
German, Grades 11 and 12	– .17	1.02	.11	.97	.28

Notes. a. Means and standard deviations were not available, so the effect sizes were estimated via  $F$ -values and sample sizes, where Cohen's  $d = \sqrt{(F(n_1 + n_2 / n_1 n_2) / (n_1 + n_2 / n_1 + n_2 - 2))}$ . b. B-S-J-G (China) refers to the four participating provinces in mainland China: Beijing, Shanghai, Jiangsu, and

Guangdong. c. Descriptive statistics were not provided by the authors, so Cohen's  $d$ s were calculated using  $t$ -values and  $dfs$  (degree of freedom), where Cohen's  $d = 2t / \sqrt{df}$ .

With regards to the affect component, some, but not as many, scholars have discovered that girls are more motivated to learn foreign languages than boys (e.g., Carr and Pauwels, 2006; Williams et al., 2002). Some others also disclose that girls reportedly enjoy themselves more than boys in language courses (Dewaele, MacIntyre, Boudreau, and Dewaele, 2016). Thus, there seems to be a kernel of truth underlying the affect component of FALS, which describes girls' affective relationship with languages more positively than that of boys.

However, to interpret all the above-mentioned research as validating FALS's authenticity can be simplistic, considering a) the socio-culturally sensitive relationship between gender and linguistic development (e.g., Halpern, 2012), b) the likelihood that gender intersects and interacts with other identities in schools (e.g., Francis and Paechter, 2015), and c) the self-fulfilling nature of academic GSs. To determine FALS's accuracy, indeed, requires a close and thorough examination of stereotype endorsement in a specific language-learning context, *and* a search for the plausible self-fulfilling effects of FALS. Both tasks are what the current project aims to approach.

### **3.2.3 Differences in Stereotype Endorsement Due to Demographic Background**

As prevalent as academic GSs are, people do not homogeneously endorse them to the same extent. For instance, as Section 3.2.2.1 has pointed out, there is evidence that language teachers have stronger gender-reading stereotypes than parents do (Muntoni and Retelsdorf, 2019; Retelsdorf et al., 2015). Another related dimension, cultural variation or regional specificity, might also affect stereotype endorsement (see Section 3.1.4.3), although no studies have investigated the possibility regarding academic GSs.

Two other demographic features, gender and age, have received some attention from the academia. Table 3.3 here summarises the studies reporting gender and/or age differences in respect to gender-language stereotypes. The findings concerning gender differences appear inconsistent: boys express stronger stereotypical beliefs in one study (Martinot, Bages, and Desert, 2012), the opposite is claimed by some more (e.g., Epting, Rand, and D'Antuono, 2014; Nosek, Banaji, and Greenwald, 2002), and some others find no substantial difference between male and female participants' GSs (e.g., Guimond and Roussel, 2001; Heyman and Legare, 2004). Such complexity is expected, because the same baffling situation exists with male-STEM stereotypes. Sometimes, boys seem more convinced of a male advantage in math (e.g., Steffens and Jelenec, 2011; Tomasetto et al., 2015). Evidence indicating that girls were

**Table 3.3: Summary of Studies on Gender and Age Differences in Endorsement of Language-Related Gender Stereotypes**

Study			Participants, findings, and differences in stereotype endorsement due to gender and age
Epting et al., 2014	Participants	Undergraduates from two American universities, mostly Caucasian ( $n = 377$ )	
	Major findings	Four literacy activities (reading, writing, revision, and grammar) were rated feminine/girlish	
	Gender difference	Girls showing stronger stereotypes for all four activities	
Georgiou, 2008 <sup>a</sup>	Participants	Elementary school teachers ( $n = 154$ ) and pre-service teachers ( $n = 159$ ) in Greece	
	Major findings	Girls were estimated as doing better in languages <i>while</i> boys better in math (Achievement)	
	Gender difference	No statistically significant difference	
	Age disparity	Experienced teachers holding a stronger stereotype	
Guimond and Roussel, 2001	Participants	Female high school students ( $n = 61$ ), psychology ( $n = 463$ ) and science undergraduates in France ( $n = 80$ )	
	Major findings	Women were considered more gifted in language by all three groups of participants (Aptitude)	
	Gender difference	No statistically significant difference in all three groups of participants	
Heyman and Legare, 2004	Participants	Kindergartners and 1 <sup>st</sup> graders ( $n = 60$ ), and 4 <sup>th</sup> and 5 <sup>th</sup> graders ( $n = 60$ ) from southwestern United States	
	Major findings	Girls were thought to have stronger spelling ability in both implicit and explicit measures (Aptitude)	
	Gender difference	Girls showing a stronger stereotype for explicit measure only	
	Age disparity	Older children holding a stronger stereotype for implicit measure only	
Kurtz-Costes et al., 2014	Participants	4 <sup>th</sup> ( $n = 187$ ), 6 <sup>th</sup> ( $n = 141$ ) and 8 <sup>th</sup> ( $n = 135$ ) graders from south-eastern USA	
	Major findings	Girls were rated better readers/writers than boys by three groups of participants (Achievement)	
	Gender difference	Girls with a stronger stereotype	
	Age disparity	Older students holding a stronger stereotype	
Martinot et al., 2012 <sup>b</sup>	Participants	5 <sup>th</sup> graders from France, mostly Caucasians ( $n = 398$ )	
	Major findings	Girls and women more often chosen as most competent readers than boys or men. (Aptitude)	

	Gender difference	Boys with a stronger stereotype associating adult women with reading abilities
Nosek et al., 2002 <sup>a</sup>	Participants	Respondents taking the Implicit Association Test, non-preselected or targeted ( $n = 73,117$ )
	Major findings	Females were more often associated with liberal arts, but males with science (implicit & explicit measures)
	Gender difference	Female participants showing a slightly stronger gender-domain stereotype
Peacock, 2001 <sup>b</sup>	Participants	Trainee teachers of English from a university in Hong Kong ( $n = 146$ )
	Major findings	More participants regarded women as better learners of foreign languages (Aptitude)
	Age disparity	2 <sup>nd</sup> -year trainees with the strongest stereotype, and 3 <sup>rd</sup> -years with the weakest one
Plante et al., 2009	Participants	French-speaking students in Grades 6, 8, and 10 from Canada ( $n = 984$ )
	Major findings	Girls were thought to have stronger abilities and enjoyment in languages (Aptitude and affect)
	Gender difference	Females having stronger stereotypes
	Age disparity	No statistically significant difference (though the 10 <sup>th</sup> graders seemed to have stronger stereotypes than the 6 <sup>th</sup> graders judging from descriptive statistics)
Zhao et al., 2014	Participants	Students from middle schools in north-eastern China ( $n = 131$ ), their parents ( $n = 131$ ), and their teachers ( $n = 52$ )
	Major findings	Girls were rated more gifted and higher-achieving language learners by of parents and teachers, and better essay writers by students (Aptitude and achievement)
	Gender difference	Female students having a stronger stereotype about essay writing; female teachers having a stronger stereotype about verbal abilities

*Notes.* a. These two studies did not study female-language/literacy stereotypes in isolation; instead, they studied gender-domain stereotypes, which contained, simultaneously, male-STEM and female-language/literacy stereotypes. b. These two studies did not examine gender or age differences in stereotype endorsement; the results reported here were analysed by me using descriptive statistics provided by the original authors.



the more stereotypical group is also presented by some researchers (e.g., Yee and Eccles, 1992; Liu, et al., 2010), so is there a claim of a lack of gender differences (e.g. Whitehead, 1996; Guimond and Roussel, 2001). One reason for the conflicting findings might be confounding variables intersecting with gender, such as culture, and the specific domain under inspection. Thus, mono-cultural studies focusing on one subject or domain are needed to clarify gender differences in stereotype endorsement.

Variations in stereotype endorsement due to age, on the contrary, exhibits a persistent and reliable pattern, as Table 3.3 here illustrates. As learners proceed to higher grades or higher levels of education, they seem to have developed greater and greater tendencies of gender stereotyping. This seems reasonable, considering that GSs are grounded in people's observation of the social world. Since language classrooms and the language teaching profession are predominantly occupied and staffed by females (e.g., Carr and Pauwels, 2006), students naturally assume a stereotypical connection between language domains and the female gender. In addition, GSs can filter social perception and impact social behaviours, making the disconfirming of them even harder as people grow up (Sections 3.1.4.1-3.1.4.2).

### **3.2.4 Stereotype Threat Effect on Negatively Stereotyped Learners**

In Section 3.1, when entertaining how GSs generally form and what features they have, impacts of GSs on social cognition, as well as social behaviour, have been alluded to. However, with regards to academic GSs, their specific influences on learners have been identified. The impact on negatively stereotyped learners, stereotype threat (ST) effect, will take the foreground in this section.

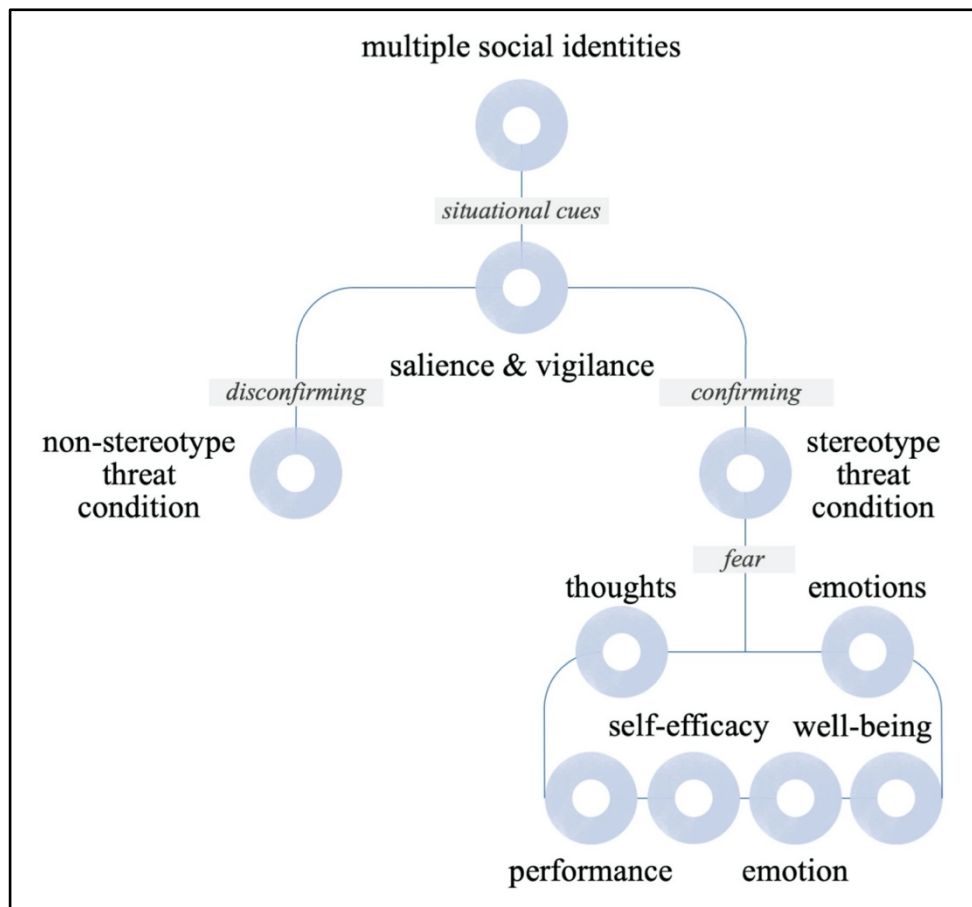
When situational cues raise the possibility that individuals might be judged or treated on the basis of their social identity, apprehensions of confirming the unfavourable stereotypes about their social group arise, negatively affecting individuals' behaviour, self-cognition, emotional status and psychological wellbeing. This process is called ST effect (e.g., Appel and Kronberger, 2012; Murphy and Taylor, 2012; Spencer, Steele and Quinn, 1999; Steele, Spencer and Aronson, 2002). ST research was initially designed to understand why ethnic minorities and women students underperform in STEM areas, especially when their academic aptitudes (assessed by SAT scores) have indicated otherwise (e.g., Steele and Aronson, 1995; Steele et al., 2002). Prior to the ST approach, academic underperformance is either attributed to biology or upbringing (Inzlicht and Schmader, 2012). ST's unique contribution, therefore, is to side-step the nature vs. nurture cliché, pointing out that the learning and/or assessment environment itself might have led to performance decrements.

Since its onset in the 1990s, the burgeoning ST research has been conducted across a wide range of contexts, yielding reliable and robust findings. For instance, ST effect has been found among African–American students sitting math tests (e.g. Steele and Aronson, 1995), senior citizens performing physical and cognitive tasks (e.g. Stone, 2002), as well as women taking selective STEM subjects (e.g. Lamont, Seift, and Abrams, 2015). In addition, ST effect remains robust across contexts (e.g., Schmader and Beilock, 2012; Appel and Kronberger, 2012; Nguyen and Ryan, 2008). Because of the scarcity of studies concerning ST effect of language-gender stereotypes, the current review will draw on findings from ethnic minorities and women in STEM to illustrate the mechanisms and consequences of ST.

### 3.2.4.1 Mechanisms of Stereotype Threat Effect

Figure 3.3 here paints how seemingly innocuous situational cues can bring out ST effect among adversely stereotyped individuals.

**Figure 3.3: Situational Cues, Vigilance Process, and Stereotype Threat**



As Murphy and Taylor (2012) have explained, individuals simultaneously hold multiple social identities, for example: gender, age, race or ethnicity. Sometimes, in a given situation,

certain signals make a particular identity more salient than others, and people's attention is directed to other situational cues to decide whether that identity may be a liability. That is, whether this particular social identity will become a source of devaluation, stigma or mistreatment (Spencer, Logel, and Davies, 2016). If the situation disconfirms the worry, people will relax and focus on the tasks at hand. But if the worry is confirmed, people will experience ST: they fear that they may confirm the negative stereotypes about their social group, and thus end up exerting extra effort into inhibiting troubling thoughts and regulating negative emotions (Inzlicht, McKay, and Aronson, 2006; Johns, Inzlicht, and Schmader, 2008).

The situational cues differ with regards to salience. In fact, ST research has identified that subtle, moderately explicit, and blatant cues can all activate ST effect among different target groups on different tasks. Table 3.4 here summarises the common cues utilised by ST research to activate negative stereotypes.

**Table 3.4: Summary of Stereotype-Activating Cues in Research on Stereotype Threat Effect**

Explicitness		Description	Examples
<b>Subtle cues</b>	Content	Indirectly convey group differences	<ul style="list-style-type: none"> <li>• Use gender ratio of participants or experimenter gender to make gender salient (Steele and Aronson, 1995)</li> <li>• Describe the test as diagnostic or evaluative (Marx and Stapel, 2006)</li> </ul>
	Timing	Prior to or during a test (in test instructions)	
	Result	Stereotype activated subconsciously	
<b>Moderately explicit cues</b>	Content	Directly state group differences	<ul style="list-style-type: none"> <li>• State that men and women usually perform differently in the area (Brown and Pinel, 2003)</li> <li>• State that men and women usually perform differently in this test (Keller and Dauenheimer, 2003)</li> </ul>
	Timing	Prior to or during a test (in test instructions)	
	Result	Stereotype activated consciously	
<b>Blatant cues</b>	Content	Directly state the stereotype about inferiority	<ul style="list-style-type: none"> <li>• State that men have been found to score higher in this test than women (Aronson et al, 1999)</li> <li>• Explain ST and ask participants to suppress associated thoughts (McGlone and Aronson, 2007)</li> </ul>
	Timing	Prior to a test (in test instructions)	
	Result	Stereotype activated consciously	

Tsui and Venator (2008) employed blatant cues to activate language-gender stereotypes among Chinese college students, but failed to find ST effect. However, the study recruited only 23 participants, which has limited statistical power. In fact, because no manipulation check was carried out, it was likely that the cues did not successfully activate stereotypes at all. According to meta-analytic works, effect sizes (Cohen's *ds*) of ST typically range between  $|.36|$  and  $|.45|$  (e.g., Nguyen and Ryan, 2008). Thus, larger sample sizes and effective cues, preferably being tried in a pilot study, are needed in order to detect plausible ST effects among language learners.

### **3.2.4.2 Consequences of Stereotype Threat Effect**

Originally, ST studies emphasised on performance decrements caused by negative stereotypes across task and test types on members of various social groups (e.g., Aronson et al., 1999; Lamont et al., 2015; Stone, 2002). Gradually, more and more consequences beyond performance impairments are located, including diminishing interest and thoughts of quit (e.g., Flore and Wicherts, 2015), lowered career aspirations (e.g., Davies, et al., 2002), decrease in self-confidence (e.g., Muzzatti and Agnoli, 2007), and long-term adverse impact on psychological wellbeing (e.g., Spencer et al., 2016). Yet, experiments appear to be the predominating data collection methods, which might risk ecological validity: in real life, academic GSs are ubiquitous, can take different forms, and may come from multiple sources; their influences, therefore, will be exerted on learners through an extended time frame and in an additive manner. Furthermore, idiosyncrasies in the interpretation of gender-stereotypical information, as well as nuanced responses to others' GSs, tend to be overlooked when experimental designs are applied.

### **3.2.5 Stereotype Lift and Boost on Non-Negatively Stereotyped Learners**

Given that ST effect unfolds the experience of individuals being adversely stereotyped, some scholars naturally probe into the side of those not targeted by the stereotypes in the same situation. Walton and Cohen (2003), in their meta-analytic review, have revealed that when individuals are exposed to negative stereotypes about an outgroup, they will engage in downward comparison with members of the outgroup, leading to elevations in self-efficacy, motivation, and performance. This phenomenon is termed stereotype lift, and has been applied to understand performance enhancements in an intellectual test by French natives in the presence of African immigrants (Chatard, Selimbegović, Konan, and Mugny, 2008), in a math test by men in the company of women (Johnson, Barnard-Brak, Saxon, and Johnson,

2012), in a motor task by men with women at their side (Chalabaev, Stone, Sarrazin, and Croizet, 2008), and etc.

A related phenomenon, stereotype boost, refers to improvement in task accomplishment after being reminded of positive stereotypes about one's own ingroup (Shih, Pittinsky, and Ambady, 1999). For example, the elderly are found to experience ST and then memory decrements after being primed of unfavourable age-related expressions, such as 'senile' and "Alzheimer's"; but when the words are positive ('learnt' and 'advise', for instance), the senior participants will perform better in memory tasks (Levy, 1996). Stereotype boost has also been replicated across situations—Asian Americans sitting math exams (Armenta, 2010), biracial Black-White students taking verbal ability tests (Gaither, Remedios, Schultz, and Sommers, 2015), and etc.

Although both stereotype lift and boost indicate positive effects of stereotypes on individuals, there is research suggesting otherwise. After all, the majority of studies on lifting and boosting are conducted without observer pressure, which is unlikely in the real world, where individuals almost always face attention and expectation from surrounding people. Such observation and anticipation by others might become sources of pressure, and lead to performance decrement, even with the presence of favourable stereotypes (Krendl, Gainsburg, and Ambady, 2012). Another factor further complicating the issue is that academic GSs tend to be complementary, depicting girls as better language learners and boys as less competent ones. Experimental methods might be able to approach ST, stereotype lift, and stereotype boost in isolation, but in reality, these effects might arise in the same classroom.

### **3.3 Limitations in Previous Work and Proposed Contributions of Current Project**

So far, a selective account of pertinent scholarly inquiry into GSs has been presented, from their formation and circulation in general to their specific roles in education. The majority of academic sources are empirical papers investigating gender (role) attitudes and/or stereotypes published between the 1970s and the 2010s. From time to time, reviews (narrative, systematic, and meta-analytic types) on gender studies are cited, especially when comparison or synthesis across diverse socio-cultural, social-temporal, and socio-spatial contexts is made: for example, on p. 18, Swim (1994) and Guimond (2008) were consulted to argue for the kernel of truth contained in cross-cultural GSs. Particularly, raw data from large-scale surveys and assessment programmes were utilised in Section 3.2.25 to identify the directions and magnitudes of gender differences in language aptitude and performance across

demographic groups. To avoid being restrictive or negligent, the author has actively sought for studies conducted in a number of language-learning contexts in different countries, such as literacy lessons (e.g. in Germany, Goetz et al., 2008; in China, Zhao et al., 2014), second or foreign language courses (e.g., in Belgium, Dewaele, 2007; in the UK, Taylor and Marsden, 2014), and more specifically EFL curriculums (e.g., in South Africa, Masqud, 2006; in China, Zhao, 2011).

However, despite the researcher's efforts to be inclusive, this appraisal of GS literature has two limitations. First, although most empirical research cited is generally up-to-date (published in 2000 and onwards), some relatively dated studies (published in the 1980s-1990s) are relied upon when showcasing parents' GSs (the second paragraph on p. 19) and summarising genuine gender differences in linguistic potentials (Table 3.1 on p. 31-2). Acknowledging that psychological research can be social-temporally specific, it is likely that findings from these papers might not reflect current thinking on gender or cognitive/developmental trends. Yet, on the grounds that 1) these studies are seminal and 2) GSs tend to be durable, the author decided to include such older research for the sake of completeness and thoroughness.

Secondly, more space is devoted to trends and patterns reported by observational and experimental studies, while only some attention is paid to individual perspectives and situated descriptions uncovered via qualitative approaches. This imbalance, in fact, also points to a methodological gap in the literature itself. Previous works usually follow the quantitative tradition, adopting questionnaires and experiments to approach the content and effect of academic GSs. The inevitable result is an overlook of the plausible idiosyncrasies in the interpretation of different dimensions of the same gender-domain stereotype, the existence of possible gender-neutral or even counter-stereotypical attitudes, and more nuanced responses to GSs beyond performance decrements. This is partly the reason that this doctoral project has decided to use a mixed-method design (see Section 4.1).

Furthermore, the existing literature on academic GSs, though extensive and marked with intellectual rigour, leaves three theoretical gaps. To begin with, compared to the amount of scholarly inquiries directed at male-STEM stereotypes, academic exploration into the female-qualitative stereotypes is inadequately developed. The language-gender stereotypes, furthermore, is under-researched. Few studies approach the two academic GSs simultaneously using explicit self-report questionnaires (e.g., Plante et al., 2009), and some more relied on implicit methods (e.g., Nosek, et al., 2002). The latter, however, is more

instrumental in the measurement of stereotype magnitude than the theoretical understanding of the nature of female-qualitative stereotypes. Besides lack of examination of the content of language-gender stereotypes, in particular, efforts to unearth how such stereotypes affect language learners from ST perspective is extremely scarce. Therefore, the current project will contribute to academic literature by uncovering the content and the effect of GSs targeting language learners.

Additionally, previous works tend to fixate on learners' views; in contrast, parents' and teachers' opinions are not fully investigated. Moreover, very few studies have administered the same instrument to more than one participant group, let alone making any cross-group comparison. The only exception is two papers by the same group of German researchers, who inspected teachers' and parents' gender-reading stereotypes using the same scale (Muntoni and Retelsdorf, 2019; Retelsdorf et al., 2015), but still, the researchers did not directly compare the amplitudes of stereotype endorsement of the two groups. A related problem regarding participants is the absence of cross-cultural contributions that reach beyond western, industrialised samples. Considering the socio-culturally sensitive relationship between gender identity development, linguistic development, and stereotype endorsement (e.g., Becker and Sibley, 2016; Halpern, 2012; Williams et al., 1999), to fully understand the female stereotyping of languages in a specific language-learning context is necessary.

The third issue needs addressing is the overemphasis on competence-based stereotypes, even though the literature suggests that stereotypical ideas associating the female gender with language domains might encompass three dimensions: aptitude, achievement, and affect (Section 3.2.2). The likelihood that the three components differ in strength, as a result, is hardly ever inspected. Hence, FALS, the trio-component construct, is proposed, depicting females as more gifted, higher-achieving, and more enthusiastic learners of language.

Combining the last two gaps discussed just now, this doctoral project aims to study the opinions of all three groups (students, parents, and teachers) in China, a country with a large population of EFL learners and a dearth of gender-related research on the English curriculum. Given that parents, unlike students and teachers, are not intensely immersed or actively engaged in EFL contexts, their stereotypes might be less nuanced. For them, FALS might only exist on a macro, more general level. In comparison, students and teachers might construct FALS along a micro level in addition to the macro level. To be more specific, the micro level will include four skills (listening, speaking, reading, and writing) and two types

of knowledge (grammar and vocabulary), the six scopes covered by EFL education in China (see Section 2.3 for more details).

### **3.4 Research Questions**

Based on the review of previous literature, as well as the gaps identified in the last section, two research questions are raised:

- 1) Do students, their guardians<sup>1</sup>, and their teachers of English endorse FALS in Chinese senior high schools? How do they interpret each FALS components?
- 2) Does FALS affect students' test performance? What are their responses to FALS?

---

<sup>1</sup> Guardians, instead of parents, are included here, because in China, especially in the relatively rural, underdeveloped regions, children might live with other relatives (typically grandparents) rather than their parents, who tend to work away in larger cities in order to earn higher income for the family. Since the project aims to recruit participants from diverse backgrounds, guardians are invited here to avoid excluding children living away from their parents from participating the research.



## **Chapter 4: Methodology**

In this chapter, I will describe how the current two-phase project was conceptualised and operationalised. Starting with a brief explanation of the entire project in Section 4.1, the two phases, each consisting of two studies, will be introduced in Sections 4.2-4.3. All four studies will be described in full, including their corresponding data collection techniques, sampling processes, pilot studies, data analysis procedures, and ethical considerations. As each phase was a mixed-methods (MM) design on its own, data integration in both will also be discussed towards the end of their corresponding sub-sections.

### **4.1 Research Design**

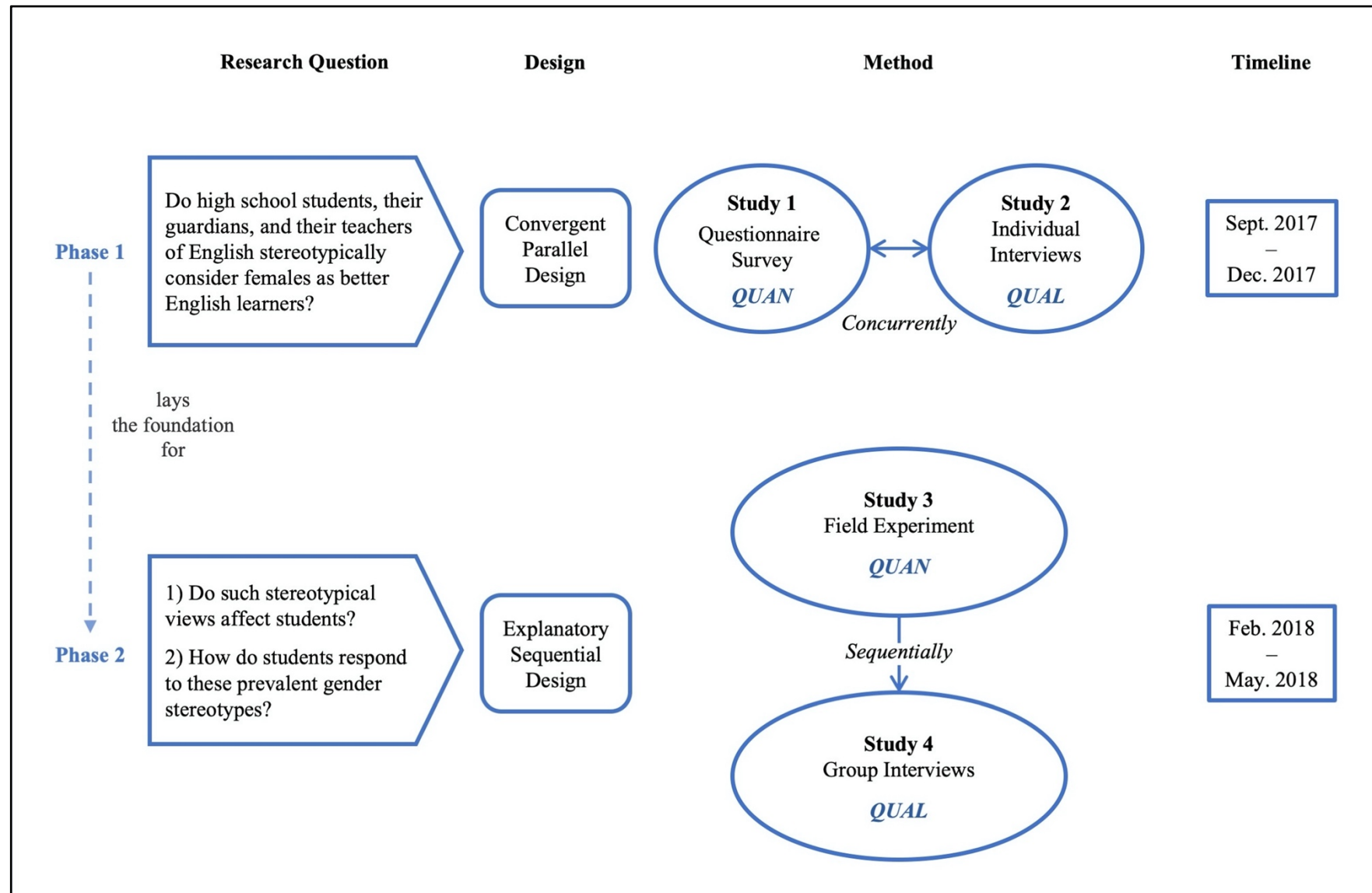
To address the proposed research questions (RQs) (see Section 3.4), this project adopted a MM design comprising two phases: a convergent parallel phase, and an explanatory sequential one. After an overview of the project, I will argue for certain methodological decisions behind the project by expounding my epistemological stance.

#### **4.1.1 Two-Phase Mixed Methods Research**

The top half of Figure 4.1 shows Phase 1, which was devoted to capturing the existence and the composition of the female-advantage-in-languages stereotype (FALS). Two studies were carried out simultaneously from September to December in 2017: Study 1 was a questionnaire survey, which gathered quantitative data from students, their guardians, and their teachers of English from eight high schools; Study 2, undertaken synchronously with Study 1, was an interview-based survey. The interviewees were also teachers of English from the same schools, though they did not participate in Study 1. The underpinning epistemological and paradigmatic considerations of choosing two methods to collect both numeric and textual data will be clarified in Section 4.1.2 later. Besides, the reasons that only teachers, instead of all three types of participants in Study 1, joined Study 2 will be presented in Section 4.2, as well as Section 4.2.2. Finally, Section 4.2.3 will reveal how results from two studies were combined, compared, and contrasted to achieve a more comprehensive understanding of gender stereotypes (GSs) targeting language learners.

Data from Phase 1 went through an initial analysis, and the results indicated that FALS was harboured by all three groups of participants on the macro level. This laid the foundation for Phase 2, which aimed at examining how FALS might have affected high school students (see the bottom half of Figure 4.1). Unlike Phase 1, this phase consisted of two studies conducted sequentially from February to May in 2018. Study 3, a field experiment, investigated whether FALS would affect students' performance in a standardised

Figure 4.1: Overview of Two-Phase Mixed-Methods Research Project



comprehensive English test. The quantitative data were analysed and followed up with Study 4, a qualitative study utilising group interviews to unearth how FALS might have permeated in schools and how learners might have reacted to it. After documenting the details of two studies in Sections 4.3.1-2, Section 4.3.3 will report how findings from both were integrated.

#### **4.1.2 Epistemological Underpinnings and Methodological Considerations**

Like most MM researchers (Johnson, Onwuegbuzie, and Turner, 2007), I endeavoured to broaden and deepen academic understanding of GSs by integrating quantitative and qualitative traditions in my research. This integration operationalised at epistemological, methodological, as well as analytical levels throughout this doctoral project.

My approach to educational research is primarily informed by pragmatism, a paradigm prioritising the research's practical need over the researcher's metaphysical allegiance (Teddle and Tashakkori, 2009). Paradigms are defined as sets of philosophical assumptions concerning the nature of reality (ontology) and knowledge (epistemology), as well as the way that knowledge is produced (methodology) (Guba and Lincoln, 2005). Paradigmatic pluralism has been witnessed in social and behavioural sciences (Shannon-Baker, 2016), where two broad, opposing traditions reside: positivism and constructivism (Creswell, 2014). Despite certain disagreements and variants, adherents of the former generally view reality as independent of consciousness, consider knowledge to be observable, quantifiable, and generalisable, and aspire to remain objective and separate from their research object (Cohen, Manion and Morrison, 2011). Scholars in the constructivist camp, on the other hand, typically see reality as socially constructed, regard knowledge as meanings and interpretations generated via human agency on social levels, and acknowledge their (inter-)subjectivity in the research process (Bryman, 2004). Paradigmatic affiliation, cautions Morgan (2014), can lead to methodological exclusivity, i.e. an either-or situation between quantitative and qualitative approaches.

Alternatively, a growing group of researchers, me myself included, attempt to integrate quantitative and qualitative approaches in our academic inquiry, because we appreciate similarities shared by the approaches and believe research to be outcome-oriented and practice-driven. Positivism and constructivism are incommensurate as they embrace contradictory worldviews and lead to divergent methodological choices, but they both aim to describe and explain phenomena by collecting observational data, and seek corroboration and elaboration when multiple data sources are gathered (Johnson et al., 2007). These commonalities make a case for methodologically eclectic research. Furthermore, we

pragmatists focus on which approach, or approaches, answers the research question more adequately and comprehensively; metaphysical concerns, as a result, give way to methodological pragmatism.

Applied to the current research, this pragmatic perspective has prompted me to adopt MM designs for their capacity to address an array of confirmatory and exploratory research questions in a single project (Creswell, 2014). Traditionally, as Teddlie and Tashakkori (2009) have pointed out, collecting and analysing quantitative data confirms how likely a hypothesis will stand, as in the case of questionnaire surveys; accumulating and interpreting qualitative data, on the other hand, explores the terrain of the phenomenon under investigation, as in the case of in-depth interviews. In the case of the current project, nevertheless, both hypothesis testing and phenomenon inspecting were its goals. The purpose of Phase 1 was to investigate GSs targeting language learners, a purpose simultaneously confirmatory and exploratory: While quantitative data collected from questionnaires could measure to what extent participants endorsed each component of FALS that I had assumed to exist, qualitative data from interviews could capture how individuals had interpreted each component, how they had perceived the relationships among the three components, as well as uncover other circulating GSs accompanying FALS. Next, in Phase 2, the primary aim was to detect the plausible impacts of FALS. As Spencer, Logel, and Davies (2016) have summarised, experimental designs have managed to find that priming negative stereotypes consistently and reliably undermines targets' performances in related domains in previous research. Thus, in Study 3, to provide evidence for the causal link between FALS activation and performance decrements, an experiment would of course be the reasonable choice. Then, the subsequent group interviews in Study 4 were implemented to help understand the broader impacts of FALS. Therefore, for such an exploratory plus confirmatory project, a multi-phase MM design was chosen.

My pragmatist perspective underpins three methodological issues of designing and data analysing. The first was the sequence of two phases. In this project, Phase 1 laid the ground for Phase 2, and the latter reinforced the pragmatic impact of the former: Only after Phase 1 had captured the existence of FALS, at least on the macro level, could Phase 2 have proceeded to examine its plausible effects. Subsequently, because Phase 2 was able to detect how FALS had negatively affected learners' academic performance and psychological wellbeing (see details of findings in Chapter 8), it actually highlighted the real-life significance of Phase 1. Linked together, the two phases speak to one common objective: to

raise awareness of establishing a gender-fair English-learning environment in Chinese high schools.

The second design issue was the relative weights assigned to quantitative and qualitative strands in MM designs (Morse, 2010). As advised by MM experts, my decision was guided by the purpose of the project and research questions (Creswell, 2014; Johnson, Onwuegbuzie, & Turner, 2007; Teddlie & Tashakkori, 2009). Phase 1 consisted of two concurrent studies collecting quantitative and qualitative data respectively. It seemed sensible to assign equal weights to these two studies because they gathered similarly large amounts of data and required equivalently complex analyses. In Study 1, for the sake of generalisability and ecological validity, I collected data in different high schools from a relatively large sample (see Sections 5.1.2, 5.2.2, and 6.1 for detailed descriptions of the participants' demographic background). To discover the general trends and group differences in FALS endorsement, I ran a series of mixed-design factorial MANOVAs. Likewise, Study 2 demanded considerable time and effort, since the goal was to explore gendered patterns among teachers' descriptions about and comments on learners. For the sake of credibility, I interviewed 20 English teachers within 12 weeks (as described in Section 4.2.2). To uncover accurate and reliable descriptions of GSs, I utilised various strategies in data analysis, including member checking and triangulation. Thus, the extensive procedures to compile and examine data in both parts call for my equal attention. In Phase 2, however, the priority was given to Study 3 during fieldwork period, because it called for more energy to organise than Study 4, considering that there were over four hundred participants (see details in Section 4.3.1). Then, during data analyses, more time and attention was devoted to Study 4, where transcription and multiple coding cycles were carried out.

The final methodological consideration was about data analysis. MM research was selected as it enabled me to draw on multiple data sources and analysis procedures to draw stronger inferences. MM research provides opportunities for triangulation (Johnson and Onwuegbuzie, 2004), the search for corroboration of one data source from another. This was particularly pronounced in Phase 1, where results from the questionnaires informed me of the extent to which people generally endorsed the FALS; meanwhile, with idiosyncrasies reported by interviewees, I also uncovered GSs prescribing gender differences alongside the education-profession continuum (as expounded in Section 7.3). No doubt this exploration of supplementary GSs through interviews, which supposedly fuelled or perpetuated FALS, will further substantiate the finding of FALS in high schools via questionnaires.

Aside from triangulation, another way by which MM research enables stronger inferences is to encourage alternative perspectives. By contrasting the two data sources in Phase 1, I gathered perspectives contrasting FALS (see relevant themes reported in Section 7.3), which would have been obscured if only questionnaires had been utilised. The contradictions between data sources have completed the picture of GSs concerning language learners, thus enabling more informative inferences. Therefore, for the current project, collecting both quantitative and qualitative data helped acquire a deeper understanding of a multifaceted and contextualised construct like FALS.

#### **4.2 Phase 1: Capturing Existence of Female-Advantage-in-Languages Stereotype**

Phase 1 was a convergent parallel MM design made of two concurrent studies: 1) a questionnaire-based survey, and 2) an interview-based survey. Study 1 investigated whether students, their guardians and teachers of English perceived males and females differently regarding affect, achievement, and aptitude in relation to the English subject. Study 2, meanwhile, explored gendered patterns in English teachers' descriptions about their students.

The reason that the participants in two studies varied were two-fold. First, as Section 3.3 has suggested, FALS harboured by students and teachers might have two levels and be more nuanced compared to other people (e.g., guardians) because of their active engagement with English education. Therefore, it was necessary to inspect their opinions. Involving guardians in the study, meanwhile, provided opportunities to 1) probe into the viewpoints of those with a less close relationship with the English subject, and 2) compare the magnitudes of FALS endorsed by different groups.

Second, among the three groups, teachers may have the most extensive and in-depth experience with FALS. After all, they were once English learners, they have been confronted by others holding FALS, and their job may have been propelling them to ponder FALS on a daily basis. Thus, in order to fully appreciate the complexity and dynamics of their perspectives, I utilised semi-structured interviews alongside questionnaires when investigating teachers' FALS endorsement. Some might argue that the same would go for students, especially high-achieving male students. This is certainly probable. However, given that I have conducted a phenomenological inquiry into this group in my MPhil project, and that the MPhil project actually was a precursor of this doctoral project, I thought it less imperative to repeat the process.

Section 4.2.1 below presents the details of Study 1, which includes the design, implementation, and analysis procedures<sup>2</sup>. Following this is Section 4.2.2, where corresponding details of Study 2 are explained. Finally, Section 4.2.3 introduces the integration of findings from the teacher datasets in two studies.

#### **4.2.1 Study 1: Questionnaire-Based Survey**

In Study 1, I used two questionnaires to measure FALS: one for students, and the other for their guardians and teachers of English. Section 4.2.1.1 gives reasons for using questionnaires as data collection method. Then, Section 4.2.1.2 explains how the questionnaires were designed and revised in order to improve face validity and content validity. Section 4.2.1.3 describes Pilot Study 1, which aimed at examining the instruments' construct validity and reliability. Finally, Sections 4.2.1.4-7 clarifies the sampling, distribution, analysis, and ethical issues with Study 1.

##### **4.2.1.1 Questionnaires as Data Collection Methods**

GS researchers have traditionally used self-report methods to investigate popular beliefs about and attitudes towards gender (e.g., Cejka and Eagly, 1999; Haines, Deaux, and Lofaro, 2016). In recent years, however, scholars advocate incorporating explicit measures with implicit ones (Greenwald, Poehlman, Uhlmann, and Banaji, 2009) to assess the magnitude of stereotypes. Common implicit measures include semantic priming (e.g. Fazio, Jackson, Dunton, and Williams, 1995) and the Implicit Association Test (IAT; see Greenwald, McGhee, and Schwartz, 1998), both using response latencies to indicate the strength of the association between two gender categories and particular attributes. Compared to the explicit ones, implicit measures are most revealing when explicit self-report methods 1) fail to discriminate between stereotype awareness and stereotype endorsement; 2) neglect idiosyncrasies in gender-stereotypic perceptions; and 3) engender intentional measurement errors caused by social desirability concerns on the participants' side. For Study 1, a direct questionnaire-based survey would be more appropriate, as the research design should have overcome the first two limitations: To ensure that responses reflected genuine beliefs about FALS instead of a mere acknowledgement of FALS existence, participants were instructed at the beginning of the questionnaire to select the options that best reflected their own attitudes. Besides, each item in the questionnaire was worded 'I believe/think/suppose/am convinced

---

<sup>2</sup> Demographic characteristics of the four groups of participants in Study 1 (guardians, students, teachers) and 2 (teachers) will be presented in relevant sections in the results chapters, because there is no space in the current chapter on methodology to do justice to the topic. Besides, demographic characteristics provided contexts for the interpretation of certain findings.

that...’, which also straightforward elicited participants’ personal opinions. Secondly, to avoid neglecting idiosyncrasies, results from the questionnaire dataset were compared to those from interviews in Study 2.

Finally, the third limitation seems out of context in China. As we have seen in Chapter 2, gender differentiation is commonplace in today’s Chinese society. Besides, both implicit and explicit methods have already been used to establish a strong stereotypical association of women with Liberal Arts and men with Science across the globe (reported earlier in Section 3.2.1). In fact, China ranks top in Nosek and colleagues IAT project measuring the magnitude of gender-subject stereotypes among 61 countries (Nosek *et al*, 2009). Similar evidence is also available in studies using explicit measures. For example, in my MPhil project, all eight participants stereotypically believed that there was a genuine difference between male and female’s average language aptitude. Not only did they give examples of their teachers’ blatant statements characterising female classmates as more gifted language learners and essay writers, but some also claimed to have read papers reporting consistently and reliably that women process verbal information faster than men because of the distinctive features in the right-hemisphere of their brains. Therefore, I would argue that in Study 1, participants would be unlikely to hide gender-differential attitudes. In addition, since the questionnaires were anonymous, participants should have had few reservations about appearing socially unacceptable even if they had expressed their gender-disparaging opinions. To sum up, self-presentational issues might not have been a concern in Study 1.

#### **4.2.1.2 Developing Questionnaires**

**Design.** The questionnaires took two stages to develop. Appendices A and B are the first drafts of questionnaires for guardians/teachers and students respectively. (Questionnaires in appendices are bilingual (Chinese and English) for illustrative purposes. In Pilot Study 1, as well as Study 1, Chinese was the only language used in correspondence.) The former contained six items measuring FALS on the macro level: three targeting female learners and another three targeting male learners. The latter, in contrast, consisted of 42 items encompassing all aspects on the macro and micro levels of FALS. Of course, it could be argued that teachers should also fill out the questionnaire measuring FALS on both the macro and micro levels, i.e. Student Questionnaire. Yet, it was deemed impractical and unnecessary. The sample size needed for Student Questionnaire (see details in Section 4.2.1.4) was 400, and it was unlikely that this many teachers of English could be recruited. Besides, since



Study 2 was designed to explore teachers' voices using in-depth interviews, applying Student Questionnaire to teachers could be redundant.

The questionnaires exhibited ecological validity and content validity: The items were grounded in the data obtained from my MPhil thesis, where male language learners had recounted how their guardians, teachers, and peers described girls as more competent language learners. Incorporating this in situ qualitative exploration, instead of approaching the topic with a predetermined list of variables, was deemed capable of leading to more generalisable findings. Besides, previous GS literature (e.g. Cejka and Eagly, 1999) were consulted so that all facets of FALS could be included, which safeguarded content validity (Cohen, Swerdlik, and Sturman, 2015).

The first drafts were employed in Pilot Study 1, where participants commented highly on their face validity. According to Onwuegbuzie and colleagues, face validity refers to the extent to which an instrument appears to measure what it claims to (Onwuegbuzie, Witcher, Collins, Filer, Wiedmaier, and Moore, 2007). In the instructions, the aim of the survey was described as investigating 'what opinions high school students and their guardians in China have about English learning'. All participants thought their perspectives and opinions were exactly the what the questionnaire purported to measure 'on the face of it'. This strong face validity helped them establish confidence in the perceived effectiveness of the research, and consequently increased their willingness to participate and motivation to answer truthfully.

After Pilot Study 1, two revisions were made. First, the rating scale was changed from an eleven-point one into a seven-point one to increase test-retest reliability while maintaining enough discriminating power. This was because over a dozen students and guardians expressed their worry over the scale's reliability<sup>3</sup>. Furthermore, Preston and Colman's research provided evidence that a seven-point scale would be a suitable choice for most questionnaires because it demonstrated strong test-retest reliability, sufficient discriminating power, and was also respondents' second most favoured scale (2000). Another improvement was the wording of items. The first drafts utilised questions ('Do you believe that girls/boys in general ...?'), but some respondents in Pilot Study 1 reported that reading identical openings for all items was monotonous. Besides, some others also suggested that questions

---

<sup>3</sup> As evidenced by this quote from a student's email, 'If I were to re-do the survey, I would choose different answers for most questions, because I chose too many 7s and 8s this time. Personally, I don't think that these two are that different from each other, so I don't think this scale can accurately measure my opinion.' (Translated from Chinese.) Although the participant used the word 'accurately', I think he/she meant 'accurate over time', i.e. 'reliably accurate'.

were ‘not as direct as statements’. Therefore, in the second drafts, all items became statements with varied openings (see Appendices C and D). These were used in Study 1.

**Format.** The format of the questionnaires aimed to minimise potential biases in three ways. First, the order of the items followed these principles (de Vaus, 2014, p. 110-111) so that respondents would not be biased in a given direction:

- 1) Items on the micro level came before those on the macro level (see Section 3.3 for discussions about the two levels);
- 2) The order of micro-level items was randomised using *Excel* software (each item was assigned a random number between 0 and 1, and then sorted in ascending order);
- 3) The order of macro-level items was also randomised in the same way.

Second, the items about girls are presented in a separate part from those about boys so that the questionnaire itself would not impose a ‘male-versus-female comparison’ upon participants<sup>4</sup>. In other words, one part of the questionnaire measures participants’ stereotypical understanding of a female English learner, and the other part their shared image of a male English learner. In this way, the questionnaire simultaneously measured two constructs, the stereotypical image of female learners and the stereotypical image of male learners (see the next section’s discussion about construct validity for more details). The comparison of them was interpreted as FALS (achieved by a series of statistical analyses, see more elaborate analyses in Chapters 5 and 6).

The third feature to reduce intentional gender differentiation propelled by the instrument was to randomise the order of the section for male learners and another for female ones. In Pilot Study 1, this was realised by *Qualtrics.com*’s built-in randomisation function. In Study 1, which utilised paper-and-pencil questionnaires to reduce careless responses (see the section below for explanations), the randomisation was done manually (see Section 4.2.1.5).

#### 4.2.1.3 Pilot Study 1

---

<sup>4</sup> In the classic studies, scholars have always assessed GSs by presenting a given characteristic (such as ‘gifted in English’) that is immediately followed by two group labels, ‘typical man’ and ‘typical woman’, each with their own Likert-type rating scale (Deaux, 1984). With this procedure, it may be argued that when participants rate the typical man, they already compare and think about the typical woman. That is, though participants are not led to rate one group as better than the other, they may still be motivated to look for differences between the two genders. Thus, by separating questions about girls and boys, I wish to find if without juxtaposing males and females in the research instrument, participants would still consider gender a differentiating variable when predicting English abilities.

The participants (students and their guardians) were recruited from a private high school in a northern Chinese city<sup>5</sup>. In this report, I only utilised responses for participants that completed the entire questionnaire (79% of the guardian sample and 67% of the student sample), with a final  $N_{guardian} = 288$  and final  $N_{student} = 143$ . In the guardian sample, there were 210 females and 78 males, and in the student sample, there were 62 girls and 81 boys. Of course, because all participants were recruited from a single school, the pilot sample was not representative of the target population at all. Yet, as the data were interpreted with caution, the sample served the purpose of validation well enough.

**Procedures.** The teachers in charge of Grades 10 and 11 sent out and collected parental consent on my behalf. Students who gained parental consent were given a link of Student Questionnaire (i.e. Appendix B) and asked to fill out the questionnaire after school. Later, all guardians of the students in Year 10 and Year 11 were sent a link of Guardian/Teacher Questionnaire (i.e. Appendix A) and asked to finish the questionnaire if they wished to join the survey themselves.

**Data Screening.** Data screening refers to deleting cases with careless responses in order to get a dataset that reflects respondents' genuine answers, as is recommended by many scholars (e.g. Cohen *et al*, 2015). In addition, Meade and Craig (2012) pointed out that careless responses often arise in anonymous online questionnaires completed by student samples, with a base rate ranging from 3.5% to 10.6%. Therefore, before analysing the sample dataset, I used three screening methods to identify and delete three types of careless responses, as displayed in Table 4.1 (Case 36 appeared twice, and was marked in bold type font). The screening process left 240 and 86 effective cases in the guardian and student samples respectively. Because the rate of careless responses in both samples are alarmingly high (17% for guardians and 40% for students), modifications had to be made in Study 1 to encourage more effective responses.

**Construct validity.** Principal Component Analysis (PCA) was used to assess construct validity. As introduced by Owuegbuzie *et al* (2007), construct validity is concerned with whether an instrument is a meaningful and appropriate measure of the construct it claims to measure. Before running PCA, sampling adequacy was assessed by Kaiser-Meyer-Olkin (KMO) measure and correlations between items by Bartlett's Test of Sphericity.

---

<sup>5</sup> All teachers of English in this school were native speakers of English. As the research was conceived to measure FALS held by Chinese people, none of the teachers participated in the pilot study.

**Table 4.1: Data Screening Process in Pilot Study 1**

Type of careless responses	Reasons for screening and methods Used	Cases identified	
		Parent	Student
<b>Extremely long response time</b>	<p>According to Meade and Craig (2012), participants who responded too slow might have been distracted, thus giving inattentive responses.</p> <ul style="list-style-type: none"> <li>For the parent sample, descriptive statistics revealed that 95% of the participants spent fewer than 32m47s on the survey. Besides, the value of the 5% trimmed mean was very different from the mean, <math>\bar{x}_{.05} = 5m25s</math> and <math>\bar{x} = 17m20s</math>, indicating that there were some extremely slow responses. As a result, 14 cases where response times were longer than 32m47s were identified as careless responses.</li> <li>For the student sample, descriptive statistics revealed that 95% of the students spent fewer than 8m15s on the survey. Given that 5% trimmed mean and the mean were relatively close, <math>\bar{x}_{.05} = 3m07s</math> and <math>\bar{x} = 4m52s</math>, there seemed to be no extremely low answers.</li> </ul>	#26, #34, #36, #95, #101, #138, #175, #191, #207, #214, #231, #238, #253, #268.	/
<b>Uniform response pattern</b>	<p>Participants selecting the same answer for all items were likely to be careless.</p> <ul style="list-style-type: none"> <li>There were 22 cases in the parent sample where uniform response patterns emerged, among which Case 36 appeared once already in the previous category (Extremely Long Response Time).</li> <li>There were 41 cases in the student sample where uniform response patterns emerged.</li> </ul>	#6, #25, #36, #40, #50, #53, #85, #247, #62, #100, #108, #112, #120, #126, #171, #195, #198, #219, #227, #247, #255, #277.	#1, #3, #6, #7, #8, #13, #14, #16, #18, #19, #20, #21, #22, #24, #35, #37, #39, #42, #48, #49, #54, #56, #58, #62, #66, #75, #76, #80, #81, #88, #89, #81, #82, #106, #108, #111, #121, #124, #131, #141, #142.
<b>Random response pattern</b>	<p>Participants who randomly chose answers may have rushed through the questionnaire, thus giving unengaged responses (Johnson, 2005). As assessed by Mahalanobis distance (<math>p &gt; .001</math>),</p> <ul style="list-style-type: none"> <li>there were 13 multivariate outliers in the parent sample;</li> <li>there were 16 multivariate outliers in the student sample.</li> </ul>	#20, #33, #57, #128, #142, #143, #152, #181, #193, #209, #240, #269, #280.	#4, #12, #25, #32, #45, #53, #72, #73, #83, #85, #95, #98, #100, #116, #135, #139.

For the guardian sample, the overall KMO measure was .744 with all individual KMO measures greater than .680, classifications of 'mediocre' to 'meritorious' according to Kaiser (1974). Bartlett's Test of Sphericity was  $\chi^2 (15) = 763.37, p < .001$ . For the student sample, the overall KMO measure was 0.897 with all individual KMO measures greater than 0.80, classifications of 'meritorious' to 'marvelous' according to Kaiser (1974). Bartlett's Test of Sphericity was  $\chi^2 (861) = 5605.64, p < .001$ . These results indicated that the data were suitable for PCA.

For Guardian/Teacher Questionnaire, PCA revealed two components that had eigenvalues greater than one and which explained 48.14% and 31.094% of the total variance, respectively (see Appendix E). The scree plot shown in Appendix F also indicated that two components should be retained (Cattell, 1966). In addition, a two-component solution met the interpretability criterion (Field, Miles, and Field, 2012). As such, two components were retained. This two-component solution explained 80.07% of the total variance. The Varimax with Kaiser Normalisation method was employed here to provide a rotated solution, which exhibited 'simple structure' (Tabachnick and Fidell, 2013). The interpretation of the data was consistent with theoretical composition of FALS on the macro level: Guardian/Teacher Questionnaire was designed to measure with strong loadings of typical-female-learner items on Component 1 and typical-male-learner items on Component 2. This provides strong evidence for construct validity of Guardian/Teacher Questionnaire.

The factorial construct of Student Questionnaire was not as straightforward. PCA revealed four components that had eigenvalues greater than one, explaining 49.24%, 28.06%, 2.64% and 2.42% of the total variance, respectively. However, the scree plot (Appendix F) indicated that only two components should be retained (Cattell, 1966). In fact, different rotation methods have been used but no solution could meet the interpretability criterion (Field, Miles, and Field, 2012).

Reviewing the ambiguous results above, I thought it appropriate to run a forced factor extraction of two components for the following reasons: 1) Student Questionnaire was conceptualised as a two-factor instrument; 2) the scree plot indicated two components; 3) the eigenvalues of the last two components hovered close to one (1.11 and 1.02) respectively. The forced extraction yielded a two-component solution explaining 77.07% of the total variance. A Varimax orthogonal rotation was employed here and the rotated component matrix is shown in Appendix G.

The interpretation of the data was consistent with theoretical composition of FALS in the eyes of students: Student Questionnaire was designed to measure with strong loadings of typical-male-learner items on Component 1 and typical-female-learner items on Component 2. This provides satisfactory evidence for construct validity of Student Questionnaire.

**Reliability Analysis.** I used internal consistency coefficients (Cronbach's  $\alpha$ ) as the indication of reliability for both questionnaires. According to Tavakol and Dennick (2011), it is the most common reliability estimate for questionnaires using multiple Likert-scales to measure a single unidimensional latent construct. Since the questionnaires asked participants about their views on female and male students separately, each questionnaire simultaneously measured two components: the stereotypical image of female learners and the stereotypical image of male learners. Therefore, two internal consistency coefficients were calculated for each questionnaire. As shown in Table 4.2, all items demonstrated good to excellent reliability.

**Table 4.2: Internal Consistency of Items in Two Questionnaires in Pilot Study 1**

Questionnaire	Component	Items included ( $N = \text{no. of items}$ )	Cronbach's $\alpha$ ( $n = \text{no. of respondents}$ )
Guardian/ Teacher Questionnaire	Typical female learners	Appendix A, Part 1. ( $N=3$ )	.890 ( $n = 240$ )
	Typical male learners	Appendix A, Part 2. ( $N=3$ )	.857 ( $n = 240$ )
Student Questionnaire	Typical female learners	Appendix B, Part 1. ( $N=21$ )	.982 ( $n = 86$ )
	Typical male learners	Appendix B, Part 2. ( $N=21$ )	.987 ( $n = 86$ )

**Modifications.** The above analysis suggested that both questionnaires were valid and reliable, but given the high rate of careless responses, three techniques had to be employed to tackle this issue in Study 1: first, since it is typical of online surveys to have more careless responses (Johnson, 2005), questionnaires utilised in Study 1 changed to paper-and-pencil format. Second, because response time could not be gathered for paper-and-pencil surveys, I employed two additional ways to detect careless responses: two bogus items and a scale of self-report attention level. Two bogus items were inserted to the Student Questionnaire, items with unambiguous correct answers ('In my high school, we use at least one textbook in English lessons' and 'My high school does not offer English classes'). Thus, if a respondent chooses the wrong answer, it will be safe to assume that he/she is not paying attention. In order to avoid false negatives (the cases where a respondent choose the right answer while

not paying attention), the bogus items are reverse coded. Also, because the Parent/Teacher Questionnaire was relatively short, no bogus item was created. A self-report item was also added to the end of both questionnaires (Meade and Craig, 2012), asking participants to indicate the amount of attention they had exerted to the questionnaire.

#### **4.2.1.4 Sampling**

Drawing representative samples randomly from the population under inspection is the cornerstone of high quality survey research. To achieve this takes three steps: to define the population, to obtain a complete and unbiased sampling frame, and to select a large enough sample. This section will go through each step taken in Study 1, delineating measures to reduce coverage error and sampling error.

The population of the present survey is senior high school students (usually aged 16-18 years old), their guardians and their teachers of English (who are native Chinese) in China. Therefore, a complete sampling frame for the current study would be a list of all senior high schools in China, but it is impractical to compile such a list: According to the Ministry of Education ([www.moe.gov.cn](http://www.moe.gov.cn)), there were 13,947 senior high schools in China in 2017, when Study 1 was designed. Besides, China has such a vast regional and cultural diversity that GS endorsement may vary across the country. Thus, to reduce coverage error, a pragmatic approach was to find schools with varied characteristics in terms of geological locations (north and south), socio-economic statuses of the locations (developed cities and less developed counties), and education sectors (public-funded schools and private ones). Altogether, over a dozen schools were contacted, and eight joined Study 1 (see Table 4.3).

The sample size for Study 1 was originally set as about 500 families (which would mean 500 students and around 1,000 guardians) and 60 teachers. This was decided on two considerations. First, as de Vaus (2014) has advised, the accuracy and the confidence interval a researcher desires for the generalisations from the sample sets the lower limit for sample size. For the current study, as I wished to tolerate 5% of sampling error at 95% confidence level, a minimum of 400 samples was required for both guardian and student samples. Second, given that participants may give careless answers as in the case of Pilot Study 1, I imagined that inviting 500 families to join the survey would suffice while still remain manageable. Luckily, all participating schools were more cooperative and enthusiastic about the survey than I had hoped for, and in the end, I was able to collect 1298 responses from students, 2498 from guardians, and 62 from teachers.

**Table 4.3: Demographic Information of Participating Schools in Study 1**

NO.	Region	Location	Education Sector	Questionnaire (Distributed; Collected)		
				Student	Guardian	Teacher
1	North	County S	Public	360; 312	720, 624	24; 17
2	North	County S	Private	360; 287	720, 574	18; 10
3	South	County Z	Public	360; 304	720; 608	24; 10
4 <sup>a</sup>	South	City G	Private	80; 67	160; 134	12; 7
5 <sup>a</sup>	South	City CD	Private	80; 12	160; 24	12; 0
6	South	City CQ	Public	200; 167	400; 334	24; 7
7	North	City T	Public	120; 100	240; 200	24; 11
8 <sup>b</sup>	South	City X	Public	50; 49	N/A	N/A
<b>Total</b>				1,610; 1,298	3,220; 2,498	138; 62

*Notes.* a. School 4 and 5 were both private schools from the south. Both schools were included because they were relatively small and only inviting one school would have led to too few participants from this particular kind of schools. b. Only Student Questionnaires were distributed in School 8. This was because the Student Questionnaires collected from the first seven schools did not contain enough cases in two cells of the design: Year 11 and Year 12 boys in southern schools. Therefore, School 8 was contacted so that more questionnaires could be collected from students fitting this particular demographic profile.

#### 4.2.1.5 Distribution and Collection Procedures

The majority of the fieldwork for Study 1 took place between September and December, 2017<sup>6</sup>. In each school, the distribution process was more or less similar: First, the headmasters informed the class-teachers<sup>7</sup> of the questionnaire survey. The class-teachers then invited students and their guardians to join the survey on my behalf. Meanwhile, the headmaster introduced the survey to teachers of English at the school, and took a note of the number of teachers who were willing to participate. Families who were interested in joining signed up, and the headmaster notified me of the number of all potential participants (families as well as teachers).

<sup>6</sup> Questionnaire distribution and collection in the first seven schools ended in December, 2017. School 8 was contacted in April and joined in May, 2018.

<sup>7</sup> In Chinese schools, if a teacher teaches a subject (e.g., math, physics), he/she will be called a course-teacher. If, at the same time, the teacher also takes administrative duties alongside teaching work, he/she will also be called a class-teacher. The administrative duties include but are not limited to setting up rules for classroom sanitation, coordinating the relationships of course-teachers teaching the same class, contacting parents and discussing their children's academic and social performances. A class-teacher is usually responsible for only one class, hence the term *class-teacher*.



Then, I arrived at each school with printed questionnaires and distributed them to class-teachers. Each participating student was to receive a brochure from his/her class-teacher, a brochure containing two copies of consent forms (one for the participants to return to me, and the other for them to retain; for an English translated version, see Appendix H), two copies of Guardian/Teacher Questionnaire (one for female guardians and another for male ones in the same family), and one copy of Student Questionnaire. The students then took the brochure home, shared it with his/her guardians, and returned the brochure in a week's time after its completion. In terms of the teachers, each participating teacher also received a brochure containing two copies of consent forms (one for the teachers to return to me, and the other for them to retain; Appendix H) and a copy of Guardian/Teacher Questionnaire. They were also asked to hand in the brochure within a week.

When developing the questionnaires, to reduce potential biases, it seemed necessary to randomise both the order of items and the sequence of the two parts about male and female learners. In Study 1, randomisation was achieved by the following process for the teachers:

1. Randomising the order of items in Guardian/Teacher Questionnaire:
  - I. The order of 3 items describing females on the macro-level was randomised using *Excel* software (each item was assigned a random number between 0 and 1, and then sorted in ascending order);
  - II. The order of 3 items describing males on the macro-level items was also randomised in the same way;
2. Generating 2 versions of the brochure, V1 with the part describing female learners preceding the part describing male learners, and V2 with the reversed sequence;
3. Printing  $n$  copies of V1 and V2 brochures when visiting a school with  $2n$  participating teachers;
4. Shuffling  $2n$  copies of brochures manually;
5. Distributing the copies to teachers.

For participating families, a similar process of randomisation was carried out:

1. Randomising the order of items in Student Questionnaire:
  - I. The order of 18 items describing females on the micro-level and one reverse-coded bogus item was randomised using Excel software (each item was assigned a random number between 0 and 1, and then sorted in ascending order);

- II. The order of 3 items describing females on the macro-level items was also randomised in the same way;
- III. Items in Step II were placed after those in Step I;
- IV. Steps I to III were repeated for the 21 items describing males;
- 2. Randomising the sequence of parts describing male and female learners in the brochure by generating 8 versions of the brochure with the sequences shown in Appendix I;
- 3. Printing 8*m* copies of V1—V8 brochures when visiting a school with 8*m* participating families;
- 4. Shuffling 8*m* copies of brochures manually;
- 5. Distributing the copies to families.

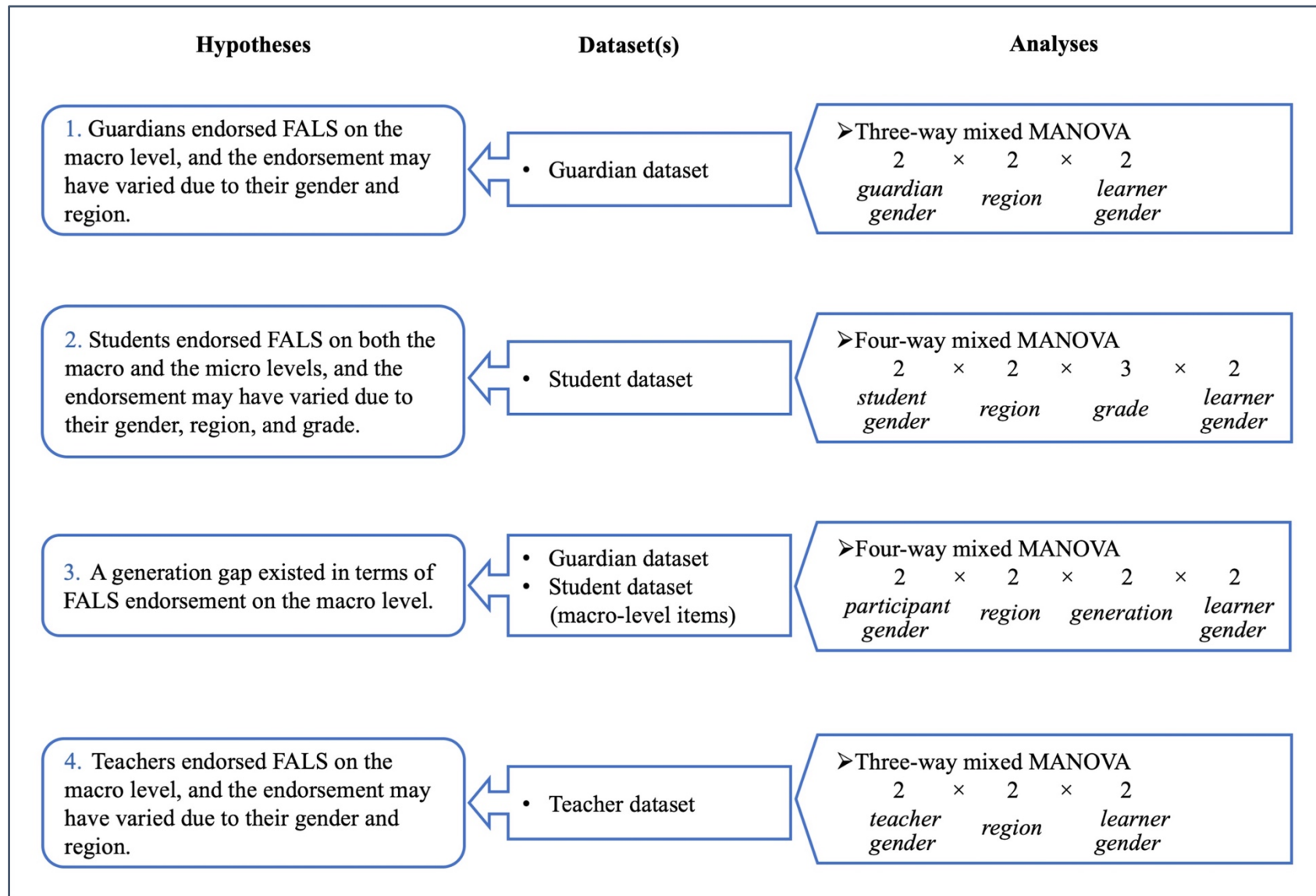
#### **4.2.1.6 Data Screening and Analysis Procedures**

The data collected in Study 1 were used to 1) examine the magnitude of the FALS endorsed by three groups of participants, and 2) compare group differences in stereotype endorsement. Figure 4.2 shows the analytical framework of Study 1. The left column identifies the four hypotheses Study 1 aimed to validate, and the column in the middle lists the particular datasets relevant to each hypothesis. Hypotheses 1, 2, and 4 were concerned with whether three groups of participants—guardians, students, and teachers of English—endorsed FALS. Hypothesis 3, instead, involved the merge of the guardian dataset with the student one (macro-level items only) to examine the existence of a generation gap in stereotype endorsement. The right column summarises the analyses processes. Of course, before data analyses, screening procedures similar to those in Pilot Study 1 were run for each dataset (see details in Section 5.1.1, 5.2.1, and 6.1)

#### **4.2.1.7 Ethical Considerations**

When conducting Study 1, I followed the ethical guidelines of BPS (2010) and BERA (2011); the equivalent was unavailable in China. Before collecting data, I had fully informed the headmasters about the research aim and discussed with them appropriate time and manners to administer and collect questionnaires. The headmasters, in the process, acted as gatekeepers to supervise the survey. A consent form for students and their guardians were sent out by class-teachers on my behalf, which contained information about the aims and procedures of the research and my contact information (see Appendix H). In the consent form, participants were notified that they retained the right to withdraw from the study at any phase without having to clarify their reasons. Students and guardians were asked to return a

Figure 4.2: Analytical Framework of Study 1



copy of consent form with completed questionnaires. Teachers also received two copies of consent forms together with their questionnaires. Similarly, they were asked to return a copy with their completed questionnaires.

Confidentiality and anonymity have been strictly preserved. The questionnaires did not ask for any personal information that could lead to the identification of any participant, and all the data collected were carefully protected. Only my supervisor and I saw the data. In addition, all findings will be presented in the following chapters with school names omitted. I ensure that the schools and participants will not be personally identified by any demographic information provided in the report.

It is unlikely that the research has had any negative consequences on participants. The questionnaire merely asked students to express their GS endorsement. Although the statements may have been somewhat self-relevant to students, this is inevitable with opinion surveys. Of course, I have tried to avoid any unwanted influences by assuring participants the right to withdraw. Participants also had access to my contact details if and when they had questions about the research.

#### **4.2.2 Study 2: Interview-Based Survey**

The objective of Study 2 was two-fold: 1) to find whether teachers of English would describe male and female learners of English divergently, and if so, 2) to build explanations about how these gender-stereotypical accounts relate to each other and/or to other GSs prescribing gender differences. Study 2 collected narrative data from 20 interviewees between September and December, 2017, and the data went through thematic analysis and Fisher's probability test. Figure 4.3 here shows the complete design of Study 2. The first column summarises the theoretical framework, which has been discussed fully in the previous chapter. The remaining three columns illustrate the design of the interview schedule, and steps taken to collect and to analyse data. The specifics will be expounded in Sections 4.2.2.2-4.

Before diving into the details of Study 2, I should explain the reasons why teachers, instead of guardians or students, were decided the most suitable interviewees. The primary reason was because of teachers' prolonged and intense engagement with FALS and potential other GSs about language learners. Compared to guardians, who tended to have the least contact with English education among the three groups, teachers' perspectives would be grounded in nuanced daily observation and personal experiences. Students, likewise, had the

same advantage as the teachers, but it was not necessary to include them in Study 2 because I have carried out a phenomenological investigation into students' side of the story in my MPhil project, though with undergraduates majoring in English, not high school students per se. In addition, given that Study 4 was designed to characterise high school students' encounters with FALS (see details of its design in Section 4.3.2 and findings in Chapter 8), their interpretations and endorsement of it would be explored by then. Therefore, teachers were selected as ideal interviewees in Study 2.

#### **4.2.2.1 Interviews as Data Collection Methods**

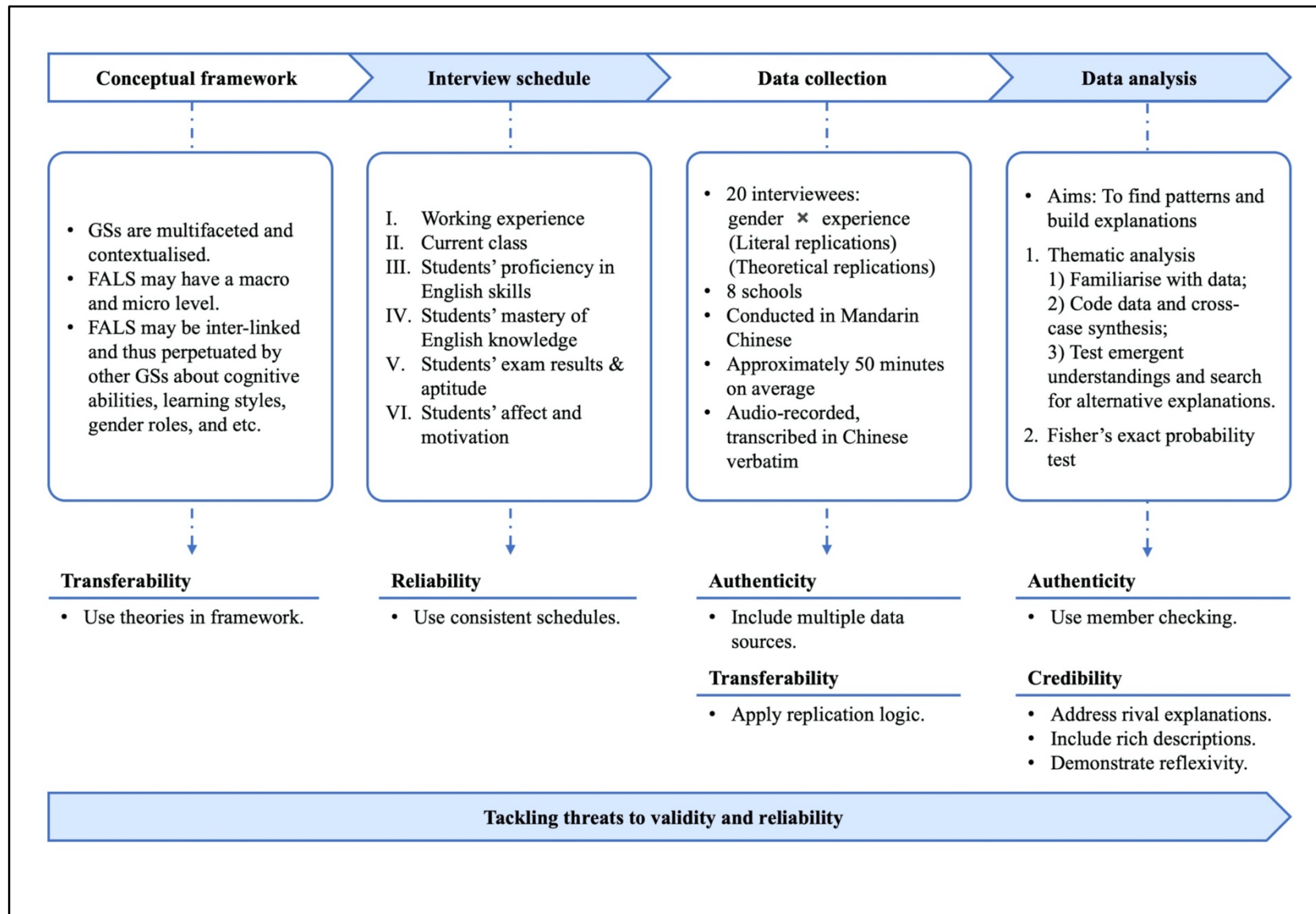
Conducted concurrently with Study 1, Study 2 employed in-depth interviews. It was a necessary companion of Study 1 in two ways. First, it enabled convergent validation. During the interviews, participating teachers were asked to describe their students' aptitude, affect, and achievement regarding English, and the gendered patterns in their accounts were comparable to the tripartite FALS investigated using questionnaires in Study 1. Through the analysis of these overlaps, stronger inferences could be drawn based on both data sources. Furthermore, Study 2 compensated Study 1 via its potential to elicit rich and thick descriptive data. Since Study 1 confined scientific evidence to quantifiable data, it would inevitably neglect elements of the context. Study 2 neutralised this weakness by focusing on 'the cultural, everyday, and situated aspects of human thinking, learning, knowing, acting, and ways of understanding ourselves as persons' (p. 15, Brinkmann and Kvale, 2015). Thus, in order to capture a more holistic and contextual portrayal of FALS, Study 2 was carried out alongside Study 1.

Threats to the validity or quality of Study 2 will be addressed accordingly in Sections 4.2.2.2-4. For example, measures were taken to ensure transferability, the concern about whether one can extend findings in the research setting to a wider range of similar settings, and credibility, the issue of whether participants judge the results to be credible (Teddlie & Tashakkori, 2009). Another common threat, failure to reach data saturation (Mason, 2010), was also tackled.

#### **4.2.2.2 Developing Interview Schedule and Pilot Study 2**

Appendix J is the first draft of the interview schedule. It began by introducing the purpose of the interview, how data would be recorded, and issues about confidentiality and anonymity. Then, questions about participants' demographic information, their working experience and their current students were listed. As these questions were retrospective and descriptive, it was anticipated that participants could build rapport with me, the interviewer,

Figure 4.3: Design of Study 2



in a relaxing manner. Following the introduction part, four major topics were included: 1) students' performance in English, 2) students' exam results and aptitude in English, 3) students' affect or motivation for English learning, and 4) teacher-parent interactions. The purpose was to seek for descriptions and comments that indicate whether teachers consider gender a differentiating variable among learners.

This draft was used in Pilot Study 2, where Lipeng, an experienced male English teacher, and Chen, a novice female teacher, joined (both names here are aliases). Based on their feedback, three adaptations were made in Draft 2 (Appendix K). The first was to revise the questions about students' performance in English. During the pilot study, both participants thought that the original questions were to assess their pedagogical knowledge. As a result, the questions failed to elicit interviewees' teaching experiences or their beliefs about English learners. Thus, the adapted questions in Draft 2 directly asks interviewees how they assess students' achievement and what qualities they thought were ideal for learning success. Another problem arose in questions about students' affect or motivation. Participants raised that Q14 and Q15 were somewhat too similar (see Appendix J), and they were thus combined in Draft 2.

Finally, the last topic, teacher-parent interactions, was deleted in Draft 2. The topic was designed to elicit how teachers treat guardians' gender-stereotypical expectations about English results (if any), but it was found unnecessary in Pilot Study 2. Lipeng said he had had many experiences dealing with guardians, because he has been a class-teacher for almost a decade. But as a class-teacher, he mostly contacted guardians to discipline the inappropriately behaved or help the lower-achieving. The conversations he had with guardians were not exclusive to the English subject, and he was the one who steered the talks. Chen, on the other hand, did not deal with guardians at all, because as a young, course-teacher, she had no such responsibility. According to Chen, guardians would go to class-teachers if they were concerned about their children's academic performance, because class-teachers were expected to communicate with guardians about any issue happened in school. Given the above results, I decided to delete this topic in Draft 2.

To sum up, the revised interview schedule, Draft 2, has four topics: proficiency in English skills, mastery of English knowledge, students' exam performances and aptitude, and students' affection/motivation.

The interview agenda was designed to increase transferability and reliability. To enhance transferability, I selected the topics on the basis of existing GS theories (see details in the left column in Figure 4.3). To strengthen reliability, the agenda remained consistent for all participants so that the effects of changes in research settings on interviewees' responses could be minimised (Yin, 2014). In addition, consistency is also a requirement for reaching data saturation, as purported by Fusch and Ness (2015): Only when the same questions are answered by multiple participants can their replies be meaningfully analysed through the same coding process. Then, when the researcher find it no longer feasible to further code the data, he/she can claim that the saturation point is achieved.

#### **4.2.2.3 Sampling and Interviewing**

Study 2 employed a purposive sampling method: All interviewees were teachers of English in participating schools in Study 1. Before arriving at each school, the headmaster, on behalf of the researcher, extended an invitation to all teachers of English to join a face-to-face individual interview about their teaching practices. Those willing to participate would directly contact me, and I would select potential interviewees from each school. The goal was to recruit 20 interviewees in total, ranging on the basis of gender (male or female) and experience (veteran or novice<sup>8</sup>), i.e. 5 female veterans, 5 female novices, 5 male veterans, and 5 male novices. This sampling scheme was chosen to help reach saturation, because the two selection criteria and the participating schools' diverse backgrounds (varying regarding geological locations, socio-economic statuses, and educational sectors) were deemed sufficient to collect inclusive data. Besides, through gathering data from multiple interviewees and later integrating results from Study 2 with those from Study 1, the authenticity of interview data is increased. This form of triangulation can also help reach the saturation point.

However, only 1 male novice was found in all 8 participating schools. In fact, I kept looking for male novice teachers until the end of the fieldwork (May, 2018) from all

---

<sup>8</sup> After consultation with the two participants in Pilot Study 2 and headmasters from participating schools in Study 2, 'novice teachers' were operationally defined as those with three or fewer years of teaching experience. This was because in a public high school, a new teacher typically starts with teaching students in Grade 10. After one academic year, the teacher will then stay with the same cohort, which will be in Grade 11, and a year after that, Grade 12. Usually a teacher is considered novice until he/she has completed this three-year-cycle, i.e. having taught a cohort across three grades. In some private schools, teachers work on a one-year-cycle, where they specialise in teaching one grade. But according to the headmasters, the teachers typically change positions (to teach a different grade, to transfer to a new school, or to quit teaching, etc.) after two to three years. Thus, for the sake of clarity and consistency, teachers with three years or fewer years of experience were regarded novices in the current project. Those with more than three years of experience would be considered experienced ones.



secondary schools I had contacts with, but the second never showed up. Implications of the lack of male novice English teachers will be discussed in Sections 10.3 and 10.6.

Transferability was also addressed in the sampling procedures. The interviewees sharing similarities were literal replications of each other, as it was more likely to yield similar patterns among them. For example, Ren<sup>9</sup>, Jian and Yun were all female novices, so it could be argued that the convergences in their accounts might be extended to a wider context, to other female novices working in similar settings. Meanwhile, the teachers that differed on one particular characteristics were theoretical replications of each other. The divergences in their descriptions, according to Yin (2014), could have resulted from their contrasting features. As in the case of Fen and Zhe, who differed only in teaching experience: Fen was a novice but Zhe was an expert. A similar difference existed between Yuehan and Zengque. Therefore, if Fen's opinions contrasted Zhe's the same way as Yuehan disagreed with Zengque, the contrasting views here might have arisen due to experience. Such a replication logic is analogous to repeating multiple experiments on the same topic, where similar conditions predict consistent findings and different conditions expect disparate results. Applying this replication logic in interviewee-selection, therefore, safeguarded transferability, because it increased the probability that inferences from the interviewees may transfer to a wider context.

The interviews took place at comfortable and convenient settings chosen by interviewees, usually in a meeting room or private office in the school they worked for. After gaining interviewees' consent, the interviews began and were audio-recorded. All interviews were conducted in Mandarin Chinese. Of course, it was anticipated that, as interviewees were teachers of English, fragments of English appeared in their accounts of teaching practices, materials, and quotations of themselves and colleagues.

During the interview, I used simple and direct probes to further build rapport, to ensure accuracy of participants' recollections, to ask for justifications, and to extend their narratives (Gillham, 2005). For example, by saying 'That must have been hard.', I showed my understanding of teachers' burden and thus built rapport. By paraphrasing 'Let me see if I've got things right. You said that ...', I could check how accurately interviewees have remembered an event. By asking 'What makes you say that?', I could ask for participants to rethink their judgmental statements and unpack the gender-stereotypical assumptions

---

<sup>9</sup> For a list of all interviewees and their demographic information, see Table 6.1 in Section 6.1.

underlying their comments. And by demanding ‘Tell me a bit more about xxx...’, I could extend the narratives into more detailed ones. Finally, I brought the interview to a closure by summarising topics covered and expressing my gratitude.

All interviews were transcribed verbatim by me. Each transcript was a comprehensive semantic record of the interview. Additionally, any identifiable non-verbal utterances like laughter, in-/out-breath, pauses and hesitations were also noted by symbols displayed in Table 4.4 (adapted on the basis of the appendix in Silverman, 2013, p. 465-466).

**Table 4.4: Transcription Notes in Studies 2 and 4**

Symbol	Example	Explanation
(.)	get (.) treatment	A dot in parentheses indicates elapsed time in silence.
<u>word</u>	I’ve got <u>enough</u>	Words with wavy lines indicate some form of stress, via pitch and/or amplitude.
<i>word</i>	It’s called <i>XXX</i> .	Italicised words were originally in English (in a Chinese transcript).
::	O::Kay?	Double colons indicate prolongation of the immediately prior sound.
*hh	I feel that *hh	A row of h’s prefixed by an asterisk indicates an in-breath; without the asterisk, an out-breath. Double h’s indicates a longer breath.
( )	Future and ( )	Empty parentheses represent unidentifiable utterances.
(word)	see (positive)	Parenthesised words are possible hearings.
(( ))	but ((continues))	Double parentheses contain researcher’s descriptions rather than transcriptions.
[]	[Friday] was hot	Words in square brackets are researcher’s additions to make the account logically coherent after being translated.

#### 4.2.2.4 Data Preparation and Analysis Procedures

The rightmost column of Figure 4.3 outlines the aims and analytical procedures of Study 2. The data were first subject to the thematic analysis approach with the help of *MAXQDA2018*, a software specialising in treating qualitative data.

**Familiarisation.** After transcribing and reading each transcription three times while listening to audio recordings, I, the researcher, managed to familiarised myself with interview data. During this time, analytic memos of six types were noted down:

1. Personal experience: how the researcher personally relates to the interviewees themselves and their stories;

2. Research questions: how some segments or large chunks of data may answer research questions;
3. Emergent patterns: how certain segments of data may represent similar concepts or converge to display particular patterns;
4. Related literature: how some segments of data reminds the researcher of related existent theories;
5. Reflexivity: how some segments of data may reflect potential problems with the study, and how personal or ethical dilemmas arose during the interview process;
6. Future directions: how particular missing elements or a need for additional data may merit future investigation.

**Coding and cross-case synthesis.** The interview data then went through three cycles of coding and cross-case synthesis procedures (see Table 4.5 below for a list of coding methods used in all three cycles).

**Table 4.5: Coding Methods and Example Codes in Study 2**

Cycle	Method	Notes	Example codes
1	<b>Attribute Coding</b>	<ul style="list-style-type: none"> <li>Demographic characteristics of interviewees;</li> <li>Usually noted at the beginning of a transcript, but may also be embedded within it.</li> </ul>	<ul style="list-style-type: none"> <li>Personal attributes: gender, region, administrative unit, marital status</li> <li>Professional attributes: experience, sector</li> </ul>
	<b>Provisional coding</b>	<ul style="list-style-type: none"> <li>Researcher-generated codes based on previous literature and research questions;</li> <li>Usually applied to longer chunks of data.</li> </ul>	<ul style="list-style-type: none"> <li>Stereotype awareness (female-language)</li> <li>Stereotype endorsement (female-language)</li> <li>Gendered patterns (male-competence)</li> </ul>
	<b>Structural coding</b>	<ul style="list-style-type: none"> <li>Content-based phrases that segment and categorise the data according to the topics of interviews;</li> <li>Collected for more detailed coding.</li> </ul>	<ul style="list-style-type: none"> <li>Working experience: transfer, class teacher, colleagues, hours per week</li> <li>Current students: grade, branch, gender ratio</li> </ul>
2	<b>Value coding</b>	<ul style="list-style-type: none"> <li>Conceptual phrases that reflect interviewees' values or attitudes that are gender-related and/or concerned with English learning.</li> </ul>	<ul style="list-style-type: none"> <li>Underscoring 'student-teacher relationship'</li> <li>Downplaying aptitude</li> </ul>
	<b>Versus coding</b>	<ul style="list-style-type: none"> <li>Conceptual phrases that identify groups of people, or certain behaviours or activities in dichotomous terms.</li> </ul>	<ul style="list-style-type: none"> <li>Boys VS. girls</li> <li>Career-driven VS. Home-oriented</li> </ul>

3	<b>Process coding</b>	<ul style="list-style-type: none"> <li>• Gerunds (nouns formed by attaching ‘-ing’ to the end of a verb) used to connote observable or conceptual action.</li> </ul>	<ul style="list-style-type: none"> <li>○ Subtyping exceptional male learners</li> <li>○ Self-handicapping</li> </ul>
	<b>Causation coding</b>	<ul style="list-style-type: none"> <li>• Causal beliefs about learning outcomes;</li> <li>• Combined with versus coding to identify gender-specific attributions.</li> </ul>	<ul style="list-style-type: none"> <li>○ Affect → Achievement</li> <li>○ Aptitude → Achievement</li> <li>○ Achievement → Affect</li> <li>○ Specific to English/Global to all subjects</li> </ul>
	<b>Emotion coding</b>	<ul style="list-style-type: none"> <li>• Labels that reveal the emotions experienced by interviewees, possibly inferred by the researcher;</li> <li>• Provides insights into the gender-related perspectives of interviewees.</li> </ul>	<ul style="list-style-type: none"> <li>○ Surprise (when describing exceptional male learners)</li> <li>○ Bemoaning (when describing male learners’ self-handicapping behaviours)</li> </ul>

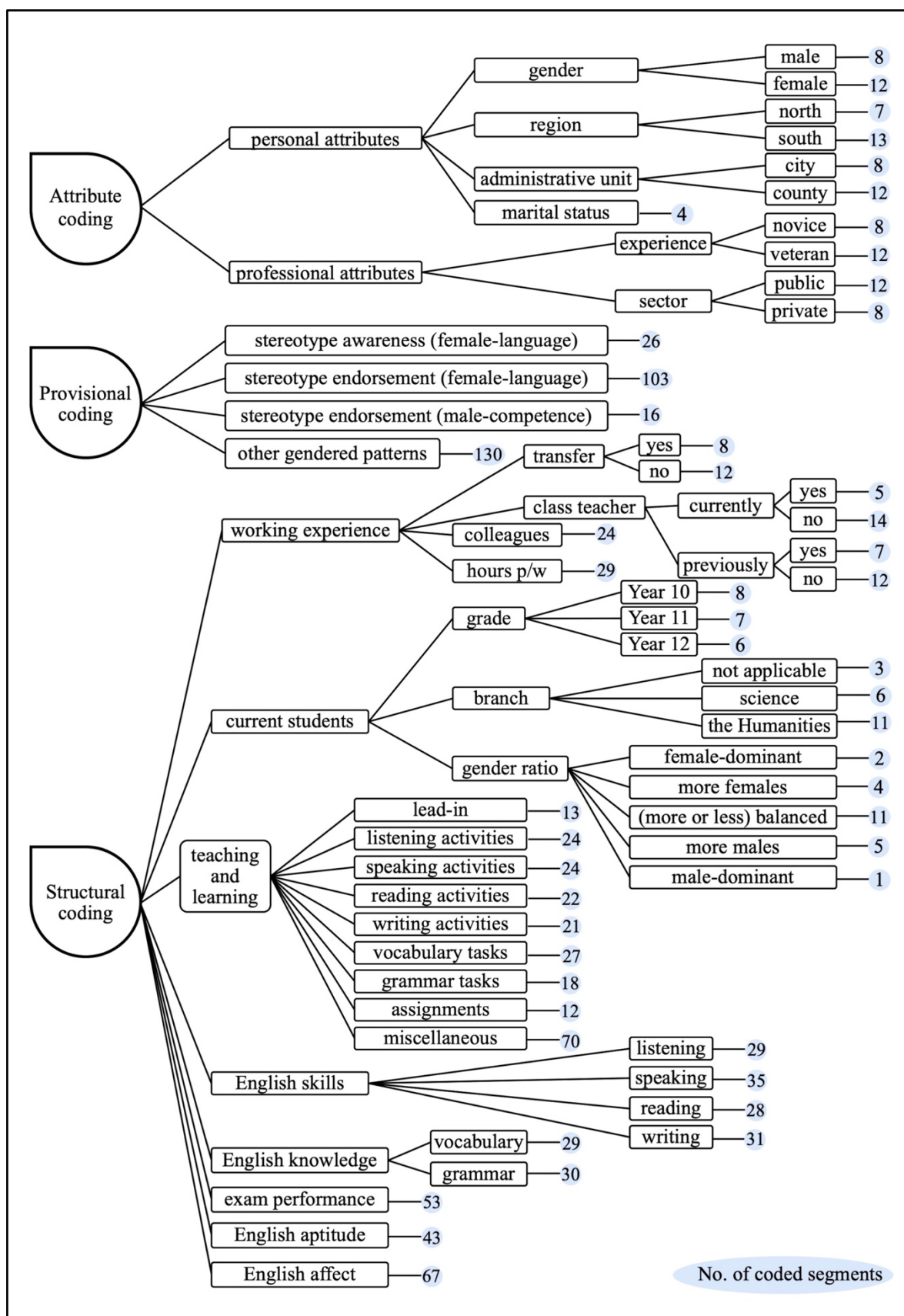
*Notes.* The eight coding methods listed in the table were applied to the interview data after consulting reference books on coding and analysing qualitative data (e.g., Bernard, Wutich, and Ryan, 2016; Saldaña, 2013).

The first cycle managed to segment and categorise the interview data according to researcher-generated codes based on the topics covered in interviews, previous literature, and research questions. Once all interviews were coded, the codes were reviewed for common and contrasting themes among interviewees and across interviewees. The aim of such cross-case synthesis was to removed less pertinent codes and combined similar codes. As a result, the first cycle coding yielded a total of 1,130 coded data entries (see Figure 4.4). Following this, the second and third coding cycles aimed to refine the current codes so that gendered patterns in FALS endorsement and other stereotypes concerning learners could be identified.

**Testing emergent understandings and search for alternative explanations.** The refined codes were then reviewed again with the whole transcript to generate some understandings of the data. This process generated the overarching theme of “teachers’ overall conservative attitudes towards *gender*” emerged with three inter-related superordinate themes: 1) the widespread endorsement of FALS (Chapter 6 and 7); 2) GSs accompanying FALS (Chapter 7); 3) critical perspectives about FALS (Chapter 7).

Precautions against potential threats to authenticity and credibility were taken in the data analysis process. As Seale (1999) has summarised, authenticity refers to the research’s ability to report different realities, to help participants appreciate others’ viewpoints, and to improve participants’ understanding of the researched phenomenon. According to Creswell (2014), member-checking was an effective strategy: I took back the themes and major findings,

Figure 4.4: First-Cycle Coding System in Study 2



reported to participants, and asked them to decide whether they have found these analyses accurate (six interviewees engaged in the process). Two additional strategies contributed to credibility when reporting findings, including entertaining alternative explanations and citing rich descriptions (see details in Chapters 6 and 7).

Aside from the above thematic analysis, Fisher's exact probability test was run, too. This analysis was utilised to examine whether gender and regional differences existed in FALS endorsement among interviewees in Study 2, whose results were then compared and contrasted with findings from the MANOVA carried out on the teacher questionnaire dataset in Study 1. The details will be reported in Sections 6.4-5.

#### **4.2.2.5 Ethical Considerations**

A consent form (for an example, see Appendix L) with explanations of how confidentiality and anonymity would be protected was signed before the interview. During the interview, I asked for participants' permission before I audio-record the interviews. The safety and convenience issues of the participants were taken into account when deciding the time and places for interviews. All interviews were transcribed in Chinese verbatim, and participants were invited to review the transcriptions. In the transcript, any information the participant mentioned that could be identified with him/her, such as names and places, were changed to protect confidentiality. The analysis was conducted in Mandarin so that I could avoid losing any cultural nuances in translation. Interview excerpts used in the thesis were translated by me and reviewed by a professional translator so that they could truly reflect interviewees' perspectives.

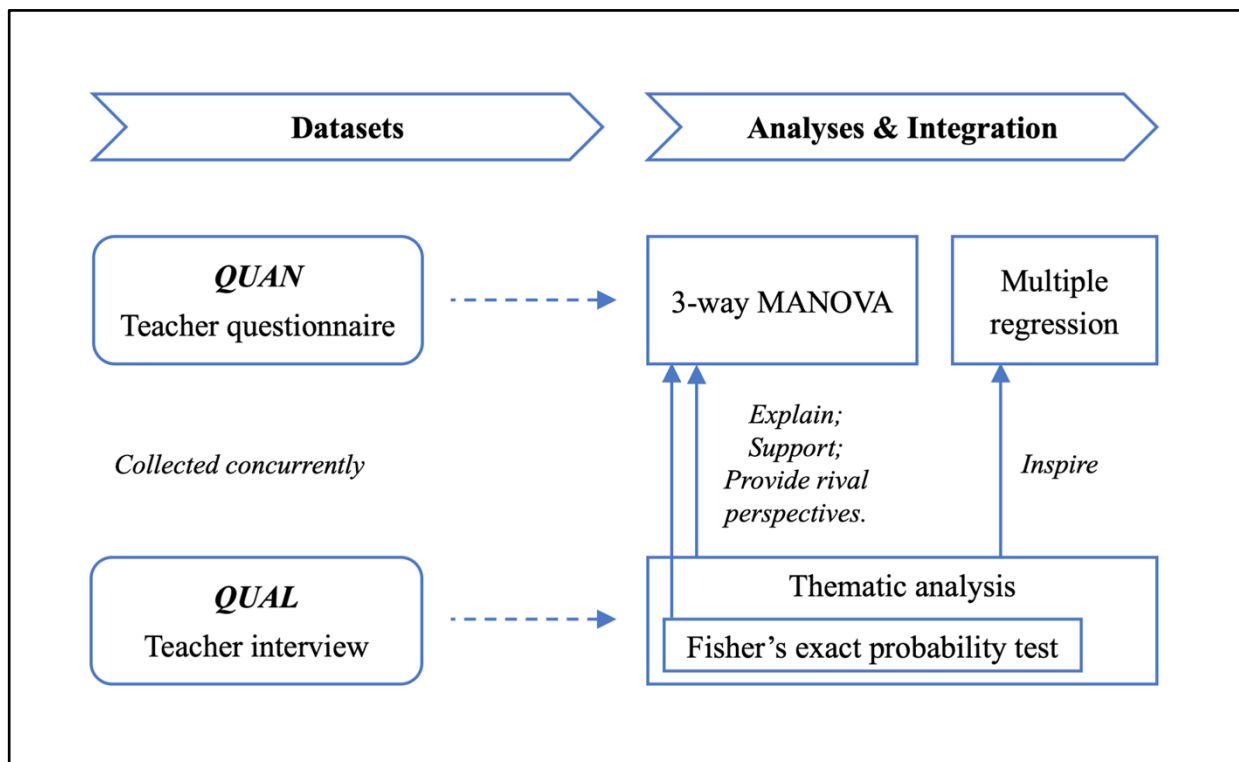
#### **4.2.3 Integrating Data from Studies 1 and 2**

Though data from Study 1 and 2 were analysed separately, integration of the two sources was necessary in order to achieve an overall interpretation of GSs targeting learners of English. More specifically, numerical data from the teacher dataset in Study 1 and textual data from Study 2 (interviews with teachers) were combined, compared and contrasted.

Figure 4.5 here demonstrates the analyses procedures and how findings from two strands were integrated. As planned, a three-way MANOVA was performed on the questionnaire dataset. Meanwhile, thematic analysis and Fisher's test were conducted on the interview data. Then, the findings were compared to identify convergence, divergence, contradictions, as well as relationships between the two strands.

This process led to three outcomes. To begin with, the first theme emerged from the interview dataset (‘the widespread endorsement of FALS’) aided the interpretation of MANOVA results from the questionnaire dataset: it explained how FALS was construed by teachers and provided rich textual evidence for their endorsement. Second, results from the Fisher’s test echoed the relevant findings from MANOVA, both failing to find gender or regional differences in FALS endorsement. Finally, the thematic analysis suggested causal associations among FALS components as perceived by teachers. Inspired by this interesting finding, a multiple regression was carried out to verify whether such patterns were evident in the questionnaire dataset. The above integration process will be reported in a point-by-point manner in Chapter 6.

**Figure 4.5: Data Integration of Teacher Datasets in Phase 1**



Additionally, two more emergent themes from the thematic analysis (‘GSs accompanying FALS’ and ‘critical perspectives about FALS’) provided rival perspectives concerning FALS that Study 1 was incapable of detecting. Chapter 7 is devoted to these additional themes.

### 4.3 Phase 2: Detecting Effects of Female-Advantage-in-Languages Stereotype

Data from Phase 1 went through an initial analysis, and the results indicated that FALS was harboured by all three groups of participants (see detailed findings in Chapters 5-6). Progressing from these preliminary results, Phase 2 endeavoured to examine how FALS

might have affected high school students through an explanatory sequential design. In this phase, Study 3, a field experiment, was conducted first, and it was followed up by Study 4, an interview-based study. The participants were from the same participating school, because the primary goal of Study 4 was to aid explanation of specific quantitative results in Study 3.

Section 4.3.1 below presents the details of Study 3, which includes issues with designing, sampling, procedures and data analysis procedures. Following this is Section 4.3.2, explaining corresponding specifics of Study 4.

#### **4.3.1 Study 3: Field Experiment**

Study 3 was employed to determine, primarily, if stereotype threat (ST) effect might have existed among male learners of English: if activating FALS prior to a standardised comprehensive English test affected male students' performance, this should have provided evidence for ST effect. Of course, stereotype boost might also occur among female learners, because FALS depicts them positively.

By deliberately controlling and manipulating independent variables (participant gender and FALS priming in this study), experiments are the most powerful method to falsify the null hypotheses (Cohen *et al*, 2011), to rule out rival explanations (Creswell, 2012), and thus to establish causal relationships between primed stereotypes and performance decrements. Particularly, lab experiments are often preferable to field experiments, because they offer stronger inferences through randomisation procedures in sampling and in experimental/control group allocation (Shadish, Cook, and Campbell, 2002). But for the current topic, lab experiments are neither feasible nor even desirable. To begin with, randomly sampling in the population pool (all high school students in China) was not possible. Furthermore, as Spencer and colleagues (2016) have argued, field experiments may be better positioned to detect ST effect in academic contexts for two reasons: 1) the manipulations taken might have better represented how stereotypes are activated in the real world; 2) the research was conducted on school campus, a real-life setting where ST already exists, which might not have caused participants additional harm.

A common critique of the experimental method, proposed by Hage and Meeker (1988, cited in Cohen *et al*, 2011), was that the experimental settings, instead of the interventions, may be acting causally, so the results 'are largely a function of their context'. This limitation can be overcome by the current designs given the topic under inspection. First, in Study 3, participants performed the exams in the school's computer rooms, so the research setting



would not be significantly different from their actual, familiar English-testing or English-learning environment<sup>10</sup>. Second, ST occurs precisely because the situational cues prime certain negative stereotypes, as shown in Figure 3.3. That is, context itself is a valid determinant of ST. In fact, the manipulation in the experimental group moderately mimicked explicit cues participants may experience in school (see details in Section 4.3.1.1 below).

#### **4.3.1.1 Pilot Study 3**

Pilot Study 3 was conducted to find valid situational cues to activate FALS, which took three stages to develop. 31 participants joined the study. Although demographic information was not collected, they were either graduate students aged between 22-26 (seven females and six males in Stages 1 and 2) or high school students aged between 16-18 (seven females and 11 males in Stage 3).

In Stage 1, I tried to use subtle cues (test instructions) to activate FALS: under the experimental condition, 3 participants read an instruction page containing a gender difference statement and a picture showing three girls and one boys (Appendix M). Under the control condition, on the other hand, another 4 participants read a test instruction displaying a gender-fair statement and a picture with two girls and two boys in it (Appendix N). The rationale was that the gender difference statement and the skewed gender ratio in the picture would have activated FALS, and the gender-fair statement and the gender-balanced picture in the control condition would have nullified any concerns about FALS.

However, in the manipulation check procedure, I found that the gender difference test instruction had failed to active FALS among two participants (as shown in the upper half of Table 4.6). Participant 1 said that she only noticed people's ethnicities in the picture, but completely missed the skewed gender ratio. Similarly, Participant 2 also commented that for Chinese participants, ethnic identities must have been far more salient than gender identities.

In light of this, in Stage 2, I changed the pictures in both types of instructions so that only Asian faces were included. However, this modification failed on a greater scale: all four participants in the experimental group thought that the cues were too subtle to be noticed (see in the bottom half of Table 4.6).

---

<sup>10</sup> According to the school teachers, students in that province are required to take computerised exams of oral English in Grade 12. Thus, participants have had extensive practice with learning and taking tests of English in the computer rooms before taking part in Study 3.

After reviewing relevant ST literature, I decided that more blatant cues should have been utilised to activate FALS. Because men usually feel identity-safe in their social surroundings (Franceschini, Galli, Chiesi, and Primi, 2014), the subtle cues that produces the strongest ST effect on women might not work on men precisely due to their subtlety. Thus, I adopted a more explicit way to prime FALS in Stage 3, where 18 participants joined. Eight participants underwent the experimental condition this time, under which they watched a video named *Dragons and Phoenixes High School*<sup>11</sup>. The video was introduced as an excerpt from a sitcom, and it contained 5 FALS-related notions emerged in Study 2:

**Table 4.6: Results of Manipulation Check in Stages 1 and 2, Pilot Study 3**

Stage	Situational Cues	Participants	FALS Activation
1	a) Gender-different instruction; b) Picture with unequal gender ratio (multiple ethnicity).	1 Female	×
		2 Male	×
		3 Female	√
2	a) Gender-different instruction; b) Picture with unequal gender ratio (Chinese only).	1 Male	×
		2 Female	×
		3 Male	×
		4 Female	×

1. Boys usually do not enjoy English lessons.
2. Boys are not that willing to spend time on tasks that require extra attention and patience, such as practicing English pronunciation.
3. Compared to boys, girls are more gifted in tasks involving reading, writing, or memorising English words.
4. Boys generally do not do well in subjects dealing with textual materials, such as English.
5. English teachers are more likely to be females.

After the video had finished, participants were asked to answer the following question, *The video shows some popular assumptions about boys'/men's behaviour and preferences. Can you write down ALL the assumptions from the video that you can remember?*

<sup>11</sup> I created the video with help from some friends. I wrote the screenplay myself and invited six friends to play in it. The story happens with five high school students and one English teacher, during which FALS is mentioned and discussed in Scene 2 and Scene 3. The link provided has an English subtitle, but the version that participants watched in Pilot Study 4 was with a Chinese one.

Respondents' answers were used as manipulation check. Those in the control group watched a 4-minute excerpt from *Koyaanisqatsi*<sup>12</sup>. Similarly, after the video had finished, they were also asked to answer a question about the video. Their answers were not recorded here because only answers from the experimental group would serve as manipulation check.

**Table 4.7: Results of Manipulation Check in Stage 3, Pilot Study 3**

Participant	Gender	Manipulation Check (Activation of FALS-Related Notions)
1	Male	1, 2.
2	Male	1, 2, 3, 4, 5.
3	Male	1, 2, 3, 4, 5. <i>'Girls usually perform well in English.'</i>
4	Female	X
5	Male	1, 5. <i>'Boys should not learn English.'</i>
6	Male	1, 2, 3, 4. <i>'Girls usually perform well in English.'</i>
7	Female	5. (and two more sentences irrelevant to FALS.)
8	Male	1, 2, 4, 5.

As Table 4.7 here illustrates, FALS-related notions were activated for seven out of eight participants. Participant 2 actually wrote, 'I agree with the two boys' comments about high school subjects. I was just like them.' Participant 4, for whom the manipulation failed, only wrote one sentence, and it was irrelevant to FALS. Given that all other participants wrote down at least one FALS-related notion, it was likely that Participant 4 simply ignored the question's instruction and answered rather randomly. Therefore, it was determined that the video would be a valid cue to activate FALS in Study 3. Yet, to avoid situations similar to Participant 4, in Study 3, the manipulation check was changed into a multiple choices questionnaire (see as the left part of Appendix O).

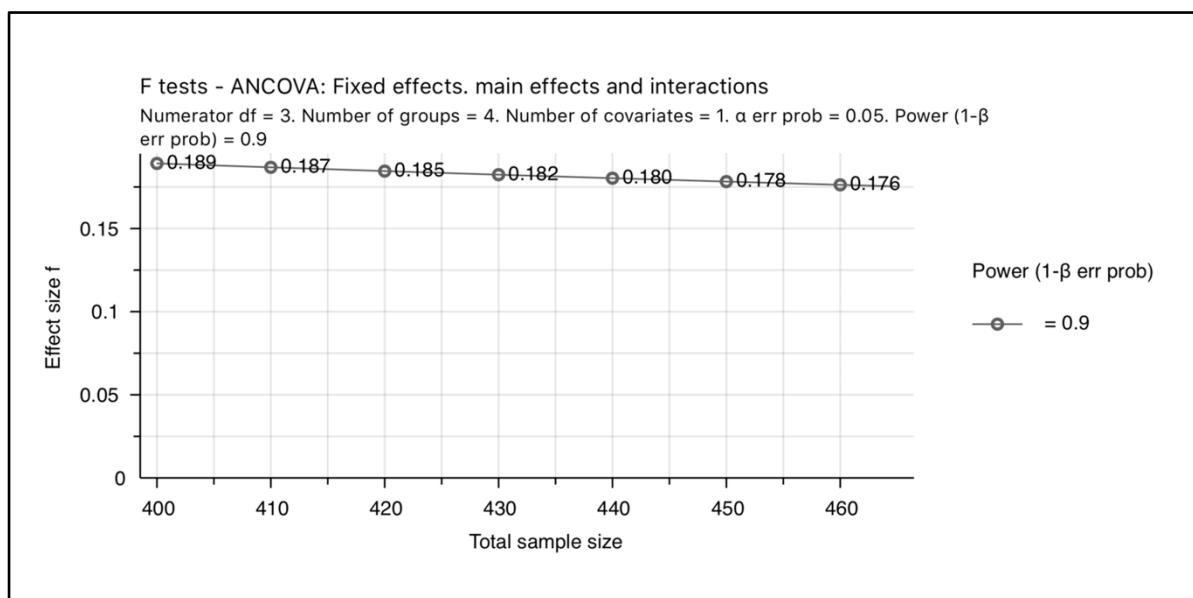
#### 4.3.1.2 Sampling

The sample size required in Study 3 was calculated by G\*Power software introduced by Faul, Erdfelder, Lang, and Buchner (2007) given effect size and  $\alpha$  error probability. The

<sup>12</sup> According to IMDB, *Koyaanisqatsi* is 'a collection of expertly photographed phenomena with no conventional plot. The footage focuses on nature, humanity and the relationship between them'. (<http://www.imdb.com/title/tt0085809/>) It is chosen because it has no plot or characters at all, which will be unlikely to prime any thoughts related to GSs.

effect size was  $f = .18$ , as summarised in Nguyen and Ryan's meta-analysis (2008)<sup>13</sup>. As Figure 4.6 shows, to control  $\beta$  level and maximise statistical power, a minimum sample size would be 440.

**Figure 4.6: Sample Size Needed in Study 3**



#### 4.3.1.3 Procedures

A high school in southern China participated in Study 3. The head-master asked the class-teachers to invite all students in Grade 11 to do an English test (Oxford Online Placement Test<sup>14</sup>, OOPT) on my behalf, while he himself was fully informed of the experiment's genuine purpose and procedures. 464 students signed up (they were told that this was an English placement test) and signed a confidentiality agreement. The agreement ensured them that the test administrator (the researcher) would not disclose their personal information (name, gender, date of birth, and test score) to any third party, and it also asked them to 1) not search for practice materials for the test before they took it and 2) not talk about anything related to the test among themselves before all students finished taking the

<sup>13</sup> According to the meta-analysis, effect sizes of ST vary based on social group membership and stereotype-activating cues used. The meta-analysis examined ST effects of racial stereotypes about overall academic aptitude and social behaviours, as well as those of gender stereotypes about female's inferior quantitative skills. For the current research, I think that the average effect size,  $d = .36$ , found on females performing math-related tasks should be a suitable indicator of possible effect size expected to be found on males performing English-related tasks, because racial stereotypes is not as domain-specific as gender stereotypes.

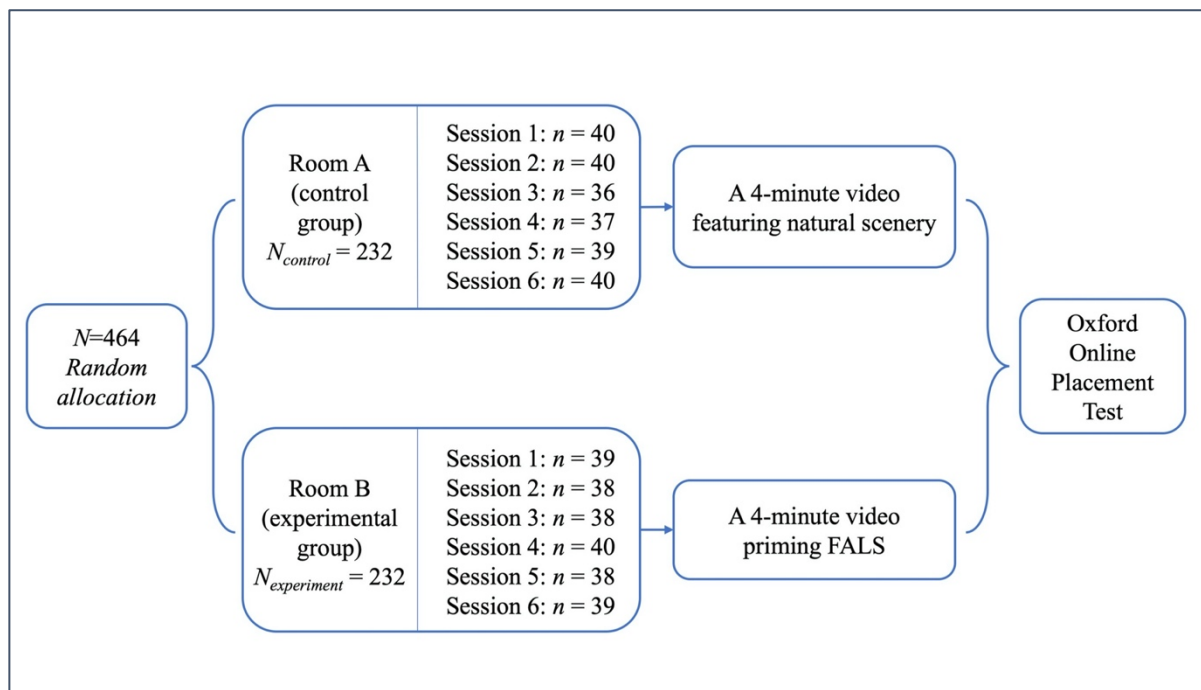
<sup>14</sup> The test was chosen for its proven validity and reliability: according to its developer, Oxford University Press, it 'has been pretested and validated by more than 19,000 students in 60 countries' (<https://www.oxfordenglishtesting.com/defaultmr.aspx?id=3048>). More explanations of the test's scoring scale and its interpretation will be discussed in Section 8.1, where findings from Study 3 are presented.

test. A consent form was distributed to participants' parents after they had signed the agreement, and class-teachers collected the consent form for me.

Participants were randomly assigned to either the experimental or the control group: each student was randomly assigned either 0 or 1 by *Microsoft Excel*. Those with 0s would be in the control group, and those with 1s the experimental group. Participants in the control group were told to go to Room A on the day of the test, and those in the experimental group were told to sit their test in Room B.

The experiment took place in six sessions during two days (due to limited availability of computer rooms and headphones). Figure 4.7 illustrates the procedures of Study 3. On the test day, students went into the designated computer rooms under the instructions of two IT teachers (one in each room), who worked as experimenters. The participants were asked to sit down in front of a computer and watch a 4-minute long video clip. They were told that this was to give them time to relax themselves before the test. Just like Stage 3 of the Pilot Study 3, participants in the experimental watched *Dragons and Phoenixes High School*, and those in the control group watched *Koyaanisqatsi*.

**Figure 4.7: Procedures of Study 3**



Then, participants were told to give their feedback on the video clip by filling out a questionnaire placed on their desk. For those in the experimental group, the questionnaire would be used as manipulation check later. In the questionnaire, participants were asked to tick all the “popular assumptions about boys’ and girls’ differences” presented in the video

(see Appendix O). If none was ticked, the manipulation would be considered a failure. A similar procedure was done for the control group, but only answers from students in the experimental group were recorded, because these would serve as manipulation check.

After the experimenters had collected the questionnaires, students were instructed to take the test and finish within 50 minutes. The researcher waited in the IT teachers' office to be handed over all the questionnaires collected from participants. Since neither the experimenter nor the participants knew which condition they were subject to, potential risk for the experimenter effect was decreased, thereby increasing the internal validity. Finally, when all six sessions had finished, a debriefing session was held by the researcher, informing students of the aim of the study and providing them with information about ST.

#### **4.3.1.4 Data Preparation and Analysis Procedures**

To determine whether ST has affected participants test performance, a 2 (gender)  $\times$  2 (FALS activation) ANCOVA was performed on the test scores participants attained. The covariate was the scores students had obtained in an English exam three weeks ago.

Before performing the ANCOVA, the dataset had to be prepared. Although 464 students volunteered to join the experiment, 6 failed to show up: 1 boys from the experimental group, 4 boys from the control group, and 1 girl from the experimental group. According to the head teacher, one boy was on sick leave and the other five (four boys and a girl) were in another city participating in the National Biology Competition.

Therefore, a total of 458 participants took the test. Among them, 31 were excluded from data analysis for the following three reasons:

- 1) Invalid test score: 2 participants (a boy and a girl) in the experimental group did not finish the test on time and thus did not receive a score.
- 2) Withdrawal: 13 participants withdrew after the briefing session, which included 4 boys from the experimental group, 6 girls from the experimental group, and 3 more girls from the control group. The briefing took place in the evening when the experiment was over. In both the briefing and the confidentiality agreement participants signed when they signed up for the experiment (which was described as an English test), participants were informed that they could ask the researcher more about the experiment if they wanted to and that they could withdraw at any point. Two boys came to me right after the briefing and said that they had wished to withdraw, and another 11 participants waited for almost all other students had left

the auditorium to express their intentions to withdraw. In fact, about a dozen students gathered around me after the debriefing, possibly with an intention to withdraw from the study. Some asked me more about ST and whether this experiment would exert a long-lasting effect on them (although they had been assured that it would be extremely unlikely *during* the briefing). I also tried to explain how the experiment might benefit them (knowledge about ST helps to deal with negative stereotypes, as discussed in Spencer *et al*, 2016). These clarifications and explanations managed to persuade some students to remain in the study, because some students seemed satisfied and left the room. But in the end, still, 13 participants decided to withdraw.

- 3) Failed manipulation: For 16 participants (10 boys and 6 girls), the experimental manipulation failed, so their data had to be excluded from the analysis. There were supposed to be 232 participants in the experimental group, but because 2 did not show up for the experiment, I was supposed to collect 230 questionnaires. In fact, I only got 220: 10 participants told the IT teachers (who were acting as experimenters) that they forgot to fill out the questionnaire prior to the test. And among the 220 questionnaires, 5 were blank: the participants only wrote their names on the questionnaire but did not answer any questions. Additionally, 1 participant wrote on the questionnaire, 'the video made me speechless', without answering any of the questions. For these 16 ( $10 + 5 + 1 = 16$ ) participants, it could be assumed that the manipulation did not work for them, and thus their scores should not be used in data analysis.

To sum up, after deleting all these 31 cases from the original dataset, 427 remained in the dataset and were analysed later.

#### **4.3.1.5 Ethical Considerations**

Measures similar to those taken in Study 1 was used to protect confidentiality and anonymity. In addition, Study 3 took place on campus so that participants' safety and convenience were guaranteed. A debrief was held on the evening that Study 3 had been conducted in the school's auditorium, during which participants were revealed the true intention of the study (to examine ST effect on male language learners), in addition to information about gender stereotypes, ST, and how to manage negative stereotypes. In the end, participants were reassured that they could withdraw from Study 3, and 13 did. Participants' questions and concerns were answered during and after the debrief.

There were two major ethical concerns. The first was the slight deception involved in the instructions participants received prior to tasks. Participants were told that they were invited to take an English test, but it was in fact to determine ST effects. This level of deception was justified for two reasons. First, Research (Spencer *et al*, 2016) has shown that knowledge about ST, explained in the debrief, can effectively reduce the negative effects stereotypes have on targets, so participants may receive long-term benefits in their academic performances after participating in the research. Second, the head-master have already been informed of the true intent and been acting as the gatekeeper of the research. If he/she deemed that there was more harm than good, he/she would have demanded that the study be stopped.

Another ethical concern lies with the FALS video. Since the video implies that the male gender is somewhat incompatible with the English subject, it was possible that unpleasant feelings were caused among male participants who highly identified with English. This concern was addressed in three ways. First, the FALS video largely resemble the stereotypical assumption students already know of, so the experiment could not have caused much additional harm. Second, the participants had been told that they could withdraw at any point of the experiment, so if they did feel hurt by the video, they could have opted out. Third, participants were fully debriefed about ST after the experiments had ended, so the possibility of longer-term harm would have been reduced.

#### **4.3.2 Study 4: Group Interviews**

The primary goal of Study 4 was to aid interpretation of results from Study 3. Semi-structured interviews with eight groups of three students, instead of individual interviews, were utilised to address the issue with power asymmetry between the interviewer and the interviewees. This asymmetry looms large particularly in one-to-one interviews, as many qualitative researchers have warned, because the interview process can sometimes become a one-directional, instrumental, or even exploitative dialogue where conversations are reduced to a means for providing narratives whose interpretation are monopolised by the interviewer/researcher (e.g., Brinkmann, and Kvale, 2015; Creswell, and Planto Clark, 2011).

For the current research, the interviewees were students from the same high school in Study 3. They had an average of 8.96 years' experience with English learning (see more demographic information in Table 8.2 in Section 8.2). Contrastingly, the interviewer, me myself, is a PhD. candidate in an elite English university with over 16 years of English learning and three years of teaching experiences. As a result, during the interview,



interviewees might have had qualms about disclosing their feelings and experiences to me, a somewhat authoritative figure regarding English from their perspectives. If individual interviews had been held, the risk of an off-balance power relationship would have been exceedingly high—interviewees could have blindly followed my pre-set interview guide, leading to trivial and predictable findings. Thus, in order to relieve students' concerns, to allow them to be actively engaged, and to provoke them respectfully to raise potentially conflicting opinions, group interviews were employed. During the interviews, students were able to express their thoughts in the company of their peers. Meanwhile, they were also encouraged to share, discuss, and possibly debate among themselves. This way, epistemological issues about the knowledge produced in interviews could be handled by allowing the researcher to follow the polyvocal and even contradictory meanings emerging through interviewees' accounts.

Sections 4.3.2.1-2 will explain how the power asymmetry was approached during the design of the interview schedule, the sampling procedures, and the interview process. Then, Sections 4.3.2.3-4 will focus on data analysis and ethical considerations of the study.

#### **4.3.2.1 Developing Interview Schedule**

The interview schedule took two steps to develop. Version 1 (Appendix P) was tried in Pilot Study 4, where a group of three<sup>15</sup> girls and another of three boys joined. It contained four major topics: in Part 1, interviewees were asked to describe their general experiences with and feelings about English learning. As the questions here were mostly descriptive and retrospective, it was anticipated that a rapport could be built between the interviewees and the interviewer. The next two parts invited students to discuss what qualities/characteristics a successful learner of English might possess, and how they would compare boys with girls as English learners. These two parts were used to explore their FALS awareness and endorsement. Finally, Part 4 asked about the circulation and influence of FALS. Based on feedback from the pilot interviews, the following changes were made in Version 2 (Appendix Q).

1) Format. A designated place for note-taking was created for each interviewee in the same group in Version 2, because it was found hard to keep track of who said what using Version 1 during Pilot Study 4, which had only one column for notes.

---

<sup>15</sup> The reasons for the group size will be discussed in the next section.

2) Introduction. In both versions, the interview started with explaining the aim of the research, the ways interview data would be recorded and transcribed, and measures to protect interviewees' confidentiality and anonymity. The following sentence was added in Version 2,

*'During the discussion, if you agree with someone else, you can give examples and reasons to support your shared opinion; but if you do not have a shared opinion about a certain topic, it is okay to have a debate among yourselves.'*

The aim was to encourage discussions and debates. The revision was made because in Pilot Study 4, turn-taking, instead of group discussion, became the dominating pattern. It seemed common for interviewees to neglect their peers' speech while focusing on preparing their own when it was their turn. It might have been that they treated the group interview as three one-to-one interviews compressed together, instead of an opportunity to share, challenge, or even contradict each other.

3) Sequence of questions/probes in Part 3. In Version 1, interviewees were asked if they thought there were gender differences in English achievement, aptitude, and affect. They also were encouraged to list ways they thought boys and girls differ. In the feedback session after Pilot Study 4, one boy also said that the former question limited his answer for the latter, 'It was clear from the first part that you were interested in those three aspects. We've answered your questions about them, and then when you ask us to freely comment on how boys and girls differ, we just do not know what else to say.' In addition, two girls and one boy felt that the latter question was made rather redundant given the former one. Based on these remarks, the sequence of the questions was reversed.

#### **4.3.2.2 Sampling and Interviewing**

Purposive sampling method was used in Study 4, where 24 interviewees took part. When making decisions about sampling, literature on focus groups was consulted on the premise that issues about group size and composition, and group number would have been similar for group interviews and focus groups. Regarding group size, Morgan's suggestion (1998) was six to ten members, which seemed neither feasible nor desirable for the current study. According to the head teacher (the same one from Study 3), students had seven classes during daytime and the interviews could only be scheduled between 18:20 and 22:00, during which time the evening self-study sessions were held. The class teachers of the school insisted that no more than five students would be allowed to skip the session for around an hour, considering the disruptions for the rest of the class when they exited and entered the

classrooms. In addition, it would be even harder to maintain an engaging and lively conversation for everyone if the group size was too big. In fact, Sharp (2009) advised that smaller groups would work better for teenagers. Thus, it was decided that a group of three students and one interviewer would be reasonable. The idea was tried in Pilot Study 4, and all six participants from two groups agreed that the group size felt appropriate and comfortable for them.

A total of eight groups were recruited, with equal numbers of boys and girls: the head teacher announced to students in Grade 11 that they were invited to a roundtable discussion on English learning with an expert (which referred to me). Students were asked to sign up as groups of three, and all three should have been of the same gender. A consent form similar to the one used in Study 2 (Appendix L) was distributed to the parents of volunteering students. After the parents had signed the forms, class-teachers collected the consent form on my behalf.

Gender was used as the stratifying criterion to form homogenous groups. This was an obvious decision, considering that the research was about gender stereotypes. This sampling procedure made sure that boys' and girls' voices were represented well-balanced. Additionally, by comparing and contrasting interviewees' perceptions of their own gender and perceptions of them from the opposite gender, an opportunity for triangulation emerged.

The interviews took place in a meeting room in the school. After gaining interviewees' approval, the interviews began and were audio-recorded. All interviews were conducted in Mandarin Chinese. During the interviews, similar probes in Study 2 were used to build rapport and ask for clarifications. All interviews were later transcribed verbatim by the researcher. The same system of transcription notes from Study 2 was used (see Table 4.4 in Section 4.2.2.3).

#### **4.3.2.3 Data Preparation and Analysis Procedures**

Similar to Study 2, thematic analysis was applied to the data in Study 4 using the software programme *MAXQDA2018*. The analysis began with reading the transcripts while re-listening to the audio recordings. After sufficient familiarisation with the data, the data was analysed mainly through structural and provisional coding methods (see definitions of the two methods in Table 4.5 in Section 4.2.2.4). Figure 4.8 illustrates the total of 513 coded segments generated from this initial coding cycle. Two more coding cycles were conducted, using process coding, causation coding, emotion coding, value coding, and versus coding

methods (details in Table 4.5, too). Once this was done for all interviews, the third procedure was cross-case synthesising—reviewing each code for commonalities and contrasts across interviewees. During the process, less irrelevant codes were removed from the analysis and ones alike were combined (Brinkmann and Kvale, 2015).

Next, these codes were reviewed so that an initial understanding of the data was achieved. The final two steps were to test the emergent understanding and to search for rival explanations. This process generated the over-arching theme of “students’ emergent understanding of *gender*” with three inter-related superordinate themes: 1) the widespread FALS; 2) approaching FALS critically; and 3) GSs accompanying FALS. Chapter 8 will explain the findings in full.

In order to enhance authenticity and credibility, member-checking was proposed to the interviewees after the interview had been done: they agreed to review the major findings and left their contact details. As Seale (1999) has summarised, authenticity refers to the research’s ability to report different realities, to help participants appreciate others’ viewpoints, and to improve participants’ understanding of the researched phenomenon. However, when I sent back the themes and major findings to the participants, none responded. As a result, two additional strategies were employed to enhance credibility when reporting findings, including entertaining alternative explanations and citing rich descriptions.

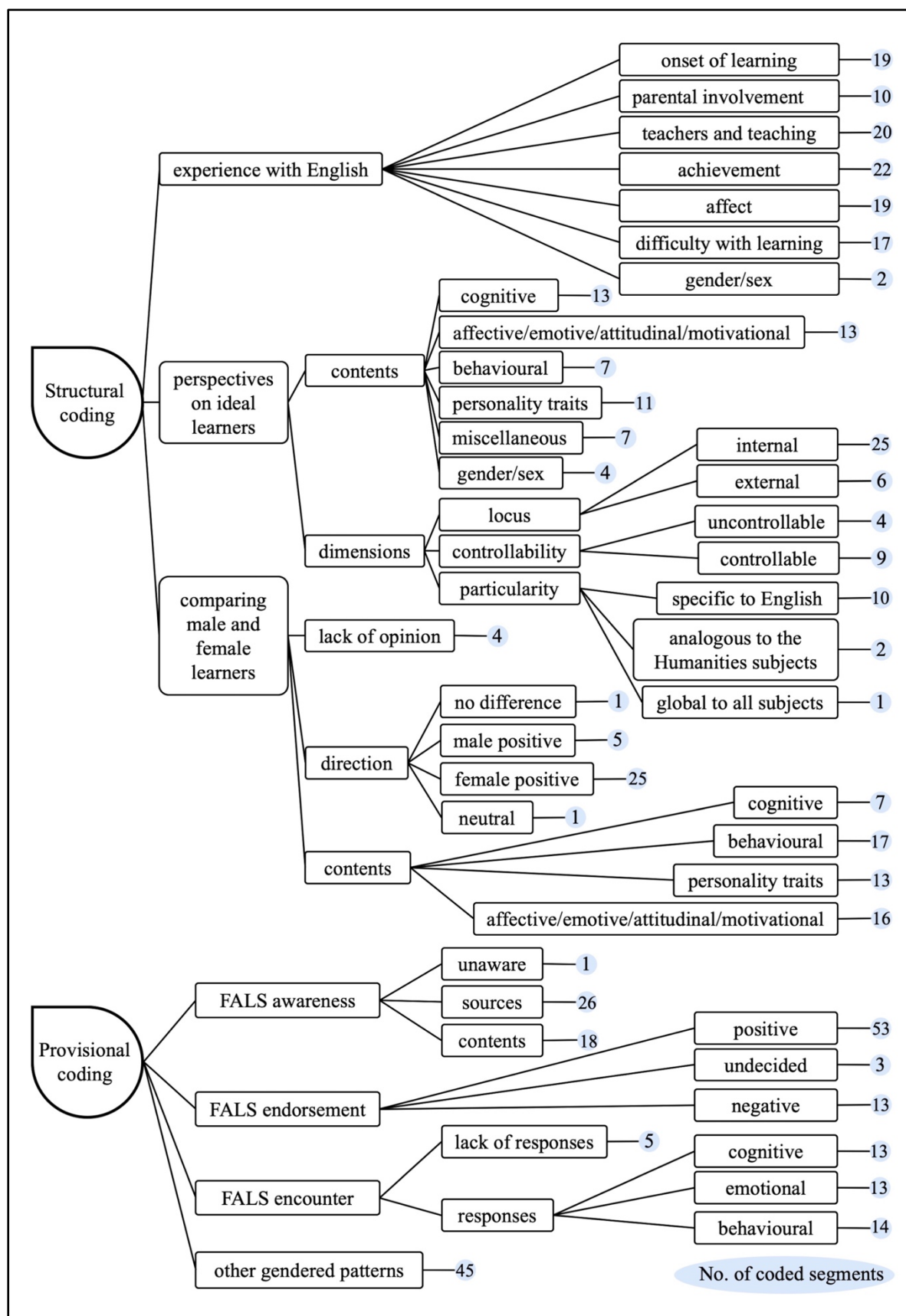
#### **4.3.2.4 Ethical Considerations**

At least a week prior to the interview, interviewees and their guardians were asked to sign two copies of the consent form and return one copy right before the interview started. In the form, measures to protect confidentiality and anonymity were explained for the interviewees. In addition, the safety and convenience issues of the participants were taken into account when deciding the places where interviews were conducted. After each interview session, a debrief was held immediately for the group of interviewees. The information provided was similar to the debrief in Study 3: gender stereotypes, ST, and how to manage negative stereotypes, reassurance that they could withdraw if they had wanted to. Interviewees’ questions were also answered in the debrief.

#### **4.3.3 Integrating Data from Studies 3 and 4**

Data in Studies 3 and 4 were analysed separately and presented sequentially in Chapter 8. In addition, Section 8.3 compared and contrasted findings from both studies to further illustrate the role of gender stereotyping in English classrooms.

Figure 4.8: First-Cycle Coding System in Study 4



## Chapter 5: Findings and Discussions I —

### Perceiving Girls as Better Learners of English by Guardians and Students

This chapter, together with the following two chapters, will present results from Phase 1. The current chapter will focus on findings based on the guardian and student datasets in Study 1, and Chapters 6 and 7 will proceed to the findings from two teacher datasets—the questionnaire survey in Study 1 and the interviews in Study 2.

Combined together, Studies 1 and 2 answered the first research question: female learners *were* stereotypically regarded as better learners of English in Chinese high schools. It was found that the female-advantage-in-languages stereotype (FALS) was endorsed to varying degrees by all three groups of participants: the guardians, the students themselves, as well as their teachers of English. Sections 5.1-2 below will delineate the analyses of the guardian and the student datasets from Study 1 separately, testing the two hypotheses that guardians and students endorsed FALS. Then, Section 5.3 will examine the guardian and the student datasets jointly to identify if generation gaps existed regarding FALS endorsement. Given that MANOVA procedures were performed on guardian and student datasets (macro level items) twice here, in order to control over the Type I error rate, a lower  $\alpha$  level (.025) for the multivariate *F*-tests was adopted (Tabachnick and Fidell, 2014). Finally, Section 5.4 will summarise guardians' and students' FALS endorsement.

#### 5.1 Guardians' Perspectives from Study 1

A total of 2,498 questionnaires were collected from guardians. After a three-step screening process (see Section 5.1.1 below), 1,904 responses (76.22% of the guardian sample) were deemed genuine and thus were retained for analysis. Section 5.1.2 displays the demographic information of this guardian dataset.

A three-way mixed MANOVA was performed on the dataset, and the findings will be elucidated in Section 5.1.3. From Section 5.1.3.1 to 5.1.3.4, whether or not guardians endorsed FALS and whether their endorsement differed on the basis of participant gender and region will be documented<sup>16</sup>. Finally, in Section 5.1.4, a summary of the findings from the guardian dataset will be provided.

---

<sup>16</sup> For the three datasets in Study 1, multiple results were generated from the MANOVA procedures. Yet, only those related to the hypotheses listed in Figure 4.2 are reported in the thesis (i.e., in the current chapter and Chapter 6) due to limited space and relevance to the first research question (RQ1). For example, a 2 (guardian gender)  $\times$  2 (region)  $\times$  2 (learner gender) mixed MANOVA was performed on the guardian dataset. Because only the effects involving learner gender (e.g., the main effect of learner gender, the interaction effect between

### 5.1.1 Data Screening

Before analysing the guardian dataset, each response was assigned a case number, and a three-step data screening procedure was utilised to identify three types of careless responses, as listed in Appendix R. The first type is the responses made by self-identified inattentive respondents. Moreover, participants who answered all items with one single answer or with random answers were also considered absent-minded, which became the second and third categories. After this process, 594 responses were deemed unengaged and thus were excluded from the subsequent MANOVA procedure.

### 5.1.2 Demographic Information of Guardian Participants

This section sums up the background characteristics of guardian respondents, including gender, relationship with participating students, age, education, ethnicity, region, and occupation. To understand the representativeness of the guardian dataset, certain demographic features are compared with those of guardians participating in the China Education Panel Survey (CEPS) wherever data are available from CEPS<sup>17</sup>. CEPS was used because of its national representativeness: it recruited over 20,000 7<sup>th</sup> and 9<sup>th</sup> graders (in the 2014-15 academic year) and their guardians from 28 districts randomly chosen across China. By the time Study 1 was conducted (2017), these students from CEPS became 10<sup>th</sup> and 12<sup>th</sup> graders in high schools. That is, guardian and student participants from CEPS can be regarded as nationally representative of the corresponding cohorts involved in Study 1.

**Gender.** 965 respondents are female (50.7% of the guardian sample), and the remaining 939 are male (49.3%). The proportion of male and female guardians are similar, indicating that the survey results would be capable of reflecting voices from both sides equally.

**Relationship.** 220 participants did not specify what the relationship was between them and the participating students. Among those who did, the majority are parents (90.3%), a proportion similar to that in CEPS (93.3%). The rest are either grandparents (1.4%) or other adult relatives (7.9% in total, including: adult brothers or sisters, 6.1%; uncles or aunts, 1.5%; adult cousins, 0.3%). In CEPS, the percentage of grandparents are higher (4.9%), but that of other relatives are lower (1.9%).

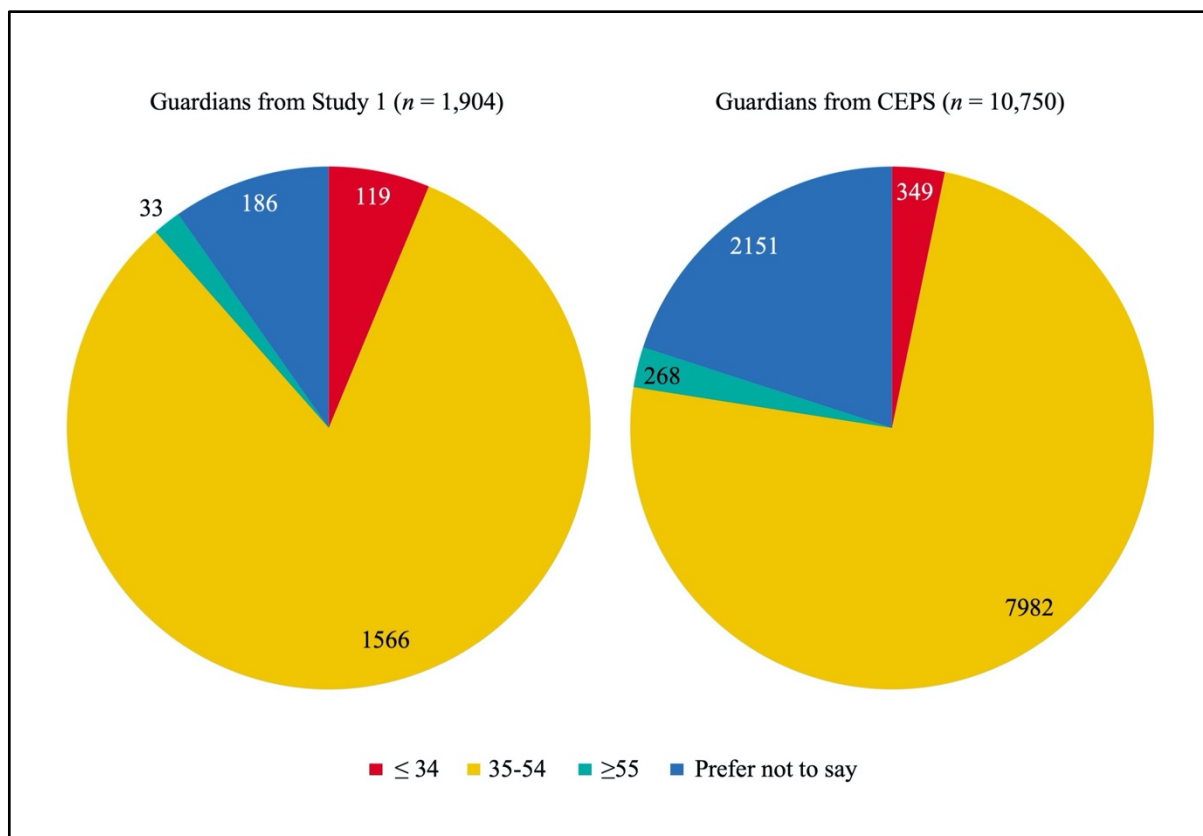
---

learner gender and region, etc.) answered RQ1, other effects (e.g., the main effect of guardian gender, etc.) are excluded from this thesis.

<sup>17</sup> Raw data were obtained from <https://ceps.ruc.edu.cn/index.php?r=index/index>. All percentages of relevant demographic information in CEPS reported were calculated by the researcher using the raw data.

**Age.** The left part of Figure 5.1 shows the age make-up of guardians in Study 1. Apparently, the proportions of age groups conform with those of different relationships documented in the previous paragraph: among those who reported their age, 91.2% were between 35 and 54 years old, an age group that is likely to be parents<sup>18</sup>, or adult uncles and aunts of high school students (around 91.8% of the guardian sample). In addition, guardians aged 55 and over take up 1.9%, which matches the proportion of grandparents (1.4%). Moreover, 6.9% of guardians are under 34 years old, which approximates the percentage of adult brothers, sisters, and cousins (6.4%). Such a correspondence between the two demographic characteristics offers evidence for the authenticity of the data by indicating that participants here answered the two questions regarding relationship and age truthfully.

**Figure 5.1: Age Make-Ups of Guardians from Study 1 and the China Education Panel Survey**



The right part of Figure 5.1, additionally, illustrates the age groups present in CEPS. Clearly, in both datasets, the majority were aged between 35 and 54. Additionally, in accordance with the disparities between the two datasets regarding relationship, the

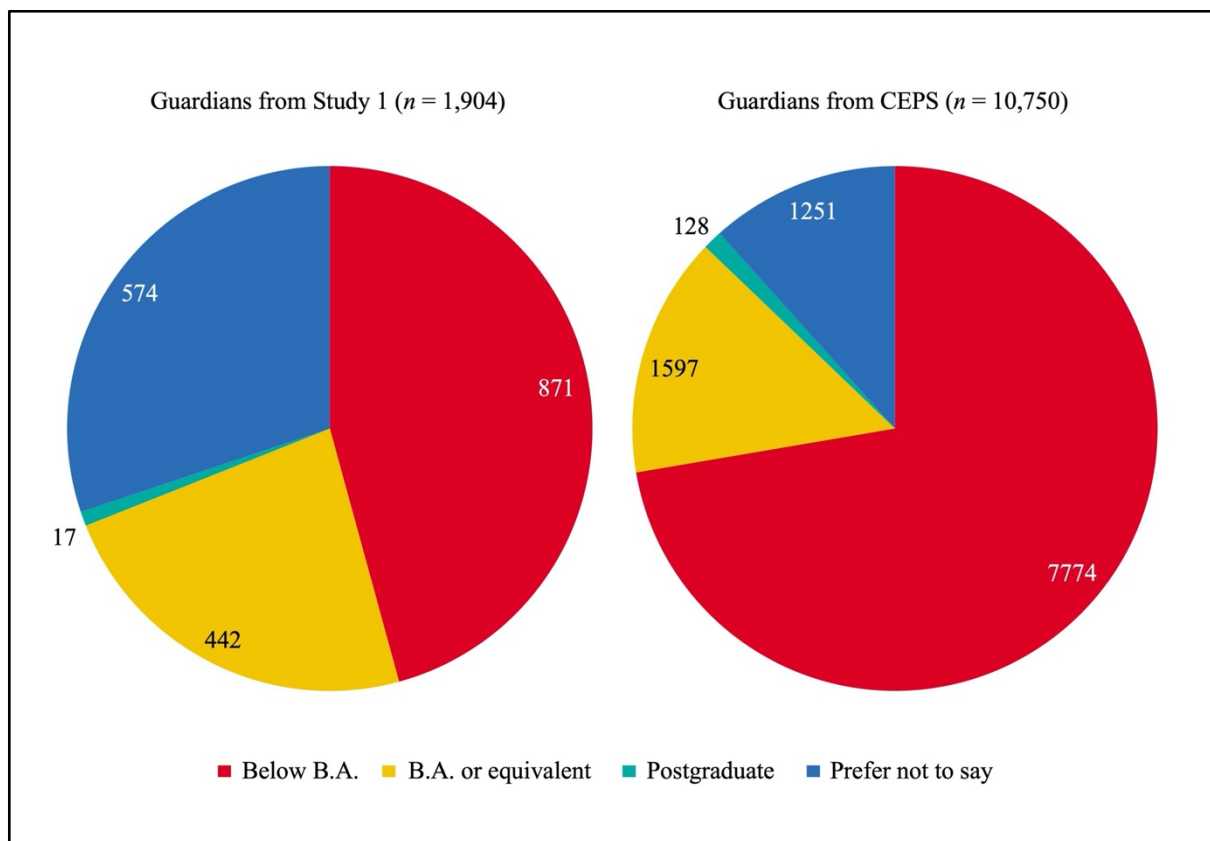
<sup>18</sup> According to the Marriage Law of the People's Republic of China (<https://www.cecc.gov/resources/legal-provisions/marriage-law-of-the-peoples-republic-of-china-amended>), the youngest legal age to get married in China is 20 years old for females and 22 for males, which means that the youngest parent of a 10<sup>th</sup> grader (typically aged between 15-16) should be 35.



proportion of guardians aged over 55 (likely to be grandparents) were higher in CEPS (3.1%); yet, that of participants under 34 years old (other adult relatives) are lower in CEPS (4.1%).

**Education.** Figure 5.2 reveals that compared to the sample of CEPS, the guardians in Study 1 seems relatively more highly educated: in Study 1, only 65.5% of those who specified their educational qualifications did not go to college, but that percentage in CEPS was higher (81.8%). Furthermore, 34.5% of participants in Study 1 are educated to at least college level, while only 18.2% in CEPS are so.

**Figure 5.2: Education Backgrounds of Guardians from Study 1 and the China Education Panel Survey**



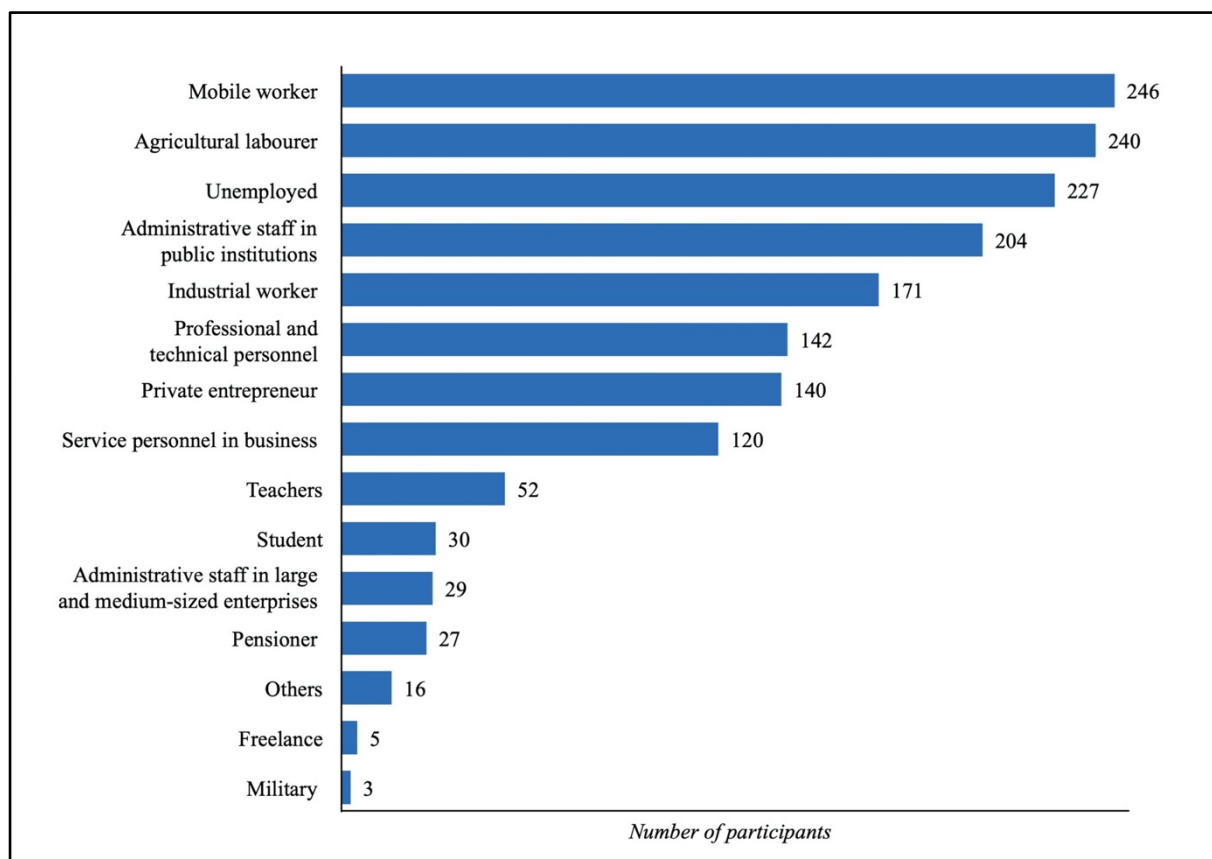
**Ethnicity.** 609 guardians (32%) refused to indicate their ethnicity. Among those who did, Han people account for 80.6%, which was representative of the Chinese population, of which 91.46% are Han<sup>19</sup>. The remaining 19.4% include Gaeml Minority (14.2%), Miao Minority (2.9%), Hui Minority (1.5%), and others (0.9%).

<sup>19</sup> The figures are obtained from the sample survey data of 1% population in 2015, published by the Chinese National Bureau of Statistics ([http://www.stats.gov.cn/tjsj/zxfb/201604/t20160420\\_1346151.html](http://www.stats.gov.cn/tjsj/zxfb/201604/t20160420_1346151.html)).

**Region.** 1,142 guardians (60.0%) are from the north, and the rest are from the south. This is somewhat different from the national profile, where 41.86% are northerners and 57.93% are southerners<sup>20</sup>. In fact, approximately similar numbers of questionnaires were distributed to the two regions (53.8% to the north and 46.2% to the south, see exact numbers in Table 4.3), but the return rate of southern guardians was lower (46%) than their northern counterparts (56.0%). This divergence might have resulted in the slight imbalance in regional distribution.

**Occupation.** Figure 5.3 here summarises the occupations of the guardians who indicated their job in Study 1. Clearly, a range of occupations are represented here, where mobile workers, agricultural labourers, and administrative staffs in public institutions are the top three categories, accounting for 43.2% of the dataset. This was similar to the national profile, where approximately 53.1% of the national population are in these fields<sup>21</sup>.

**Figure 5.3: Occupations of Guardian Dataset**



<sup>20</sup> The figures are calculated based on the 2018 China Statistical Yearbook (<http://www.stats.gov.cn/tjsj/ndsj/2018/indexeh.htm>).

<sup>21</sup> The percentage here is calculated based on the 2017 China Labour Statistical Yearbook (<http://tongji.cnki.net/kns55/navi/YearBook.aspx?id=N2018070151andfloor=1>). It is worth noting that although the yearbook was published in 2017, the most up-to-date statistics about occupations were recorded in 2002.

### 5.1.3 Findings from Three-Way Mixed MANOVA

A 2 (guardian gender)  $\times$  2 (region)  $\times$  2 (learner gender) mixed MANOVA was performed on three dependent variables (achievement, aptitude, and affect ratings), where guardian gender and region were between-subject factors, and learner gender was the within-subject factor. This statistical procedure addressed the first hypothesis in the analytical framework of Study 1 (see Figure 4.2), the hypothesis that guardians endorsed FALS on the macro level, and that the endorsement might have varied due to their gender and region.

Statistical assumptions were checked. There was a linear relationship between the dependent variables, as assessed by scatterplot, and no evidence of multicollinearity, as assessed by Pearson correlation ( $|r| < 0.9$ ). There were no univariate outliers in the data, as assessed by inspection of boxplots, but 21 multivariate outliers were detected using Mahalanobis distance ( $p > .001$ ). A MANOVA without these outliers was run and the results from both analyses were practically the same, so it could be assumed that the outliers have exerted little influence over the results. Thus, to preserve statistical power, results from the whole dataset will be reported here.

Graphical assessment of the normality by chi-square versus Mahalanobis distance plot suggested that the variables did not form a multivariate normal distribution in three cells of the design (female guardians from the north rating girls, female guardians from the south rating girls, and male guardians from the north rating girls). Considering that MANOVA is fairly ‘robust’ to deviations from normality with respect to Type I error (Bray and Maxwell, 1985), the analysis was carried out regardless of this violation. After all, Weinfurt (1995) notes that in practice MANOVAs tend to be performed even if the data is not normal due to a general consensus that MANOVA is robust to non-normality.

There was homogeneity of covariance matrices, as assessed by Box's M test ( $p = .001$ ), and homogeneity of variances, as assessed by Levene's Test of Homogeneity of Variance ( $p > .05$ ).

#### 5.1.3.1 Hypothesis 1.1: Guardians Endorsed Female-Advantage-in-Languages Stereotype

There was a statistically significant main effect of learner gender on the combined dependent variables,  $F(3, 1795) = 576.358, p < .001$ , Wilks'  $\Lambda = .509, \eta_p^2 = .491$ . Therefore, univariate main effects of learner gender were examined for all three dependent variables, and as Table 5.1 shows, the effects were statistically significant for all three dependent variables.

**Table 5.1: Univariate Main Effect of Learner Gender in Guardian Dataset**

Stereotype component	Learner gender	Mean	SD	<i>F</i> (1, 1797)	<i>p</i> -value	$\eta_p^2$
Achievement	Female	5.000	1.144	1262.130	< .001	.413
	Male	3.622	1.146			
Aptitude	Female	5.012	1.133	1124.572	< .001	.385
	Male	3.746	1.149			
Affect	Female	4.941	1.154	1206.583	< .001	.402
	Male	3.621	1.147			

**Figure 5.4: Achievement, Aptitude, and Affect Ratings as a Function of Learner Gender in Guardian Dataset**

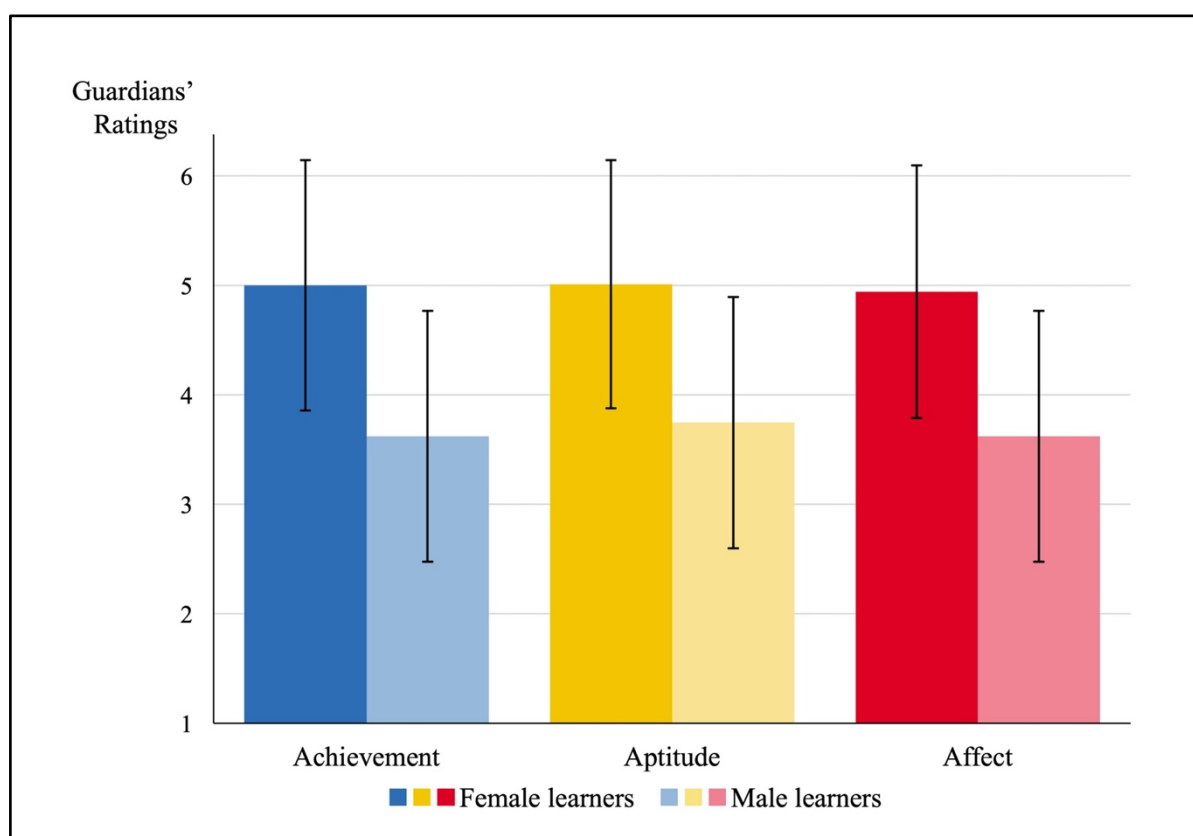


Figure 5.4 here illustrates parents' differential ratings for boys and girls. From the means in Table 5.1, as well as Figure 5.4, it is clear that guardians thought that female learners were better than their male counterparts in English achievement, aptitude and affect. Of course, the standard deviations implied that there was variability in guardians' perspectives (more discussions about this in Section 9.1.3.2), but the above-mentioned evidence did support Hypothesis 1.1.

### 5.1.3.2 Hypothesis 1.2: Guardians' Endorsement Differed Due to Gender

The two-way interaction effect between guardian gender and learner gender on the combined dependent variables was not statistically significant,  $F(3, 1795) = 1.720, p = .161$ , Wilks'  $\Lambda = .997, \eta_p^2 = .003$ . Because any interaction effect is supposed to be symmetrical, this statistically insignificant interaction might be construed as both the effect of guardian gender being dependent on learner gender, and vice-versa. Although either or both interpretations could be allowed, the choice of one or both is essentially a theoretical and conceptual issue. Because the primary concern of Study 1 was the effect of learner gender, an effect suspected to depend on whether a participant was male or female in Hypothesis 1.2. Thus, participant gender was deemed to have moderated the effect of learner gender on guardians' ratings<sup>22</sup>. In other words, the result above was viewed as evidence that male and female guardians did not differ in their endorsement of FALS. Therefore, Hypothesis 1.2 was not supported.

The univariate interaction effects between guardian gender and learner gender for each of the dependent variable were not followed, due to the insignificant multivariate interaction effect.

### 5.1.3.3 Hypothesis 1.3: Guardians' Endorsement Differed Due to Region

The two-way interaction effect between region and learner gender on the combined dependent variables was statistically significant,  $F(3, 1795) = 4.968, p = .002$ , Wilks'  $\Lambda = .992, \eta_p^2 = .008$ , which meant that guardians from the north differed in their endorsement of FALS from those from the south. In other words, Hypothesis 1.3 was supported.

Follow-up univariate two-way ANOVAs were run. These generated a statistically significant interaction effect between region and learner gender for achievement rating,  $F(1, 1797) = 5.808, p = .016, \eta_p^2 = .003$ , for aptitude rating,  $F(1, 1797) = 13.950, p < .001, \eta_p^2 = .008$ , and for affect rating,  $F(1, 1797) = 8.033, p = .005, \eta_p^2 = .004$ .

As such, a simple main effect analysis of region on each level of learner gender was conducted for three dependent variables. The relevant statistics are shown below in Table 5.2. Apparently, there was a statistically significant difference between regions for male learners' aptitude and female learners' affect (marked with an asterisk mark). On the contrary, the difference between regions was not statistically significant for female learners' achievement, male learners' achievement, female learners' aptitude, or male learners' affect.

---

<sup>22</sup> The same logic applies to all interaction effects involving learner gender from MANOVAs reported in this chapter, as well as in Chapter 6.

Therefore, simple pairwise comparisons were run for the differences between regions in mean aptitude ratings for male learners and in mean affect ratings for female learners. For male learners' aptitude, there was a statistically significant mean difference of .212, 95% CI [.105, .319],  $p < .001$  (see descriptive statistics in Table 5.2). Considering that the ratings for female learners' aptitude guardians gave were essentially the same regardless of their region, this mean difference may suggest that the perceived gender gap in aptitude was larger from the perspective of southern guardians, as depicted by the left half of Figure 5.5.

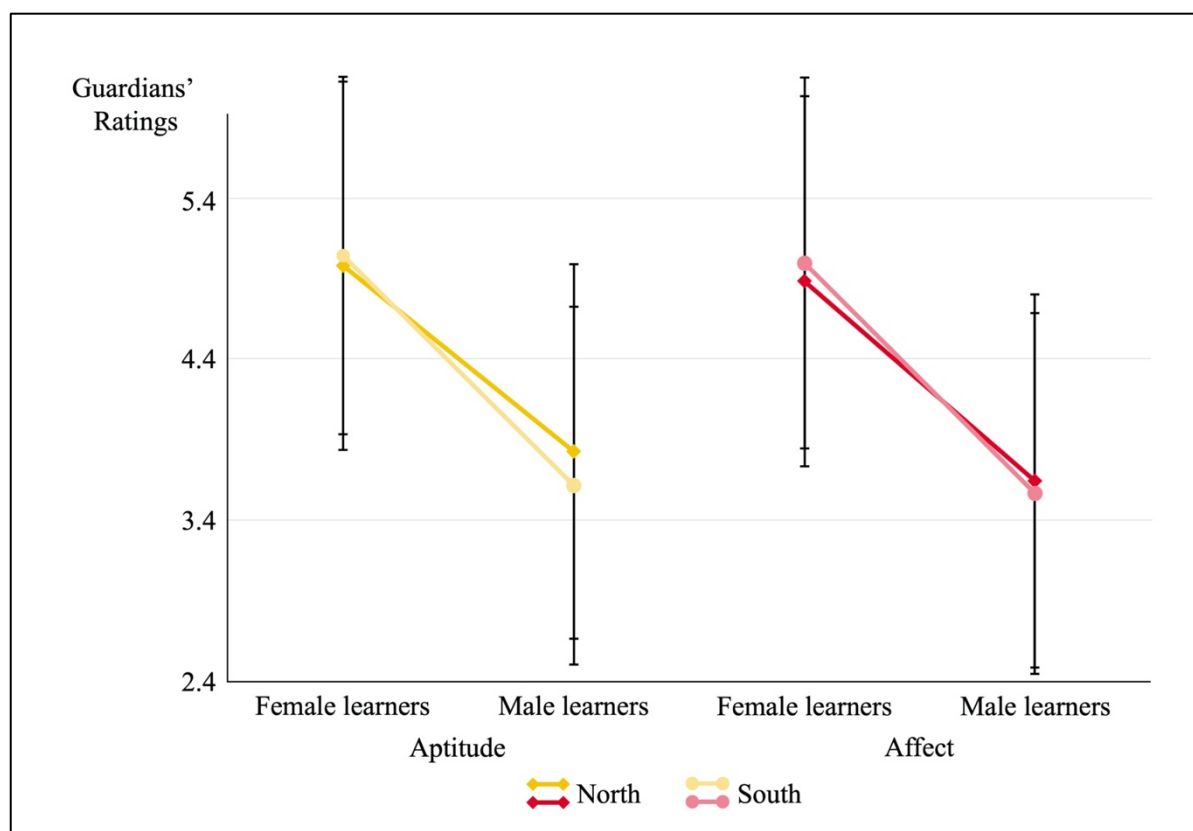
**Table 5.2: Simple Main Effect of Region on Learner Gender in Guardian Dataset**

Stereotype component	Learner gender	Mean (SD)		$F(1, 1797)$	$p$ -value	$\eta_p^2$
		North	South			
Achievement	Female	4.972 (1.160)	5.036 (1.120)	2.532	.112	.001
	Male	3.657 (1.172)	3.570 (1.103)	3.361	.067	.002
Aptitude	Female	4.987 (1.146)	5.050 (1.113)	1.620	.203	.001
	Male	3.831 (1.165)	3.619 (1.114)	16.223	< .001*	.009
Affect	Female	4.895 (1.152)	5.008 (1.154)	5.502	.019*	.003
	Male	3.652 (1.161)	3.575 (1.123)	2.518	.113	.001

For female learners' affect, similarly, there was also a statistically significant mean difference of -.113, 95% CI [-.222, -.004],  $p = .019$  (descriptive statistics also in Table 5.2). Thus, this mean difference might indicate that a wider gender gap in affect was assumed by guardians from the south. The right half of Figure 5.5 show the affect ratings as a function of region and learner gender.

So far, both pairwise comparisons indicated that the gender-stereotypical gaps in aptitude and affect for English were larger from the eyes of southerners than northerners. These regional variations in FALS endorsement might have arisen because of the more task-orientated sub-culture in the south (Huo & Randall, 1991). As Hofstede (2011) has pointed out, people in task-orientation societies tend to harbour rigid gender stereotypes more than those in person-orientation ones. Nevertheless, for both northerners and southerners, the stereotypical belief that 'English is not a masculine domain' is still strong, because both gave higher ratings for female learners in terms of English aptitude and affect. In fact, these differences between regions were marginal in practical terms, as attested by the small effect sizes ( $\eta_p^2 = .009$  for aptitude and .003 for affect) and the considerable overlaps in ratings made by southern and northern guardians (see Figure 5.5).

**Figure 5.5: Aptitude and Affect Ratings as a Function of Region and Learner Gender in Guardian Dataset**



#### 5.1.3.4 Hypothesis 1.4: Guardians' Endorsement Differed on the Basis of Guardian Gender and Region

The three-way interaction effect between guardian gender, region, and learner gender on the combined dependent variables was not statistically significant,  $F(3, 1795) = .519, p = .669$ , Wilks'  $\Lambda = .999, \eta_p^2 = .001$ . The result indicated that there were no two-way interactions between guardian gender and region that varied across the levels of learner gender, thus no follow-ups tests were done. That is, Hypothesis 1.4 was not supported.

#### 5.1.4 Summary of Guardian's Perspectives

All in all, it was found that regardless of their own gender, guardians stereotypically thought female learners were better than male ones in terms of English achievement, aptitude, and affect. Nevertheless, regional differences in FALS endorsement were discovered: in terms of aptitude and affect, guardians from the south perceived a slightly wider gender difference than those from the north.

### 5.2 Students' Perspectives from Study 1

Altogether, 1,298 questionnaires were collected from students from eight high schools. However, 35 questionnaires were blank, and thus only 1,263 went through a four-step

screening process (see Section 5.2.1 below), through which 304 careless responses were detected and thus excluded from later analysis. Section 5.2.2 displays the demographic background of the student dataset. Then, two four-way mixed MANOVAs were performed on the dataset, one on the macro level (presented in Section 5.2.3) and another on the micro level (exhibited in Section 5.2.4). Finally, in Section 5.2.5, a summary of the findings from this student dataset will be provided.

### 5.2.1 Data Screening

Similar to the case of the guardian dataset, data screening procedures were applied to the student dataset prior to data analysis. In total, four categories of careless responses were identified: the first three were identical with the ones in the guardian dataset, which were responses with low amount of attention, uniform response patterns, and random response patterns. The last category was responses that answered the bogus item wrongly (see details of data screening procedure in Appendix S). This screening process left 959 effective responses (75.93% of the student sample) for the later MANOVA.

### 5.2.2 Demographic Information of Student Participants

The demographic characteristics of student respondents, including gender, ethnicity, age, grade, and academic branch are summarised here. Unlike the case with the guardian dataset, CEPS will not be referred to here, either because of the availability of national statistics (in terms of gender and academic branch), or because of the incomparability between the senior high schoolers in this project and the junior high schoolers in CEPS (regarding age and grade).

**Gender.** 189 students refused to indicate their gender in the questionnaire. Among the remaining 770, there were 451 girls and 319 boys. Since the percentage of female students in Chinese high schools was 50.85 in 2017<sup>23</sup>, it would seem that in this student dataset, the proportion of girls was larger (58.50% of the students who specified their gender).

This lack of male respondents might not have been due to design or distribution error on the part of the researcher, because women are found to be more likely to participate in survey research than men in survey response studies (for example, Curtin, Presser, and Singer, 2000; Moore and Tarnai, 2002; Singer, van Hoewyk, and Maher, 2000). Still, it could be argued

---

<sup>23</sup> See the website of the Ministry of Education of China ([http://www.moe.gov.cn/s78/A03/moe\\_560/jytjsj\\_2017/qg/201808/t20180808\\_344694.html](http://www.moe.gov.cn/s78/A03/moe_560/jytjsj_2017/qg/201808/t20180808_344694.html)).



that the gender imbalance might have skewed the results of the student survey, a limitation that future research might try to overcome.

**Ethnicity.** 105 students (10.9%) did not indicate their ethnicity background in the survey. Similar to the situation with guardian participants, Han people account for 78.1% of the student participants. The remaining 21.9% include Gaeml Minority (17.1%), Miao Minority (2.6%), Hui Minority (1.3%), and others (0.9%).

**Age.** The majority of students (90.5%) were aged between 15 and 19 at the time Study 1 took place, an age range typical of high school students in China. The youngest in the sample are two 13-year-olds, and the eldest is a 21-year-old.

**Grade.** Aside from the 39 students who did not reveal their grade, the remaining 920 were evenly distributed among Grades 10 (29.5%), 11 (34.0%), and 12 (32.4%).

**Academic branch.** Nearly one third of the students suggested in the questionnaire that they were yet to be assigned to either academic branch, whose proportion corresponds to that of the Grade 10 students. 42.0% students belonged to the Sciences Branch, and the rest (37.8%) belonged to the Humanities Branch. In fact, according to official statistics ([http://gaokao.eol.cn/bei\\_jing/dongtai/201703/t20170329\\_1502627.shtml](http://gaokao.eol.cn/bei_jing/dongtai/201703/t20170329_1502627.shtml)), there are more Sciences students than the Humanities students in many provinces. In Heilongjiang Province, for example, the ratio of the Sciences branch to the Humanities branch was as high as 3:1 in 2017. Thus, the disproportion in our sample seems inevitable.

### 5.2.3 Findings from Four-Way Mixed MANOVA on Macro Level

A  $2$  (student gender)  $\times 2$  (region)  $\times 2$  (grade)  $\times 2$  (learner gender) mixed MANOVA was performed on three dependent variables (achievement ratings, aptitude ratings, and affect ratings) on the macro level, where student gender, region, and grade were between-subject factors, and learner gender was the within-subject factor. This addressed the second hypothesis in the analytical framework of Study 1 (see Figure 4.2), the hypothesis that students endorsed FALS on the macro levels, and that the endorsement might have varied due to their gender, region, and grade.

Statistical assumptions were checked: there was a linear relationship between each pair of the dependent variables, as assessed by scatterplot, and no evidence of multicollinearity, as assessed by Pearson's correlation ( $|r| < 0.9$ ). There were two univariate outliers and six multivariate ones in the data. A MANOVA without these outliers was run and the results

from both analyses were practically the same. Therefore, results from the whole dataset will be reported here.

There was multivariate normality, as graphical assessment of the normality by chi-square versus Mahalanobis distance plot suggested. Furthermore, there was homogeneity of variances, as assessed by Levene's Test of Homogeneity of Variance ( $p > .05$ ). However, the assumption of homogeneity of covariance matrices was violated, as assessed by Box's M test ( $p < .001$ ). Given that Box's M test may sometimes flag covariance matrices as unequal (when they are not) with large sample sizes (e.g., Rencher and Christensen, 2012), the MANOVA procedure was carried on. Of course, when interpreting multivariate results, Pillai's criterion rather than Wilks' Lambda were used as the former is more robust to unequal covariance matrices (Olsen, 1976).

### 5.2.3.1 Hypothesis 2.1.1: Students Endorsed Female-Advantage-in-Languages Stereotype on Macro Level

There was a statistically significant main effect of learner gender on the combined dependent variables,  $F(3, 726) = 222.598, p < .001$ , Pillai's Trace = .479,  $\eta_p^2 = .479$ . To follow up this result, univariate main effects of learner gender were investigated, and the effects were significant for all three dependent variables (see Table 5.3). These results show that students did rate male and female learners differently in three aspects (achievement, aptitude, and affect) on the macro level.

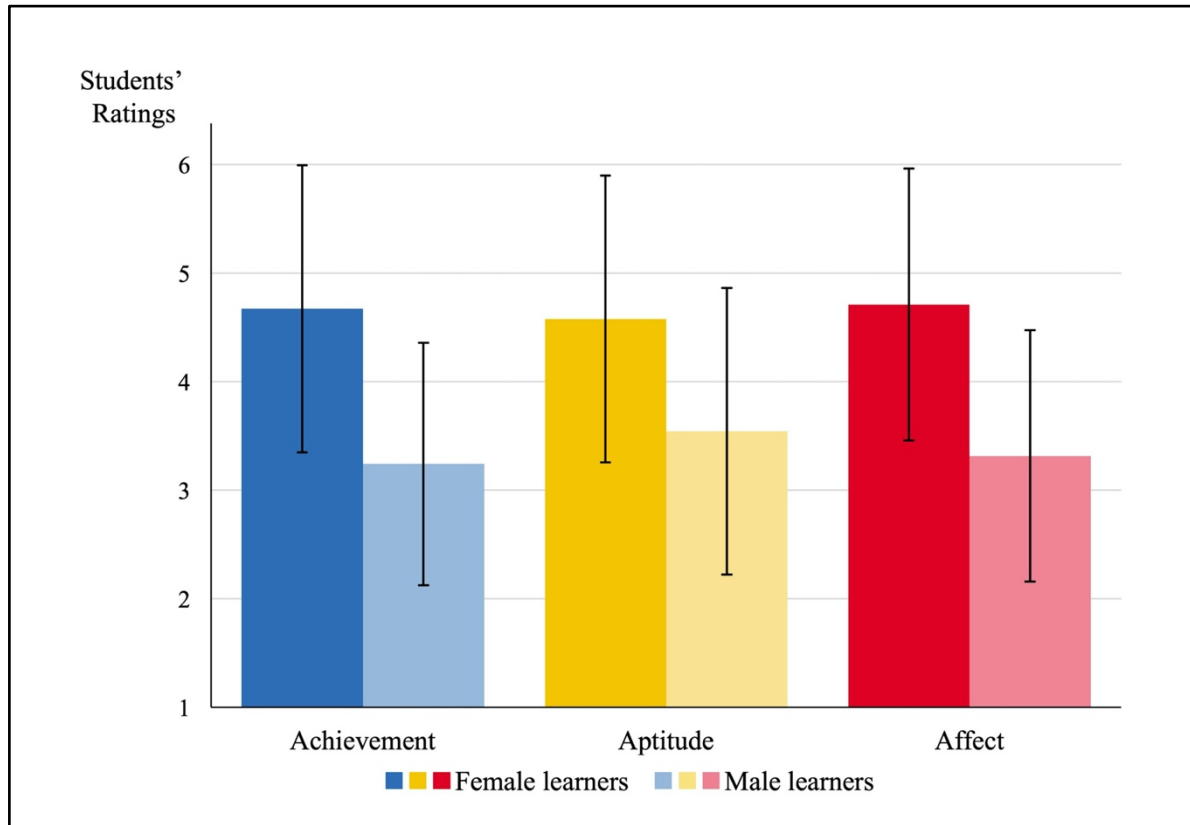
**Table 5.3: Univariate Main Effect of Learner Gender on Macro Level in Student Dataset**

Stereotype component	Learner gender	Mean	SD	$F(1, 728)$	$p$ -value	$\eta_p^2$
Achievement	Female	4.671	1.322	515.890	< .001	.415
	Male	3.241	1.119			
Aptitude	Female	4.577	1.319	216.090	< .001	.229
	Male	3.542	1.321			
Affect	Female	4.711	1.252	518.454	< .001	.416
	Male	3.316	1.157			

Figure 5.6 illustrates the mean ratings from students as a function of learner gender. From the descriptive statistics in Table 5.3, as well as Figure 5.6, it is clear that students viewed females as better learners in English achievement, aptitude and affect. That is, students did endorse FALS. This suggested that Hypothesis 2.1.1 was supported. Note that with the student dataset, large standard deviations were present, resembling the ones in the

guardian dataset (reported in Table 5.1). Therefore, students, as well as guardians, believed in FALS in general, despite variabilities among themselves.

**Figure 5.6: Achievement, Aptitude, and Affect Ratings  
as a Function of Learner Gender on Macro Level in Student Dataset**



#### 5.2.3.2 Hypothesis 2.1.2: Students' Endorsement on Macro Level Differed Due to Gender

The two-way interaction effect between student gender and learner gender on the combined dependent variables was not statistically significant,  $F(3, 726) = 2.292, p = .077$ , Pillai's Trace = .009,  $\eta_p^2 = .009$ . This result indicated that male and female students shared similar gender-stereotypical opinions about English learners on the macro level. Therefore, Hypothesis 2.1.2 was not supported.

It is worth noting that gender differences were detected, instead, on the micro level (see Section 5.2.4.2)

#### 5.2.3.3 Hypothesis 2.1.3: Students' Endorsement on Macro Level Differed Due to Region

The two-way interaction effect between region and learner gender on the combined dependent variables was not statistically significant,  $F(3, 726) = .499, p = .683$ , Pillai's Trace

= .002,  $\eta_p^2 = .002$ . This suggested that regardless of where students were from, they all harboured FALS on the macro level. As a result, Hypothesis 2.1.3 was not supported.

Interestingly, regional differences were discovered in the guardian dataset (see Section 5.1.3.3 earlier): the gender gaps in aptitude and affect were wider from the perspective of southern guardians. As there is collaborating evidence from previous literature (Section 3.1.4.3), it would be hasty to decide that no regional variations existed among students with only the current evidence. Besides, since the effect sizes in both datasets are relatively small ( $\eta_p^2 = .008$  in the guardian dataset and .002 in the student one), it could be that the statistical power was just not big enough to detect such an effect size in the student sample. After all, the observed power for the guardian dataset was .912 but only .152 in terms of the student dataset. Furthermore, there is evidence that the magnitude of FALS harboured by students was smaller than that by guardians (see comparison in Table 9.1 in Section 9.1.1). Thus, it could be argued that a larger sample size is needed to gather evidence about regional differences among students' FALS endorsement. Likewise, a similar argument can be made for the teacher dataset (for details, see Section 6.5)

#### **5.2.3.4 Hypothesis 2.1.4: Students' Endorsement on Macro Level Differed Due to Grade**

The two-way interaction effect between grade and learner gender on the combined dependent variables was statistically significant,  $F(6, 1454) = 4.494, p < .001$ , Pillai's Trace = .036,  $\eta_p^2 = .018$ . This result meant that students across grades differed in their endorsement of FALS on the macro level. In other words, Hypothesis 2.1.4 was supported.

Follow-up univariate two-way ANOVAs were run. These showed a statistically significant interaction effect between grade and learner gender for achievement rating,  $F(2, 728) = 10.358, p < .001, \eta_p^2 = .028$ , for affect rating,  $F(2, 728) = 4.649438, p = .009853, \eta_p^2 = .013$ , but not for aptitude rating,  $F(2, 728) = 1.803, p = .166, \eta_p^2 = .005$ .

As such, a simple main effect analysis of grade on each level of learner gender was conducted for achievement and affect ratings. There was a statistically significant difference between grades for female learners' achievement,  $F(2, 728) = 3.088, p = .046, \eta_p^2 = .008$ , for male learners' achievement,  $F(2, 728) = 8.531, p < .001, \eta_p^2 = .023$ , and for male learners' affect,  $F(2, 728) = 6.187, p = .017, \eta_p^2 = .017$ . On the contrary, the difference between grade was not statistically significant for female learners' affect,  $F(2, 728) = 1.231, p = .293, \eta_p^2 = .003$ .

Therefore, simple pairwise comparisons were run for the differences between grade in mean achievement ratings for female and male learners, and in mean affect ratings for male learners. The relevant statistics are shown below in Table 5.4 (all statistically significant mean differences are marked with \*). As the table exhibits, there was a statistically significant mean difference in ratings for female learners' achievement between students in Grade 10 and Grade 12, for male learners' achievement between students in Grade 10 and Grade 11, for male learners' achievement between students in Grade 10 and Grade 12, and for male learners' affect between students in Grade 10 and Grade 12.

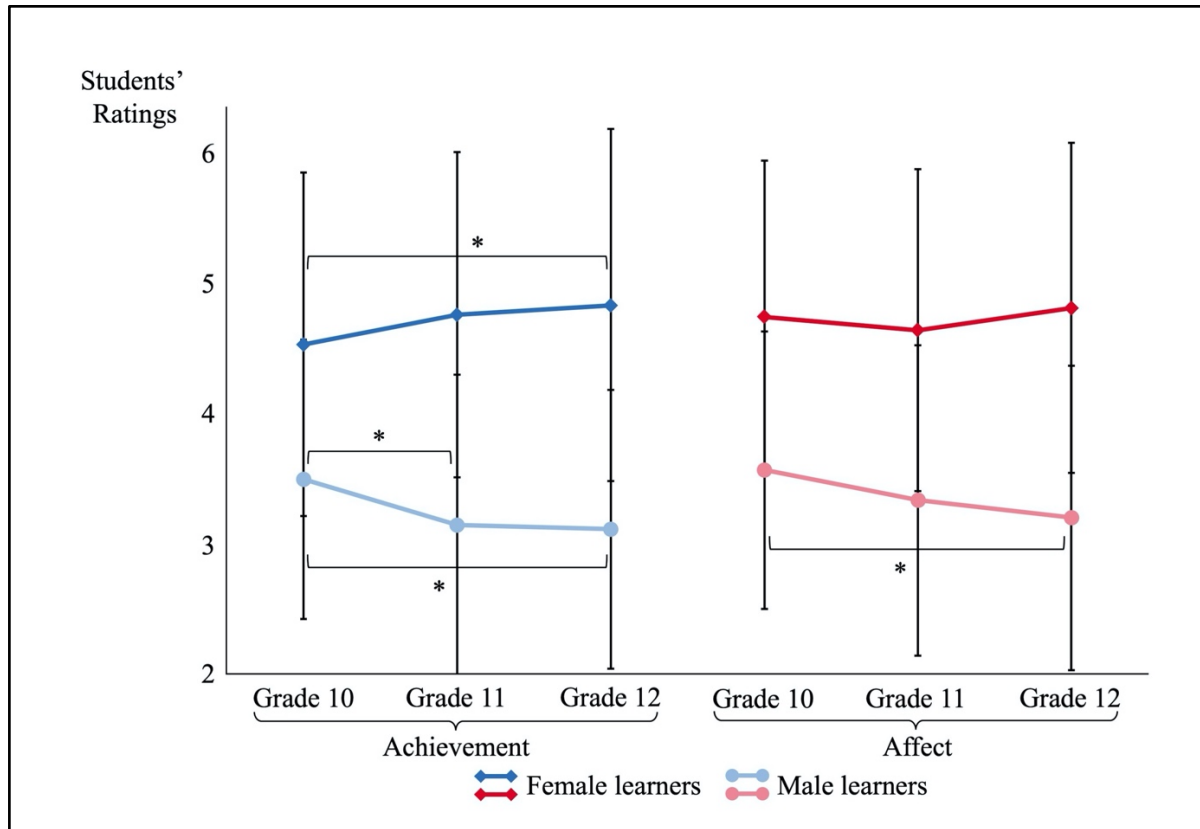
**Table 5.4: Pairwise Comparison between Grade and Learner Gender on Macro Level in Student Dataset**

Stereotype component	Learner gender	Grade	Mean	SD	Mean difference, [95% CI]	p-value
Achievement	Female	10	4.531	1.320	G <sub>10</sub> -G <sub>11</sub> -.227, [-.526, .072]	.206
		11	4.758	1.251	G <sub>11</sub> -G <sub>12</sub> -.073, [-.356, .211]	1
		12	4.831	1.353	G <sub>10</sub> -G <sub>12</sub> -.300, [-.597, -.002]	.048*
	Male	10	3.494	1.074	G <sub>10</sub> -G <sub>11</sub> .355, [.109, .600]	.002*
		11	3.140	1.157	G <sub>11</sub> -G <sub>12</sub> .031, [-.202, .264]	1
		12	3.109	1.071	G <sub>10</sub> -G <sub>12</sub> .386, [.141, .630]	.001*
Affect	Male	10	3.570	1.067	G <sub>10</sub> -G <sub>11</sub> .232, [-.021, .484]	.084
		11	3.338	1.193	G <sub>11</sub> -G <sub>12</sub> .135, [-.104, .375]	.523
		12	3.203	1.170	G <sub>10</sub> -G <sub>12</sub> .367, [.116, .619]	.001*

Figures 5.7 illustrates the achievement and affect ratings as a function of grade and learner gender. As Table 5.4 and Figure 5.7 have made clear, compared to students in higher grades, the 10<sup>th</sup> graders rated male learners' achievement in English higher. Given that they also rated female learners' achievement lower than those in Grade 12 (with the significant mean difference of .300), this would imply that the perceived gender gap in English achievement, or the strength of the gender stereotype regarding English achievement, is the weakest for the 10<sup>th</sup> graders. Similarly, Table 5.4 and Figure 5.7 showed that students in Grade 10 displayed a weaker gender stereotype regarding affect for English compared to those in Grade 12. This detection of intensified tendency of gender stereotyping as students proceed to higher grades is consistent with previous literature (Section 3.2.3). After all, stereotype formation is grounded in individuals' observation of and experience with the social world. Given the high participation rate and strong performance by women and girls in language-related fields (e.g., Carr and Pauwels, 2006), students are likely to establish

stereotypical connection between language domains and the female gender. Their stereotypical beliefs, furthermore, can be strengthened with their own extensive and intensive engagement with language subjects.

**Figure 5.7: Achievement and Affect Ratings as a Function of Grade and Learner Gender on Macro Level in the Student Dataset**



#### 5.2.3.5 Hypothesis 2.1.5: Students' Endorsement on Macro Level Differed on the Basis of Student Gender and Region

The three-way interaction effect between student gender, region, and learner gender on the combined dependent variables was not statistically significant,  $F(3, 726) = 2.047, p = .106$ , Pillai's Trace = .008,  $\eta_p^2 = .008$ . The result indicated that there were no two-way interactions between student gender and region that varied across the levels of learner gender. Thus no follow-ups tests were performed. That is, Hypothesis 2.1.5 was not supported.

#### 5.2.3.6 Hypothesis 2.1.6: Students' Endorsement on Macro Level Differed on the Basis of Region and Grade

The three-way interaction effect between region, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(6, 1454) = 1.629, p = .135$ , Pillai's Trace = .013,  $\eta_p^2 = .007$ . This result suggested that Hypothesis 2.1.6 was not supported: on the macro level, regional differences in FALS endorsement did not exist among

students across grades, nor did variations on the basis of grade happen among students from a particular region.

#### **5.2.3.7 Hypothesis 2.1.7: Students' Endorsement on Macro Level Differed on the Basis of Student Gender and Grade**

The three-way interaction effect between student gender, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(6, 1454) = .794$ ,  $p = .575$ , Pillai's Trace = .007,  $\eta_p^2 = .003$ . This result indicated that Hypothesis 2.1.7 was not supported: on the macro level, boys and girls from different grades shared similar gender stereotypes regarding English learners; meanwhile, within the same grade, boys and girls also shared FALS.

#### **5.2.3.8 Hypothesis 2.1.8: Students' Endorsement on Macro Level Differed on the Basis of Student Gender, Region, and Grade**

The four-way interaction effect between student gender, region, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(6, 1452) = .960$ ,  $p = .451$ , Pillai's Trace = .008,  $\eta_p^2 = .004$ . The result indicated that there were no three-way interactions between student gender, region and grade that varied across the levels of learner gender, thus no follow-ups tests were performed. Thus, Hypothesis 2.1.8 was not supported.

### **5.2.4 Findings from Four-Way Mixed MANOVA on Micro Level**

On the micro level, too, a  $2$  (student gender)  $\times 2$  (region)  $\times 2$  (grade)  $\times 2$  (learner gender) mixed MANOVA was performed on eighteen dependent variables, where student gender, region, and grade were between-subject factors, and learner gender was the within-subject factor.

Statistical assumptions were checked. The assumptions of linearity, no multicollinearity, and multivariate normality were all met. There were no univariate outliers, but five multivariate ones were detected in the data, as assessed by Mahalanobis distance ( $p > .001$ ). A MANOVA without these outliers was run and the results from both analyses were practically the same. Therefore, results from the whole dataset will be reported here.

The assumption of homogeneity of covariance matrices was violated, as assessed by Box's M test ( $p < .001$ ). Similar to the case with the macro level, the MANOVA procedure here on the micro level was still carried on. And when interpreting multivariate results, Pillai's criterion was used (Olsen, 1976). As assessed by Levene's Test of Homogeneity of Variance ( $p > .05$ ), the assumption of homogeneity of variances was moderately violated for

seven variables: ratings for female and male learners' reading achievement, male learners' listening aptitude, male learners' speaking aptitude, male learners' vocabulary aptitude, female learners' listening affect, and female learners' speaking affect. Thus, in order to increase control over the Type I error rate, a lower  $\alpha$  level (.01) was adopted for these variables in the univariate  $F$ -test (Tabachnick and Fidell, 2014).

#### 5.2.4.1 Hypothesis 2.2.1: Students Endorsed Female-Advantage-in-Languages Stereotype on Micro Level

There was a statistically significant main effect of target gender on the combined dependent variables,  $F(18, 470) = 23.664, p < .001$ , Pillai's Trace = .475,  $\eta_p^2 = .475$ . To follow up the above result, univariate main effects of learner gender were investigated, and the effects were significant for all eighteen dependent variables (see Table 5.5 for details).

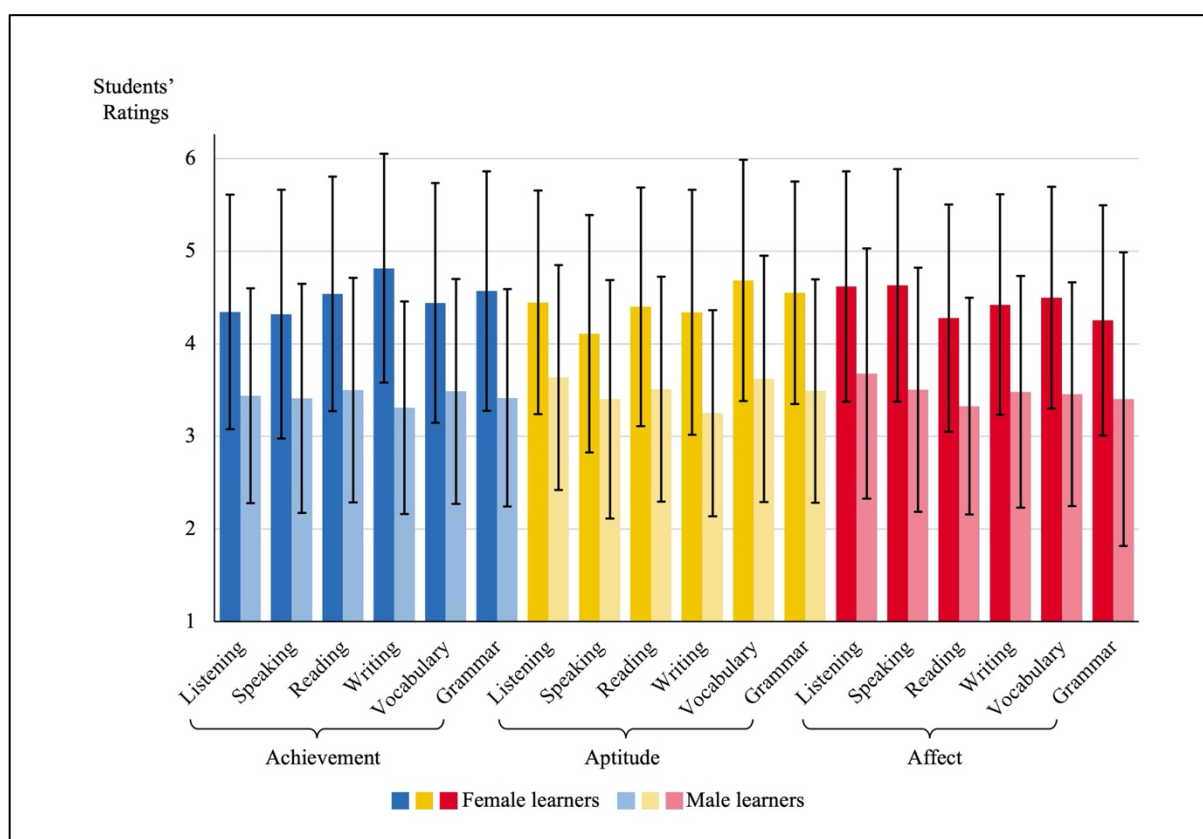
**Table 5.5: Univariate Main Effect of Learner Gender on Micro Level in Student Dataset**

Stereotype component	<i>Mean (SD)</i>		$F(1, 524)$	$p$ -value	$\eta_p^2$
	Female learners	Male learners			
<b>Listening achievement</b>	4.343 (1.264)	3.440 (1.159)	110.776	<.001	.175
<b>Speaking achievement</b>	4.319 (1.343)	3.412 (1.236)	121.103	<.001	.188
<b>Reading achievement</b>	4.538 (1.267)	3.501 (1.213)	160.392	<.001	.234
<b>Writing achievement</b>	4.816 (1.236)	3.310 (1.147)	358.155	<.001	.406
<b>Vocabulary achievement</b>	4.441 (1.295)	3.486 (1.213)	138.258	<.001	.209
<b>Grammar achievement</b>	4.57 (1.292)	3.415 (1.174)	237.597	<.001	.312
<b>Listening aptitude</b>	4.447 (1.208)	3.636 (1.213)	89.456	<.001	.146
<b>Speaking aptitude</b>	4.108 (1.281)	3.402 (1.287)	72.316	<.001	.121
<b>Reading aptitude</b>	4.399 (1.287)	3.510 (1.214)	128.78	<.001	.197
<b>Writing aptitude</b>	4.34 (1.321)	3.252 (1.113)	253.918	<.001	.326
<b>Vocabulary aptitude</b>	4.685 (1.202)	3.622 (1.329)	139.264	<.001	.210
<b>Grammar aptitude</b>	4.55 (1.201)	3.490 (1.205)	191.294	<.001	.267
<b>Listening affect</b>	4.618 (1.243)	3.677 (1.350)	153.185	<.001	.226
<b>Speaking affect</b>	4.63 (1.255)	3.503 (1.318)	211.677	<.001	.288
<b>Reading affect</b>	4.278 (1.227)	3.327 (1.169)	149.735	<.001	.222
<b>Writing affect</b>	4.422 (1.190)	3.481 (1.251)	139.051	<.001	.210
<b>Vocabulary affect</b>	4.499 (1.197)	3.454 (1.208)	206.235	<.001	.282
<b>Grammar affect</b>	4.254 (1.243)	3.404 (1.586)	125.167	<.001	.193



Figure 5.8 illustrates the mean ratings as a function of learner gender. These results show that students did endorse FALS on the micro level, which meant that Hypothesis 2.2.1 was supported. Again, the large standard deviations signalled the range of variations present in students' perspectives.

**Figure 5.8: Achievement, Aptitude, and Affect Ratings  
as a Function of Learner Gender on Micro Level in Student Dataset**



#### 5.2.4.2 Hypothesis 2.2.2: Students' Endorsement on Micro Level Differed Due to Gender

The two-way interaction effect between student gender and learner gender on the combined dependent variables was statistically significant,  $F(18, 507) = 5.646, p < .001$ , Pillai's Trace = .167,  $\eta_p^2 = .167$ . This result meant that boys and girls disagreed on FALS on the micro level. In other words, Hypothesis 2.2.2 was supported.

Follow-up univariate two-way ANOVAs were run. These showed a statistically significant interaction effect between grade and learner gender for five dependent variables: grammar achievement, writing aptitude, grammar aptitude, listening affect, and speaking affect (marked by asterisk mark in Table 5.6 below).

As such, analysis of the simple main effect of student gender on each level of learner gender was conducted for the five dependent variables. The results are summarised in Table

5.7: a statistically significant difference between student gender was found for female learners' grammar achievement, writing aptitude, grammar aptitude, listening affect, and speaking affect, as well as male learners' speaking affect (marked by asterisk mark).

**Table 5.6: Univariate Interaction Effect between Student Gender and Learner Gender on Micro Level in Student Dataset**

Stereotype English	Achievement			Aptitude			Affect		
	<i>F</i>	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>p</i>	$\eta_p^2$
Listening	2.822	.094	.005	.599	.439	.001	10.365	.001*	.019
Speaking	.598	.440	.001	1.403	.237	.003	27	< .001*	.049
Reading	1.046	.307	.002	2.035	.154	.004	2.292	.131	.004
Writing	2.98	.085	.006	12.334	< .001*	.023	0.007	.932	0
Vocabulary	2.442	.119	.005	.137	.712	0	0.013	.908	0
Grammar	12.314	< .001*	.023	4.507	.034*	.009	1.426	.233	.003

*Notes.* For each *F*-value, the numerator degree of freedom is 1, and the denominator degree of freedom is 524.

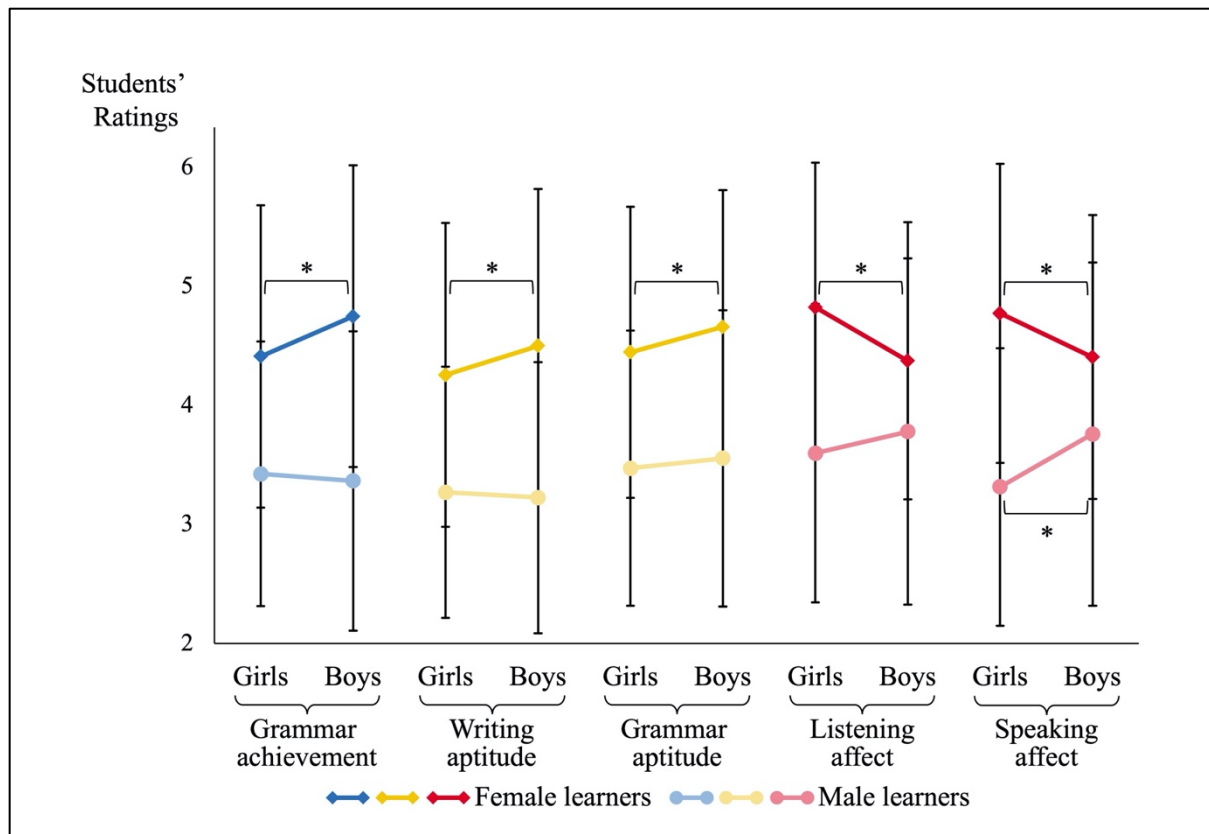
**Table 5.7: Simple Main Effect of Student Gender on Learner Gender on Micro Level in Student Dataset**

Stereotype component	Learner gender	Mean (SD)		<i>F</i> (1, 524)	<i>p</i> -value	$\eta_p^2$
		Girl	Boy			
Grammar achievement	Female	4.414 (1.271)	4.751 (1.268)	11.448	.001*	.021
	Male	3.426 (1.113)	3.368 (1.258)	1.988	.159	.004
Writing aptitude	Female	4.258 (1.277)	4.503 (1.318)	9.125	.003*	.017
	Male	3.272 (1.055)	3.227 (1.139)	2.204	.138	.004
Grammar aptitude	Female	4.448 (1.222)	4.662 (1.148)	7.663	.006*	.014
	Male	3.474 (1.156)	3.557 (1.246)	.048	.827	0
Listening affect	Female	4.826 (1.215)	4.376 (1.164)	15.744	< .001*	.029
	Male	3.600 (1.255)	3.782 (1.456)	.180	.671	0
Speaking affect	Female	4.775 (1.256)	4.408 (1.193)	14.081	< .001*	.026
	Male	3.317 (1.167)	3.761 (1.443)	11.452	.001*	.021

For the first three dependent variables (grammar achievement, writing aptitude, grammar aptitude), boys gave higher ratings for female learners' than girls did. As boys and girls gave similar ratings for male learners, it seems that the perceived gender differences in these areas were larger from boys' perspective. In comparison, in terms of listening and

speaking affect, girls gave higher ratings for female learners and lower ratings for male learners. This means that the perceived gender difference in listening and speaking affect was much larger from girls' perspective. This finding is distinctive compared to the previous three dependent variables, as Figure 5.9 here displays.

**Figure 5.9: Ratings as a Function of Student Gender and Learner Gender on Micro Level in Student Dataset**



The fact that gender differences were detected at the micro level but not the macro level is considered compelling evidence for FALS, because arguably students would have conceptualised the English subject more accurately at the micro level. It also highlight the context-sensitive nature of academic gender stereotypes.

#### 5.2.4.3 Hypothesis 2.2.3: Students' Endorsement on Micro Level Differed Due to Region

The two-way interaction effect between region and learner gender on the combined dependent variables was not statistically significant,  $F(18, 507) = 1.530, p = .075$ , Pillai's Trace = .052,  $\eta_p^2 = .052$ . This suggested that regardless of where students were from, they all harboured FALS on the micro level. As a result, Hypothesis 2.2.3 was not supported.

#### 5.2.4.4 Hypothesis 2.2.4: Students' Endorsement on Micro Level Differed Due to Grade

The two-way interaction effect between grade and learner gender on the combined dependent variables was statistically significant,  $F(36, 1016) = 2.222, p < .001$ , Pillai's Trace = .146,  $\eta_p^2 = .073$ . This meant that students across grades differed in their endorsement of FALS on the micro level. In other words, Hypothesis 2.2.4 was supported.

Follow-up univariate two-way ANOVAs were run. These showed a statistically significant interaction effect between grade and learner gender for seven dependent variables: reading achievement, vocabulary achievement, grammar achievement, writing aptitude, grammar aptitude, writing affect, and vocabulary affect (marked by \* in Table 5.8 below).

**Table 5.8: Univariate Interaction Effect between Grade and Learner Gender on Micro Level in Student Dataset**

Aspect of English	Achievement			Aptitude			Affect		
	<i>F</i>	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>p</i>	$\eta_p^2$	<i>F</i>	<i>p</i>	$\eta_p^2$
Listening	.827	.438	.003	2.165	.116	.008	.455	.634	.002
Speaking	.093	.911	< .001	.375	.688	.001	.586	.557	.002
Reading	5.231	.006*	.02	1.852	.158	.007	.102	.903	< .001
Writing	2.682	.069	.01	3.171	.043*	.012	4.932	.008*	.018
Vocabulary	5.238	.006*	.02	.206	.814	.001	4.395	.013*	.016
Grammar	9.796	< .001*	.036	3.046	.048*	.011	.957	.385	.004

*Notes.* For each *F*-value, the numerator degree of freedom is 2, and the denominator degree of freedom is 524.

As such, analysis of the simple main effect of grade on each level of learner gender was conducted for the seven dependent variables. The results are summarised in Table 5.9 below: a statistically significant difference between grades was found for female learners' vocabulary achievement, and male learners' grammar achievement, grammar aptitude, and vocabulary affect.

**Table 5.9: Simple Main Effect of Grade on Each Level of Learner Gender on Micro Level in Student Dataset**

Stereotype component	Female learners			Male learners		
	<i>F</i> (2, 524)	<i>p</i> -value	$\eta_p^2$	<i>F</i> (2, 524)	<i>p</i> -value	$\eta_p^2$
Reading achievement	3.174	.043	.012	2.334	.098	.009
Vocabulary achievement	3.581	.029*	.013	2.237	.108	.008

<b>Grammar achievement</b>	2.220	.110	.008	9.523	< .001*	.035
<b>Writing aptitude</b>	1.439	.238	.005	1.623	.198	.006
<b>Grammar aptitude</b>	.735	.480	.003	3.666	.026*	.014
<b>Writing affect</b>	2.356	.096	.009	2.176	.114	.008
<b>Vocabulary affect</b>	1.048	.351	.004	4.187	.016*	.016

Following this, simple pairwise comparisons were run, and the relevant statistics are shown below in Table 5.10 (all statistically significant mean differences are marked with \*). As the table displays, there was a statistically significant mean difference in ratings for female learners' vocabulary achievement between the 10<sup>th</sup> and 11<sup>th</sup> graders, for male learners' grammar achievement between the 10<sup>th</sup> and 11<sup>th</sup> graders, as well as between the 10<sup>th</sup> and 12<sup>th</sup> graders, for male learners' grammar aptitude between the 10<sup>th</sup> and 12<sup>th</sup> graders, and for male learners' vocabulary affect between the 10<sup>th</sup> and 12<sup>th</sup> graders.

**Table 5.10: Pairwise Comparison between Grade and Learner Gender on Micro Level in Student Dataset**

<b>Stereotype component</b>	<b>Learner gender</b>	<b>Grade</b>	<b>Mean</b>	<b>SD</b>	<b>Mean difference, [95% CI]</b>	<b>p</b>
<b>Vocabulary achievement</b>	<b>Female</b>	<b>10</b>	4.284	1.234	G <sub>10</sub> -G <sub>11</sub> -.372, [-.711, -.033]	.026*
		<b>11</b>	4.655	1.238	G <sub>11</sub> -G <sub>12</sub> .145, [-.194, .484]	.914
		<b>12</b>	4.510	1.343	G <sub>10</sub> -G <sub>12</sub> -.227, [-.551, .098]	.282
<b>Grammar achievement</b>	<b>Male</b>	<b>10</b>	3.712	1.173	G <sub>10</sub> -G <sub>11</sub> .427, [.125, .73]	.002*
		<b>11</b>	3.284	1.164	G <sub>11</sub> -G <sub>12</sub> .059, [-.244, .361]	1
		<b>12</b>	3.225	1.128	G <sub>10</sub> -G <sub>12</sub> .486, [.197, .776]	< .001*
<b>Grammar aptitude</b>	<b>Male</b>	<b>10</b>	3.717	1.131	G <sub>10</sub> -G <sub>11</sub> .264, [-.052, .581]	.136
		<b>11</b>	3.453	1.244	G <sub>11</sub> -G <sub>12</sub> .058, [-.258, .375]	1
		<b>12</b>	3.395	1.197	G <sub>10</sub> -G <sub>12</sub> .323, [.02, .626]	.032*
<b>Vocabulary affect</b>	<b>Male</b>	<b>10</b>	3.701	1.145	G <sub>10</sub> -G <sub>11</sub> .188, [-.123, .499]	.439
		<b>11</b>	3.513	1.270	G <sub>11</sub> -G <sub>12</sub> .170, [-.14, .481]	.567
		<b>12</b>	3.343	1.173	G <sub>10</sub> -G <sub>12</sub> .358, [.061, .655]	.012*

Figure 5.10 here illustrates the vocabulary achievement and affect ratings as a function of grade and learner gender. As we can tell from Table 5.10 and Figure 5.10, students in Grade 11 rated female learners' vocabulary achievement much higher than those in Grade 10 (with a significant mean difference of .372). This suggested that the assumed gender gap in vocabulary achievement was larger for students in Grade 11. Besides, since students in Grade

10 rated male learners' vocabulary affect much higher than those in Grade 12 (with a significant mean difference of .358), this would mean that the perceived gender difference in vocabulary affect, or the strength of the gender stereotype regarding vocabulary affect, is stronger for students in Grade 12. Combined together, the two findings pointed out that as students proceeded to higher grades, it is likely that they would perceive a widening gender difference in vocabulary achievement and affect.

**Figure 5.10: Vocabulary-Related Ratings as a Function of Grade and Learner Gender on Micro Level in Student Dataset**

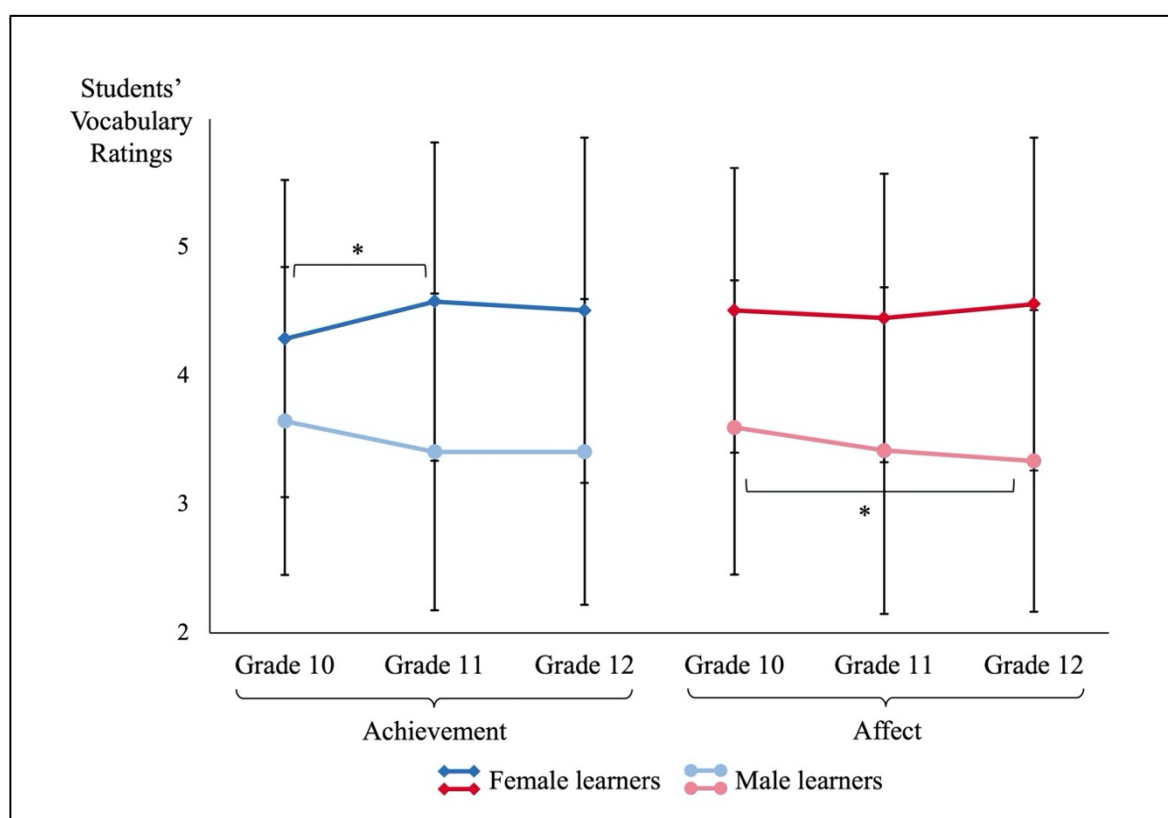


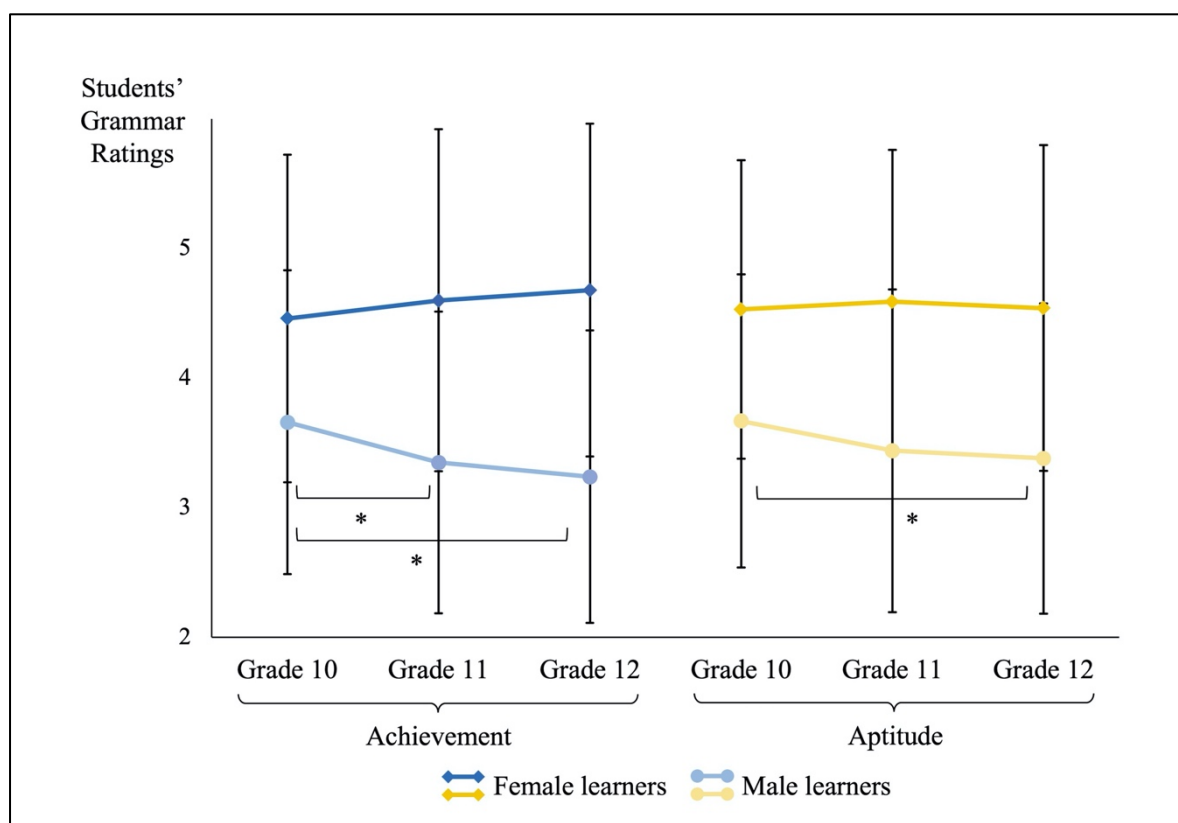
Figure 5.11 illustrates the grammar achievement and aptitude ratings as a function of grade and learner gender. As we can tell from Table 5.10 and Figure 5.11, compared to students in higher grades, those in Grade 10 rated male learners' grammar achievement much higher. In addition, students in Grade 10 also rated male learners' grammar aptitude much higher than those in Grade 12 (with a significant mean difference of .323). Combined together, these two results indicated that when students proceeded to higher grades, it is likely that they would perceive a widening gender difference in grammar achievement and aptitude.

#### 5.2.4.5 Hypothesis 2.2.5: Students' Endorsement on Micro Level Differed on the Basis of Gender and Region

The three-way interaction effect between student gender, region, and learner gender on the combined dependent variables was not statistically significant,  $F(18, 507) = 1.324, p$

= .166, Pillai's Trace = .045,  $\eta_p^2 = .045$ . The result indicated that there were no two-way interactions between student gender and region that varied across the levels of learner gender, thus no follow-ups tests were performed. That is, Hypothesis 2.2.5 was not supported.

**Figure 5.11: Grammar-Related Ratings as a Function of Grade and Learner Gender on Micro Level in Student Dataset**



#### 5.2.4.6 Hypothesis 2.2.6: Students' Endorsement on Micro Level Differed on the Basis of Region and Grade

The three-way interaction effect between region, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(36, 1016) = .944, p = .565$ , Pillai's Trace = .065,  $\eta_p^2 = .032$ . The result indicated that Hypothesis 2.2.6 was not supported.

#### 5.2.4.7 Hypothesis 2.2.7: Students' Endorsement on Micro Level Differed on the Basis of Student Gender and Grade

The three-way interaction effect between student gender, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(36, 1016) = .999, p = .471$ , Pillai's Trace = .068,  $\eta_p^2 = .034$ . Thus, Hypothesis 2.2.7 was not supported.

#### **5.2.4.8 Hypothesis 2.2.8: Students' Endorsement on Micro Level Differed on the Basis of Student Gender, Region, and Grade**

The four-way interaction effect between student gender, region, grade, and learner gender on the combined dependent variables was not statistically significant,  $F(36, 1016) = 1.05$ ,  $p = .39$ , Pillai's Trace = .072,  $\eta_p^2 = .036$ . The result indicated that there were no three-way interactions between student gender, region and grade that varied across the levels of learner gender, thus no follow-ups tests were performed. Hence, Hypothesis 2.2.8 was not supported.

#### **5.2.5 Summary of Students' Perspectives**

On the macro level, it was found that regardless of their own gender, students stereotypically thought female learners were better than male ones in terms of English achievement, aptitude, and affect, which was similar to the situation with guardians. However, contrary to the case with guardians, regional differences in FALS endorsement were not discovered, possibly due to the limited sample size and resultant insufficient statistical power. One additional finding on the macro level was that compared to students in higher grades, the 10<sup>th</sup> graders have weaker gender stereotypes regarding achievement and affect.

On the micro level, it was also found that students endorsed FALS. In addition, gender disparities in endorsement were also identified: the perceived gender differences in grammar achievement, writing aptitude and grammar aptitude were larger from boys' perspective, but the cases with listening and speaking affects showed the exact opposite pattern. Such a finding of gender differences on the micro level is not considered contradictory to a lack of gender differences on the macro level. Instead, it is additional proof that students endorsed FALS, and it signifies the contextual nature of gender stereotypes.

Again, regional differences did not surface on the micro level. The suspected reasons were identical to those on the macro level: sample size and statistical power. Furthermore, akin to the macro level, differences across grades were established on the micro level, too: as student proceeded to higher grades, it would seem that they perceived a widening gender difference in vocabulary achievement and affect, as well as in grammar achievement and aptitude.

### **5.3 Generation Gap in Endorsement of Female-Advantage-in-Languages Stereotype**

A  $2$  (generation)  $\times 2$  (participant gender)  $\times 2$  (region)  $\times 2$  (learner gender) mixed MANOVA was performed on three dependent variables (achievement ratings, aptitude



ratings, and affect ratings), where generation, participant gender, and region were between-subject factors, and learner gender was the within-subject factor. This statistical procedure addressed the third hypothesis in the analytical framework of Study 1 (see Figure 4.2), the hypothesis that generation gaps existed in terms of FALS endorsement, and the magnitude of such gaps may have varied due to participant gender and region.

Statistical assumptions were checked. There was a linear relationship between the dependent variables, as assessed by scatterplot, and no evidence of multicollinearity, as assessed by Pearson correlation ( $|r| < 0.9$ ). There were 123 univariate outliers and 25 multivariate ones in the data. A MANOVA without these outliers was run and the results from both analyses were practically the same. Therefore, results from the whole dataset will be reported here.

Graphical assessment of the normality by chi-square versus Mahalanobis distance plot suggested that the variables did not form a multivariate normal distribution in 8 cells of the design (ratings for female learners). However, MANOVA is fairly 'robust' to deviations from normality with respect to Type I error (Bray and Maxwell, 1985). In fact, Weinfurt (1995) notes that in practice MANOVAs tend to be performed even if the data is not normal due to a general consensus that MANOVA is robust to non-normality.

The assumption of homogeneity of variances was violated for one variable, female learners' aptitude, as assessed by Levene's Test of Homogeneity of Variance ( $p > .05$ ). To increase control over the Type I error rate, a lower  $\alpha$  level ( $=.01$ ) was adopted in the univariate  $F$ -test (Tabachnick and Fidell, 2014).

The assumption of homogeneity of covariance matrices was also violated, as assessed by Box's M test ( $p < .001$ ). Of course, this might have resulted from the lack of multivariate normality and relatively large sample sizes (Rencher and Christensen, 2012). Nevertheless, this means that when interpreting multivariate results, Pillai's criterion rather than Wilks' Lambda should be used as the former is more robust to unequal covariance matrices (Olsen, 1976).

### **5.3.1 Hypothesis 3.1: Generation Gap Existed in Stereotype Endorsement**

The two-way interaction effect between generation and learner gender on the combined dependent variables was statistically significant,  $F(3, 2698) = 6.215, p < .001$ , Pillai's Trace  $= .007, \eta_p^2 = .007$ . This result indicated that a generation gap did exist in FALS endorsement. In other words, Hypothesis 3.1 was supported.

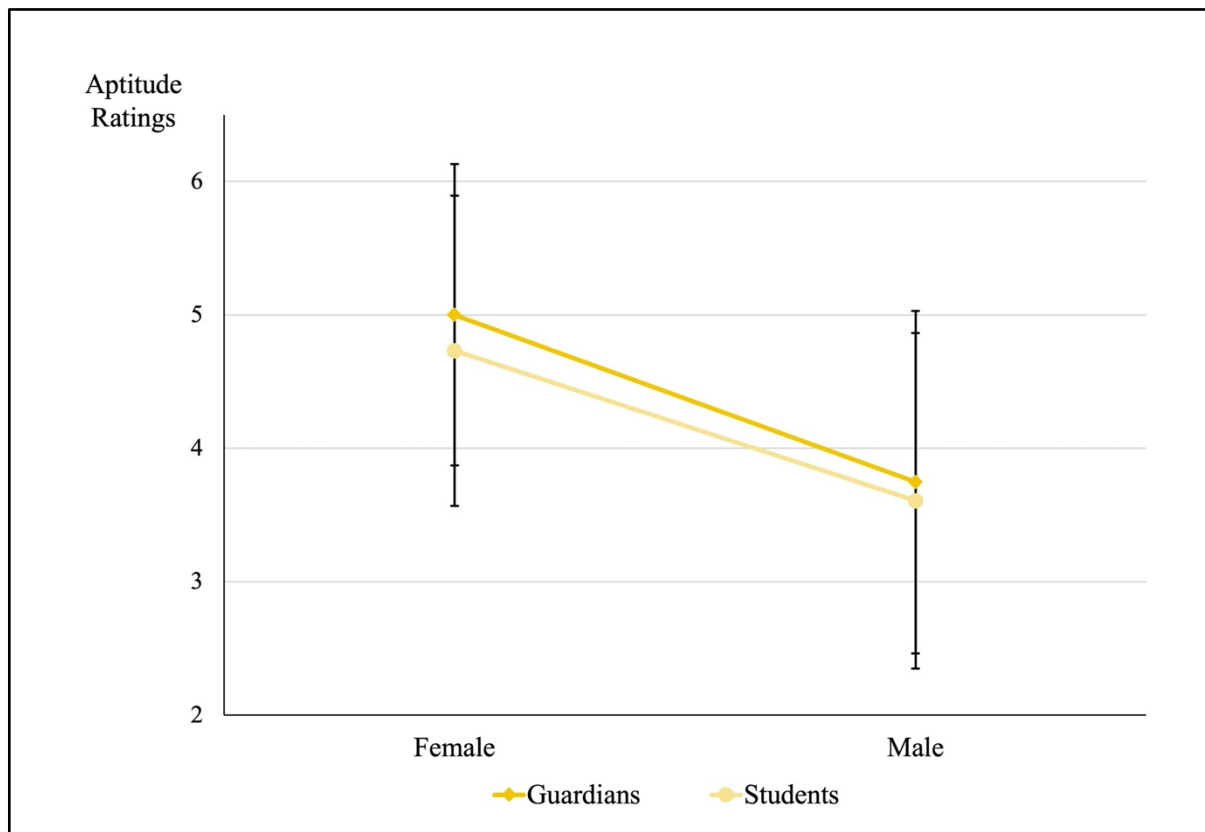
Follow-up univariate two-way ANOVAs were run. These showed a statistically significant interaction effect between generation and learner gender only for aptitude rating,  $F(1, 2700) = 5.287, p = .002, \eta_p^2 = .002$ , but not for achievement rating,  $F(1, 2700) = .448, p = .504, \eta_p^2 = .000166$ , nor for affect rating,  $F(1, 2700) = 2.765, p = .096, \eta_p^2 = .001$ .

As such, a simple main effect analysis was conducted for aptitude rating. The relevant statistics are shown below in Table 5.11. Evidently, there was a statistically significant difference between generations for both male and female learners' aptitude ratings (marked with an asterisk mark).

**Table 5.11: Simple Main Effect of Generation on Learner Gender in Guardian and Student Datasets**

Stereotype component	Learner gender	Generation	Mean	SD	$F(1, 2745)$	$p$ -value	$\eta_p^2$
Aptitude	Female	Guardian	5.000	1.130	33.619	< .001*	.012
		Student	4.731	1.285			
	Male	Guardian	3.746	1.164	7.856	.005*	.003
		Student	3.607	1.258			

**Figure 5.12: Aptitude Ratings as a Function of Generation and Learner Gender in Guardian and Student Dataset**



Therefore, simple pairwise comparisons were run for the differences between generations in mean aptitude ratings for male learners and in mean affect ratings for female learners. For female learners' aptitude, there was a statistically significant mean difference of .267, 95% CI [.179, .355],  $p < .001$  (see descriptive statistics in Table 5.11). Furthermore, for male learners' aptitude, there was also a statistically significant mean difference of .139, 95% CI [.049, .229],  $p = .005$  (also see descriptive statistics in Table 5.11).

Judging from Figure 5.12, there seemed to be a wider gender gap in aptitude from the perspective of guardians. In other words, a generation gap existed, where guardians, the senior generation, had a stronger gender stereotype about aptitude.

### **5.3.2 Hypothesis 3.2: Width of Generation Gap Differed Due to Gender**

The three-way interaction effect between generation, participant gender, and learner gender on the combined dependent variables was not statistically significant,  $F(3, 2698) = .256, p = .857$ , Pillai's Trace = .000285,  $\eta_p^2 = .000285$ . The result indicated that there were no two-way interactions between generation and participant gender that varied across the levels of learner gender, thus no follow-ups tests were performed. That is, Hypothesis 3.2 was not supported.

### **5.3.3 Hypothesis 3.3: Width of Generation Gap Differed Due to Region**

The three-way interaction effect between generation, region, and learner gender on the combined dependent variables was not statistically significant,  $F(3, 2698) = 1.300, p = .273$ , Pillai's Trace = .001,  $\eta_p^2 = .001$ . This result suggested that Hypothesis 3.3 was not supported: regional differences in generation gap did not exist.

### **5.3.4 Hypothesis 3.4: Width of Generation Gap Differed on the Basis of Gender and Region**

The four-way interaction effect between generation, participant gender, region, and learner gender on the combined dependent variables was not statistically significant,  $F(3, 2698) = .532, p = .660$ , Pillai's Trace = .001,  $\eta_p^2 = .001$ . The result indicated that there were no three-way interactions between generation, participant gender and region that varied across the levels of learner gender, thus no follow-ups tests were performed. Thus, Hypothesis 3.4 was not supported.

### **5.3.5 Summary of Generation Gap in Endorsement**

Results from the four-way MANOVA suggested that a generation gap existed, where guardians, the older generation, had a stronger gender stereotype about aptitude, compared to

the younger one. However, this generation gap did not vary due to participant gender or region.

#### **5.4 Summary of Guardians' and Students' Perspectives**

This chapter reports guardians' and students' FALS endorsement. Both groups of participants stereotypically thought female learners were better than male ones in terms of English achievement, aptitude, and affect (macro and micro levels, in the case of students). In addition, on the macro level, students in Grades 11 and 12 tend to have stronger gender stereotypes regarding achievement and affect than those in Grade 10. Differences across grades were established on the micro level, too. When student proceeded to higher grades, it would seem that they perceived a widening gender difference in vocabulary achievement and affect, as well as in grammar achievement and aptitude.

There was no gender difference in FALS endorsement among guardians or students (macro level only). But on the micro level, the perceived gender differences in grammar achievement, writing aptitude and grammar aptitude were larger from boys' perspective, while the cases with listening and speaking affects showed the exact opposite pattern. This finding further signifies the contextual nature of gender stereotypes.

Region specificity only arose in the guardian dataset: in terms of aptitude and affect, guardians from the south stereotypically perceived a slightly wider gender difference than those from the north.

When the two generations were compared and contrasted, a generation gap was found—guardians, the older generation, had a stronger gender stereotype about aptitude than the younger one. However, this generation gap did not vary due to participant gender or region.

## **Chapter 6: Findings and Discussions II —**

### **Perceiving Girls as Better Learners of English by Teachers of English**

This chapter will continue presenting results from Phase 1, focusing on the integrated findings from a) the teacher respondents in Study 1 and b) the teacher interviewees in Study 2. A total of 62 questionnaires were collected from teachers in Study 1, and 20 interviews were conducted in Study 2. No single teacher participated in both studies. A three-way mixed MANOVA and a multiple regression were performed on the questionnaire dataset, and the interview dataset was subject to thematic analysis as well as Fisher's exact probability test. The aim of this chapter is to gain a more comprehensive and nuanced understanding of teachers' perspectives on the female-advantage-in-languages stereotype (FALS) through integrating numeric and textual data from the two strands (see Section 4.2.3 and Figure 4.5 for an explanation of the integration process). In addition, to aid the understanding of certain results, some findings may be briefly discussed with pertinent literature.

Section 6.1 will briefly display and compare the demographic characteristics of the two teacher datasets. Then, Sections 6.2-5 will report the integrated findings from two studies in a point-by-point manner, drawing on results from the MANOVA performed on the questionnaire dataset, the first theme of the interview dataset ('the widespread endorsement of FALS'), and Fisher's exact probability test on the interview dataset. Section 6.2 will show how teachers were found to endorse each component of FALS: achievement, affect, and aptitude. Subsequently, Section 6.3 will explain the relationships among three FALS components as perceived by teachers. Next, Section 6.4 and 6.5 will provide evidence against gender and regional variations in FALS endorsement. Finally, Section 6.6 will offer a summary of the above integrated findings based on both quantitative and qualitative data sources.

As explained in Section 4.2.2.4, the thematic analysis actually generated three inter-related superordinate themes: 1) the widespread endorsement of FALS; 2) gender stereotypes accompanying FALS; and 3) critical perspectives about FALS. For the sake of clarity, coherence, and space, the current chapter only refers to the qualitative findings corresponding to those in the quantitative strand—namely, the first theme. The latter two themes, which further attest to how teachers of English perceive gender as a differentiating variable, will be delineated in Chapter 7.

## 6.1 Demographic Information of Teacher Participants

This section describes the demographic characteristics of teacher participants in both Studies 1 and 2, including age, gender, experience, education, region, and features of the schools they worked for. Although information about teachers' working environment and experience were only collected in Study 2, given that such information is necessary for understanding the context where FALS operates, it will also be summarised here.

**Data screening of the questionnaire dataset.** Before analysing the questionnaire dataset, each response was assigned a case number and a three-step screening procedure was utilised to identify three types of careless responses: those answered by self-identified inattentive respondents (zero response), those showed a uniform response pattern (one response), and those displayed a random response pattern (one response). Therefore, 60 responses (97% of the questionnaire sample) were retained for analysis.

**Age.** The exact age make-up of the two studies are shown here in Figure 6.1. Excluding the 5 teachers (8.3%) who refused to disclose their ages in Study 1, the two samples seem comparable and complementary regarding age. Clearly, the age ranges were the same in two data sources, both between early twenties and early fifties. Such a similarity in age ranges serves as an essential prerequisite for comparing and connecting results from two datasets. Nevertheless, there is a discrepancy between these two samples: the respondents in Study 1 were more or less evenly distributed among each age group, with slightly more teachers aged 35 and above (around 65%); the interviewees in Study 2, on the other hand, clustered around the younger end of this age scale, with 70% of them under 35 years old. This disparity was not taken as a threat to validity, because it made the two data sources complementary by allowing voices from all age ranges to be explored.

**Gender.** Except for 3 teachers who did not indicate their gender in the questionnaire, the number of female teachers (36) was 1.7 times of that of male teachers in Study 1 (21). The case with Study 2 was similar, where the ratio of female and male teachers was 1.5:1 (12 females and 8 males). This unbalanced gender ratio seemed reasonable, because the gender ratio in teachers of English in Chinese high schools was around 3:1 in 2018, according to the Ministry of Education's website<sup>24</sup>.

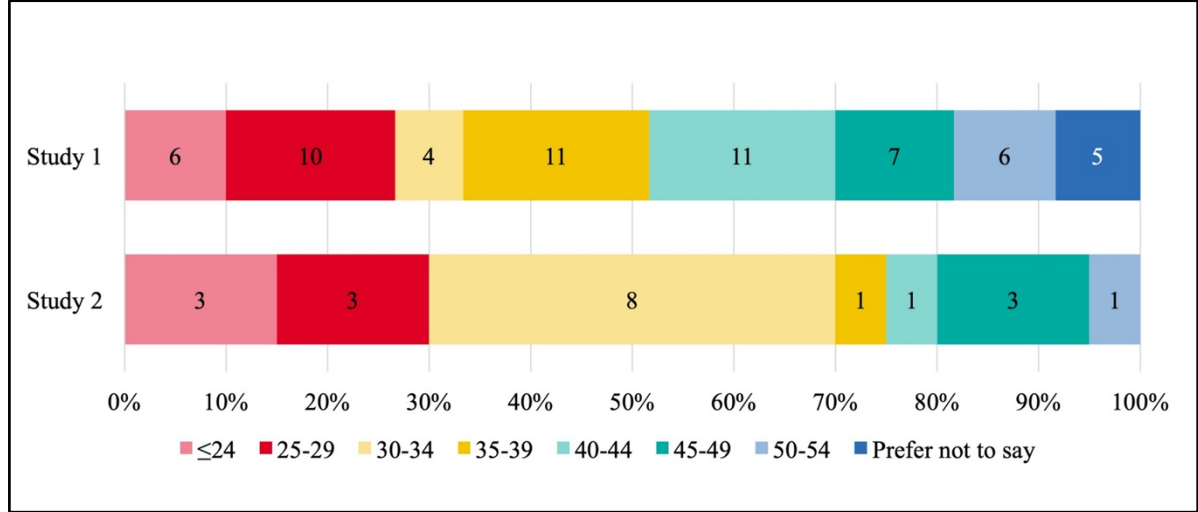
**Experience.** In Study 1, 5 teachers did not reveal for how many years they have been

---

<sup>24</sup> Official statistics about gender, education, and ethnicity among teachers of English in Chinese high schools are obtained from [http://www.moe.gov.cn/s78/A03/moe\\_560/jytjsj\\_2018/qg/201908/t20190812\\_394323.html](http://www.moe.gov.cn/s78/A03/moe_560/jytjsj_2018/qg/201908/t20190812_394323.html).

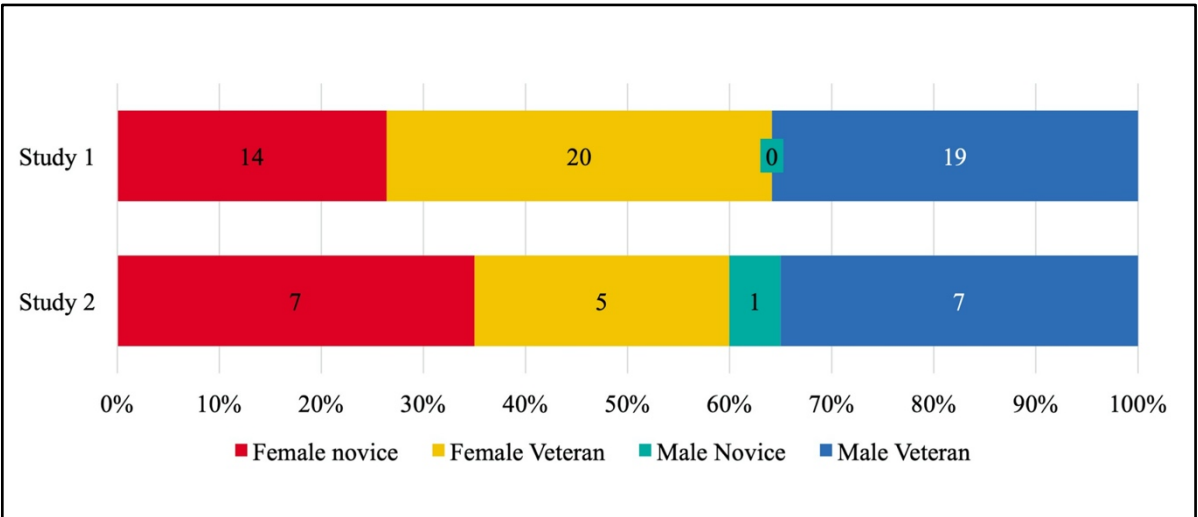
working as teachers of English. Among the remaining 55, 15 were novices with three years or fewer teaching experience, and 40 were seasoned teachers. In Study 2, similarly, more veterans (12) volunteered as interviewees compared to novices (8).

**Figure 6.1: Age Make-up of Teacher Datasets in Studies 1 and 2**



Combining the two abovementioned demographic characteristics (gender and experience), Figure 6.2 here illustrates the proportions of four types of interviewees: female novices, female veterans, male novices, and male veterans. Apparently, the only rarity in both datasets is male novices: in fact, the researcher kept on looking for male novice teachers outside the eight participating schools for months, but another male novice never showed up. Except for this type, the other three types were represented relatively equally in both studies. This not only indicates that perspectives from teachers with these demographic profiles could be explored evenly, but also solidifies the foundation upon which results from two studies could be compared—the similarity, or comparability, between the two data sources.

**Figure 6.2: Gender and Experience Make-Up of Teacher Datasets in Studies 1 and 2**



**Education.** In Study 1, among the 54 teachers who responded, 51 had a college degree or equivalent, and the remaining 3 were educated to college level and above. In Study 2, the ratio between university graduates and postgraduates is 17:3. Neither dataset is representative of the corresponding national population, though: the same ratio was approximately 8:1 in 2018. That is, degree-holders were over-represented in Study 1, but postgraduates were so in Study 2, which actually makes the two studies complementary.

**Ethnicity.** Ethnicity was only investigated in Study 1. Among the 51 respondents who indicated their ethnicity, the majority was Han people (35, 68.6%). The remaining 31.4% belonged to Gaeml Minority (21.6%), Hui Minority (5.9%), Miao Minority (2.0%), and other ethnic minorities (2.0%). In the national population of English teachers, however, the proportion of ethnic minorities was only 7.4% in 2018. The overrepresentation of ethnic minorities in Study 1 was not considered as a significant threat to ecological validity, because most of the minorities involved (Gaeml, Hui, and Miao) shares the culture, custom, and language of Han people, the dominant ethnic group in China<sup>25</sup>.

**Region.** In Study 1, 38 teachers (63.3%) were from the north, and the rest are from the south. The pattern is reversed in Study 2, where northern teachers only took up 35% of the interview dataset.

**Working environment and experience.** Table 6.1 here summarises the background information regarding interviewees' working environment and experience (ordered on the basis of gender and teaching experience). The aliases were designed to reflect participants' gender and experience: monosyllabic names are female, and disyllabic ones are male; names ending with 'n' are novice teachers, and those ending with 'e' are experienced ones. For example, 'Fen' is a female novice, and 'Zhoujie' is a male veteran.

As the table reveals, participants in Study 2 come from different types of schools, including public and private schools in counties and cities. On average, they have taught English for 9.9 years at the time of their interview, and their weekly teaching time was averagely 17.0 hours. The ratio of participants teaching three grades was approximately 1:1:1, and the ratio of the Humanities Branch to the Sciences Branch was 1:2. After all, the number of Sciences students was larger than that of the Humanities students in most places

---

<sup>25</sup> For Gaeml people, see introduction on [http://www.gov.cn/guoqing/2015-07/23/content\\_2901628.htm](http://www.gov.cn/guoqing/2015-07/23/content_2901628.htm). Similar information about Hui and Miao can be found from [http://www.gov.cn/guoqing/2015-03/18/content\\_2835611.htm](http://www.gov.cn/guoqing/2015-03/18/content_2835611.htm) and [http://www.gov.cn/guoqing/2015-07/23/content\\_2901564.htm](http://www.gov.cn/guoqing/2015-07/23/content_2901564.htm).



**Table 6.1: Professional Background of Interviewees in Study 2**

No.	Alias	School		Teaching		Current students			
		Location	Sector	Experience <sup>a</sup>	Hours per week	Grade	Branch	Number	Gender make-up
1	Jian	South city	Private	1	30	12	N/A	4	Male dominant
2	Lun	South county	Public	1	16	10	N/A	124	(More or less) balanced
3	Ren	South city	Private	1	5	12	N/A	15	Female dominant
4	Kun	North county	Private	2	22	11	SCI <sup>b</sup>	154	(More or less) balanced
5	Lan	South county	Public	2	14	10	N/A	130	(More or less) balanced
6	Fen	North county	Public	3	18	10	N/A	153	(More or less) balanced
7	Yun	South city	Private	3	16	10	N/A	29	More females
8	Zhe	North county	Public	6	18	11	SCI	156	(More or less) balanced
9	Jue	North county	Private	9	22	11	SCI	157	(More or less) balanced
10	Tie	South county	Public	17	18	12	HUM <sup>c</sup> and SCI	105	Female-dominated (HUM) More males (SCI)
11	Jie	South city	Public	20	16	10	N/A	100	(More or less) balanced
12	Ye	South city	Public	21	14	11	HUM	100	More females
13	Yuehan	South city	Private	2	4	12	N/A	25	(More or less) balanced
14	Lixie	North county	Public	6	18	10	N/A	132	(More or less) balanced
15	Sunyue	North county	Private	7	22	11	SCI	153	(More or less) balanced
16	Zengque	South city	Private	7	20	12	N/A	8	More males
17	Songqie	South city	Public	10	14	12	HUM	89	(More or less) balanced
18	Zhoujie	North county	Public	23	20	11	SCI	156	More males
19	Duantie	South county	Public	26	16	10	N/A	117	(More or less) balanced
20	Fengbie	South county	Public	30	16	10	N/A	123	(More or less) balanced

Notes. a. Years spent as teachers of English. b. SCI = Sciences. c. HUM = the Humanities.

(see Section 5.2.2 for more details). Evidently, teachers in public schools tend to have more students, which makes sense ; for most teachers, the gender ratio of students was more or less balanced.

## **6.2 Integrated Finding 1: Teachers Endorsed Female-Advantage-in-Languages Stereotype**

Both Studies 1 and 2 uncovered evidence that teachers endorsed FALS. In Study 1, multivariate, as well as univariate, main effect of learner gender was found to be statistically significant, and the mean ratings for female learners were higher than those for the male ones. These quantitative results are presented in Section 6.2.1 below. In addition, 18 out of 20 interviewees in Study 2 also appeared to harbour the same stereotype, as documented in Section 6.2.2. To be more specific, Sections 6.2.2.1-3 will a) explain how the achievement, affect, and aptitude components of FALS were interpreted by teachers, and b) provide evidence that interviewees endorsed each component. Subsequently, Section 6.2.3 combines and compares results from both datasets to discuss differences in magnitude and scope of the three components. Finally, Section 6.2.4 exhibits accounts where interviewees explicitly acknowledged their awareness of FALS, including those from Yuehan and Jian, the only two non-believers of FALS. This further reflects FALS's wide audience.

### **6.2.1 Evidence from Questionnaire Dataset**

A 2 (teacher gender)  $\times$  2 (region)  $\times$  2 (learner gender) mixed MANOVA was performed on three dependent variables (achievement, affect, and aptitude ratings), where teacher gender and region were between-subject factors, and learner gender was the within-subject factor. This statistical procedure addressed the fourth hypothesis in the analytical framework of Study 1 (see Figure 4.2), the hypothesis that teachers endorsed FALS on the macro level, and the endorsement may have varied due to their gender and region. This section will only report results indicating whether teachers endorsed FALS, and those regarding gender and regional variations will be disclosed in Sections 6.4-5.

**Statistical assumptions.** There was a linear relationship between the dependent variables, as assessed by scatterplot, and no evidence of multicollinearity, as assessed by Pearson correlation ( $|r| < 0.9$ ). There was multivariate normality, as assessed by chi-square versus Mahalanobis distance plot. No multivariate outliers were detected using Mahalanobis distance ( $p > .001$ ).

However, there were 12 univariate outliers in the data, as assessed by inspection of boxplots. Thus, two MANOVAs were run, one on the whole dataset, and another on the

dataset without these outliers. As the results were substantially similar, the results obtained from the whole dataset will be reported here.

There was homogeneity of covariance matrices, as assessed by Box's M test ( $p = .001$ ), and homogeneity of variances, as assessed by Levene's Test of Homogeneity of Variance ( $p > .05$ ).

**MANOVA results.** There was a statistically significant main effect of learner gender on the combined dependent variables,  $F(3, 50) = 19.856, p < .001$ , Wilks'  $\Lambda = .456, \eta_p^2 = .544$ . Therefore, univariate main effects of learner gender were examined for all three dependent variables, and as Table 6.2 below shows, the effects were statistically significant for all three dependent variables.

**Table 6.2: Univariate Main Effect of Learner Gender in Teacher Dataset in Study 1**

Stereotype component	Learner gender	Mean	SD	$F(1, 52)$	$p$ -value	$\eta_p^2$
Achievement	Female	5.254	1.060	47.311	< .001	.476
	Male	3.600	1.210			
Aptitude	Female	5.169	1.117	39.452	< .001	.431
	Male	3.683	1.142			
Affect	Female	5.237	.953	58.371	< .001	.529
	Male	3.500	1.200			

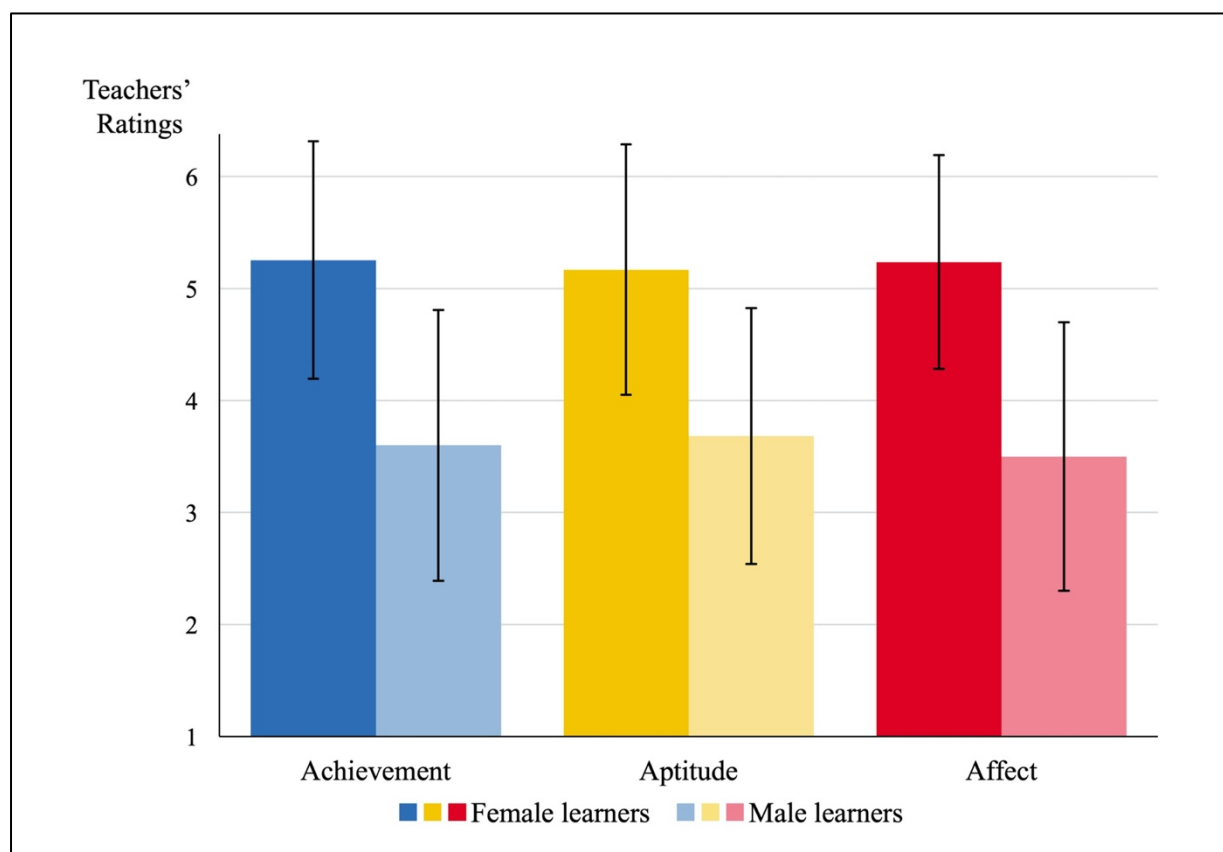
Figure 6.3 here illustrates the means and standard deviations for teachers' achievement, affect, and aptitude ratings as a function of learner gender. From both Table 6.2 and Figure 6.3, it is clear that teachers thought that female learners were better than their male counterparts in English achievement, affect and aptitude. That is, teachers did endorse FALS. The results here corresponded to those in the guardian and student datasets in Chapter 5—so far, evidence suggested that all three groups of participants in Study 1 endorsed FALS at least on the macro level. Another similarity between the three datasets in Study 1 is the relatively large standard deviations, which signals variability among participants' stereotypical images of boys and girls (Section 9.1.3.2 will discuss this in further detail).

## 6.2.2 Evidence from Interview Dataset

After the first-cycle coding, 103 data entries from 18 interviewees were coded as manifesting teachers' endorsement of FALS. In other words, out of 20 interviewees, 90% appeared to have stereotypically believed that females, instead of males, were better learners

of English. Interestingly, the only two disbelievers, Yuehan and Jian, explicitly acknowledged that they were aware of FALS (see Section 6.2.4).

**Figure 6.3: Achievement, Affect, and Aptitude Ratings as a Function of Learner Gender in Teacher Dataset in Study 1**



To fully understand the scope of FALS held by teachers, each FALS coded datum was subcoded according to three features: content, function, and salience (see a list of each subcode, including their definition and exemplars, in Appendix T). This section will focus on textual evidence regarding FALS content (one sub-section for each component), and the remaining two qualities will be discussed in Sections 7.1-2. Interview segments from the non-believers, Yuehan and Jian, will be presented in Section 6.2.4 to further expose the ubiquity of FALS.

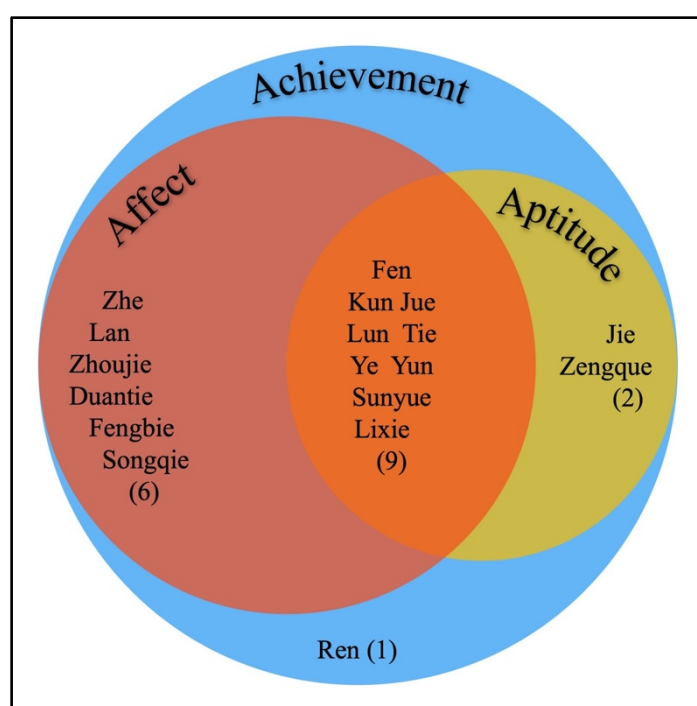
Corresponding to the dependent variables in Study 1, the subcodes about FALS content are ‘achievement’, ‘affect’, and ‘aptitude’<sup>26</sup>. Data entries were subcoded ‘achievement’ if they characterised females outperforming males in language skills (listening, speaking, reading, and writing), language knowledge (vocabulary and grammar), tests and exams, and

<sup>26</sup> For detailed definitions of these subcodes and exemplars, please refer to Appendix T.

other language-related tasks or activities. Those subcoded ‘affect’, on the other hand, described female learners’ more positive affective, emotive, motivational or attitudinal profiles. Finally, those subcoded as ‘aptitude’ depicted female’s biological superiority, cognitive advantage, and other ‘gift’-related strength in contrast to males’ weaknesses. As could be expected, the three subcodes overlapped because a single datum may include narratives about more than one content.

All 18 subscribers of FALS thought females exceeded males in English achievement. In fact, 79 coded segments were about achievement, an overwhelming 76% of all FALS segments. Aside from achievement, teachers in Study 2 also stereotypically presumed gender differences in affect and aptitude: 15 thought that girls had a more positive affective relationship with the English subject, and 37 coded segments were about affect, approximately 36% of all FALS segments; 11 considered girls as more gifted English learners, but only 17 coded segments, a mere 17%, were related to aptitude.

**Figure 6.4: Venn Diagram of Believers of Female-Advantage-in-Languages**



#### **Stereotype Components in Study 2**

Figure 6.4 here uses a Venn diagram to indicate the endorsers of each FALS component and combinations of FALS components. The blue circle represents the ‘achievement’ component, and the red and the yellow ones stand for ‘affect’ and ‘aptitude’ respectively. The nine interviewees in the region where three circles overlap are those endorsing all three FALS aspects. The six participants locate where the red and

the blue circles intersect simultaneously endorsed the affect and achievement components<sup>27</sup>, and the two in where the yellow and blue circles meet believed in the aptitude and achievement components. And Ren, sitting only in the blue circle, believed in the

<sup>27</sup> Zhe’s interview was curtailed before the interviewer could have inquired about her opinion on language aptitude, because she had to leave early to teach. Since Zhe did not consent to continue her interview at another time, her views on aptitude remained unclear. Nevertheless, this does not change the finding that achievement and affect aspects were endorsed by more participants compared to the aptitude aspect.

achievement component solely. In Sections 6.2.2.1-3, each component will be discussed accordingly.

### 6.2.2.1 Teachers' Agreement and Disagreement on *Achievement*

A supposed female excellence in English achievement emerged among all 18 FALS endorsers, making it the most popular component. However, teachers' perspectives varied regarding what 'achievement' actually entailed, as Table 6.3 here illustrates.

Apparently, all 18 subscribers of the achievement component agreed that girls tended to get higher grades in exams than boys. In fact, as quotations<sup>28</sup> here would demonstrate, not only did teachers consider girls as higher-achievers on average, they also thought that the top and middle tiers were likely to be female-dominated while the bottom tier is clustered with males.

**Table 6.3: Coded Segments Describing Female Excellence in English Achievement in Study 2**

Area of achievement	Coded segments	Endorsers (number)
English skills	Listening	15 Fen, Zhe, Kun, Jue, Tie, Lan, Ye, Sunyue, Lixie, Duantie, Songqie (11)
	Speaking	18 Fen, Zhe, Kun, Jue, Tie, Lan, Ye, Yun, Sunyue, Lixie, Duantie, Zengque, Songqie (13)
	Reading	14 Fen, Zhe, Jue, Tie, Lan, Ye, Zhoujie, Sunyue, Lixie, Duantie, Songqie (11)
	Writing	17 Fen, Zhe, Kun, Jue, Tie, Lan, Ye, Zhoujie, Sunyue, Lixie, Duantie, Songqie (12)
English knowledge	Vocabulary	15 Fen, Zhe, Kun, Jue, Lun, Tie, Ye, Jie, Yun, Zhoujie, Sunyue, Lixie, Songqie (13)
	Grammar	9 Fen, Zhe, Kun, Lun, Ye, Jie, Sunyue, Songqie (8)
Exam performance		39 Fen, Zhe, Kun, Jue, Lun, Tie, Lan, Ye, Ren, Jie, Yun, Zhoujie, Sunyue, Lixie, Fengbie, Zengque, Songqie (18)
Others		11 Zhe, Kun, Jue, Tie, Lan, Yun, Lixie, Songqie (8)

<sup>28</sup> Please refer to Table 4.4 for symbols representing identifiable non-verbal utterances, such as laughter, in-/out-breath, pauses and hesitations.

*Generally, you know, girls, compared to boys, I mean on average, are better at English. That is what typically happens. And of course, ... there are boys who have relatively good grades in English. But usually, boys just don't do English as well as girls. [Fen, Paragraph 118]*

*You know, generally we may, after we've had an exam, we'll have some high-achievers in a class, such as those ranking top ten, or those getting the highest marks in their classes. Basically, the majority of these will be girls. You know, boys who can do well in English exams, and who can rank top, are relatively much fewer. [Zhe, Paragraph 147]*

*... Um.: I think, 130, that would be a pretty good grade [out of 150], right? I think from that score and above, there will be more girls. ... In general, whether it is the top or the middle tier, I think there are always more girls. [Lan, Paragraph 146-148]*

*... Those with high marks in English exams ... are absolutely girls, absolutely girls. ... Those with low marks ((a short pause)), as far as I am concerned, ... those will definitely be mainly boys. Even on the same academic level, I mean, if you look at the overall academic results, and compare a boy and a girl on similar academic levels, that boy's performance in English will definitely be weaker, much weaker. ((She sighs)) [Tie, Paragraph 132-134]*

Of course, some cautious readers might object to construing the quotations above as manifestations of FALS endorsement. As they may well maintain, such quotes could be fact-based, i.e., teacher may have merely described gender differences they have been observing based on their current and/or past teaching experience. In other words, the question here is whether or not these accounts can be considered gender-stereotypical if they genuinely reflect the 'truth' as perceived by teachers. And the answer may arguably still be affirmative, if we review the very definition of 'gender stereotypes': Since the term refers to generalist images about males and females shared by a certain culture or community, the source of these images and the extent to which they are seen as factual do not disqualify them from being gender-stereotypical. Naturally, the same logic works for all similar quotes that follow<sup>29</sup>.

---

<sup>29</sup> Of course, teachers may still render gender stereotypical comments even if their current students contradict FALS, as Fengbie's quotation on p. 168 (Paragraph 142-148) about affect will show. This phenomenon was not observed with the achievement component, though.

Curiously, the assumed female superiority in exam performance was the only aspect of achievement that all 18 FALS believers agreed upon. In fact, they disagreed on whether girls outperformed boys in other areas to varying extent. For example, approximately similar proportions of endorsers also expressed that girls were better at two productive skills, speaking and writing (72% and 67% respectively). Those speaking of girls' outperformance in receptive skills, listening and reading, were slightly fewer (61%).

*... And one more thing, the majority of students [who are good at listening] are girls. ... I suppose among those five or six students, four are girls. Yes. One of them is boy. ((He chuckles.))* [Songqie, Paragraph 40]

*... There ... is [gender] difference among those who speak well. Those who can speak well are still:: Those girls [we talked about earlier] are good enough. [By 'speak well',] I mean the ability ((a short pause)) to speak longer English chunks in a relatively fluent manner, and the way the words is organised is also quite sensible.* [Jue, Paragraph 72-76]

*... Regarding reading performance, ... there seems to be more girls with good results.* [Ye, Paragraph 116]

*Students who write well, I think, they are adept at accumulating words, phrases, and sentence structures. ... Such accumulation we are talking about, typically involves these students using a designated notepad, a notebook. Usually, we [teachers] assign students to recite paragraphs from a text. These students will note down the eloquent sentences from those paragraphs in that notebook. Or, when they work on a practice exam paper, they may encounter well-formed sentences, and they will also copy them down. Or else, like words. Any language item is likely [to be noted down]. ... Most of these students are girls.* [Lixie, Paragraph 104-108]

In contrast to the four abovementioned skills, English knowledge was where a greater divergence in FALS endorsement was discovered, with 72% of believers thinking that girls had better knowledge of vocabulary but just 44% concerning grammar. The consensus about vocabulary was that, girls, compared to boys, tended to grasp English vocabulary more solidly in quantity and quality: FALS endorsers were convinced that girls knew more English vocabulary and that their retention was better.



*Uh:: Honestly, I think in my class, girls' surely have a better knowledge of vocabulary than boys. ... Even from the routine exercises they do, I mean the sort of practice reading exams they routinely do, I can see that boys, in their copies of the exam papers, write down far more notes of English words' Chinese meaning. ((He chuckles)) So, it is obvious: [boys] just lack knowledge about English vocabulary. [Songqie, Paragraph 76-78]*

*... I give them a test [on vocabulary]. I named it, which is held once a week, 'Wordsman Test'<sup>30</sup>. And then we have Diamond, Gold, and Silver<sup>31</sup> [Prizes]. This is supposed to be a fair test for them, because I'll only ask about the words I've just taught during the week. ... We just started last week, and all prizes went to girls. ((She giggles)) ... As for this week's Wordsman Test ... we'll see. In the end, tonight, we'll definitely see girls winning again. [Yun, Paragraph 144-152]*

With grammar knowledge, the picture becomes more complicated. Eight interviewees were convinced that girls could outperform boys in analysing grammatically complex sentences, and producing grammatically correct items in compositions and tests.

*Speaking of mastering grammar, boys, I feel that their knowledge is not as detailed as girls'. ... Girls, they are more rational, and they will analyse whether a component might be inversion<sup>32</sup>, or an adverbial clause<sup>33</sup>. Boys, they can say mechanically, 'the start of a sentence is just its subject.' Somewhat less detailed than girls. [Sunyue, Paragraph 148]*

*... There's about twenty-something [good writers in my class]. ... In the sentences they write, there will be relatively fewer grammatical errors. ... For instance, they won't make mistakes about when to use the adjective form or the adverbial form [of a word]. ... And among the twenty students and more, girls may take up more or less twenty. I mean boys, ((She giggles)) boys, ((She giggles again)) are the 'something', the oddments. [Zhe, Paragraph 136-141]*

---

<sup>30</sup> The phrase 'Wordsman Test' is English in the original transcript.

<sup>31</sup> The three prizes are also English in the original transcript.

<sup>32</sup> Inversion is a literary technique through which normal word orders are reversed to achieve emphasis. e.g., *What a beautiful picture this is!*

<sup>33</sup> An adverbial clause is a dependent clause that functions as an adverb, used to modify a word, a phrase or another clause that serves as a noun, an adjective, or an adverb. e.g., *I remembered, after I'd left the house, that I'd forgot to lock the door.*

*Grammar is where all my students are relatively weak. The best they can do would be to know a lot about a certain [grammar] item. But as for some other items, they would be lost. They wouldn't know a lot. ... But if you take a look at the whole grade, in terms of proportions, there're more girls who master grammar well. Averagely speaking, girls are indeed better in grammar. [Jie, Paragraph 120-124]*

But, among 18 FALS endorsers, five participants actually insisted upon a lack of gender difference in grammar knowledge (Jue, Yun, Lan, Zhoujie, and Lixie). The remaining five participants claimed ignorance on the topic, either because their teaching did not involve grammar (Duantie, Fengbie) or because they did not examine students' grammar knowledge in isolation (Tie, Ren and Zengque). In other words, more data are needed in order to determine if FALS about grammar achievement was really endorsed by fewer teachers compared to other achievement-based FALS.

#### **6.2.2.2 Teachers' Congruous Understanding of *Affect***

During the interviews, interviewees were invited to talk about the affective, emotive, motivational or attitudinal profiles of learners. After the first coding cycle, 68 data segments were coded "students' English affect" in 20 interviews, and it was found that teachers concurred in their perception of affect. Almost all teachers conceptualised *affect* as a two-element construct, the first of which was personal (an interest in English language, the English subject, and/or the relevant culture), and the second was relational (a pleasant relationship with the English teacher). The following quotes exemplify this finding:

*I mean, when you first start to learn English, at the very beginning, the first and foremost thing needed is your interest in the subject. [Ye, Paragraph 168]*

*Interest. I feel that it is genuinely interest [that contributes to learning English]. Interest is the most important [factor]. A student might feel (.) that he/she<sup>34</sup> is also quite close to me. If a student likes the [English] teacher, [he/she] will like the English subject. Or perhaps he/she likes English, and so he/she grows fond of the teacher, right? [Fengbie, Paragraph 112]*

---

<sup>34</sup> In Chinese Mandarin, the third person pronouns, 'he', 'she', and 'it', have the same pronunciation: /ta:/. Thus, in the interview excerpts quoted in this thesis, the term 'he/she' will be used when the gender of the person being referred to is not specified.

*Um:: First and foremost, he/she:: needs to be interested [in order to learn English well]. And this is not just in English—he/she should also be rather interested in me. Otherwise, I mean, I myself managed to learn English quite well due to my being fond of my English teacher. ... For example, if a student finds the teacher amusing, or that the teacher is impressive. Students want teachers to be humorous, funny, and become a model for them to look up to or something. [Lan, Paragraph 125-128]*

In the three excerpts above, teachers were talking about affect without comparing male and female learners. When they did compare female and male learners, 15 thought that girls were more enthusiastic learners of English, as illustrated by the following two quotes:

*... Compared to girls, boys are not very willing to learn words. This is related to interest in learning, and intrinsic motivation. ... Boys' motivation—because of boys' way of thinking, [boys] should be more inclined to ((a short pause)) do abstract thinking, and be rather strong in logical thinking. And then, they lack certain interest in, well, language stuff. Chinese, English, are what they lack interest in. ... So by comparison, they are just not interested in English. [Lixie, Paragraph 163-169]*

*Another problem is that, students' motivation is just not strong, ... especially boys. Many boys tell me, 'Ms. Yun, don't get me to memorise vocabulary. I'll forget. I memorise them for once, and then I forget once and for all. Next time I see the words, I still won't know them. That's the way it is.' They just keep denying themselves. It's mainly boys. [Yun, Paragraph 52]*

For three believers, the affect component only not portrayed girls in a more positive light, but also paints boys as more liable to hold negative feelings towards English. For instance, Fen here said that at the beginning of the term, she asked students to express what concerns or difficulties they had for the English subject:

*... I asked every student to stand up and speak up, to [say their problems], and some just said, 'Miss, honestly, I am just not interested in English.' ... Typically these students are not well grounded in English. They have failed to learn it [in the past]. The more difficult for them to learn, the less interest they have for it. They fear English. ... These students are generally boys. ... Girls, they don't say such things. [Fen, Paragraph 66-80]*

Such sweeping stereotypes may be an over-simplification of the affective relationships boys and girls actually have with English. For example, Dewaele and colleagues (2016) surveyed over 1,700 foreign language (FL) learners across the world, and they found that not only did girls experience more excitement and enjoyment, they also reported higher levels of anxiety. This wider range of emotions from girls also appeared in Study 4, which are reported and discussed accordingly in Section 8.2.1, as well as Section 9.2.2.

Obviously, all 20 interviewees reached a consensus on the meaning of affect. Among 18 FALS endorsers, 15 subscribed to the affect component. Therefore, in comparison with the achievement component, affect was less prevalent, yet consensual.

### 6.2.2.3 Teachers' Diversified Interpretations of *Aptitude*

Before analysing the aptitude aspect of FALS, it was necessary to investigate how teachers understood the concept. After the first coding cycle, 42 data segments from 19 interviewees<sup>35</sup> were coded 'students' English aptitude', and it was found that teachers' understanding of aptitude was diversified, ranging from biological mechanisms in the brain to specific cognitive faculties. As Table 6.4 here will illustrate, a total of nine interpretations were proposed in Study 2.

**Table 6.4: Teachers' Diversified Interpretations of *Aptitude* in Study 2**

Interpretation		Notes
1 Memory	Definition	The innate ability to remember linguistic items
	Endorsers	Fen, Kun, Jue, Ye, Ren, Duantie, Yuehan (7)
	Example	<i>I meant an aptitude for language. ... It might be like, for some students, like girls, they memorise words really fast.</i> [Jue, Paragraph 118-120]
2 Language savvy	Definition	The ability to notice and understand language faster than others
	Endorsers	Fen, Kun, Lun, Lan, Jian, Lixie (6)
	Example	<i>Aptitude, it seems to me, is a relatively strong language savvy. Sometimes when I teach grammar, I feel that girls just (.) find it easier to get the message, to understand things.</i> [Fen, Paragraph 130]
3 Language intuition	Definition	The ability to know or understand language based on feelings instead of facts or evidence
	Endorsers	Tie, Jie, Yun, Sunyue, Zengque (5)

<sup>35</sup> Since Zhe's interview could not manage to include questions about language aptitude, her views on the issue remained unclear.

4 Verbal ability (especially for production or expression)	Example	<i>It means, for example, if someone has a strong sense of a certain language, he/she may not be able to say why, but he/she just knows that it is how it works. So [for me 'language centre'] means a sense of languages. [Jie, Paragraph 134]</i>
	Definition	The ability to use language to understand and express information
	Endorsers	Lun, Yun, Lixie (3)
5 Imitation capacity	Example	<i>Well, to determine if one's language aptitude is high, one should examine his/her, well, verbal, I mean, the ability to understand [others] and to express [oneself] with words. [Lixie, Paragraph 184]</i>
	Definition	The ability to copy language produced by others
	Endorsers	Zengque, Jian (2)
6 Sensitivity to linguistic stimuli	Example	<i>Well, perhaps my interpretation would be, whether or not someone has an aptitude for language will depend on if he/she, when he/she listens to a new language, can imitate the language fast, in terms of features like pronunciation or intonations. [Jian, Paragraph 106]</i>
	Definition	The ability to react to subtle features of linguistic items
	Endorsers	Sunyue, Zengque (2)
7 Smartness	Example	<i>I just think that girls probably are relatively sensitive to this; they are relatively sensitive to language input. [Sunyue, Paragraph 132]</i>
	Definition	The ability to think quickly or intelligently
	Endorsers	Zhoujie, Fengbie (2)
8 Thinking capacity	Example	<i>I feel that gift is not important [for learning English]. More often than not, those who are gifted can't learn English well. The smarter a student is, the worse his/her English will be. [Zhoujie, Paragraph 130]</i>
	Definition	The ability to process information in order to make connections, find patterns, or create new ideas
	Endorsers	Songqie (1)
9 Tone quality	Example	<i>Gift, gift does not come into English learning. I think compared to math, physics and chemistry, the thinking capacity it [English] requires, ((he clicks his tongue)), is quite minimal. [Songqie, Paragraph 98]</i>
	Definition	The quality of a speaker's voice being intelligible and pleasant
	Endorsers	Duantie (1)
	Example	<i>I don't think that boys lack [language aptitude]. ... Boys' voices sound attractive, especially when they reach Grade 12. When boys turn 17, 18, there will be voices rich as a baritone, or even a bass. [Duantie, Paragraph 76]</i>

As shown in the table above, the three mainstream interpretations of ‘language aptitude’ encompasses memory, language savvy, and language institution. Furthermore, Table 6.4 here also reveals aptitude as the most contentious component, since even the most popular understanding, memory, was only raised by seven interviewees. Compared to achievement and affect components, aptitude was endorsed by the least number of interviewees—a mere eleven. Interestingly, among the remaining eight who talked about aptitude, one was ambivalent, two insisted that there was no gender difference, and five believed in a male advantage in aptitude (see Section 7.3.1).

### 6.2.3 Differences in Magnitude and Scope of Three Components

According to Figure 6.4, the smallest number of interviewees stereotypically believed in gender difference in aptitude (11 out of 20, 55%). Similarly, among the three effect sizes ( $\eta_p^2$ ) in Study 1, the smallest was found with the aptitude component (.431). Thus, it would appear that not only was aptitude the least commonly endorsed aspect of FALS, it was also the component with the smallest magnitude<sup>36</sup>. Two reasons might have led to its lack of popularity and small magnitude: disagreement over an operational definition and a supposed male biological/cognitive distinctiveness. As reported in Section 6.2.2.3, teachers could not even agree upon the definition of ‘aptitude’ in Study 2. Moreover, five teachers even thought that male learners had gender-specific biological/cognitive advantage over female ones when learning English (see Section 7.3.1).

The situations with the other two components, achievement and affect, are rather intricate. In Study 1, the effect size of achievement ( $\eta_p^2 = .476$ ) was smaller than that of affect (.529). Yet, in Study 2, the achievement aspect was endorsed by more teachers (18, 90%) than affect (15, 75%). Thus, combining the results from both data sources, achievement would appear to be most widely endorsed among three components of FALS, while affect was the largest of the three with regards to magnitude. Such a seeming divergence, nevertheless, does not necessarily mean that the two studies are in conflict. First, the wide scope of FALS achievement seemed reasonable due to the salience and directness of English achievement: teachers of English would inevitably be more concerned with students’ achievement, including test scores, and mastery of language skills and knowledge; in addition, achievement, unlike the other two FALS components, could be measured and compared straightforwardly and more frequently. Moreover, the fact that achievement’s

---

<sup>36</sup> This exact pattern repeated itself among students in Study 4, as Section 8.2.2.2 will reveal. The implications of such convergence between the teacher and student datasets will be discussed in Section 9.X.X.X.

magnitude is smaller than that of affect can be supported by another finding: in both studies, the stereotypically assumed gender differences in aptitude and affect were thought to have led to the postulated gender gap in achievement (see Section 6.3). Given teachers' disagreement over aptitude, their gender-stereotypical opinions about achievement would be likely to be moderated despite their strong conviction about girls' relative advantage in affect.

#### **6.2.4 Teachers' Awareness of Female-Advantage-in-Languages Stereotype**

17 interviewees acknowledged that they were aware of at least one aspect of FALS circulating among colleagues (by seven participants), students (by three), guardians (by Lixie), and media (by Jue), including the two non-believers—Jian and Yuehan.

For example, Lixie maintained that in general, girls tended to outperform boys in four English skills. He then commented that his opinion was shared by colleagues and guardians:

*Through my conversations with colleagues, I've discovered that many a person hold the same opinion as I do, including colleagues, and the guardians of students, too. They also hold this view. ((Interviewer asks, 'what specifically do you and your colleagues talk about?')) Specifically, well, in fact all people may think that it is due to the innate aptitude for language—girls' [aptitude] is just better than that of boys. So this difference in aptitude will manifest itself in skills, in listening, speaking, reading, and writing. And another point is, about interest in learning [English]. There is also a [gender] difference, so this is another factor. ... Yeah, guardians have such ideas, too. Because some guardians, especially those of male students, they would often say that their children's English are rather bad. And they would say, 'boys all suck at English. How do they improve?' They just worry like this ((He chuckles)), often. [Lixie, Paragraph 112-118]*

Another example is Jian, who attested to the durability of FALS throughout both her parents' and her own generation. As Jian recalled, when she was an undergraduate majoring in German, the gender ratio in her class was strikingly imbalanced: three males and 25 girls. According to her description, she, together with her fellow classmates, was obviously informed by FALS:

*In university, when we first started, we all thought ... 'boys just shouldn't do this.' Why would they? ((She giggles)) At that times, we girls would also say, 'So weird! Why have these boys come here?' [Jian, Paragraph 82]*

Later on, Jian expressed that the same gendered pattern was perceived in the teaching team, which consisted of one male teacher and six female ones. She further explained that the phenomenon could have arisen due to the same stereotype.

*Um:: ((She clicks her tongue)) It is still relevant [to the same stereotype]. What's more, if we take a look at the past, probably, such as in the 1950s and 1960s, among those who could go to college, still, many a girl would choose languages, and the Humanities domains. Many a boy would choose sciences and engineering ones. ... So if you think about it, I mean, from our parents' generation to our generation, the situation remains the same. And the trend is getting more and more ((a short pause)), it hasn't changed in the opposite direction. It's still, still, the same trend persists, actually. [Jian, Paragraph 118-120]*

Except for the above-mentioned 17 FALS-aware interviewees, the remaining three can be divided into two sub-groups: Duantie and Fengbie did not clarify whether they thought other people endorsed FALS; Zhoujie, on the other hand, was the only interviewee completely unsuspecting of FALS's broad audience.

*((Interviewer asks, 'so the view you just expressed, 'girls get better results in English than boys', do you think that it is held by other people?') Perhaps not. ... ((Interviewer asks, 'so if generally, if you ask some random person, 'who do you think can do English better, boys or girls?' What will the answer be?)) Of course, it will be 'boys'. Boys' ability to understand, usually people think that boys' ability to understand is stronger, so they would all say that boys can learn English better. [Zhoujie, Paragraph 83-86]*

Clearly, Zhoujie was aware of the stereotype that males are more competent than females, and he was convinced that compared to this stereotype, FALS must have been much less popular. In fact, similar to Zhoujie, seven more participants harboured both stereotypes simultaneously, too (see more detailed discussions in Section 7.3.1). Coincidentally, one boy in Study 4 denied in the group interview that he had ever heard of FALS (see Section 8.2.3 for more). Nonetheless, the fact that only Zhoujie and the boy were heedless of FALS's wide circulation provided evidence that FALS was regarded commonplace by the majority of teachers and students (Section 8.2.3).



### 6.3 Integrated Finding 2: Causal Links among Three Components of Female-Advantage-in-Languages Stereotype as Perceived by Teachers

Section 6.2 above has introduced that after the second coding cycle, it was found that among 18 FALS endorsers, 11 considered girls as more gifted English learners, and 15 thought that girls had a more positive affective relationship with the English subject. Further analysis revealed that the majority of interviewees believed that it was girls' stronger language aptitude and more intense liking for English that contributed to their relative learning success, though to varying degrees. This qualitative finding will be presented in Section 6.3.1 below. It was thus suspected that a similar pattern might have existed in the questionnaire dataset, where the assumed gender differences in aptitude and affect may both have predicated the differences in achievement. A multiple regression was utilised to explore this hypothesis, and the result will be reported in Section 6.3.2 below<sup>37</sup>.

#### 6.3.1 Evidence from Interview Dataset

Figure 6.5 presents evidence from Study 2. The inner darker circles hold believers of FALS components, where the blue, red, and yellow colours represent achievement, affect, and aptitude components respectively. The outer lighter circles, instead, contains non-believers of FALS<sup>38</sup>. The numbers in filled arrows stand for the quantity of FALS believers assuming a causal connection between two FALS components, while those in hollow arrows show non-believers presuming the same relationship. The hollow arrows, in addition, mean that the links are conditional or limited. The positive or negative signs of the arrows signify the nature of the causal associations as conducive or hindering.

The arrows linking affect to achievement indicate that both the 15 FALS believers and the five non-believers thought that a favourable attitude towards English could contribute to higher attainment. Besides, three believers (Zhe, Lan, and Ye) also thought that good performance might even result in a stronger interest in English, forming a virtuous circle for female learners. The two quotes below will exemplify such two sorts of relationships:

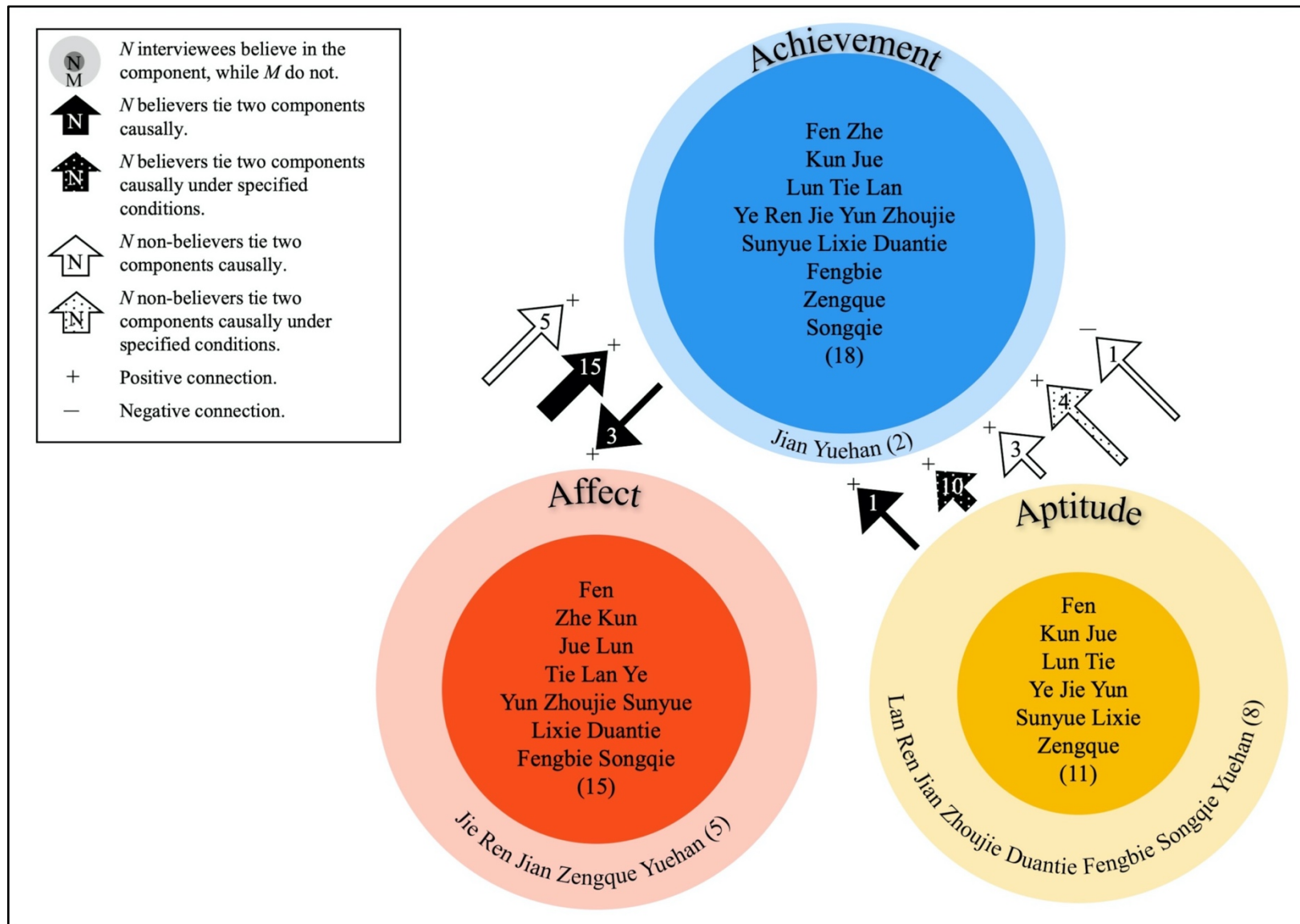
*With regards to speaking performance, [using English] to express ideas, girls generally tend to do better. Because, how to put it, for girls, even with their*

---

<sup>37</sup> The same causal links between FALS components also emerged in Study 4, group interviews with students, which led to a multiple regression analysis on the student dataset in Study 1. See Section 8.2.2.2 for details.

<sup>38</sup> Again, Zhe is missing from the yellow circles because her interview did not cover aptitude.

**Figure 6.5: Perceived Causal Links between Aptitude and Achievement, and Affect and Achievement by Teachers in Study 2**



*attitude, with their willingness to learn English, girls already have an advantage [over boys]. [Tie, Paragraph 98]*

*Such proactivity, I mean, speaking of being proactive in directing their learning problems to teachers, boys still differ from girls. Because in the end, in fact, it is the scores that determines [the degree of] their proactivity. Because when students get higher grades, they will feel that their English is good, that they are probably quite competent learners of English. So, they will be willing to learn. If they learn, they will encounter problems, they will be willing to ask [the teacher]. But, if their grades are not that good, perhaps it is not because of a lack of competence: they might have just not spent enough time [on English]. But they will just think, 'anyways, my English result is not good, my competence is not strong, so I'll never learn. So I'd better quit learning. This way, I won't have any problems, and I don't have to ask any.' So they'll avoid the teacher altogether. [Lan, Paragraph 166]*

The patterns of causal relations between aptitude and achievement are rather complicated. Only one FALS believer (Yun) and three non-believers (Jian, Duantie, and Yuehan) thought that the stronger the aptitude, the better the performance. For the ten remaining FALS believers and four other non-believers, the positive relation is either limited or conditioned, thus such a relationship is symbolised with dotted arrows. Three examples of the abovementioned types of causal links between aptitude and achievement are presented here: 1) a positive one, 2) a positive one with a limitation, and 3) a positive one with a condition.

*Um, those performing well in reading tasks, they, I think in everyday life, perhaps their perception powers are rather strong. ... Students who get high grades in reading comprehension, they will think, and they will understand. ... But some students are like, even after you've explained everything, they still won't get it. It is so frustrating. Their perception powers are not quite enough. ... In my classes, it is always boys who argue ... about ((she giggles)); in my classes, when we deal with cloze texts<sup>39</sup>, those who wrangle with me over why a word is chosen instead of another are always boys. They just can't get it. I mean, seriously, they really can't understand why. ... Girls just understand*

---

<sup>39</sup> A cloze text is an exercise, test, or assessment consisting of a portion of language with certain items, words, or signs removed, where the learners are asked to replace the missing language item.

*[English texts] faster. In terms of language, girls probably have a relatively stronger faculty of understanding. [Kun, Paragraph 112-122]*

*((Interviewer asks, 'in terms of English, if a learner wants to learn it well, how important do you think is aptitude?')) The proportion of [the contribution of] aptitude, 20%? ((Interviewer inquiries, 'then what about the rest 80%?')) The rest 80%, I think interest is very important. Interest is extremely important. I mean, because this [learning English] is something you must keep on working. ... I mean this, [learning English], if you are not interested in it, it will be very hard to do it repeatedly. [Zengque, Paragraph 163-166]*

*I think, ... if it's simply to reach the standard set by the College Entrance Exam, I suppose aptitude is not that important. The contribution is not big enough. More importantly, I think the results will depend on students' willingness to work hard, to keep on working. ... I just think, I mean, if someone wants to master a language that is different from the one he speaks in a short period of time, or if someone wants to master many languages, that will need aptitude to a very large extent. [Lun, Paragraph 186-188]*

Downplaying aptitude, exemplified by Zengque and Lun's quotes above, was common among 14 interviewees (see the dotted arrows in Figure 6.5). Eleven participants said that the contribution of aptitude was limited, with a lowest 5% (by Ren) and a highest 60% (by Fengbie). Among them, seven teachers put the proportion between 20% and 30%. The remaining three, Lun, Ye, and Sunyue, set specific conditions for aptitude to be significant, such as Lun's quote above.

Curiously, Zhoujie was the only teacher who established an adverse relationship between aptitude and affect, which is represented by a negative sign in Figure 6.5. A close reading of the following two excerpts will further clarify his ideas:

*I feel that gift is not important [for learning English]. More often than not, those who are gifted can't learn English well. The smarter a student is, the worse his/her English will be. ((Interviewer further questions, 'what do you mean by 'smart'?')) I mean the capacity to learn fast. They also forget fast. Such learners lack perseverance. And if there is a lot to learn, they will lose interest in English. As their English gets worse and worse, they will become less and less interested in English. [Zhoujie, Paragraph 130]*

*((Interviewer asks, 'in terms of aptitude, who do you think is stronger? Boys or girls? Or is there no difference?')) It should be boys who have stronger aptitude. But boys don't work hard enough. They lack perseverance. And boys' personality, I think, smart students tend to favour fast food. Fast food, subjects that they can learn once and for all, no need to work on it any further. So, but English is, more often than not, you learn the vocabulary, you forget, and you try to remember again. A mere reliance on aptitude is far too limited. You also need perseverance, and you need resilience. These areas are where girls typically do better than boys. Boys are just playful. [Zhoujie, Paragraph 175-176]*

Apparently, Zhoujie thought that boys' smartness actually impeded their learning outcome, because they tended to be less resilient in the face of setbacks. For him, girls would keep on learning diligently despite their innate disadvantage in intelligence, and therefore, they could receive better results. Besides, the above quotes also implies Zhoujie' endorsement of another gender stereotype, one that portrays males as more intellectual learners, whose details will be disclosed in Section 7.3.1.

Based on the types and widths of arrows in Figure 6.5, it is reasonable to suggest that the causal link between aptitude and achievement could be less strong than that between affect and achievement, a finding that is supported by quantitative evidence in the questionnaire dataset (see the following section).

### **6.3.2 Evidence from Questionnaire Dataset**

Unlike the MANOVA procedure documented in Section 6.2.1, the multiple regression was not planned. It was inspired by the finding in Study 2 that aptitude and affect aspects of FALS might have contributed to the achievement aspect. The aim was to predict perceived gender difference in achievement from perceived gender differences in aptitude and affect.

**Statistical assumptions.** Statistical assumptions for multiple regression were checked before the procedure was applied. The assumption of normality was met, as assessed by a Q-Q Plot. There was linearity as assessed by partial regression plots and a plot of studentised residuals against the predicted values. There was independence of residuals, as assessed by a Durbin-Watson statistic of 2.405. There was homoscedasticity, as assessed by visual inspection of a plot of studentised residuals versus unstandardized predicted values. There was no evidence of multicollinearity, as assessed by tolerance values greater than 0.1. There

were no studentised deleted residuals greater than  $\pm 3$  standard deviations, or values for Cook's distance above 1. Two cases had leverage values greater than 0.2 (.23 and .22 respectively), but as their residuals were relatively small (2.45 and -1.23 respectively), their influences were considered small and thus were retained in the dataset for later analysis.

**Multiple regression results.** The multiple regression model statistically significantly predicted perceived gender difference in achievement,  $F(2, 56) = 47.266, p < .001$ ,  $\text{adj. } R^2 = .615$ . Both variables added statistically significantly to the prediction, and the regression coefficients and standard errors can be found in Table 6.5 below. The regression coefficient of affect is larger than that of aptitude, which again corresponds to the findings in the previous section.

**Table 6.5: Multiple Regression Analysis of Teacher Dataset in Study 1**

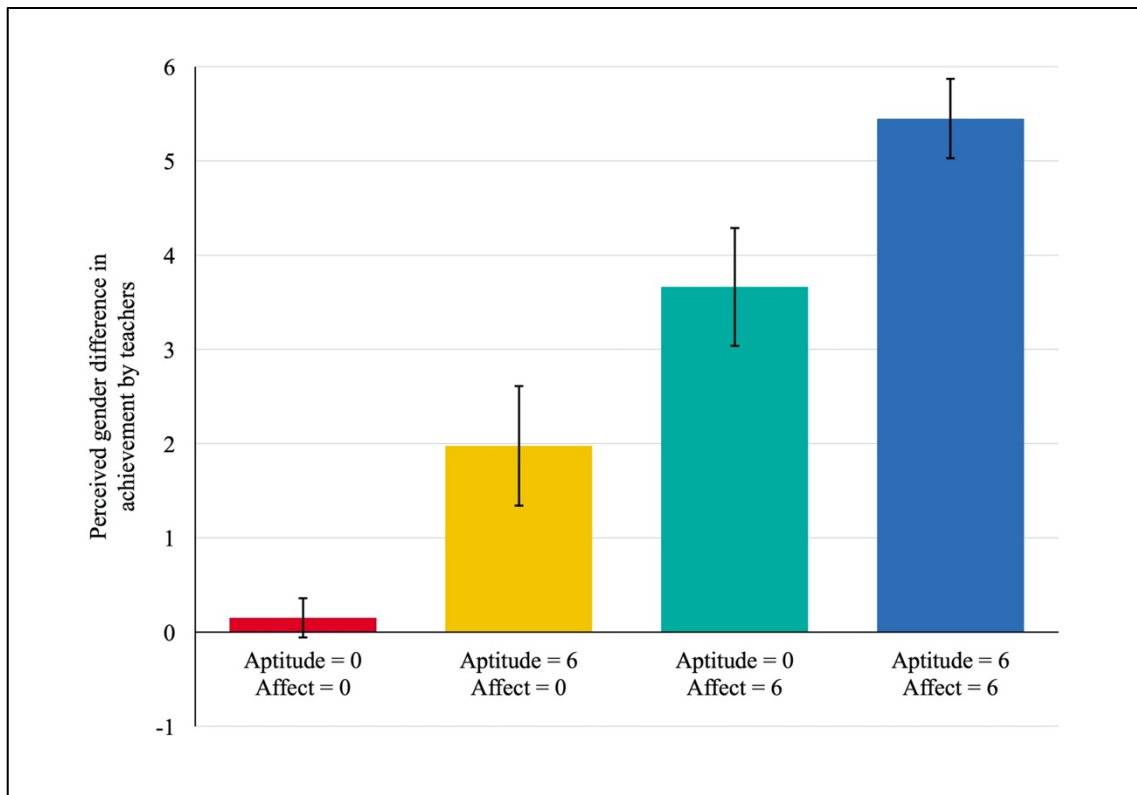
Variable	$B$	$SE_B$	$\beta$	$p\text{-value}$
Intercept	.152	.207		.467
Aptitude	.304	.103	.319	.005
Affect	.585	.116	.545	< .001

*Notes.*  $B$  = unstandardised regression coefficient;  $SE_B$  = standard error of the coefficient;  $\beta$  = standardised coefficient.

Predictions were made to determine the mean of perceived gender difference in achievement for teachers who assume maximum differences in aptitude and affect (both equals six). The mean was predicted as 5.448, 95% CI [4.645, 6.332], as the blue bar in Figure 6.6 shows. On the contrary, the mean of perceived gender difference in achievement by teachers observing no gender differences in aptitude and affect at all was predicted as .152, 95% CI [-.263, .567], as demonstrated by the red bar in the same figure. In other words, if a teacher does not endorse the aptitude and affect aspects of FALS, he or she is unlikely to endorse the achievement aspect either.

Predictions were also made to determine the mean of perceived gender difference in achievement for teachers who presume no difference in aptitude but maximum difference in affect, as illustrated by the green bar in Figure 6.6. The mean was predicted as 3.663, 95% CI [2.413, 4.913]. Then, the mean of perceived gender difference in achievement by teachers viewing maximum gender differences in aptitude but zero difference in affect at all was predicted as 1.977, 95% CI [.705, 3.249], seen as the yellow bar in the figure below.

**Figure 6.6: Predictions of Means of Perceived Gender Difference in Achievement Based on Four Combinations of Perceived Gender Differences in Aptitude and Affect from Teacher Dataset in Study 1**



#### **6.4 Integrated Finding 3: Gender Differences in Endorsement of Female-Advantage-in-Languages Stereotype**

In the questionnaire dataset, evidence supporting gender differences in FALS endorsement could not be found, similar to the cases of the guardian and student datasets (macro-level only). The same could be said for the interview dataset, too.

##### **6.4.1 Evidence from Questionnaire Dataset**

The two-way interaction effect between teacher gender and learner gender on the combined dependent variables was not statistically significant,  $F(3, 50) = .923, p = .437$ , Wilks'  $\Lambda = .948, \eta_p^2 = .052$ . This result reveals that male and female teachers did not differ in their endorsement of FALS. The univariate interaction effects between guardian gender and learner gender for each of the dependent variable were not followed, due to the insignificant multivariate interaction effect.

##### **6.4.2 Evidence from Interview Dataset**

In order to examine if gender differences existed in FALS endorsement among interviewees in Study 2, Fisher's exact probability test was run on three cross tabulations, each between gender and one aspect of FALS. Of course, in cases where participants' opinion

about aptitude and/or affect appear undecided or ambivalent, their attitudes were excluded from the contingency tables, and were thus discarded when conducting Fisher's exact probability test. Table 6.6 shows the summary data. As the table indicates, of the male teachers recruited, 7 (87.5%) considered that girls could outperform boys in English. For females, 11 (91.7%) also shared the same view. There was no statistically significant difference between male and female participants in their endorsement of the achievement component of FALS, as assessed by Fisher's exact test,  $OR = .636$ , 95% CI [.034, 11.909],  $p = 1.000$ . Similarly, no gender difference was found for aptitude,  $OR = .281$ , 95% CI [.038, 2.079],  $p = .332$ ; nor for affect,  $p = .515$ . Odds ratio for affect could not be defined because there were no non-believers among the male participants.

## **6.5 Integrated Finding 4: Regional Variations in Endorsement of Female-Advantage-in-Languages Stereotype**

It was found that regional variations in teachers' FALS endorsement were unlikely to have arisen in the questionnaire dataset, similar to the case of student dataset (both micro- and macro-levels). Furthermore, the same pattern persisted in the interview dataset, too.

### **6.5.1 Evidence from Questionnaire Dataset**

The two-way interaction effect between region and learner gender on the combined dependent variables was not statistically significant,  $F(3, 50) = 1.303$ ,  $p = .284$ , Wilks'  $\Lambda = .927$ ,  $\eta_p^2 = .073$ . This suggested that regardless of where teachers were from, they shared similar FALS.

Interestingly, regional variations were also missing from the student dataset (see Section 5.2.3.3), but present in the guardian dataset (see Section 5.1.3.3). It would seem inadvisable to decide that no regional variations existed among teachers with only the current evidence for two reasons: 1) there is collaborating evidence from previous literature suggesting regional differences in stereotype endorsement (Huo and Randall, 1991); and 2) the observed power was only .32. In fact, it could be argued that a larger sample size is needed to gather evidence about regional differences among teachers' FALS endorsement. In spite of this, converging evidence from the interview dataset does suggest a lack of regional variations, as the section below will soon unveil.

### **6.5.2 Evidence from Interview Dataset**

To uncover whether regional variations were present in FALS endorsement among participants in Study 2, Fisher's exact probability test was run on three cross tabulations, each between region and one aspect of FALS. Table 6.7 shows the summary data. As Table 6.7



Table 6.6: Cross Tabulations between Gender and Three Aspects of Female-Advantage-in-Languages Stereotype in Study 2

		Achievement			Total	Affect			Total	Aptitude			Total
		Against	For	Against		For	Against	For					
Gender	Female	Count	1	11	12	Count	2	9	11	Count	3	8	11
		% gender	8.3	91.7	100.0	% gender	20.0	81.8	100.0	% gender	27.3	72.7	100.0
		% achievement	50.0	61.1	60.0	% affect	100.0	60.0	64.7	% aptitude	42.9	72.7	61.1
	Male	Count	1	7	8	Count	0	6	6	Count	4	3	7
		% gender	12.5	87.5	100.0	% gender	0.0	100.0	100.0	% gender	57.1	42.9	100.0
		% achievement	50.0	38.9	40.0	% affect	0.0	42.9	37.5	% aptitude	57.1	27.3	38.9
Total	Count	2	18	20	Count	2	15	17	Count	7	11	18	
	% gender	10.0	90.0	100.0	% gender	12.5	88.2	100.0	% gender	38.9	61.1	100.0	
	% achievement	100.0	100.0	100.0	% affect	100.0	100.0	100.0	% aptitude	100.0	100.0	100.0	

Table 6.7: Cross Tabulations between Region and Three Aspects of Female-Advantage-in-Languages Stereotype in Study 2

		Achievement			Total	Affect			Total	Aptitude			Total
		Against	For	Against		For	Against	For					
Gender	Female	Count	0	7	7	Count	0	7	7	Count	1	5	6
		% <i>region</i>	0.0	100.0	100.0	% <i>region</i>	0.0	100.0	100.0	% <i>region</i>	16.7	83.3	100.0
		% <i>achievement</i>	0.0	38.9	35.0	% <i>affect</i>	0.0	46.7	41.2	% <i>aptitude</i>	14.3	45.5	33.3
	Male	Count	2	11	13	Count	2	8	10	Count	6	6	12
		% <i>region</i>	15.4	84.6	100.0	% <i>region</i>	20.0	80.0	100.0	% <i>region</i>	50.0	50.0	100.0
		% <i>achievement</i>	100.0	61.1	65.0	% <i>affect</i>	100.0	57.1	62.5	% <i>aptitude</i>	85.7	54.5	66.7
Total	Count	2	18	20	Count	2	15	17	Count	7	11	18	
	% <i>region</i>	10.0	90.0	100.0	% <i>region</i>	12.5	88.2	100.0	% <i>region</i>	38.9	61.1	100.0	
	% <i>achievement</i>	100.0	100.0	100.0	% <i>affect</i>	100.0	100.0	100.0	% <i>aptitude</i>	100.0	100.0	100.0	

indicates, all of the northern teachers recruited thought that girls could outperform boys in English achievement. Among the southerners, 11 (84.6%) also shared the same opinion. There was no statistically significant difference between teachers from the north and the south in their endorsement of the achievement component of FALS, as assessed by Fisher's exact test,  $p = .521$ . Similarly, no regional difference was found for aptitude,  $OR = .200$ , 95%  $CI [.018, 2.265]$ ,  $p = .316$ ; nor for affect,  $p = .485$ . Odds ratio for the achievement and affect aspects could not be defined because there were no non-believers among the northern participants.

## **6.6 Summary of Integrated Findings about Teachers' Endorsement**

It was found that teachers endorsed FALS in three aspects, achievement, affect, and aptitude, in both datasets. Even if two participants did not personally endorse FALS in Study 2, it was evident that FALS has accessed an extensive audience, including guardians, students, and mass media. Furthermore, the two data sources provided converging evidence that among the three components, 1) achievement was most widely harboured; 2) affect was the component with the largest magnitude; and 3) aptitude ranked lowest in both scope and magnitude. Secondly, it was found in both datasets that teachers established a stronger link between affect and achievement than the one between aptitude and achievement. Finally, evidence for gender difference and regional variations in FALS endorsement was not found.

## **Chapter 7: Findings and Discussions III —**

### **Perceiving Gender as a Differentiating Variable in English Classrooms**

#### **by Teachers of English**

This chapter, as well as the previous one, foregrounds the voices of teachers. By integrating numeric and textual data in Studies 1 and 2, Chapter 6 has delineated how teachers interpreted the content and structure of FALS. The primary goal of this chapter, therefore, is to further unfold the complexity of teachers' gender-stereotypical views concerning learners. In Sections 7.1 and 7.2, the focus will continue to be the female-advantage-in-languages stereotype (FALS). The former will analyse salience, i.e. whether teachers endorsed FALS blatantly, moderately explicitly, or subtly; the latter will decipher function, i.e., whether teachers used FALS in descriptive, prescriptive, or evaluative manners.

Furthermore, two other themes emerging from the thematic analysis, 'gender stereotypes (GSs) accompanying FALS', and 'critical perspectives about FALS', also awaits clarification: in Section 7.3, evidence of other circulating GSs will be presented: 1) the competent boys stereotype (CBS), 2) the quantitative-orientated boys and qualitative-orientated girls stereotypes, 3) the slack boys and diligent girls stereotypes, and 4) the career-driven males and home-bound females stereotypes. Relevant details of how they fuelled the spread of FALS will also be shown. Following these is the last theme, 'critical voices about FALS', in Section 7.4. Finally, Section 7.5 will summarise the findings from Study 2 by synthesising all three themes, including the two in this chapter and 'the widespread endorsement of FALS' theme in Chapter 6.

#### **7.1 Salience of Female-Advantage-in-Languages Stereotype among Teachers**

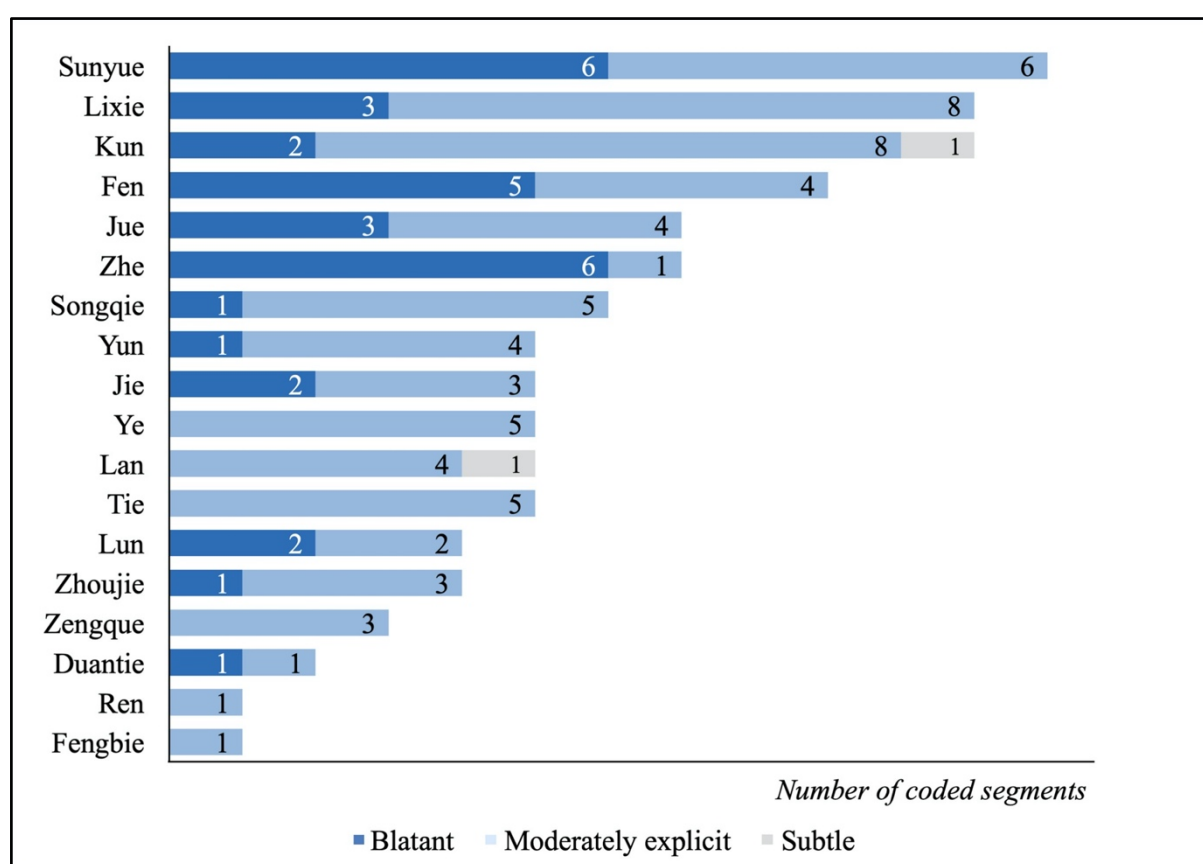
As Section 4.2.2.4 has introduced, 103 data entries were identified as indicating teachers' sweeping endorsement of FALS, and each datum was subcoded on the basis of three features—content, salience, and function—to achieve a more comprehensive and nuanced understanding of FALS (see Appendix T for a list of each subcode, including their definitions and examples). Results concerning the content of FALS have been delineated in accordance with pertinent findings from the teachers' questionnaire survey in Sections 6.2.2-4. Thus, the current section and the next one will turn to the remaining two features accordingly.

FALS entries were categorised as 'blatant', 'moderately explicit' or 'subtle' on the basis of salience. When interviewees voluntarily expressed FALS, the corresponding data segments would be subcoded 'blatant', because these segments reflected interviewees' inclination to acknowledge gender differences without any prompt from the researcher. That is, interviewees

here were capable of directly communicating FALS to an audience. On the other hand, when interviewees made FALS claims as a result of being prompted by the interviewer to compare learners of both genders, these claims would be labelled ‘moderately explicit’: given that participants were willing to comment on how boys and girls could differ when opportunities arose, the gender-stereotypical comments they made would indicate that they subscribed to FALS fairly straightforward. The weakest form of explicitness was subcoded as ‘subtle’, which were accounts that hinted participants’ endorsement of FALS without them asserting that female learners surpass male ones directly.

Figure 7.1 here shows the salience of all FALS-coded segments articulated by the 18 FALS-believers. As it shows, they all made moderately explicit FALS comments—68 out of 103 FALS segments were of this type, taking up approximately 66%.

**Figure 7.1: Salience of Stereotypical Comments Made by Endorsers of Female-Advantage-in-Languages Stereotype in Study 2**



Take Zhoujie as an example. In the following interview excerpt, he was describing the common characteristics among students who are interested in English.

*Well, generally students who get good exam results in English tend to be interested in the subject. They memorise words fast, and they learn fast. And they don't find*

*it a strain learning English. They seem quite relaxed. ((Interviewer asks, ‘and do you think that there is a gender difference in interest? Do you think that boys have a stronger interest in English, or girls, or is there no gender difference?’)) These students ((a short pause)), I suppose, there should be more girls like this. Girls are definitely the majority. Girls perhaps are more interested in it. And also what I feel is going on is that, the majority of those students who can offer certain methods are always girls. ... By ‘methods’ I mean ways to memorise English words. I ask students to summarise [ways that work for them]. Those who can remember words well, remember words fast, and offer methods are always all girls. [Zhoujie, Paragraph 106-110]*

When being inquired, Zhoujie proposed that girls should have been more enthusiastic learners. Zhoujie’s reply reveals that to him, boys, as a group, tended to show less interest in English. Given that he did not voluntarily label girls as more devoted learners without the interviewer’s prompt, his gender-stereotypical account here was considered moderately explicit.

However, a prudent audience might question the appropriateness of employing gender-related prompts due to validity concerns. They may argue that the prompts themselves could have led participants to express gender-stereotypical thoughts, thus reducing the credibility of participants’ answers. After all, when being confronted with a query about gender differences, even if participants themselves could have been potentially not gender-stereotypical, they might still have felt a ‘need’ to specify any trivial gender differences so that they could have ‘pleased’ the interviewer, or appeared reasonably knowledgeable about the topic. Such worries can be relieved for two reasons. First, since the purpose of the interview was introduced as understanding practices of English teaching in Chinese high schools, questions in the interview schedule revolved around teachers’ teaching activities and their opinions about learners. As a result, interviewees would not necessarily have felt an urge to centralise gender issues even with gender-related prompts. Second, in an attempt to avoid misleading participants with leading questions, all gender-related prompts were phrased as neutral questions: ‘do you think there is a gender difference here? Are boys better, or girls better? Or is there no gender difference?’ In this way, participants were unlikely to have been intrigued to exaggerate or fabricate gender differences with these prompts<sup>40</sup>. Thus, their answers should have been considered credible and authentic. Of course, participants’ replies still need to be interpreted with caution. Take

---

<sup>40</sup> In fact, 13 interviewees did maintain that there was no gender difference in certain areas of English learning in 30 data entries. The areas included grammar knowledge, vocabulary retention, listening skills, and *etc.*

Zengque's answer here as an example.

*((Interviewer asks, 'so if you are comparing boys against girls regarding their English skills, do you think there will be any gender differences, or is there no gender difference?')) Um, this is a good question. ((He pauses for seconds)) Girls typically, I'm saying, but this is just my, I haven't got statistically processed data, but I, when I look back at students I've taught, I feel that girls appear to have fewer problems in pronunciation. I mean critical problems in pronunciation. Those are relatively rare among girls. I don't know if this is because girls' sensibility makes them more sensitive [to unique features of English pronunciation], or it may have enhanced their ability to imitate [native English speakers]. I just feel, well, relatively speaking, there are fewer girls with critical problems in pronunciation. And, um. ((He inhales)) Boys perhaps learn grammar faster. I don't have statistics to support me. ((He laughs)) But I am trying hard to, I think about the students I've thought, and here is a very good example. ((He describes a male student good at grammar quizzes)) About this, I:: can't reach a conclusion. Just, judging from my observation, my experience, there are two [gender differences]. [Zengque, Paragraph 131-136]*

Zengque's response to the interviewer's question is construed as a moderately explicit manifestation of his FALS endorsement. Here, Zengque's need for time to formulate an answer might be due to his ignorance to gender issues in language classroom, or his discomfort with blatantly expressing FALS without prompts. With either cause, the reply Zengque formed after contemplation would indicate that he was willing to compare male and female learners when being queried, making his comment a moderately explicit FALS segment. This interpretation is further supported by the hesitation/reluctance signalled by his frequent use of qualifiers: '*I haven't got statistically processed data*', '*I don't have statistics to support me*', '*About this, I can't reach a conclusion*', and '*Just, judging from my observation, my experience*'. These remarks, via evincing Zengque's discretion and disposition to not pass his subjective opinions as scientifically sound proof, discloses his resistance to misleading his audience, not to conveying his gender-stereotypical thoughts.

In addition, 12 out of 18 endorsers made blatant FALS comments. 33 FALS segments were of this kind, taking up 32%. Blatant FALS claims often emerged naturally in participants' accounts without the interviewer's prompts, such as the following two quotes from Fen and Duantie.

*What [students who are good at speaking] have in common is that their exam scores in English are relatively high, and their overall performance in English is relatively good. Their pronunciation of English words is good. All of them are girls. ((She laughs)) [Fen, Paragraph 100]*

*I tell the students, 'if you want to study English well, you try to get close, as close as possible to me'<sup>41</sup>. 'That's what I tell them. You must get as close to me as possible, 'if you want to learn English well'<sup>42</sup>, right? ... Like today, I was in Class 2, and I said, 'you boys, the four Yangs, why do you never want to get close to me? Is it because we are both males and like charges repel? Girls all like me so much. They are all willing to get close to me, and they always ask me questions after class.' I said, 'boys never come forward to me. Is it because you think males, like charges repel?' They responded, 'No::, sir. We are just timid.' So I said, 'Never mind your timidity, you have to get close to me to learn English well. If you keep me at arm's length, how can you learn English well? Right?' [Duantie, Paragraph 52]*

The last category, subtle FALS comments, was only expressed by two interviewees, with one data segment from each. The segments were coded as such because FALS was implied, not articulated, as the following quote will demonstrate. Kun had just described in detail a boy struggling with English due to his lack of language aptitude. According to Kun, the boy was excellent in math and sciences subjects, but his English was helpless. The interviewer then asked if this was a special case, and here is her answer:

*In fact, you won't call him a special case, [because] basically there are always boys who have uneven performances across subjects in any class in the Sciences Branch, boys who have gone overboard with the sciences subjects. ... Girls [of this kind also] exists. There is one in my class. She's also gone overboard with the sciences subjects. She is actually the opposite of that boy. This girl's English is not good, but her results in sciences subjects are still not, relatively speaking, very good. But perhaps with regards to her English results, she's gone overboard with the sciences subjects. [Kun, Paragraph 210-212]*

What signifies Kun's subtle FALS endorsement was her labelling a girl as 'the opposite of that boy' when the girl *really* was not. According to Kun's depiction, the girl was better at the

---

<sup>41</sup> Originally in English.

<sup>42</sup> Originally in English.

science subjects than she was at English, just like the boy. Thus, the girl's academic record was more like a less extreme version of the boy's: the girl was 'relatively not very good' in sciences, but the boy was excellent; the girl was 'not good' in English, yet, the boy was helpless. Therefore, it made no sense to call the girl an exact opposite of the boy. A plausible explanation behind Kun's reasoning was that, she was convinced that any boy could have done English badly, while girls regularly should have done English well. The boy followed her expectation, but the girl actually challenged her stereotypical view. And this was why the girl was regarded a contrast to the boy.

Another example is from Lan.

*((Interviewer asks, 'you've just talked about many students. Some of them are doing well, but some need to improve. So generally speaking, do you think that boys and girls differ in their English achievement? Or do boys and girls have similar results?')) I think, if a boy is willing to learn, he will learn better than girls, faster than girls. ((Interviewer asks a follow-up question, 'can you elaborate on that?')) I mean:: generally, don't boys put their minds into the science subjects, or in playing games? ((She giggles)) Cell phones are a big problem among students. Our school does not allow cell phones, but where there is a rule, there is a strategy getting around that rule. You know, parents would always prioritise their children's needs, [so they buy cell phones for their children.] Yet, they don't know that the students want cell phones for games. ((She pauses)) So you know, if a boy can put his mind into English, then he will learn it rather well. [Lan, Paragraph 91-94]*

In the above excerpt, Lan expressed that she believed in boys' potentials to outperform girls so long as they were committed to learning English. But she also suggested that this was usually not the case with boys. Therefore, the quote here implies that she subtly believed that typically, girls do better than boys in English.

## **7.2 Function of Female-Advantage-in-Languages Stereotype among Teachers**

FALS entries were categorised as 'description', 'prescription' or 'evaluation' due to function<sup>43</sup>. When interviewees used FALS to depict what male and female learners were like generally, and/or what male and female learners typically did, the corresponding data segments would be subcoded 'description'. This was found to be the most frequent usage of FALS, being

---

<sup>43</sup> For definitions of the subcodes about FALS function, please refer to Appendix T.



adopted by all 18 FALS believers. In fact, 99 out of 103 FALS segments were of this type, a predominant proportion of 96%. For example, Jue here sketched girls as more gifted English learners:

*I think girls, when they study, they have this gift for learning English. ((She giggles.)) Boys' [aptitude] may be, relatively speaking, worse. But there are boys who do quite well [in English], relatively speaking, though not as well as girls can do. ((Interviewer asks, 'and what do you mean by 'gift'?' )) I meant an aptitude for language. ... It might, some students, like girls, they memorise words really fast. For instance, in my class, there is a girl. She can recite a text in a couple of minutes, really fast. ... Maybe girls are just somewhat better at memorising stuff like this, but maybe boys just rather enjoy thinking, [so they] can do better at the science subjects. [Jue, Paragraph 116-124]*

What seems particularly interesting is how Jue portrayed girls and boys in a contrasting yet complementary way, where girls were associated with language studies while boys with science subjects. A similar pattern persists in other participants' description, as the following excerpts from Sunyue's interview illustrates.

*I think something else [that distinguishes boys from girls] would be interest. I feel that boys and language, sometimes when they chat with me, they would say that language stuff, like Chinese and English, make them feel somewhat bored. They think that those are relatively boring. They think that, like physics, after they've learnt something new [from that course], they will be able to explain a phenomenon. Or after learning, they'll be able to build something through their own efforts. They just feel that this sort of courses, and possibly also Chemistry, are quite interesting. So relatively speaking, language courses won't be able to get them there. ... Because when I introduced many great apps to students, those who showed interest, like those who joined my extra-curriculum group, are mainly girls. Yes. I looked at the group which twenty-something students joined, there was no boys. ... And one more example would be, let me see, when we first started, in Grade 10, students voluntarily asked me to recommend some English books suitable for high school students. Novels and such. So I did recommend some, and I saw some girls did buy them, and I knew that they did read the books. But boys, in terms of this, they did not seem quite interested when I made the recommendation. Yeah. What they were mostly interested was like, 'Mr., tell us*

*some English movies.' But I don't think films would be a great help [for English learning]. ((He smiled reluctantly)) ... And also, they tend to come forward and talk to me; more girls are like this. Those who come forward and talk to me about their English learning experiences are mainly girls. [Sunyue, Paragraph 138-184]*

Actually, juxtaposing FALS with the idea that male learners outshine females in STEM subjects is a common theme emergent in teachers' accounts in Study 2, and more in-depth discussions can be found in Section 7.3.2.

Aside from description, participants were also found to have used FALS for two more purposes: prescription and evaluation. For example, participants might define what male and female learners should do based on FALS, as the following excerpt from Zengque's interview shows. When speculating on why girls outperform boys in English, he said,

*I think girls, as I see it, maybe it is because girls are not that ((a short pause)) fidgety, compared to boys. I mean, girls are more likely to be capable of calming down. Because English, to learn a language like this, there will be a lot of repetitive work, and you'll need to be, you must be willing to persist. Because girls are more patient, and also, considering something like language, anyways, girls would always appear, how to put it... ((He laughs)) Language just remind you of the Humanities, right? And the Humanities, the Humanities feel like something girls should be good at. [Zengque, Paragraph 138]*

Besides prescribing English as a feminine domain, FALS may also be adopted as a standard against which individual learners are compared, giving rise to gender-differential expectations or treatments. For instance, Kun here commented on a boy who was counter-stereotypical regarding English achievement:

*And in my classes, students who are relatively good at English:: tend to be mostly girls. Boys are quite rare. Boys, but there is one exception, a boy who is not quite good at any other subjects, but his English is very outstanding. ... Because his English can get as high as, I mean last term, the highest score in our school was more than 120, and he got as high as 117. ... But in terms of other subjects. ((She laughs)) Because that was the final exam, I took a peek at his academic report, and he got less than 20 for the remaining subjects, out of 100. The full mark was 100, and he got less than 20. So it was a dramatic contrast. ... But if you look at, only look at his English results, his score was top-notch. ... He has gone overboard with*

*just English. And this is a boy. Isn't this astonishing? ... I mean, boys who are relatively bad at English are pretty large in number, pretty common. So this, this boy, is very peculiar. When I'm in class ((she laughs)), I always want to, ((She clicks her tongue)) I just want to tell him that he needs to work on other subjects, too. It is such a pity, you know, to have a boy whose English is so good is something once in a blue moon [but at the same time his performance in other subjects is dragging him down]. [Kun, Paragraph 80-88]*

In the quotation above, Kun spoke of a boy who, unlike typical boys, is actually good at English. The boy's excellence apparently exceeded her FALS-fuelled expectations for ordinary boys, and thus in her account, the boy appeared 'exception', 'astonishing', and 'peculiar'. Interestingly, in Study 4, a male participant was a similar high-achiever, who did not regard himself as an exception—for him, male linguistic competence actually matches a type of ideal masculinity commonly found in classic Chinese romance (Song, 2004). Section 8.2.4 will elaborate on this point.

As might be expected, neither prescription nor evaluation was used as frequently as description: only five data segments were prescriptive, and nine were evaluative, both accounting for less than 10%. However, this does not necessarily mean that FALS is not widely used for prescriptive or evaluative purposes. The first reason lies with the design of the interview schedule in Study 2: it was devised to uncover gendered patterns in teachers' descriptions of learners, so naturally, FALS was revealed to be extensively used when depicting male and female learners. Second, although prescriptive and evaluative functions were not high in quantity, their coverage was not narrow: 5 out of 18 participants used FALS prescriptively and that number for evaluative use was 7. In other words, future studies are in need to reveal how FALS is used for multiple purposes.

### **7.3 Additional Gender Stereotypes Accompanying Female-Advantage-in-Languages Stereotype**

Although FALS was the GS under investigation in Study 2, other circulating GSs were also discovered. The one most closely related to FALS was the 'competent boys' stereotype (CBS), which deems males as more intelligent and/or eminent learners. Seemingly in conflict with FALS, it was in fact harboured by eight FALS endorsers. Section 7.3.1 below will explain what CBS entails and how it co-existed with FALS among the believers.

Three sets of complementary GSs also emerged in Study 2: 1) quantitative-orientated boys and qualitative-orientated girls, 2) slack boys and diligent girls, and 3) career-driven males and

home-bound females. The first set situates the English subject within the larger scope of the Humanities and Liberal Arts subjects, which are contrasted with the STEM ones (science, technology, engineering and math). With its portrayal, the former appeared feminised while the latter masculinised. The second set, ‘slack boys and diligent girls’, describes male and female students in general without reference to any particular academic subjects. And the last set, ‘career-driven males and home-bound females’, spans from students to professionals. Sections 7.3.2-4 will provide evidence for each set, with details of how they fuelled the spread of FALS.

### 7.3.1 Competent Boys Stereotype

The competent boys stereotype (CBS) was endorsed by eight FALS believers. Its existence was predicted by the Stereotype Content Model (SCM, see details in Section 3.1.3), where males are typically painted as more intelligent, rational, and competent individuals than females (e.g., Cuddy, Fiske, and Glick, 2007). This section will now address the two components of CBS and two types of its relations to FALS.

#### 7.3.1.1 Contents of Competent Boys Stereotype

Among the eight CBS subscribers, two interacting aspects were discovered: male biological/cognitive superiority and super-eminent achievement by male learners. The former presents males as more competent learners due to their supposedly unique biological characteristics and advanced cognitive faculties, such as intelligence, perceptions, rationality, and logical thinking. The latter, though acknowledging that girls can be decent English learners on average, still perceives males as more likely to reach the summit of achievement.

**Figure 7.2: Venn Diagram of Believers of Competent Boys Stereotype Components in Study 2**

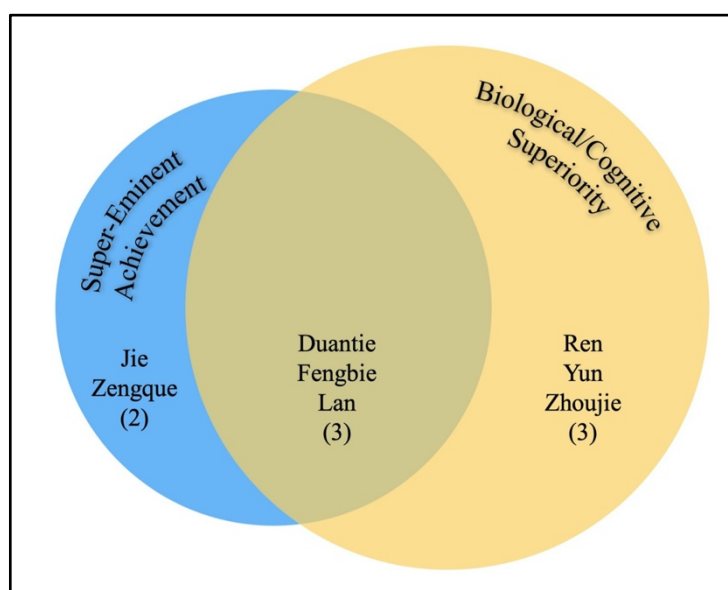


Figure 7.2 here, with the help of a Venn diagram, illustrates the endorsers of both CBS components and combinations of the two. The blue circle on the left stands for the achievement component, where two interviewees are located, and the right, yellow one represents the biology/cognition component, where another three sits. The three interviewees residing in the middle

are those endorsing both components. Among the eight believers, four were male veteran teachers, three were female novices, and only one was a female veteran. Since it is unclear whether or not the remaining twelve interviewees endorsed CBS, a Fisher's exact probability test could not be performed to determine whether gender or regional differences were present. However, it does seem that higher proportions of male veterans (57%) and female novices (43%) subscribed to CBS.

An example of the biology component is from Ren's interview, where she explained why she objected to the claim that girls were more gifted language learners:

*I think that, perhaps, speaking of a general capacity for various tasks, based on the students I've made contact with, perhaps boys' [capacity] is slightly stronger than [that of] girls. But this does not mean that girls are just, in terms of intelligence, or under certain circumstances, inferior to them [boys]. Instead, [the difference] might lie in the [biological] foundation, or, if I may say, in the logical thinking and the capacity to learn. ((Interviewer pursues the topic, 'and what do you mean by these, 'logical thinking' and 'capacity to learn'?' )) Logical thinking is probably not specific to English. Rather, [it] is applied to other courses. Because we also have modules like business studies. With regards to business studies, perhaps if you simply rely on memorising definitions for most of the time, [you'll feel] quite miserable. Especially so as our modules are taught in English. So if [a student] can start with understanding and familiarising themselves with the logic behind, they will be more likely to learn, grasp [the knowledge], and achieve mastery through a comprehensive study of the subject. Perhaps this [logical thinking] will make their learning relatively better. And girls, um, they are perhaps slightly weaker in this. But in fact, the difference won't be very big. Because after all, judging from either our recent exam results or their participation in class, in fact, girls have been making progress. Yes. Indeed, we have the [biological] foundation on one hand. But on the other hand, perhaps girls, concerning self-control, are slightly better than boys. Because, after all, they [girls] don't get together and play 'Arena of Valour' every day. ((She chuckles)) [Ren, Paragraph 101-104]*

As Figure 6.4 in Section 6.2.2 has illustrated, Ren only subscribed to the achievement component of FALS, i.e., she thought girls tended to outperform boys in English. Furthermore, Figure 6.5 in Section 6.3.1 revealed that Ren thought that both a favourable attitude towards

English and an aptitude for language could lead to learning success, although she assumed that aptitude's contribution might be a mere 5% (Paragraph 110). Therefore, combining these two views and CBS, Ren's gender stereotypical opinions about language learners can be summarised as follows: 1) compared to girls, boys have an advantage in biology; 2) but under two premises (that boys lack self-control and that biological distinctiveness does not assist English learning significantly), girls turn out to be higher-achievers in English. Clearly, FALS and CBS worked in harmony for Ren.

For an example of the achievement component of CBS, here is an excerpt from Jie's interview, where she conveyed CBS and FALS at the same time:

*... Girls' English scores, the proportion of good female students is obviously larger, yeah. But personally I feel that, perhaps, although in terms of proportion, indeed, this is what's happening, [the proportion of] girls is larger, still, boys, I think, once they learn well, even if on the same level, boys who learn well learn better than girls. ((Interviewer asks, 'what do you mean by 'once they learn well'?')) I mean ((she pauses)), how do I put it? For instance, you get two students in, a boy and a girl, when it was Grade 10. Their English performances are at the same starting line, equally good, equally excellent. In the end, it is still the boy that will learn better. ((Interviewer asks, 'and what reasons do you think have led to this?')) (She pauses.) I wouldn't know. ((She chuckles.)) Anyways, over all these years, I kept seeing instances like this for several times. [Boys and girls] with similar starting points, both rather excellent, and as they learn, in the end, boys still become more outstanding. You, if you are talking about [boys and girls] in general, girls definitely learn better than boys. But any boy who learns well must be very good, very good. Absolutely top-notch. [Jie, Paragraph 100-104]*

Unlike Ren, Jie actually endorsed both the aptitude and achievement aspects of FALS. As she saw it, girls had a stronger aptitude and performed better in English. Then, Section 6.3.1 indicated that Jie insisted that both an interest in English and an aptitude for language could aid learning. Nevertheless, Jie also downplayed aptitude's contribution, to only 30% (Paragraph 156), just like Ren. Thus, given the quote above, it would appear that Jie was juggling FALS and CBS simultaneously, although Jie failed to explain how these two could have co-existed. In fact, Ren and Jie represented two types of relationships between FALS and CBS, which will soon be expounded in the next section.

### 7.3.1.2 Relationship between Competent Boys Stereotype and Female-Advantage-in-Languages Stereotype

As the previous section has introduced, two kinds of relationships between CBS and FALS were discovered: nested hierarchy (such as Ren), and parallel procession (such as Jie). The former can be exemplified in Russian matryoshka dolls, where CBS is the outer doll that encompasses FALS, the inner doll. For Ren, and five others sharing the same logic (Lan, Yun, Duantie, Fengbie, and Zhoujie), the male biological advantage is fundamentally global to all academic subjects; the reasons that cause girls to outperform boys in English are, instead, specific to this particular subject. For example, according to Lan, Duantie, and Fengbie, boys' underachievement resulted from a lack of interest in English, considering that they all believed in the affect and achievement aspects of FALS (see details in Section 6.2.2 before). And they all said that if a boy is devoted to learning English, he will surely outperform girls (Paragraph 92 in Lan's interview, 84 in Duantie's, and 124 in Fengbie's). Zhoujie made the same point (Paragraph 190), but he also raised that boys' smartness could actually impede their learning, because it might made them too impatient to persist in English (as Section 6.3.1 before has presented).

Interestingly, Yun's argument was somewhat different from the four abovementioned teachers, perhaps due to her acceptance of all three aspects of FALS. After all, Ren, Lan, and Fengbie only endorsed the affect and achievement components. Her opinion will be unfolded through the two quotes below:

*In our exams, including the students I used to teach, who took TOEFL iBT<sup>44</sup>, the majority of those who get 100 and more marks are girls. That's for sure. Yeah. And including that thing, that, but SAT is a different situation. It is not simply [a test of] language, because SAT, it also has that Mathematics section. And there is another section, the part we call reading. But the reading part also tests what we call critical thinking. So boys, in this test, they will obviously ((a short pause)) outperform girls. Yeah. I mean, when we look at this [SAT] test, there is logical [thinking capacity], so you need to take both [logical thinking ability and language aptitude] into account. ((Interviewer asks, 'do you mean to say that boys are better at logical [thinking], compared to girls?')) Yes, I think this is what we have to*

---

<sup>44</sup> TOEFL iBT, Test of English as a Foreign Language (Internet-based Testing), is a computerised test developed by Educational Testing Service, which measures learners' ability to use and understand English at the university level (<https://www.ets.org/toefl/ibt/about>). It also evaluates how well learners combine their reading, listening, speaking and writing skills to perform academic tasks. The total score is 120.

*admit. Boys are more gifted than girls in logic. ((Interviewer asks for clarification, 'And you think that in terms of language score, it is girls who do better?')) Yes, yes. I mean, if it is simply a test of pure language [knowledge/skill], girls would still appear more excellent. [Yun, Paragraph 120-124]*

*I don't think there is any [gender] difference here [in grammar]. Honestly, no. Yeah, this actually depends on whether or not you are smart. For example, in my class, the boy who was the worst [in English], the one who got twenty something in an exam. ... I just found that this guy, he, actually is very smart. Like, I just explained to him about the classic, the basic grammar, such as subjects, predicates. He grasped them really fast. But he will definitely forget, he will forget. So I'll have to repeat, right? But if I explained the same content to some girls, they wouldn't be able to understand as fast as he did. ((Interviewer asks, 'so are you saying that boys, this boy, is smarter than girls?')) Um:: With regards to the level of smartness, I cannot say it's boys who are straightforwardly smarter than girls. I suppose, um:: Perhaps we can call it, what should [we] call it, [let's] just call it intelligence. Because smartness is a broad concept, as far as I'm concerned. If we are talking about intelligence in isolation, I think that boys, they definitely have advantage in some areas. And also, I think this advantage comes from their parents. I think it is inherited. Yeah. I don't think it is related to any postnatal influence. To me, it is just what gets inherited from parents. It is the power of genes. ((Interviewer follows up, 'so as you see it, do you find that students with higher intelligence can understand grammar faster?')) Yes. Because I felt very confused, at that time. Because that boy, that boy who get really low marks, he is supposed to be, according to my judgement, he isn't supposed to be someone with a high emotional intelligence. Yes. But more often than not, your understanding of something [depends on] what we call 'sympathy'<sup>45</sup>, right? It should be empathy. I think females tend to perform better [in tasks or situations that require empathy]. Female obviously feel empathy more strongly, right? So speaking of emotional intelligence, girls should crush boys. So, I used to believe that, for example, if a student could see, could understand what I have said very well, that would reflect his/her empathy. Because he can understand examples you give. And girls' reaction to this [situation] is more noticeable. But this boy's reaction, eh, his*

---

<sup>45</sup> This word was said in English in the original interview.



*[reaction] was very impressive. He could grasp very well, understand very fast. So I overruled myself: empathy couldn't have been the most significant factor. I think intelligence must have also contributed somehow.* [Yun, Paragraph 132-136]

It would appear that Yun may have harboured some version of the general intelligence theory, which refers to the existence of an underlying mental capacity that influences performance on a range of cognitive tasks (Cohen, Swerdlik, and Sturman, 2015). At the same time, Yun raised that specific mental faculties were needed for discrete cognitive tasks, such as language learning and logical thinking. As far as she was concerned, 1) boys' general intelligence is higher; 2) girls' language aptitude is extraordinary; 3) boys' underlying biological superiority may override girls' language aptitude in some aspects of English.

The second type of relationship, parallel procession, manifests itself in Jie and Zengque. This relationship is marked by teachers' attempt to use FALS and CBS jointly for the same purpose: explaining the gendered achievement patterns they've witnessed among students. FALS is suitable for girls' higher average score, and CBS for boys' super-eminence. This is best illustrated by the excerpt from Jie's interview (Paragraph 100-104) introduced in the previous section. Zengque, likewise, also held similar opinions. For these two teachers who hold CBS and FALS in parallel, reasons for boys' occasional super-eminence remained a mystery: Jie and Zengque admitted that they were baffled, because they actually assumed that girls have stronger aptitude for and higher achievement in English.

Two plausible ways for this CBS–FALS parallel are proposed here. First, the exceptional boys could be treated as exceptions that prove the rule. That is, boys' irregular excellence supported the claim that girls are generally better at English than boys. Second, research has found that more males than females tend to appear at the higher and lower ends of distributions of different cognitive ability (Priess and Hyde, 2010), which resembles the pattern described by Jie above. Thus, it is suspected that CBS and FALS may have reconciled in Jie and Zengque's minds through either or both of the two mechanisms here, but more evidence is needed before a conclusion can be arrived at.

### **7.3.2 Quantitative-Orientated Boys and Qualitative-Orientated Girls Stereotypes**

Aside from FALS and CBS, two more domain-specific GSs emerged from the interviews: the quantitative-orientated boys and the qualitative-orientated girls stereotypes. But unlike FALS or CBS, the two are complementary and thus often appeared together in teachers' accounts.

Among the 130 data segments coded ‘gendered patterns’, 50 (38%) were subcoded as such, and a total of 15 interviewees in Study 2 endorsed the set.

In Study 2, teachers stereotypically prescribed quantitative domains as masculine because they were convinced that boys were gifted and therefore could perform well in these subjects. They also proposed that compared to girls, boys tended to show interest and thus cluster in these fields. Meanwhile, qualitative domains, such as Chinese and English, as well as the Humanities subjects (politics, history and geography in Chinese high school curriculum), were attached with feminine labels. For example, Zengque here is talking about how boys and girls are prewired for different academic fields:

*I think ((a short pause)), of course this is not entirely politically correct ((he chuckles)). But to be honest, I do think that different sexes have different advantages. It seems that a previous principle of Harvard was reproached for saying things like this. ((He laughs)) But I do agree with this idea. I think different sexes just have sex-differential advantages or disadvantages. I think the Humanities subjects are just what females are capable to master faster and better. Regarding sciences subjects, it is males who can master better. This is the result of sex differences in biological makeup. [Zengque, Paragraph 144]*

Some quotes from previous sections have already touched upon these stereotypes (Jue, Paragraph 116-124, on p.157; Sunyue, Paragraph 178-184, on p. 157-158). In addition, teachers were discovered to persuade their students to make GS-congruent academic choices, as the following quote will imply:

*Speaking of the two academic branches, I had a particularly vexing experience with the students graduated last year. ... When they were asked to choose between the Sciences Branch and the Humanities Branch, a few girls, they really wanted to join the former out of many concerns. I suppose the biggest concern was university, the types of majors they can choose<sup>46</sup>. Well, I just thought, ... girls in particular, it seems that studying the Humanities subjects is a safe choice. ... And also, how do I put it, if a girl chooses the Sciences Branch, especially when she gets to Grade 12, it will be hard for her. Because in previous years, the exam papers on Sciences*

---

<sup>46</sup> When applying for universities, students from the Sciences Branch can choose most majors in many universities. But some majors, especially the STEM ones, won’t take students from the Humanities Branch. So in a way, if a student selects the Humanities Branch in high school, his/her choices of university majors will be limited, compared to a peer in the Sciences Branch.

*subjects and Math, was extremely difficult. Sometimes students leave the examination room with tears, and sometimes this could happen to good students. So at that time, I, based on this reason, I tried to persuade [the girls to choose the Humanities Branch]. ... Out of the top ten in my class, there were seven who wanted to join the Sciences Branch. All girls. So I persuaded them to do choose the Humanities. ... As a result, I found that in Grade 12, when they choose university majors, Uh-oh, I became dumbfounded. I found that even if they could get into a good enough university, their choices of major were very limited: tourism management, or something about language studies, English, or some minor languages. ... They just couldn't get to where they had wanted. ... I was really defeated. So this time, I won't ever try to persuade anyone this year. [Tie, Paragraph 94]*

As Tie explained, she had only done the persuasion out of genuine concerns for her female students, though the concerns were formed on the basis of the stereotype that girls could study qualitative fields easier and better. In fact, according to Ye and Lun below, some teachers believed that such a stereotype could be internalised by some students themselves, and even become self-fulfilling prophecies.

*In the Humanities Branch, there are definitely more girls, while in the Sciences Branch, there are definitely more boys. They [students] all have their preference. I mean, I just feel that this is the way it is, boys and girls have such preferences when they choose academic branches. [Ye, Paragraph 38]*

*... Besides, another factor among girls is, in fact, in their hearts, they know that others would all think girls are supposed to be good in English. So they will also tell themselves, they would do, how to put it, they would attach a label to themselves, saying, 'as a girl, I just ought to be good in English.' So she will be relatively willing to spend time on English, to learn it. And also, they will identify with this, gender identity. I feel that this is somewhat true. And also, they, what else, because they've spent more time on time, they've had more contact with English, they've written more and read more, they will definitely do better. And sometimes they may also feel, during the process of their study, if they are no match to boys in other subjects, she can feel, oh, she can find a sense of accomplishment in English. ... ((Interviewer asks, 'other subjects?')) Math, perhaps they don't react to math*

*problems as fast as boys. And also the Sciences ones, such as Physics. [Lun, Paragraph 220-224]*

In addition, teachers were found to apply the qualitative-orientated girls stereotype to males in qualitative domains, as the quote below will show.

*Oh, it's girls who have the better results. In my classes, there are more boys at the lower end of the list. More boys [than girls]. And those boys, if they don't like this course, they just honestly don't like it, and their results won't get better. ... Um:: for now, [the boys and girls in] the two classes I teach don't differ [in their interest in English]. But in the past, I mean generally speaking, girls tend to be more interested. Haven't you noticed that us English teachers are all:: quite tender? ((He chuckles)) ((Interviewer asks, 'is it so?')) Haven't you noticed? Boys who major in foreign languages, because they often hang out with girls, they become like Baoyu Jia. Foreign language departments are full of girls. So you can see, Duantie is also quite, well, you know. [Fengbie, Paragraph 142-148]*

Here, a character in literature, Baoyu Jia, is likened to boys who major in language studies. As the leading protagonist in one of China's most famous classic novel, *Dream of the Red Chamber*, Baoyu is known for his rebellion against traditional masculinity. Instead of rote-learning Confucian classics or committing himself to gaining a position in the government, he enjoyed poetry and idled away with his female relatives. In fact, he is considered feminine or even effeminate due to his compassion and thoughtfulness. Thus, to compare Baoyu with male foreign language majors reflects Fengbie's logic that male language learners could become feminised due to their entrance to a female-dominated field. Not only so, he also implied that male English teachers, such as his colleague, Duantie, could also become effeminate. An almost identical argument is made by Songqie, too.

*In the Humanities classes, there are more girls. Boys are a minority. And the characteristics of the boys, well, are quite introvert. They don't talk. It seems that they are assimilated by females, become feminised. Because there are many girls. ((He chuckles)) This is what happens. [Songqie, Paragraph 16]*

Clearly, the two GSs, together with FALS, prescribe English as a feminised domain and the learning of English a feminine pursuit. Additionally, in Fengbie and Songqie's account, the qualitative-orientated girls stereotype seems to have overridden gender norms, a tendency also observed among students in Study 4 (see Section 8.2.5).

### 7.3.3 Slack Boys and Diligent Girls Stereotypes

During the interviews, a pair of complementary GSs painted boys as slack learners and females as diligent ones. Yet, unlike all previous GSs, this pair is not domain-specific. Out of the 130 data segments coded ‘gendered patterns’, 45 (35%) were subcoded as such, with a total of 16 believers. According to the stereotypes, female learners were better at self-control, tended to work hard on their own, and were willing to follow teachers’ instructions and practise answering exam questions according to scoring standards. Compared to their female counterparts, boys were thought to lack self-control, muck about in school, and fail to follow instructions and meet standards. Interestingly, this pair of GSs was also present among students in Study 4 (see Section 8.2.5).

In interviews, teachers typically resorted to these two stereotypes in order to explain the perceived gender gap in English achievement, as the following three quotations will show:

*... Girls, if they have set a goal, they will know how to overcome some ((a short pause)) distractions in the surrounding, those things that might distract them. So we can say that girls are more committed than boys. In terms of boys, if this looks interesting, they’ll come here; if that looks somehow [different], they’ll just go there, won’t they?* [Yun, Paragraph 156]

*... Girls, well, they learn on their own. Boys, for examples, in the self-study sessions in the evening, sometimes we give them assignments. But sometimes there are not many assignments, and they are expected to find things, exercises, to do. And girls tend to be more proactive in these situations. If you look at boys, well, if they’ve finished the assigned tasks, they will just sit there, as if they’ve got nothing to do.* [Sunyue, Paragraph 196]

*... But boys, one of their weaknesses is handwriting. ... Boys’ average scores in writing are always lower than that of girls, with a gap of three to four marks. Girls’ handwriting is better. [Even if] the content is gibberish, girls can get 15 marks and more. But boys’ handwriting is horrible, so they get probably 11-12 marks. Generally, when we mark the compositions, if the handwriting was awful, we may simply give a fail. ((Interviewer asks, ‘so is this a requirement in English exams, that the handwriting needs to be good?’)) Um:: generally speaking, yes. If you want to get high marks, you need to have a good handwriting.* [Duantie, Paragraph 84-86]

In fact, in order to improve boys' performance in school, teachers reported that they were more attentive to boys' behaviours, and more responsive to boys' needs. For example, Lixie here talked about how he managed to distribute more of his energy to male students who typically appeared less competent and less motivated.

*... Usually, when I teach in class, and when I collect homework from students, I pay some extra attention to boys. For example, when you are facing, when you are grading a boy's homework, or when you are listening to a boy's answer to a question in class, you should be alert. ((He chuckles)) If this is a girl, her ability to learn language on her own will indeed be stronger. ((He chuckles)) ((Interviewer inquiries, 'so when you say 'be alert', what are you alert for, and how?')) Such alertness is like, you try to increase the number of times you question boys, increase the frequency. What else, like when you deal with homework, if a boy makes a mistake, or if there is a problem in his homework, we need to speak to him directly, to point out the mistake. That is, whether it is about homework or during class, we should pay specially attention to notice boys' English learning. Because the problem is, if a boy's English learning, at the beginning, if his results is slightly bad, and you don't grant him the attention, perhaps for him the learning will get harder and harder. In the future, he may simply give up this subject. ((He chuckles))*  
[Lixie, Paragraph 130-132]

So far, academic GSs uncovered from Study 2 include FALS, CBS, the quantitative-qualitative set and the slack-diligent pair. Together, they portray females as ideal learners of English and identify a mismatch between the male gender and the English subject. This explained two prevalent gendered patterns in teachers' accounts: When being asked to describe factors conducive to English learning, teachers would always use females as examples. For instance, Songqie, Sunyue, and Zhoujie's quotes in previous sections all highlighted the importance of interest, and they went on praising girls as more enthusiastic learners of English (Songqie, Paragraph 40, on p.132; Sunyue, Paragraph 178-184, on p.157-158; Zhoujie, Paragraph 106-110, on p. 152-153). Aside from aptitude and affect, interviewees also characterised girls as ideal learners in terms of perseverance (Zhoujie, Paragraph 175-176, on p. 145), and learning behaviours, as the example below will indicate. Zhe here was outlining behaviour patterns she found among avid learners of English.

*This kind of students usually learn English quite proactively, and they ask me questions. And also, I tell my students, 'if you want to practice writing compositions,*

*because many students are quite bad at English writing, if you want to improve, you can do writing exercises on your own. You can go and find exercises, write compositions by yourself, and you can bring your writing to me so that I can correct them for you. I'll provide you with comments and suggestions.' And those who can do this, off their own bat, there are boys who can do this, but still, the majority is girls. There are more girls. [Zhe, Paragraph 163]*

Contrary to the overlap found between ideal English learners and typical female learners, typical boys were characterised as remedial learners who are not pre-wired for language learning (for an example, see Jue's quote on p.136), liable to hold negative feelings toward English (e.g., Fen's quote on p.135), and even adopting self-handicapping strategies to avoid their failure in English from hurting their self-esteem:

*Boys in my class are reluctant to memorise English vocabulary. I just discovered that, it's not like they are against memorising vocabulary. They are actually afraid of denying themselves. Yes, they are afraid of this. Honestly. A boy in my class told me this. ... So I thought to myself, boys and vocabulary, perhaps it was not memorising vocabulary that they rebel. But perhaps what they rebel more is that feeling of denying themselves. Probably boys think differently from girls. Boys may be more concerned with their, this so-called 'face'. ... They'd rather not do it than having to admit that they failed after having done it. [Kun, Paragraph 170]*

Sunyue talked of another form of self-handicapping, where some boys gave up on English.

*... For example, last time, last semester, there was a boy. I was concerned about his bad grades in regular quizzes, and during my classes, he was also very disruptive. So I invited him to my office, and I asked him, 'why are you behaving like this in classes?' He said, 'I'm just not bothered with English. I'm not interested.' Then, I said, 'English accounts for exactly 150 marks [in the College Entrance Exam], whose total marks are just 750. If you quit English, how are you supposed to do well in the Exam?' And he simply said ((he smiles reluctantly)), 'Sir, there's only so much energy I have.' ... Then I said, 'how about you just take a look [at English] when you have time? Just try to finish the assignments?' And he said ((while smiling reluctantly)), 'what if it tires me?' so I just feel that ((a short pause)) this type of students, there is no way you can motivate them. I tried*

*to talk to him for a few times, in the end he simply told me, 'Sir, just leave me alone.'*

[Sunyue, Paragraph 190-196]

### 7.3.4 Career-Driven Males and Home-Bound Females Stereotypes

Although less widespread, the set of stereotypes prescribing males as career-driven and females as home-bound were also evident in teachers' interviews. Out of the 130 data segments coded 'gendered patterns', nine (7%) were subcoded as such, and a total of six interviewees endorsed this pair. For example, Fengbie, Tie, and Lan brought up how the two stereotypes, alongside FALS, shaped the learning and teaching of English, making it a female-dominated one.

*Now there are fewer and fewer male English teachers. I suppose there is such a tendency. Boys, like our classmates, when we were English majors in university, there were few boys to begin with. And, their ambitions were not becoming teachers. ((She giggles)) Perhaps they all had other plans. They can transfer to another field when they do a postgraduate degree, or they may start their own training school. That's also a possibility. And also, when a girl graduates, her family will have this expectation, they want her to come back home. Boys' families don't have such expectations for them.* [Lan, Paragraph 132]

*... Not many males are doing English in college in the first place. Boys may think, 'Em, English, Nah.' Or perhaps it just doesn't feel right for them ... And secondly, the good graduates usually won't come here, to be a teacher in our town. In terms of this profession, it just, um, it kind of repels them. Or maybe they are not interested? I mean males, they just, few will take an interest in the profession.* [Fengbie, Paragraph 34]

*We just cannot find any [young male graduates]! No, I mean more and more likely so, male teachers don't want to teach in schools in a small town like ours. In normal universities, there are boys majoring in English, but he might not be willing to come here. Even if someone did come, he would eventually leave to look for new career prospects. For example, if he could, he would typically [choose] cities like Changsha, in Hunan Province. Or perhaps in Guangzhou, in those places. ... Female teachers have this, 'return' thing. How do I put this? I mean, most of them, their parents are around, in ZJ<sup>47</sup>. Their parents must have wanted them to stay, tell*

---

<sup>47</sup> The name of the town where the school is located.



*them, 'you are a girl. Don't stay outside home; it's too difficult. Just come back, and be around.' And of course, girls want to be thoughtful. [Tie, Paragraph 62-64]*

The interview excerpts above paint a picture where males are pushed away from English learning and teaching while girls are pulled into the field of English: because of FALS, boys show little interest in learning English in universities. The few who do become English majors either choose to teach English in large cities or opt out the teaching profession, because they, as males, want and need better, more masculine careers. The gendered feature of English teaching was also perceived by students, who even described male English teachers as 'a bit unacceptable' (see the last subsection, 'teachers and teaching' in Section 8.2.1).

Their female counterparts, however, can select English as academic and professional pursuits without violating any GSs. Moreover, female teachers are expected to work in their hometown to stay close to their parents, as it is appropriate for them to be less ambitious and remain home-bound.

The two stereotypes also emerged when teachers talked about their own professional life. For example, both Fen and Songqie talked about how male teachers were supposed to take on more teaching and/or administrative responsibilities in schools while female teachers were expected to assume their roles as homemakers.

*[In our grade,] there is only one male [English] teacher. All the rest are females. ((She laughs)) ... And this teacher is the leader, the leading teacher in the English teaching team in our year. ... He's been a teacher for about a decade. Because female teachers ((a short pause)) probably need to take care of their families or so, so this male teacher just ((a short pause)) becomes the leader. [Fen, Paragraph 46-48]*

*In the English teaching team, in Grade 12, we have 11 teachers. Five of them are males, and six females. ... ((Interviewer was curious, 'is this gender ratio common in your school? Because I haven't seen similar things in other schools.')) Yeah? Because this year, honestly, for such a long time since I've been here, this is the most ((he laughs)) special year. Because in previous years, in our school, a particularly large number of female teachers got pregnant due to the family planning policy that allows all couples to have more than one child. So in each grade, there were probably two or three female teachers who took maternity leave. We were so understaffed that we had to draw male teachers from other grades.*

*Because male teachers don't have to [get pregnant], ((He chuckles)), they don't need to take maternity leave. So all teachers got drawn are males, and so male English teachers just gathered in our grade. ... It was quite tiring for male teachers. ... For some, they may have to teach three classes. [Songqie, Paragraph 20-28]*

Interestingly, among the twelve female teachers, two mentioned that their career choices were affected by marriage (Jue) and pregnancy (Tie). Yet, none of the eight male teachers spoke one word of how private life might have affected their jobs during the interviews.

#### **7.4 Critical Perspectives about Female-Advantage-in-Languages Stereotype**

So far, teachers' endorsement of a variety of GSs was uncovered. Nonetheless, there is still evidence, though scarce, that some teachers are critical about FALS. For example, Jian, currently a non-believer of FALS, said that she used to endorse FALS (see Section 6.2.4). But after she had learnt about stereotypes in college, she was no longer willing to accept sweeping images of males and females, nor would she apply them to her students.

In addition, three male veteran teachers, Sunyue, Lixie, and Zengque, claimed that they would not generalise the gendered pattern they witnessed to all language learners, as the following quotes will illustrate:

*((Interviewer asks, 'did you mean that girls typically learn grammar better?'))  
Well, I can't say for sure that this is [the pattern for all individuals of the] two genders. But in my classes, judging from my students, this is the case. This is the situation with the two classes I teach. [Sunyue, Paragraph 149-150]*

*Well, I think this saying [that girls are more gifted learners of English] ((he clicks his tongue)). Because, well, in fact, I haven't found relevant evidence to prove it. I mean, with regards to the biological difference between males and females, I haven't found enough evidence to support [the existence] of such a difference. I simply came up with this argument based on some scores [of my students], I just analysed the exam results. So about this saying, I haven't reached a conclusion. [Lixie, Paragraph 176]*

*About this [gender differences in English learning], I can't reach a conclusion. ... Um, I feel that the database I have is not big enough, [so] I can't get a conclusion. ... It is hard for me to summarise any specific rules. [Zengque, Paragraph 136]*

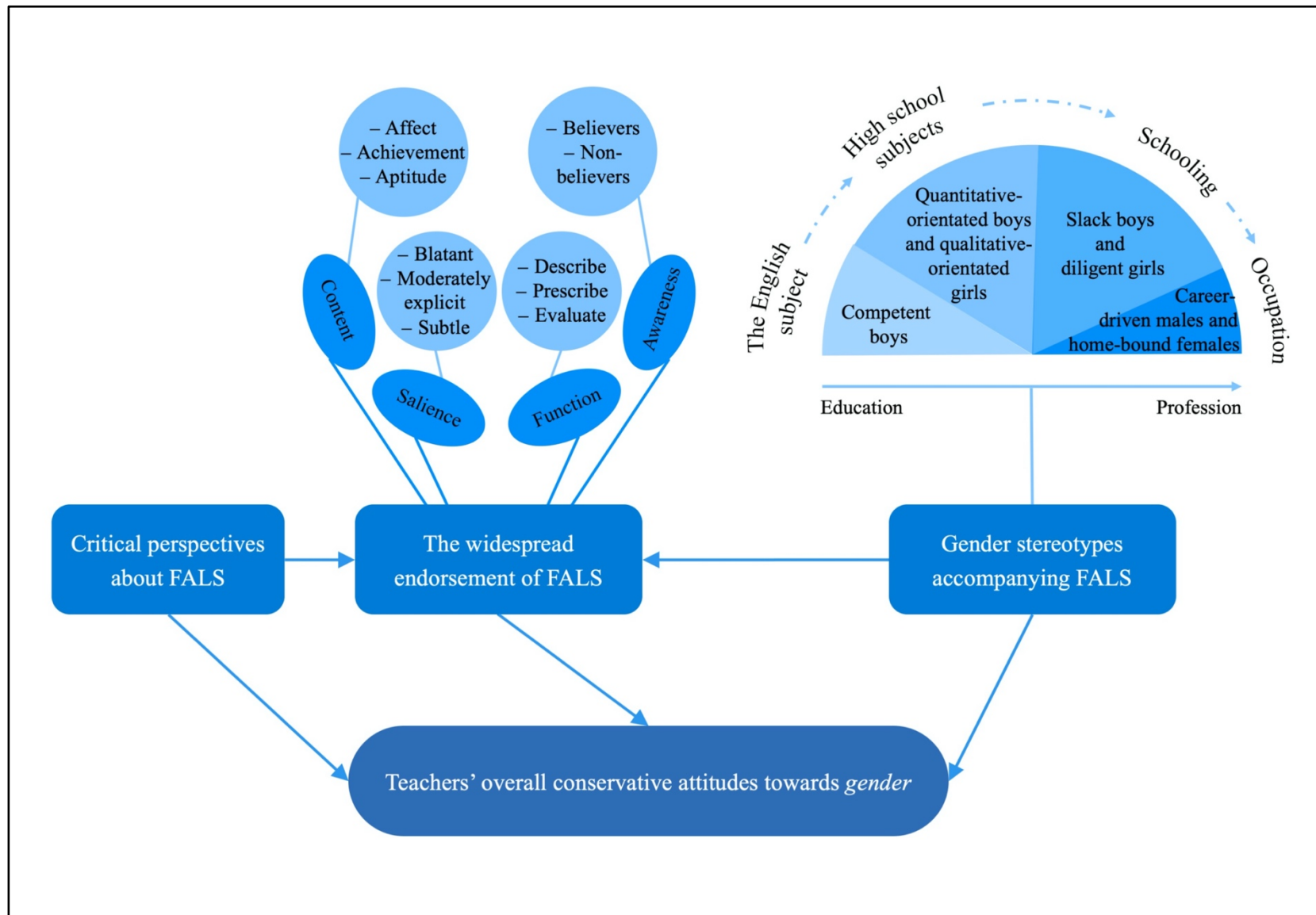
## 7.5 Summary of Teachers' Perspectives in Study 2

To sum up, three superordinate themes emerged via thematic analysis in Study 2—the widespread endorsement of FALS, GSs accompanying FALS, critical perspectives about FALS—as Figure 7.3 here shows. The majority of interviewees (90%) endorsed at least one component of FALS. All believers expressed FALS moderately explicitly in various instances, but some also did so either blatantly or occasionally, subtly. Additionally, among the believers, FALS was most frequently used descriptively. FALS was sometimes used in prescriptive or evaluative ways, but more research is needed to understand these two functions. Besides, all but one interviewee, Zhoujie (see Section 6.2.4), was aware of FALS's circulation in schools. The top middle part of Figure 7.3 illustrates this theme.

Meanwhile, eight FALS believers actually held another gender stereotype, CBS. In addition, three sets of GSs were found among teachers, ones that depicting males as quantitative-orientated, slack, yet career-driven, while females as qualitative-orientated, diligent, nevertheless home-bound. In fact, these GSs distributed along the education-profession continuum, as the top right part of Figure 7.3 shows. Each sector represents a (set of) GS(s), and the size of a sector indicates the frequency of the particular GS(s). Among all emergent stereotypes, FALS and CBS were directly targeting learners of English. The quantitative-qualitative division was domain-specific, sorting high school subjects into either masculine or feminine categories. Moreover, the slack-diligent comparison labelled all boys and girls in schools contrastingly. Taken holistically, these academic GSs characterised females as ideal learners while boys as remedial ones. Then, the career-home distinction assigned gender-stereotypical occupational roles to men and women respectively.

Finally, there are traces of critical voices regarding FALS, where teachers either rejected FALS or refused to overgeneralise FALS to all learners. However, as these critical perspectives are small in number, it would seem that in Study 2, the teachers' attitudes towards *gender* were still overall conservative, hence the overarching theme of “teachers' overall conservative attitudes towards *gender*”.

**Figure 7.3: Summary of Themes Emerged from Thematic Analysis in Study**



## Chapter 8: Findings and Discussions IV —

### Effects of Female-Advantage-in-Languages Stereotype on Learners

This chapter will present results from Studies 3 and 4 in Phase 2. Since the previous three chapters have illustrated how the female-advantage-in-languages stereotype (FALS), as perceived by students, guardians, and teachers of English, has portrayed females as more enthusiastic, higher-achieving, and more gifted learners than males, this chapter will turn its focus onto the effects of FALS on learners. In Section 8.1, findings from Study 3, the field experiment examining whether FALS would have influenced learners' test performance, will be displayed and discussed. Then, Section 8.2 will elaborate on findings from Study 4 to explore students' experiences with FALS and their cognitive, emotional, and behavioural responses to it. During the process, certain associations and comparisons between students' accounts and those of teachers in Study 2, and the student questionnaire dataset in Study 1, are also analysed. Finally, to further illustrate the role of FALS in English classrooms, Section 8.3 will integrate findings from Studies 3 and 4.

#### 8.1 Findings from Study 3

Study 3 was a field experiment investigating whether stereotype threat (ST) effect might have existed among learners of English. Section 8.1.1 below will briefly recap the participants and the study design (see Section 4.3.1 in Chapter 4 for a full description). Section 8.1.2 will then report the findings.

##### 8.1.1 Participants and Design

Participants were 458 students (*mean age* = 16.83 years, *SD* = 0.47) in Grade 11 from a high school in southern China. Approximately equal numbers of boys and girls were randomly assigned to a 2 (gender)  $\times$  2 (FALS activation) factorial design. 31 participants were excluded from data analysis for three reasons: failure to finish the test in time (2), withdrawal (13), and failed manipulation (16). Among the remaining 427 participants, 213 were boys and 214 were girls.

A 2  $\times$  2 between-groups ANCOVA was conducted to assess the effect of stereotype activation on male and female participants' performance in a subsequent English test. The independent variables were participant gender (male or female) and stereotype activation (stereotype threat or non-stereotype threat). The dependent variable was test score from the Oxford Online Placement Test (OOPT), with a full mark of 120. Scores from a paper-and-

pencil English exam administered two weeks prior to the experiment were used as a covariate to control for individual differences in English competence.

Preliminary checks were conducted to detect possible violations of the assumptions. The assumption of normality was violated for two sub-groups of the dataset, as assessed by Kolmogorov-Smirnov test (for male students in the control group,  $p = .024$ , and for male students in the experimental group,  $p = .013$ ). Despite this violation, the two-way ANCOVA was carried on, as the procedure is considered fairly robust to deviations from normality when the sample sizes in each group are similar. In addition, because there were two outliers in the data (assessed by two cases with standardized residuals greater than  $\pm 3$  standard deviations), another two-way ANCOVA without the outliers was run to see if the outliers would materially affect the result. Given that the results from two ANCOVAs were practically the same, the one performed on the whole dataset will be reported here to conserve statistical power.

### 8.1.2 Results from Two-Way ANCOVA

After adjusting for English competence, there was a statistically significant interaction effect between stereotype activation and gender,  $F(1, 422) = 5.667$ ,  $p = .018$ , with a small to medium effect size,  $\eta_p^2 = .013$ . Table 8.1 here records the descriptive statistics from the two-way ANCOVA, and Figure 8.1 illustrates the estimated means of OOPT scores as a function of participant gender and experimental conditions.

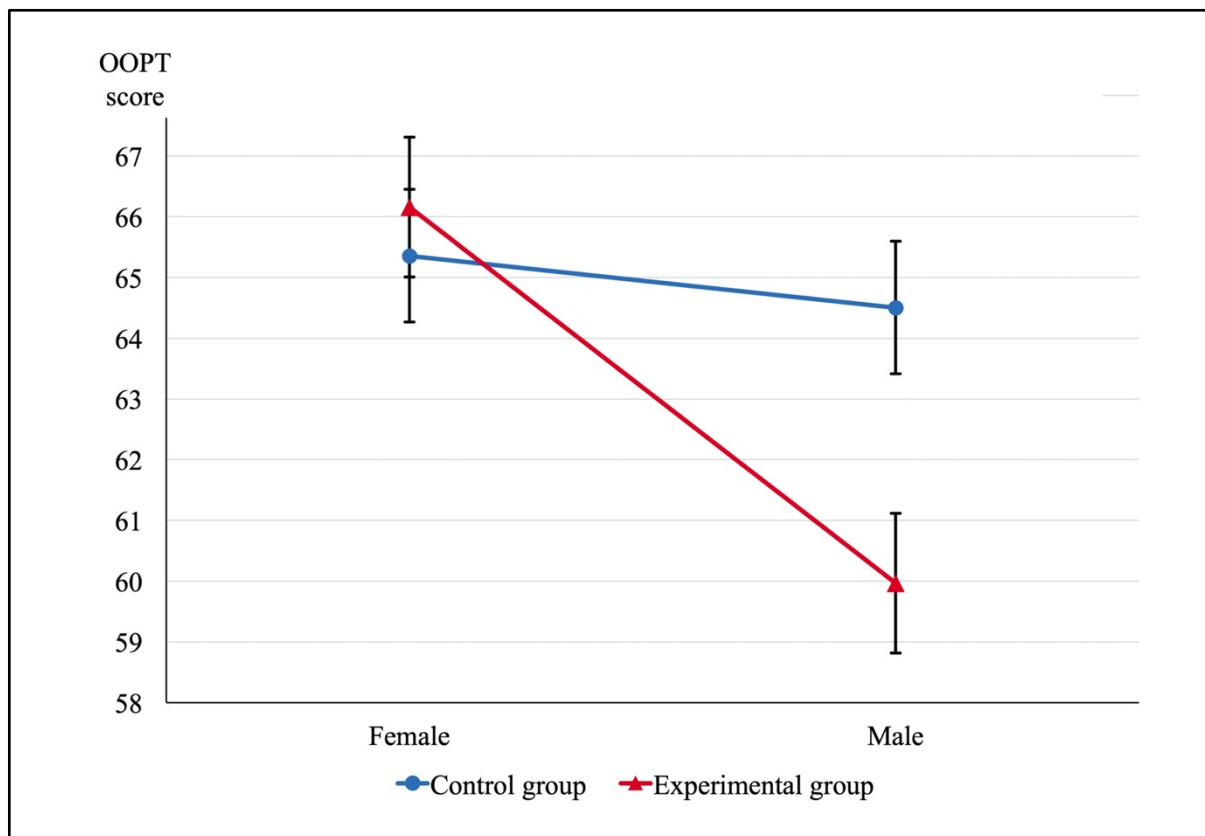
**Table 8.1: Mean Performance Scores by Gender and Condition in Study 3**

Gender	Condition	Mean	SE	95% Confidence interval	
				Lower bound	Upper bound
Female	Control ( $N = 113$ )	65.359	1.089	63.219	67.500
	Experiment ( $N = 101$ )	66.162	1.155	63.892	68.432
Male	Control ( $N = 112$ )	64.506	1.094	62.355	66.658
	Experiment ( $N = 101$ )	59.963	1.153	57.697	62.230

As Table 8.1 and Figure 8.1 have shown, boys and girls in the control group performed equally well, with girls demonstrating a slight advantage of .85 mark. However, the gender difference became statistically significant in the experimental group, where boys were outperformed by girls by 6.2 marks. In addition, among participants of the same gender, there were differences between groups, too. Girls in the experimental group achieved a marginally higher mean score than those in the control group, with a mean difference of .80. Boys, on

the other hand, experienced ST after having been exposed to FALS. These results suggest that male and female learners were affected differently by FALS. Female participants' performance in English was somewhat elevated after they had watched a video reminding them of FALS, indicating the plausibility of stereotype boost. However, male students in the same condition showed a substantial decrease in performance compared to their peers in the control group.

**Figure 8.1: Mean Performance Scores as a Function of Gender and Condition in Study 3**



The main effect of gender was also statistically significant,  $F(1, 422) = 9.799, p = .002$ ,  $\eta_p^2 = .023$ . However, the main effect of stereotype activation was not statistically significant,  $F(1, 422) = 2.777, p = .096$ ,  $\eta_p^2 = .007$ .

## 8.2 Findings from Study 4

In Study 4, 24 students (*mean age* = 16.58 years) in Grade 11 from the same high school as Study 3 attended group interviews. Each group consisted of three students of the same sex. Table 8.2 here summarised the demographic information and metadata of the group interviews. 15 students were from the Humanities branch, and the remaining 9 the Sciences branch. There were equal numbers of boys and girls. Overlooking sex differences,

interviewees averagely studied English for eight years and 11 months. In addition, the interviews lasted for 53 minutes and 38 seconds on average, and the transcripts had an average of 9,709 words. As exhibited in Table 8.2, compared with the boys, girls in Study 4 had learnt English for one month longer on average, and their interviews were slightly longer, so were the transcripts.

**Table 8.2: Demographic Information of Interviewees and  
Details about Group Interviews in Study 4**

NO.	Branch	Age	Years of English learning	Interview and transcript		
				Duration	Page count	Word count
Boy 1	HUM	17	9	00:56:48	17	9,620
Boy 2		17	9			
Boy 3		16	11			
Boy 4	SCI	17	6	00:51:12	21	12,107
Boy 5		17	9			
Boy 6		17	9			
Boy 7	SCI	17	9	00:48:30	10	6,214
Boy 8		16	9			
Boy 9		17	9			
Boy 10	HUM	16	9	00:56:48	15	7,845
Boy 11		16	9			
Boy 12		16	9			
Average (boys)		16.58	8.92	00:53:20	15.75	8946.50
Girl 1	HUM	16	11	00:58:30	22	13,196
Girl 2		16	9			
Girl 3		17	9			
Girl 4	HUM	17	6	00:53:57	16	8,881
Girl 5		17	9			
Girl 6		17	9			
Girl 7	HUM	16	9	00:46:57	21	10,447
Girl 8		17	9			
Girl 9		17	9			
Girl 10	SCI	16	9	00:56:24	17	9,366
Girl 11		17	9			
Girl 12		16	10			
Average (girls)		16.58	9.00	00:53:57	19.00	10,472.50

Notes. HUM = the Humanities. SCI = Sciences.



During the interviews, students were first invited to describe their experiences with English learning in general. Section 8.2.1 below will report the relevant details to provide a background for subsequent discussions. Then, students were prompted to compare male and female learners, during which they revealed FALS and other GSs circulating in their lives. Thematic analysis generated the over-arching theme of “students’ emergent understanding of *gender*” with three inter-related superordinate themes: 1) the widespread FALS; 2) approaching FALS critically; and 3) GSs accompanying FALS.

In the remainder of the chapter, Section 8.2.2 will delineate how FALS was construed by students and whether they endorsed it. Sections 8.2.3-4 will proceed to characterise students’ encounters with FALS, including how they became aware of it and how they responded and/or avoided responding to it. In both these sections, students’ critical approaches to FALS will also be presented. Subsequently, Section 8.2.5 will identify other gender stereotypes (GSs) circulating among the students. Finally, Section 8.2.6 will review these prevalent GSs with regards to the roles they played in shaping students’ emergent gender attitudes. Such an arrangement parallels the organisation of Chapters 6 and 7, where findings from Study 2 (interviews with teachers) were described, so that meaningful connections, convergences, and divergences between students’ perspectives and those of teachers can be drawn upon. In Sections 9.1.3-4, further connections between the two studies will be elucidated to depict the role of FALS and gender stereotyping in high schools.

### **8.2.1 Descriptions of Experiences with English Learning**

**Onset of learning.** 19 interviewees (79.17%) started learning English since Grade 3 in primary schools, when they were 8 or 9 years old. Three students (Boy 3, Girl 1 and 12) had attended extracurricular English courses signed up for them by parents before English was officially introduced in schools. Boy 4 and Girl 4 started the last, until they were in Grade 6 in primary schools. To sum up, by the time students participated in Study 4, all had been learning English for at least six years (see Table 8.2 above for exact numbers).

**Parental involvement.** Three boys and six girls recalled their parents’ involvement in their English study. Boy 3 contended that his father, an English teacher, was the reason that he got sent to extracurricular English courses at only five years old. Boy 7 noted that both his parents cared about his academic performance, and pushed him to study English harder because they thought English skills would benefit him in future. The remaining seven interviewees all spoke of only their mothers, who either exposed them to English at an early age (Boy 12, Girl 2 and 9) or sent them to extracurricular English courses (Girl 1, 5, 8, and

12) since primary schools. Of course, this does not necessarily suggest that mothers would be more involved in children's English learning. After all, 15 interviewees did not mention their parents at all.

**Affective relationships with English.** Students either volunteered or were prompted to describe their affective relationships with English. Five boys conveyed an interest in the subject—Boy 2, 3, 8, 9, and 12. Interestingly, the number of girls claiming to be fond of English was only four (Girl 3, 4, 5, and 6). That is, despite teachers' stereotypical belief that girls tend to have a more positive relationship with English (see Section 6.2.2.2), students in Study 4 stated otherwise. Actually, all four girls, after having acknowledged their general liking for English, proceeded to recounting the distress (Girl 3, 5, and 6) and stress (Girl 4) English had caused for them. For example, Girl 5 elaborated on how her unsatisfactory performance in a past English contest hurt her feelings:

*I am:: I was also quite fond of English, and I thought that my English teacher liked me very much. And in junior high school, we had that, English speaking contest thing. And I entered once when I was in Grade 8. But I was—each class had one contestant, and there were three classes—I was the one got the lowest score. ... [During the award ceremony], on the front line, there was a first prize winner who was a student of my teacher's from another class. I was standing there in the back, looking at her ((refers to the teacher)) holding her ((refers to the student)), laughing and talking extremely happily. I was, then and there, just hurt. ((She clicked her tongue.)) I felt that this experience actually affected me a bit. After this, [I] just never liked English that much. [Girl 5, Paragraph 36-40]*

Yet, such pattern of mixed feelings was absent among boys. Worse still, four girls explicitly complained that they did not enjoy English (Girl 1, 10, 11, and 12), while only three boys did so (Boy 1, 5, and 6). The findings above cast doubt on whether teachers' stereotypical image of English-loving girls, as well as English-indifferent/-disliking boys, was an over-simplified interpretation of the complex reality. In fact, empirical research provided evidence that girls did experience a wider range of emotions overall in foreign language (FL) classrooms (e.g., MacIntyre, 2007; Dewaele et al., 2016), ranging from pleasure to anxiety. This point manifested itself in the interviews, too: when students sometimes disclosed how certain aspects of English were challenging for them, only Boy 1 commented how such difficulties made him feel (panic). Girls, on the contrary, reflected

more frequently upon their emotional responses: Girl 1 explained that her teacher's temper made her upset, Girl 11 pointed out that rote learning English vocabulary annoyed her, and Girl 12 even said vocabulary retention was painful for her. In addition, girls more frequently spoke of how their relationships with English teachers affected their feelings for English, as the following subsection, 'Teachers and teaching', will soon unveil. In any case, the descriptions in Study 4, as well as findings from real-world research, seemed to suggest that the affect component of FALS was only telling one side of the story.

**Teachers and teaching.** 17 interviewees expounded how teachers, their teaching methods and styles shaped their experience with English, except for Girl 6, Boy 6, 8, 9, 10, 11, and 12. Among these 17 students, boys seemed more concerned with the impact teachers had on their attainment, while girls with the impact on their affective status. For example, five boys (Boy 2, 3, 4, 5, and 7) and two girls (Girl 7 and 9) believed that certain incompetent or irresponsible teachers left them struggling with English.

*In primary school, [the teacher] just, just you how to pronounce the words. And (.) you had no idea what they meant. The teacher would just:: leave it there.*  
[Boy 4, Paragraph 28]

*((He clicks his tongue.)) I feel that, my problem [with English] started early ((he chuckles)), got buried deep. ((he chuckles)) Because I used to study in, ((he clicks his tongue)), you might not know it Miss ((refers to the interviewer)) ((he chuckles)), a town. So, anyways, ever since the third grade, because the English teacher in our town was, really, not good at particularly speaking and listening, [my English] has fallen behind since then.* [Boy 7, Paragraph 15]

*Our, our English in Grade 10 ... gave lectures at a really fast pace, ... So I just felt that this teacher could not, could not take care of every student's need. Perhaps it was only those, whose English grades were already good [who could have followed the lectures].* [Girl 7, Paragraph 56]

On the other hand, Boy 1, together with six girls (Girl 1, 2, 3, 4, 5, and 8), expressed that their perceptions of or relations with their teachers bore upon their emotional engagement with the English subject.

*The teacher in Grade 10 was quite, took quite good care of me, and treated me well. Unlike our current teacher ((he chuckles embarrassingly)), who's a bit hot-tempered. So I just, ... I was somehow repelled by him. ... And even my*

*English grade now is, is only just average, or slightly better than average. [Boy 1, Paragraph 7]*

*In junior high school, we had self-study sessions in the morning. We could voluntarily recite passages from the textbook, and had our teacher check up on how well we did. For me, if, um, when the teacher encouraged me, told me that I did well this time, I'd be more willing to do the reciting again. [Girl 4, Paragraph 43]*

*Sometimes, if I meet an English teacher that I like, I will enjoy English very much. But if (.) I meet a teacher I dislike, I'll just (.) get average grades. [Girl 8, Paragraph 15]*

Curiously, a group of girls (Girl 10, 11, 12), raised that they felt more comfortable with female teachers than with male ones, even if they had had only one male English teacher for one month throughout their school life.

*... I think, ((she giggles)), I suppose, male, male teachers ((she giggles while exchanging looks with Girl 10 and 12; they all giggle)). I suppose that male teachers who teach English, are just a bit unacceptable. And I think, at least, generally speaking, teachers who teach languages, Chinese and English, are always females. And I think female, female teachers who teach English, well, are relatively easy to accept. [Girl 12, Paragraph 57]*

When Girl 12 was asked why she had felt this way, Girl 10 and 11 jumped in, adding that it was because they had a substitute teacher (male, obviously) for one month, when their female teacher took maternity leave. They went on describing instances where this male teacher spoke English in a muffled voice while keeping a stern face in classes. After having confirmed that this was the one and only male teacher they ever met, the interviewer inquired, 'why do you think that female teachers are just more easily accepted than males then?' The girls drew more vivid pictures about this particular male teacher with their words, including having students standing on their feet in class as a punishment for not knowing answers to his questions (Girl 10), making her nervous with his grim stare (Girl 11), and being stingy with compliments (Girl 12).

It would appear that they had fallen prey to the availability heuristic, where immediate examples coming up in mind are relied upon when evaluating a specific topic (Kunda, 1999). This time, the first example happened to be the only example, and the girls apparently

overgeneralised the teacher's features to all male teachers. This provides additional evidence that the English teaching profession has become a gendered field, where female teachers are the acceptable norm, a phenomenon echoed in Fengbie, Tie, and Lan's resounding comments that males are pushed away from English education (see the first three quotes in Section 7.3.4).

## **8.2.2 Students' Endorsement of Female-Advantage-in-Languages Stereotype**

Section 8.2.2.1 reports students' endorsement of FALS and analyses their emerging gender attitudes reflected in their accounts. Note that with teachers' FALS in Study 2, the analysis covered three features: content (Section 6.2.2), salience (Section 7.1), and function (Section 7.2). The difference occurred because Study 4's primary goal was to explore students' experiences with FALS and their cognitive, emotional, and behavioural responses to it, not to dissect FALS in interviewees' minds. That was the aim of Study 2. Students' endorsement of FALS was a pertinent topic, not the emphasis. As the tripartite composition of FALS is the theoretical foundation of the whole project, the content of FALS among students naturally warrants attention. Nevertheless, certain issues with salience and function will also be examined wherever relevant.

In addition, like the case with Study 2, further analysis indicated that some students believed that a strong language aptitude and a fondness of English contribute to higher learning attainment. It was then hypothesised that the same trend might have existed in the student questionnaire dataset from Study 1, where the perceived gender differences in aptitude and affect may both have predicated the differences in perceived gender gap in English achievement. Thus, a multiple regression was performed. Both findings from the interviews and the questionnaire dataset will be presented in Section 8.2.2.2 below.

### **8.2.2.1 Students' Endorsement of Female-Advantage-in-Languages Stereotype: From Conformity to Critical Thinking**

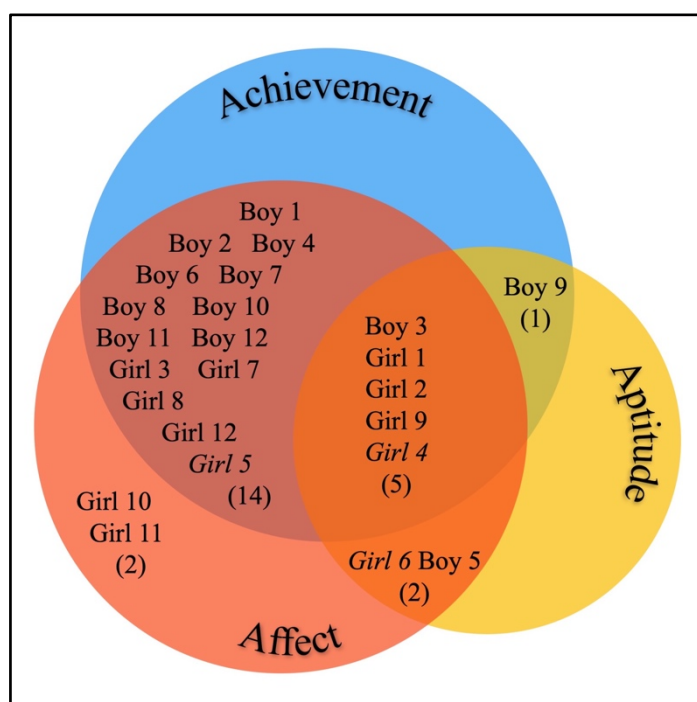
At the beginning of the group interviews, interviewees were encouraged to 1) sketch their experiences with English and 2) contemplate on the characteristics of ideal English learners (see Appendix Q for the complete interview schedule). The aim was to elicit FALS, but both techniques failed on all participants but Boy 7. That is, all but one students did not explicitly associate English learning/learners with gender, a common trend found in similar research conducted in England (e.g., Pritchard, 1987; Barton, 1997). Boy 7, an exception to his peers, blatantly expressed his endorsement of the achievement component of FALS when responding to both prompts.

... when I got to junior high school, I felt that there was a wider gap between [me] and girls. Even though we were all from the same town, but girls, ((he clicks his tongue)), they learnt very well, and their knowledge foundation was very solid. Even if we were taught by the same teacher. But in my grade, there were many boys, all like me, all felt English—((he clicks his tongue))—I cannot say that none of us was interested in it, because we had to learnt it. I mean, even if we did not spend less time on it than:: girls, ((he chuckles)), our English was (.) full of problems. Problems with listening, with vocabulary, with writing, and with speaking. [Boys] were good at nothing. [Boy 7, Paragraph 15]

I had a feeling that, people who learn English well, ... their personality, I find personalities very important. Some people's personality—because the majority of girls are quite attentive, you know, and details, well::, they pay extreme attention to details, so they learn [English] better. [Girls are] unlike boys, who are careless. [Boy 7, Paragraph 38]

The remaining 23 interviewees endorsed at least one FALS component moderately explicitly. Figure 8.2 is a Venn diagram showing the believers of FALS. The colouring scheme is the same as Figure 6.4 (FALS endorsement by teachers in Study 2): blue, red, and yellow circles represent the achievement, affect, and aptitude components, respectively.

**Figure 8.2: The Venn Diagram of Believers of the Female-Advantage-in-Languages Stereotype Components in Study 4**



According to Figure 8.2, interviewees regarded females more favourably than males in terms of affect (by 23 interviewees, 95.8%), achievement (by 20 interviewees, 83.3%), and aptitude (by 8 interviewees, 33.3%). These findings converged with the effect sizes from the student dataset in Study 1: the affect component had the largest effect size ( $\eta_p^2 = .416$ ), the achievement component had a slightly

smaller one ( $\eta_p^2 = .415$ ), and the aptitude component had the smallest effect size— $\eta_p^2 = .229$ .

The five interviewees in the region where three circles overlap are those endorsing all three FALS aspects. The 14 participants reside where the red and the blue circles meet simultaneously endorsed the affect and achievement components. Boy 9, sitting in the intersection of the yellow and blue circles, believed in the aptitude and achievement components; Boy 5 and Girl 6 only endorsed the affect and aptitude components. Girl 10 and 11 endorsed the affect component solely.

A comparison of the two Venn Diagrams (Figure 6.4 and Figure 8.2) will show that among both teachers and students, the aptitude component was the least popular, a trend also found in Study 1: the effect size of aptitude was the smallest of three FALS aspects in both teachers and students,  $\eta_p^2 = .435$  and  $.229$ , respectively. Yet, two differences between the teachers in Study 2 and the students in Study 4 were detected. First, there were two non-believers among teachers, Yuehan and Jian, but none among students. Second, affect was the most widespread component among students (23 believers), but among teachers, it was achievement that was most popular (18 believers). Due to lack of space here, these two disparities and their implications will be discussed in Sections 9.1.4.2-3.

Curiously, Girl 4, 5, and 6, whose aliases are in italics in Figure 8.2, displayed ambivalent views. They all acknowledged that in general, girls were more interested in learning English than boys, but they then had an extended discussion about this gender difference in interest:

*Girl 5: ... I think boys are more interested in:: playing hard ((she giggles)). Their minds are not put onto learning. ((She giggles)). And girls, come to think of it, I suppose not many are actually interested in English itself. Typically, girls study English for the grades, not, well, out of genuine interest.*

...

*Girl 6: I agree that, not many girls are genuinely interested [in English]. For me, I am more interested in what it brings for me, the sort of admiration from others for learning it well. ((she laughs)) ... And this sense of achievement, yes, I don't think that this counts as genuine interest. It's not like I am genuinely interested in English.*

*((An argument about what constitutes 'genuine interest' continues.)) ...*

*Girl 4: in terms of interest, I think, nobody seems to have a particular interest in English. Because, we are facing exams, pressures from exams. With high pressure, it's like running: when you are sprinting, you cannot enjoy it, you don't see the views along your way. [Girl 4, 5, 6, Paragraph 189-224]*

During their discussion, this group of girls re-examined the affect component of FALS while subscribing to the GSs 'slack boys and diligent girls', which are prevalent among teachers (see Section 7.3.3). According to them, boys just mucked about in school. Girls, on the other hand, worked hard in English because they were more inclined to improve their grades, even though both boys and girls suffered from exam stress that prevented them from enjoying learning. Their reluctance to blindly accept FALS reflects an ability to critically think about the static and essentialist construction of *gender* by cultural norms such as prescriptive GSs. The same tendency also appeared in Boy 10, 11, and 12, when their group came to acknowledge that with regards to language aptitude, there was only individual differences, instead of gender differences.

Of course, an alternative explanation for Girl 6's speech would be outgroup homogeneity (Park and Rothbart, 1982), the tendency for ingroup members (in this case, a female English learner) to see themselves as more heterogeneous than those from outgroups might assume (teachers, male English learners, etc.). Since Girl 6 relied on her own experience to clarify why not all girls were genuinely interested in English, it would seem that her intricate knowledge about herself has led her to perceive greater diversity in her social group—female English learners. Nevertheless, however the case with Girl 6 might be, her rejection of swallowing FALS still stands.

#### **8.2.2.2 Causal Links among Three Components of Female-Advantage-in-Languages Stereotype as Perceived by Students**

**Evidence from the Interview Dataset in Study 4.** During the interviews, students were encouraged to deliberate the qualities needed to learn English well, which yielded 55 data segments later coded as 'perspectives on ideal learners'. 23 of the coded segments revolved around how an interest in English and a congenial relationship with the English teacher would lead to good results. In these accounts, female learners were sometimes referred to as examples of enthusiastic learners, and male ones were described as rather detached.

*Girl 4: You have to well::, be very knowledgeable about the culture [of the English world], and like it very much.*



*Girl 6: Yes. We have a girl like that in our class. ... She likes Taylor Swift very much. She's heard every song of Taylor's. She also knows how to sing them. And, even the gossip, she also knows a lot about her ((Taylor Swift)) relationships. So her English grades are very good. Always in the first place.*  
[Girl 4 and 6, Paragraph 207-213]

*... To learn English well, you must be, be relatively fond of English. Look at some boys, like my younger brother, who dislike English. He can't even (.) take in a single word, or memorise one vocabulary, nothing. Nor can he finish the listening exercises. If they just don't want to memorise stuff, they sure can't commit anything in their minds.* [Girl 8, Paragraph 62]

A total 13 coded segments raised that a strong memory and language savvy were essential for ideal learners, features that represented 'language aptitude' for teachers in Study 2 (see Table 6.4). Boy 4, 5, and 6 further stated that it was girls who had a superior cognitive prowess, as the following excerpt demonstrates:

*A passage, a passage always has its (.) thesis, isn't it? I mean the gist that it is supposed to express. Sometimes I don't get the theses. Girls, I feel that they, it is much easier for them to get [theses]. They just grasp the theses more effortlessly. And this, I feel that this is just the way it is.* [Boy 6, Paragraph 380]

The above analysis indicated that 1) some students stereotypically perceived a stronger causal link between affect for English and achievement in English, compared to that between language aptitude and achievement, and 2) both girls' stronger linking for English and their cognitive edge in language aptitude were thought to have contributed to their relative learning success, both of which were already found among teachers in Study 2. It was then hypothesised that the same trend might have existed in the student questionnaire dataset from Study 1, where the perceived gender differences in aptitude and affect may both have predicated the differences in perceived gender gap in English achievement. The next subsection reports the relevant findings.

**Evidence from the Questionnaire Dataset in Study 1.** A multiple regression was performed to predict perceived gender difference in achievement from perceived gender differences in aptitude and affect. Statistical assumptions were checked first. There was linearity, normality, independence of residuals (Durbin-Watson statistic = 1.985), homoscedasticity, and no evidence of multicollinearity. There were no leverage values

greater than 0.2, nor were there values for Cook's distance above 1. Although one studentized deleted residual was greater than  $\pm 3$  standard deviations (-3.021), its influence was considered small and thus was retained in the dataset for later analysis.

The multiple regression model statistically significantly predicted perceived gender difference in achievement,  $F(2, 907) = 352.754, p < .001, \text{adj. } R^2 = .661$ . Both variables added statistically significantly to the prediction, as shown in Table 8.3. The regression coefficient of affect was larger than that of aptitude, which corresponds to the findings in the group interviews with students, as well as the findings from the teacher participants in Study 1 and 2.

**Table 8.3: Multiple Regression Analysis of Student Dataset in Study 1**

Variable	<i>B</i>	<i>SE<sub>B</sub></i>	$\beta$	<i>p-value</i>
Intercept	.484	.058		< .001
Aptitude	.340	.028	.363	< .001
Affect	.420	.032	.392	< .001

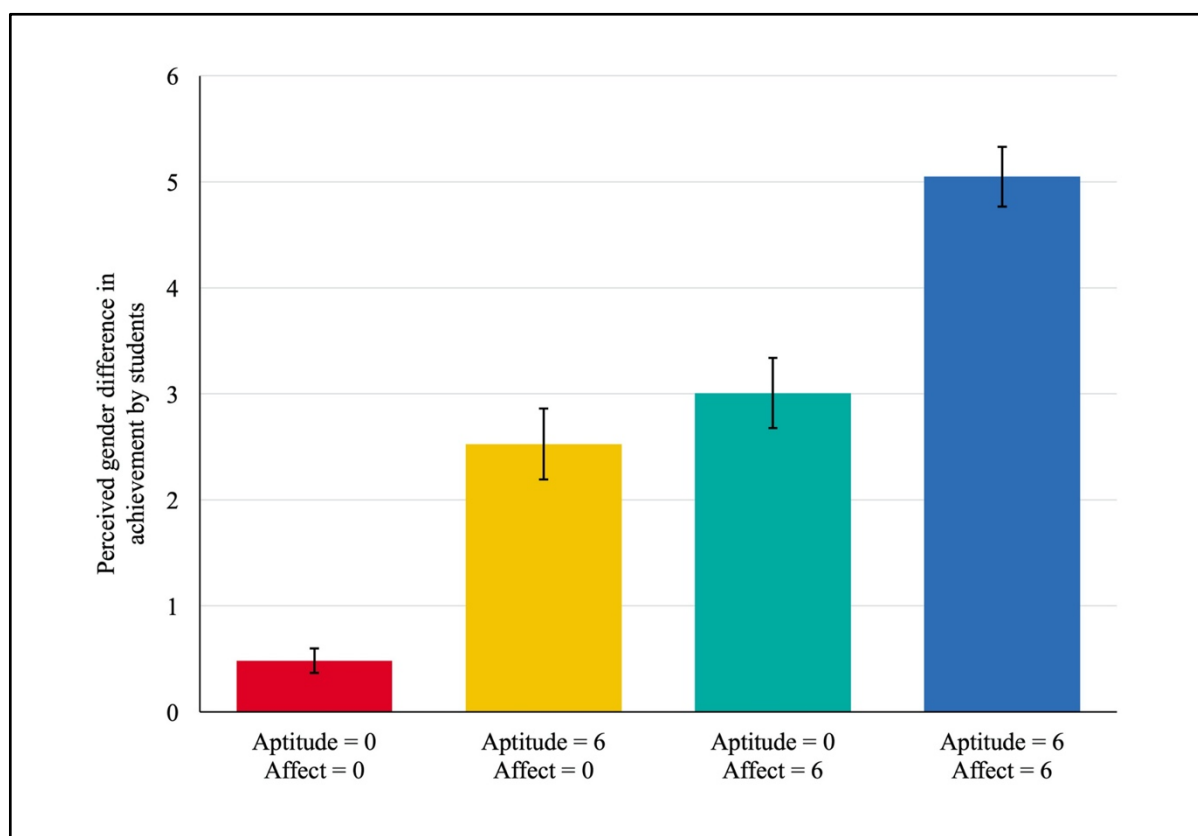
*Notes.* *B* = unstandardised regression coefficient; *SE<sub>B</sub>* = standard error of the coefficient;  $\beta$  = standardised coefficient.

Predictions were made to determine the mean of perceived gender difference in achievement for students who assumed zero difference in aptitude and affect. The mean was predicted as .484, 95% CI [.371, .598], as the first bar in Figure 8.3 below shows. On the contrary, the mean of perceived gender difference in achievement by teachers assuming maximum differences in aptitude and affect (both equals six) was predicted as 5.049, 95% CI [4.768, 5.329], as demonstrated by the fourth bar in the same figure.

Predictions were also made to determine the mean of perceived gender difference in achievement for teachers who assume no difference in affect but maximum difference in aptitude, as illustrated by the second bar in Figure 8.3. The mean was predicted as 2.526, 95% CI [2.191, 2.862]. Then, the mean of perceived gender difference in achievement by teachers assuming maximum gender differences in affect but zero difference in aptitude at all was predicted as 3.007, 95% CI [2.676, 3.338], seen as the second bar in the figure here. Furthermore, the four bars in Figure 6.6 in Section 6.3.2 and the ones in Figure 8.3 below followed the same pattern: the stereotypical gender achievement would be the smallest when teachers and students regarded male and female learners similarly in aptitude and affect, larger when they perceived females more positively in aptitude alone, even larger when they

viewed females more favourably in affect alone, and the largest when they rated females higher in both aptitude and affect.

**Figure 8.3: Predictions of Means of Perceived Gender Difference in Achievement Based on Four Combinations of Perceived Gender Difference in Aptitude and Affect from Student Dataset in Study 1**



### 8.2.3 Students' Awareness of Female-Advantage-in-Languages Stereotype

According to the interviewees, FALS had been commonplace since primary school years. The sources cited included family members (mostly parents, mentioned by 10 students), teachers (14), peers (10), and media (paper and digital forms, 3). When a specific source could not be attributed to, FALS was referred to as a 'social notion' laid out for boys and girls (by Girl 6), a saying 'heard from people around us' that hangs 'in the subconscious' (by Boy 6), and a 'common understanding in the society' with which 'everyone agrees' (by Girl 7). The quotes here provide evidence for students' social learning of the gender norms disseminating in their immediate and wider social environment. By acquiring the stereotypical association between the female gender and the English subject, students conformed to the social constructions of *gender*, as Girl 10 here said:

*My impression of this saying seems to have stayed in junior high school, because at that time it was my own belief. At that time, seriously, during every*

*lesson, our teacher paid more attention to girls. She was like, 'this girl, answer this question. That girl, answer that question.' Perhaps at that time, in our class, the gender ratio was like, one to one. But when the teacher asked questions, only one or two boys would attempt to answer. There were far more girls [who tried to answer the questions]. [Girl 10, Paragraph 341]*

Against this stated ubiquity of FALS, Boy 2's claim that he had 'honestly never heard of it' seemed like a rarity. Coincidentally, Zhoujie, a male veteran teacher was also uninformed about FALS's wide circulation (in Section 6.2.4). Considering that Zhoujie was aware of the stereotype that males are more competent than females, and deemed that it was more popular that FALS, Boy 2 might as well have shared a similar view, except that he did not express such opinions in the remainder of the interview.

A persistent pattern across interviewees is that FALS was thought to have been spreading more frequently in early school years than now (in senior high school), as demonstrated by Boy 7's experience below:

*This idea, I think that it was more frequent in primary school and junior high. Especially in primary school. My English was not very good in primary school, and sometimes my parents would say, 'why don't you learn from XX? Look at her. That girl learns English really well. What are you doing in school?' They were always comparing me with girls who did English well. [Boy 7, Paragraph 137]*

Two reasons can explain the decline of FALS in high school. First, as previous literature has noted, children's knowledge about GSs tended to develop between ages 6 to 12 and reach a peak towards the end of that period (e.g., Banse, Gawronski, Rebetez, Gutt, and Morton, 2010). Second, interviewees noted that the quantitative-orientated boys and qualitative-orientated girls stereotypes were gaining popularity as they approached high school, which perhaps overshadowed FALS (see Section 8.2.5 for more details).

#### **8.2.4 Students' Responses to Female-Advantage-in-Languages Stereotype**

After acknowledging FALS's presence in their lives, students were encouraged to comment on the effects of FALS on them and how they had responded to it. A total of eight types of responses were recorded in Table 8.4, each with an operational definition and a list of users. Of course, a student might have more than one responses.

**Lack of responses.** The most widely adopted category, ironically, is to not react to FALS at all, as the quotes from Boy 1 and 4 indicate.

*I've become indifferent to it ((refers to FALS)) because I've heard a lot. [Boy 1, Paragraph 268]*

*I, I don't feel anything. I just, whatever they say, I feel that it's none of my business. So I just don't care about it. [Boy 4, Paragraph 290]*

Similar to Boys 1 and 4, seven more interviewees also simply shrugged off FALS, either being used to it or dismissing it as irrelevant. In fact, the second reason was exclusive to boys (Boy 4, 10 and 11), perhaps as a way to construct independent self-construal (Markus and Kitayama, 1991). As identified by Cross and Madson (1997), males were more likely to separate representations of other in-group members from their self-images than females. Thus, when the boys regarded FALS as inapplicable to them, they managed to maintain independent self-concepts in order to protect their self-esteem.

**Emotional responses.** Negative emotions were expressed by three students, Boy 3, Girl 1 and 12. Boy 3 was unhappy because he endorsed FALS, and his self-esteem took a blow:

**Table 8.4: Eight Type of Responses to Female-Advantage-in-Languages Stereotype among Students in Study 4**

Type of responses and definition	Users
<b>Lack of responses</b>	
Interviewees denied having any reactions when encountering FALS either because they have grown accustomed to it and/or because they did not care about it.	9 in total 6 boys: 1, 4, 7, 8, 10, 11 3 girls: 1, 3, 5
<b>Emotional responses</b>	3 in total
Interviewees, having been judged or treated by FALS-believing people, felt uneasy or hurt.	1 boy: 3 2 girls: 1, 12
<b>Cognitive responses</b>	
1. Critical examination	
Interviewees contemplated on the validity of FALS either by re-examining the source or after having seen examples of counter-stereotypical individuals	5 in total 2 boys: 10, 11 3 girls: 6, 10, 11
2. Self-stereotyping	8 in total
Interviewees integrated FALS-congruent traits or behaviours into their own self-image.	4 boys: 3, 4, 5, 9 4 girls: 1, 4, 6, 8

3. Witnessing self-stereotyping	
Interviewees reported that they had witnessed others incorporating FALS-consistent traits or behaviours into their self-image.	4 in total 1 boy: 1 3 girls: 7, 8, 9

### Behavioural responses

1. Witnessed self-handicapping	3 in total
Female interviewees described how boys reduced their effort to learn English, accepting their failure doomed by FALS.	N/A 3 girls: 7, 8, 9
2. Actively managing	3 in total
Male interviewees described how they continued to learning English despite FALS's negative portrayal of them.	3 boys: 5, 6, 12 N/A
3. Applying FALS to others	1 in total
Interviewees judged or treated counter-stereotypical individuals negatively on the basis of FALS.	0 boy 1 girl: 2

*Notes.* Boy 2 is not in the table: after he had explicitly denied ever heard of FALS, he did not participate in Boy 1 and 3's subsequent discussion about how FALS affected them.

*I was, at that time I was unhappy with it ((refers to FALS)). ... Our teacher went on with it, and I was booing. But in fact, still, at the bottom of my heart, I still know that, or you can say that I acquiesced, I acknowledged that girls were just more gifted than us in English. [Boy 3, Paragraph 259-263]*

Girl 1 and 12, on the other hand, were hurt by their FALS-believing parents:

*My mum, it seemed like that she still feels that girls should (.) learn English well. She's always held this hope that my English grades will go up. She would say, 'they will go up. they will.' It is this inexplicable hope that makes me feel (.) a bit bad. I keep thinking, 'when will this gift thing fall upon me?' I really want to make my mom happy. [Girl 1, Paragraph 407]*

*My parents are like that. They think that girls should just, be good at language stuff, such as Chinese and English. It's only fitting girls should do well in them. Once in an exam, I got 97 for English, and my elder brother got 85. My parents told me, 'how come you only got this many marks? Why did you lose marks?' But when they spoke to my brother, they were like, 'you did quite well.' It was like, as a girl, I should have just got a full mark. ... I felt:: unhappy, you know.*

*I mean, I worked hard. ((She sighs.)) [They were] not exactly, exactly, being fair.* [Girl 12, Paragraph 353-356]

Girl 1 was afraid that she had let her mom down; Girl 12 was sad that her parents took her success in English as granted and treated her and her brother unfairly. Some might argue that FALS could have been self-serving for girls, thus stimulate positive feelings among them. Yet, despite the fact that some girls in Study 4 were high-achievers, no such emotions were reported.

**Cognitive responses.** Three types of cognitive responses also emerged during the interviews: critically examining the validity of FALS, self-stereotyping, and witnessing others (of the opposite sex) self-stereotyping. Two boys and three girls questioned whether FALS reflected the reality based on two diverging reasons: Boy 10 and 11 called FALS an ‘arbitrary way of defining people’ on the basis of biological sex routinely adopted by their parents’ generation. They recognised the prescriptive nature of such gender-stereotypical notions, and accused them of ‘neglecting individuality’. The three girls, Girl 6, 10 and 11 took a different approach: having observed counter-stereotypical examples (boys who excelled at English and girls who were remedial learners), they realised that FALS was an over-generalisation. Their perspectives are best illustrated in the quote below.

*I, of course I don't agree with it ((refers to FALS)) now. I think, if boys and girls put their minds into English, their results would be similar. There is a boy in our class, who is committed to English. He works hard, and his grades are good. There is no difference between boys and girls.* [Girl 10, Paragraph 349]

Equal numbers of boys and girls internalised FALS into their self-concepts, and four others also witnessed individuals of the opposite sex self-stereotyping. These two types of cognitive responses corresponded to each other, demonstrating students’ tendency to conform to gender norms prescribed by FALS. Boys claiming that they would take comfort from FALS, using it to explain away their lack of interest in English or bad exam grades, as Boy 3 here said:

*When I did terribly in exams, my parents would criticise me, a lot. ‘Anyhow, this ((refers to English)) is something I don't have a gift for.’ And then I compare myself against others using a subject that I am good at. You know, to refute [my parents].* [Boy 3, Paragraph 280]

Interestingly, girls actually noticed this self-stereotyping tendency: a group of girls recounted how boys idled about during English classes or avoided doing English homework because they've accepted FALS's prescription of failure. The following quote from Girl 7's description shows this pattern of girls observing boys lying back on FALS.

*I suppose it's like this: the more people say that (.) girls, um, um, girls do English better—perhaps we all share a similar starting line—the more boys hear of such sayings, the more comfortable they feel (.) about doing bad in English. [Girl 7, Paragraph 295]*

Like the case with boys, three girls also engaged in self-stereotyping. For example, Girl 8 felt that she *had* to learn English better because FALS prescribed so.

*And us girls, for example, if sometimes, when the teacher praises a list of students who do well in an exam, you wait until [the teacher] gets to the end and realise, 'hey, I was not mentioned.' You will feel that, 'hey, my English needs improving. There are so many girls in that list; I should also be there. I'd better work harder.' [Girl 8, Paragraph 309]*

Although her thoughts were only echoed by Girl 4 and 6 during the interviews, according to Boy 1, such self-stereotyping among girls was commonplace:

*I think that boys and girls see themselves in that [stereotypical] ways. Perhaps, for example, if people keep saying that girls are gifted in Chinese and English, they will probably feel confident. They believe that they have an advantage, so when they learn English, they will (.) have a more positive attitude, and they'll become more willing to learn. [Boy 1, Paragraph 142]*

Curiously, Girl 1 was different from all others—when Girl 2 and 3 commented on boys' general indifference to English, she interjected:

*Oh this, I am the best person to understand them ([refers to the boys]). Because I am also, I feel that I am a boy. ((She giggles)) I mean, honestly, with English, it's not like I don't want to listen [to the teachers]. I listen, and I listen, and I just ... lose track of what the teachers are saying. [Girl 1, Paragraph 201-205]*

Evidently, Girl 1 viewed a disinterest in English as a masculine trait. Since she was by no means an English-enthusiast either, she internalised the male aloof attitude and literally called herself a boy.



**Behavioural responses.** Three kinds of behavioural responses were documented, witnessed self-handicapping, actively managing, and applying FALS to others. The first two categories were in seeming conflicts. A group of girls suggested that they had observed boys reducing effort in English in an attempt to deflect any personal responsibilities for the bad grades:

*Sometimes boys think, because they are boys, they are entitled to be bad at English. ... And even if they get bad exam results, they don't feel like making an effort.* [Girl 8, Paragraph 301-304]

This form of self-handicapping, however, was refuted by Boy 5 and 6. They insisted that they managed to continue their study in spite of FALS; it is the constant failure that made them lose interest in English, like any FALS-congruent boy. Boy 5's following quote characterises their struggle to actively manage to learn English despite FALS.

*You (.) take a look at yourself, and the gap between you and girls. And you find that no matter what, there are boys who do well in [English] exams after all, right? So you, you build your confidence, you know. Girls get high scores, but boys, sometimes boys do well, too. So you, you will try and make an effort yourself. ... But after many exams, gradually, the little improvement you get seems marginal. And gradually, you are right back where you started. So you lose interest in it.* [Boy 5, Paragraph 292]

Another boy that managed FALS well, Boy 12, was unique in Study 4. As a high-achieving, counter-stereotypical English learner, he resorted to an unusual GS from traditional Chinese novels to expound his achievement:

*I think, boys who like the Humanities, literature, and words, are just (.) normal. There are boys like this around me. And another point is that, speaking of the talented, the literary, I think:: there are a lot of males of this kind. I mean that men, 'Wen wu shuang quan'. So 'Wen' is one aspect of [masculinity].* [Boy 12, Paragraph 308]

The expression, 'Wen wu shuang quan', 'be adept with both the pen and the sword' in English, is a Chinese modifier for a type of ideal masculinity typically found in classic Chinese romance (Song, 2004). As far as Boy 12 understood it, his success in English conformed to the 'Wen' aspect, which referred to being well versed in literature. It was unlikely that this expression simply sprung to his mind during the interviews: in Section 7.2,

Kun, a novice female teacher characterised a similarly high-performing boy as ‘exception’, ‘astonishing’, and ‘peculiar’. Perhaps Boy 12 was so familiar with such narratives that he had to find a suitable model of masculinity to fit in. Additional evidence attests to Boy 12’s readiness to defend himself is in previous literature. Carr and Pauwels (2006) interviewed Australian boys doing elective foreign language courses, and these boys recounted being laughed at or even teased by their FALS-endorsing male peers. Boy 12 might also have suffered from such peer pressures and had to explain away his counter-stereotypical achievement with this literary representation of masculinity.

The final category of behavioural response was to apply FALS to others, which was adopted by Girl 2 alone:

*I’ve got a good friend who is terrible at Chinese and English. This is a girl. And she’s really good at the science subjects. Yeah. So, this theory [FALS], every time she hears it, she feels uneasy. ((She giggles.)) Sometimes I just tease her, ‘why don’t you go ahead to be a boy?’ And she just feels uncomfortable. [Girl 2, Paragraph 395]*

All in all, the eight categories of responses demonstrated how FALS shaped students’ emergent attitudes toward *gender*: some conformed to existing gender norms, and some critically examined them.

### **8.2.5 Beyond Female-Advantage-in-Languages Stereotype**

Three additional sets of GSs were mentioned by students in Study 4—the quantitative-orientated boys and the qualitative-orientated girls stereotypes, the slack boys and diligent girls stereotypes, and the career-driven males and home-bound females stereotypes. It is no coincidence that the first two sets of stereotypes also emerged among teachers in Study 2 (Sections 7.3.2-4). After all, they are all cultural GSs. What distinguishes them from FALS is students’ sweeping endorsement of them. As previous sections have indicated, some students were able to critically ponder the validity of FALS. But throughout the interviews, students just accepted the other two sets of academic GSs. For example, Boy 5, when recalling the source of FALS, immediately thought of the quantitative/qualitative divide in Chinese senior high schools:

*[FALS] is also a form of self-recognition. Sometimes, you hear things in schools. ... For example, boys, boys, in general, when you see a boy, you just feel that he should be taking the sciences branch, right? And when you see a*

*girl, you just think that girls ought to take the Humanities branch. That's the way it is. [Boy 5, Paragraph 275-278]*

In fact, all eight groups of interviews digressed from FALS to the quantitative-orientated boys and the qualitative-orientated girls stereotypes at some point, which further illustrated their prevalence and power. Additionally, not only did they overshadow FALS (Section 8.2.3), they also overbore gender norms: according to some interviewees, boys and girls in counter-stereotypic domain would pick up traits from the opposite sex. The following quotes shows such thinking.

*Um, there are boys who do well in English. But, such boys, ((he clicks his tongue)), mostly they are that sort of boys. They are attentive, and, how do I put it, effeminate. [Boy 7, Paragraph 37]*

*Learning quantitative subjects limit girls' expressiveness. I think that, girl in the Humanities branch and those in the Sciences branch, if (.) they both leave school after classes, you will feel that they have different dispositions. [Girl 8, Paragraph 352]*

The slack boys and diligent girls stereotypes were mentioned by Girl 4, 5, and 6 during their extended discussion about gender differences in affect for English (see Section 8.2.2.1). In that conversation, all interlocutors gave total and uncritical acceptance of the stereotypes, just like the case with the quantitative-orientated boys and the qualitative-orientated girls stereotypes above.

The career-driven males and home-bound females stereotypes were raised by Girl 1 and Girl 2, which was causing confusion for Girl 1.

*Girl 2: ... [I've heard people] say that females should just, should just stay at home. And in terms of career, males should be outstanding in that aspect. ...*

*...*

*Girl 1: I don't—I can't say that it's wrong ((refers to the saying Girl 2 gave)), but it's existence sometimes make me feel bad. ... Take my mom as an example. She is perhaps quite traditional. Although she is successful with regards to her career, she still puts family first. For instance, if my parents are required to work extra hours, my dad will just go, but my mom will have to go back home. This has always been the way. And my mom keeps lecturing on me, 'women*

*need to be gentle. They have to take their responsibilities as a mother at home.’ ... I mean, on the one hand, females have to value families, be considerate, stay at home, and be gentle. But on the other hand, ... it is obvious to me that if I get interrupted from expressing who I am, it will be difficult for me to grow as a person. So, this conflict between individuality and family, makes me feel a bit painful. [Girl 1, Girl 2, Paragraph 416-425]*

Clearly, for Girl 1, her need to be true to herself and restrictions laid out by circulating GSs caused her perplexity and pain. Note that she actually felt the need to identify with boys as an English learner earlier in Section 8.2.4. For both instances, Girl 1’s emergent understanding of *gender* was featured by a sense of disorientation.

### **8.2.6 Students’ Emergent Attitudes towards *Gender*: Conformity, Confusion, and Critical thinking**

To sum up, three superordinate themes emerged from the thematic analysis in Study 4—the widespread FALS; 2) approaching FALS critically; and 3) GSs accompanying FALS. Figure 8.4 depicts how these themes unite under the overarching theme: students’ emergent understanding of *gender*. During the interviews, all but one were aware of FALS, and all endorsed at least one component of FALS. When being judged, or treated on the basis of FALS, some internalised FALS, incorporating it with their self-beliefs and behaved accordingly. Some others examined its validity critically. That is, some students conformed to FALS’s prescription of gender norms, while others challenged them. However, in the case of two additional pairs of academic GSs, the tendency to critically examine them vanished, and their overbearing power affected interviewees’ understanding of *gender* in stereotype-congruent ways. Finally, the career-driven males and home-bound females stereotypes caused confusion for one girl, who struggled to form her sense of self. Apparently, GSs in the environment were shaping their emergent understanding of *gender*, as well as their gender identities (see Section 9.1.4.3 for a discussion).

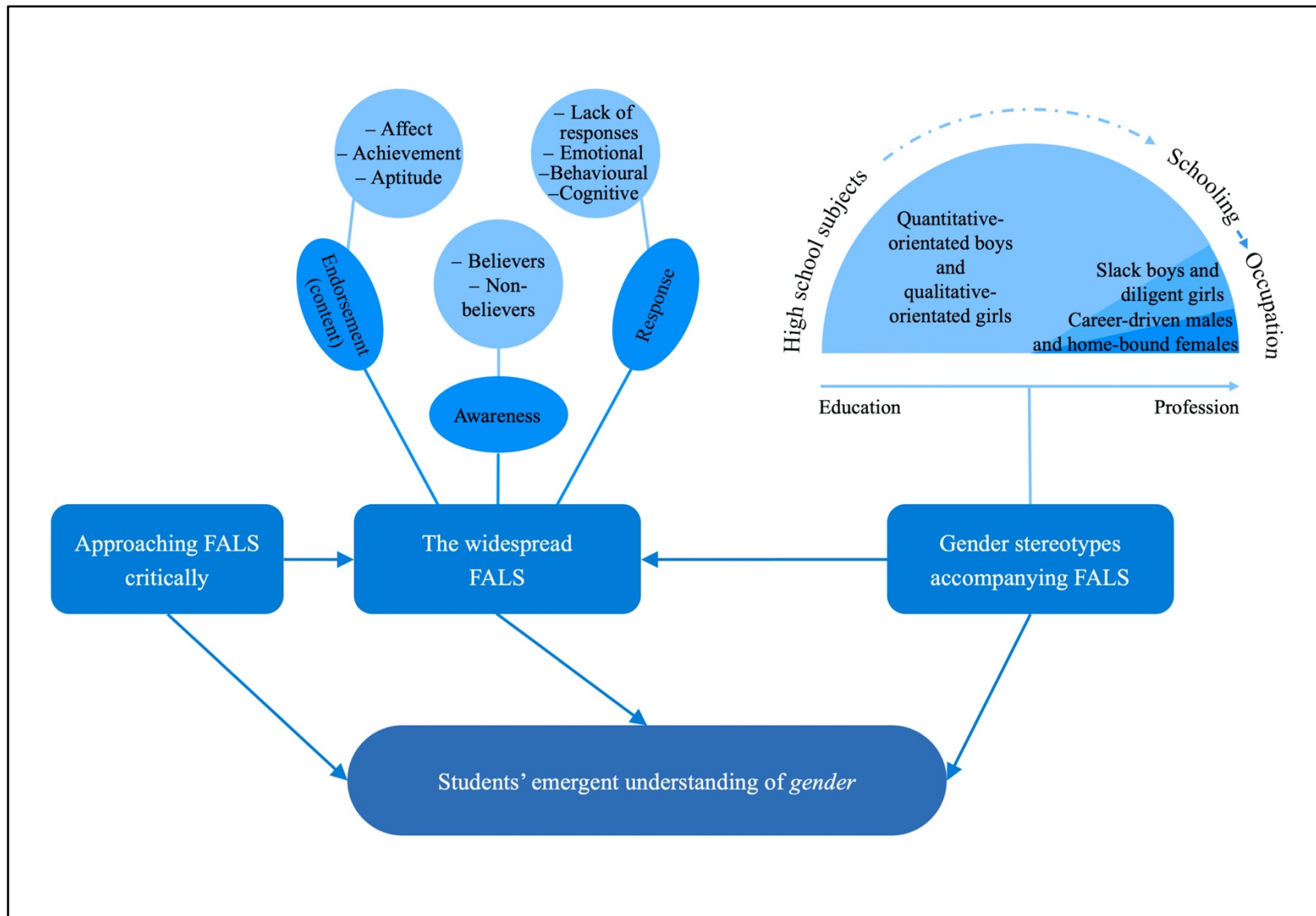
### **8.3 Integrated Findings from Studies 3 and 4**

In Study 3, it was found that boys and girls were affected differently by FALS. While girls’ test performance was somewhat boosted after having watched a video reminding them of FALS, boys under the same circumstance displayed a substantial decrease in performance compared to their peers in the control group. Some might take the former as evidence of girls benefiting from FALS. In addition, four girls in Study 4 did report that they felt obliged to learn English better because FALS prescribed so. However, the extent of this motivating

effect, even if it does translate into performance enhancement, does not seem impressive at all: in Study 3, girls in the experimental group only received an average of .80 mark more than those in the control group. For a test whose full mark is 120, such gain in performance appears trivial. In fact, it was found in Study 4 that FALS had negatively influenced girls' self-concept and emotional wellbeing. Due to FALS, girls were made to question their gender identity (e.g., Girl 1), or to feel uneasy or hurt (e.g., Girl 12).

In the case of boys, the negative effects encompassed cognitive, emotional, and behavioural aspects: boys internalised FALS's portrayal of them as remedial English learners (e.g., Boy 4), felt upset due to FALS (e.g., Boy 9), and reduced their effort to learn (e.g., Boy 6). These findings in Study 4 also explained how having been primed of FALS prior to OOPT could have led to such a considerable decrease in test results: during the test, boys had to exert extra effort into suppressing troubling thoughts and regulating negative emotions while keeping on track of the task at hand. These mechanisms of ST have been proposed and examined by previous studies (e.g., Inzlicht, McKay and Aronson, 2006; John, Inzlicht and Schmader, 2008). Thus, both Studies 3 and 4 pointed out that FALS had adversely influenced boys and girls, though via different manners.

Figure 8.4: Summary of Themes Emerged from Thematic Analysis in Study 2



## **Chapter 9: General Discussion**

The aim of the current chapter is to contextualise, evaluate, and interpret the results and findings reported in Chapters 5-8. Section 9.1 will draw on findings from Phase 1 (presented in Chapters 5, 6, and 7) to address the endorsement of the female-advantage-in-languages stereotype (FALS). Then, Section 9.2 will rely on results from Phase 2 (documented in Chapter 8) to ponder the effects of FALS. In both sections, inferences will be drawn from the results, and similarities and differences between the findings and pertinent previous work will also be identified.

### **9.1 Endorsement of Female-Advantage-in-Languages Stereotype**

The primary goal of the current project was to detect the existence of FALS, a tripartite gender stereotype (GS) depicting females as more gifted, more enthusiastic, and higher-achieving language learners than their male counterparts. To achieve this aim, two studies were conducted simultaneously in Phase 1: Study 1, a questionnaire survey among high school students, their guardians, and teachers of English, and Study 2, individual face-to-face interviews with teachers. Additionally, Study 4 (group interviews with students) also managed to explore students' endorsement of FALS, although it was carried out principally to understand learners' responses to FALS.

Section 9.1.1 will contextualise FALS unearthed in Studies 1 and 2 with regards to the work of others. Subsequently, Section 9.1.2 will evaluate the accuracy of each FALS component in reference to gender differences in language aptitude, affect, and achievement. Then, Sections 9.1.3 and 9.1.4 will consider three convergences and five divergences among three groups of participants by integrating pertinent findings about their FALS endorsement and gender attitudes in Studies 1, 2, and 4. Finally, Section 9.1.5 will consider gender stereotyping in English classrooms with more emphasis on GSs accompanying FALS that emerged from Studies 2 and 4.

#### **9.1.1 Contextualising Female-Advantage-in-Languages Stereotype**

The results from three groups of participants in Study 1 are summarised in Table 9.1. The multivariate effect sizes are comparable to those reported in previous studies on academic GSs, as Table 9.2 here summarises. For example, Plante et al. (2009), utilising multiple-item scales, assessed the female-language stereotype to range between .68 and .97 (Cohen's *ds*). These effect size indicators share a substantially similar range to the

**Table 9.1: Female-Advantage-in-Languages Stereotypes in Guardian, Student, and Teacher Datasets**

		Guardian ( <i>n</i> = 1,904)		Student ( <i>n</i> = 959)		Teacher ( <i>n</i> = 60)			
Multivariate results	<i>F</i> (df <sub>n</sub> , df <sub>d</sub> )		576.358 (3, 1795)		222.598 (3, 726)		19.856 (3, 50)		
	<i>p</i> -value		< .001		< .001		< .001		
	Test statistic		Wilks' Λ= .509		Pillai's Trace = .479		Wilks' Λ = .456		
	Effect size	η <sup>2</sup> <sub>p</sub>	.491		.479		.544		
		Cohen's <i>d</i>	.816 , 95% CI [ .788 ,  .845 ]		.729 , 95% CI [ .684 ,  .773 ]		.973 , 95% CI [ .817 ,  1.129 ]		
Univariate results	Achievement	<i>F</i> (df <sub>n</sub> , df <sub>d</sub> )		1262.130 (1, 1797)		515.890 (1, 728)		47.311 (1, 52)	
		<i>p</i> -value		< .001		< .001		< .001	
		Effect size	η <sup>2</sup> <sub>p</sub>	.413		.415		.476	
			Cohen's <i>d</i>	.840 , 95% CI [ .790 ,  .891 ]		.817 , 95% CI [ .741 ,  .892 ]		.968 , 95% CI [ .697 ,  1.238 ]	
	Aptitude	<i>F</i> (df <sub>n</sub> , df <sub>d</sub> )		1124.572 (1, 1797)		216.090 (1, 728)		39.452 (1, 52)	
		<i>p</i> -value		< .001		< .001		< .001	
		Effect size	η <sup>2</sup> <sub>p</sub>	.385		.229		.431	
			Cohen's <i>d</i>	.783 , 95% CI [ .733 ,  .833 ]		.554 , 95% CI [ .473 ,  .635 ]		.843 , 95% CI [ .572 ,  1.113 ]	
	Affect	<i>F</i> (df <sub>n</sub> , df <sub>d</sub> )		1206.583 (1, 1797)		518.454 (1, 728)		58.371 (1, 52)	
		<i>p</i> -value		< .001		< .001		< .001	
		Effect size	η <sup>2</sup> <sub>p</sub>	.402		.416		.529	
			Cohen's <i>d</i>	.819 , 95% CI [ .768 ,  .869 ]		.854 , 95% CI [ .780 ,  .928 ]		.982 , 95% CI [ .711 ,  1.253 ]	

*Notes.* All Cohen's *ds* in the table were calculated using means, standard deviations, and correlations, based on the procedure suggested by Morris and DeShon (2008, p.111). All Cohen's *ds* are positive, indicating a higher rating for females than males.



multivariate Cohen's  $d$ s exhibited in Table 9.1. Moreover, Retelsdorf and colleagues (2015) estimated the gender-reading stereotype harboured by language teachers in Germany at Cohen's  $d = 1.52$ , one of the largest effect sizes witnessed in literature on academic GSs, which is close to the upper limit of the 95% confidence intervals of teachers' FALS (1.129). Additionally, Nosek and colleagues (2002) measured the magnitude of the stereotypical connection between gender and domain (male-science/female-liberal arts) to be Cohen's  $d$ s = .73 and .72, using a single-item explicit scale and Implicit Association Test (IAT). These effect sizes correspond to the amplitude of FALS uncovered among students (Cohen's  $d = .729$ ), but smaller than those unearthed from guardians (.816) or teachers (.973).

**Table 9.2: Summary of Effect Sizes of Language-Related Gender Stereotypes in Previous Work**

Study and gender-domain stereotype	Participants	Cohen's $d$
Nosek et al. (2002)		
Gender-domain (IAT)	Non-preselected or targeted	.72
Gender-domain (explicit)	Non-preselected or targeted	.73
Heyman and Legare (2004)		
Gender-spelling	5- to 6-year old boys	– .29
Gender-spelling	9- to 10-year old boys	.18
Gender-spelling	5- to 6-year old girls	.47
Gender-spelling	9- to 10-year old girls	.35
Plante et al. (2009)		
Female-language	11- to 12-year old boys	.70
Female-language	13- to 14-year old boys	.68
Female-language	15- to 16-year old boys	.87
Female-language	11- to 12-year old girls	.90
Female-language	13- to 14-year old girls	.86
Female-language	15- to 16-year old girls	.97
Epting et al. (2014)		
Masculinity/femininity-reading	Full-time undergraduates	.20
Masculinity/femininity-writing	Full-time undergraduates	.09
Masculinity/femininity-revision	Full-time undergraduates	.27
Masculinity/femininity-grammar	Full-time undergraduates	.22
Boyishness/girlishness-reading	Full-time undergraduates	.41
Boyishness/girlishness-writing	Full-time undergraduates	.37

Boyishness/girlishness-revision	Full-time undergraduates	.37
Boyishness/girlishness-grammar	Full-time undergraduates	.27
Retelsdorf et al. (2015)		
Gender-reading	Language teachers	1.52
Koenig (2018)		
femininity-interest in languages/arts	Undergraduates and non-preselected	.86
Muntoni and Retelsdorf (2019)		
Gender-reading	Parents of boys	1.02
Gender-reading	Parents of girls	1.10

*Notes.* This table only contains pertinent studies that provided necessary descriptive statistics for the computation of Cohen's *ds*.

The discrepancies between GS literature and Study 1 here might have resulted from four sources, among which the two pivotal ones were the instruments used and the constructs measured. Study 1 employed a two-section multiple-item questionnaire where participants rated male and female learners' language aptitude, affect, as well as achievement. The extent to which participants determined boys and girls to be different was construed as the GS regarding English, i.e., FALS, stereotype characterising females more positively than males. Likewise, Plante et al. (2009) utilised multiple-item scales with two parts, one for students to evaluate how masculine they take language studies to be, and another to appraise how feminine it is. They also used the difference between two estimations to represent the GS involving languages. Plante et al. (2009) and Study 1 share their first common ground in scale design: the separate while simultaneous application of masculinity and femininity scales. This feature effectively reduces biases by 1) allowing respondents to identify how much they consider a domain to favour boys or girls and 2) avoiding imposing the "boys vs. girls" comparison upon respondents. The two studies are also analogous in that they both conceptualise the gender-language stereotype as multi-component constructs. Perhaps it is due to the two resemblances that the effect sizes in these works largely overlap.

The same logic partly explains the disparities in findings between Study 1 and Nosek and colleagues' paper (2002). The researchers relied on IAT and a single-item questionnaire, neither of which comes close to the instrument in Study 1. The two studies also diverge in terms of the GSs they investigate. Unlike Study 1, Nosek and co-researchers did not approach a stereotype between gender and one specific subject. Instead, they aimed to understand simultaneously the male stereotyping of science and the female stereotyping of liberal arts. In

their study, science included a total of nine subjects: chemistry, biology, physics, and etc. For liberal arts, there were eight fields, such as philosophy, history, Spanish, and so on. Given previous research revealing considerable variations in the amplitude of gender stereotyping from subject to subject (Tenenbaum and Leaper, 2003), Study 1, which focused on female-English stereotype, would certainly have yielded different results from Nosek et al. (2002).

Furthermore, two more reasons help clarify the different findings from Study 1 and Nosek et al. (2002): participants and cultures. Study 1 took place in China, where English is taught as a foreign language and where academic GSs were endorsed to a stronger than average level among 34 participating countries/regions in another study (Nosek et al., 2009). On the other hand, Nosek et al.'s study (2002) involved non-preselect, non-targeted adult native English speakers in the United States. Of course, the contrasts in participant and culture also exist for Plante and colleagues' paper (2009), which recruited French-speaking 11- to 16-year-olds in Canada, and for Retelsdorf et al. (2015), which invited German language teachers in German secondary schools. Due to their divergent experiences and engagement with schooling and language education, the academic GSs they held would certainly have differed. If anything, these studies actually collaborated with one another by providing converging evidence that language studies were stereotypically regarded feminine by diverse age groups in various cultures.

### **9.1.2 Accuracy of Female-Advantage-in-Languages Stereotype**

The extent to which stereotypes reflect reality has always been a hot issue in and beyond the academia, especially with GSs (Kite, Deaux, and Haines, 2008). One means to tackle the question is to compare the magnitude of GS against genuine gender difference in the given dimension. Nevertheless, the following three caveats have been acknowledged and proposed by scholars interpreting sex/gender difference in brains or hormones (e.g., Fine, 2013), cognitive abilities (e.g., Hyde, 2005), and in achievement (e.g., Voyer and Voyer, 2014): first and foremost, differences do not necessarily denote advantages or disadvantages (Halpern, 2012). For example, men and women were found to prefer gender-specific problem-solving strategies (e.g., Liu and Wilson, 2009), but problems can usually be approached by multiple ways. Even if one or more might be more efficient than others, individuals can always hone their skills by learning about multiple strategies. Second, the strengths of gender differences presented in literature typically result from comparing averages of a group of men and a cohort of women, but group averages tell little about individuals of either gender (Priess and Hyde, 2010). In fact, when standard deviations are taken into account, considerable intra-

group variability can arise. Furthermore, when the amplitudes of gender differences (commonly calculated as Cohen's *ds*, which represents differences between group means) are transformed into overlapping coefficient (Reiser and Faraggi, 1999) and probability of superiority (Ruscio, 2008), sizable inter-group overlaps are detected. That is, although females and males *are* different, they are also substantially similar, a feature that can be easily overlooked (Hyde, 2005). Bearing these in mind, in this section, accuracy of each component of FALS will be appraised in relation to gender differences in language aptitude, affect, as well as, achievement.

### 9.1.2.1 Aptitude Component as Exaggeration

Studies examining gender differences in language aptitude, or verbal ability, have yielded fascinating yet slightly confounding results (Priess and Hyde, 2010; see Table 3.1 for a summary). Despite some slight inconsistencies, the general trend discernible from Table 3.1 is a female lead ranging between Cohen's *ds* =  $|.02|$  and  $|.45|$ , which is small to medium (Cohen, 1988). The largest effect size,  $|.45|$ , means an overlap of 82.2% between males' and females' ability distributions (calculated using the formula in Reiser and Faraggi, 1999). In addition, it also signifies that the likelihood of a girl having a stronger language aptitude than a boy, both selected randomly from their groups, is only 62.5% (arrived at based on the formula raised by Ruscio, 1999). The overlap and probability of superiority for a Cohen's *d* of  $|.02|$ , instead, are 92.0% and 55.6%. Therefore, although females may have a stronger linguistic competence, the difference is minor in practical terms.

However, the participants in Study 1 appeared to have stereotypically overestimated this gender difference, as indicated by the sizes of the aptitude component of FALS. As showcased in Table 9.3, the overlaps between boys and girls are lower as perceived by participants in Study 1, but the probabilities that girls exceed boys by chance are higher. Therefore, the aptitude component of FALS would appear to have exaggerated the genuine difference between male and female's verbal ability, especially from the perspectives of guardians and teachers.

**Table 9.3: Practical Interpretations of Aptitude Component Measured in Study 1**

Participant	Cohen's <i>d</i> , 95% CI	Overlapping coefficient	Probability of superiority
Guardians	.783, [.733, .833]	69.5%	71.0%
Students	.554, [.473, .635]	78.2%	65.2%
Teachers	.843, [.572, 1.113]	67.3%	72.4%

Critical readers might challenge the above argument on the ground of limitations of the cited ability tests in Table 3.1. They are right to assume that these tests inevitably measure achievement, as well as ability. Besides, it is also reasonable to raise doubts about the comparability of these tests and FALS in the current project, which looks into stereotypical gender differences in English learners in China. Still, considering that psychologists have always struggled to separate ability from achievement when devising ability tests (Halpern, 2012), the cited references are thought to be reasonably accurate approximants of gender differences in linguistic competence. In addition, due to the unavailability of relevant statistics in the Chinese context, multiple sources from various age groups and cultural background are relied on when compiling Table 3.1 to minimise bias and error.

### 9.1.2.2 Achievement Component as Self-Fulfilling Prophecy

Table 9.4 exhibits descriptive data from academic works comparing male and female test-takers' scores in English exams in mainland China. In all these published studies, girls outperform boys. A comparison between the Cohen's *ds* in Study 1 (see Table 9.1) and the ones in Table 9.4 will illustrate that the achievement component of FALS is also an exaggeration of the extant gender difference in English attainment: while the magnitude of the achievement component ranges from .817 to .968, the sizes of genuine gender performance gaps are no larger than .66. The achievement component is correct in predicting the direction of gender difference though, because all empirical studies reported higher performance among girls.

**Table 9.4: Summary of Research on Gender Differences in English in China**

Data source, test and participants	Male students		Female students		Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Liu, 2003					
Mid-term exam, Grade 10	331.9	66.25	367.6	57.34	.58
Gan, 2006					
High School Entrance Exam, Grade 9	107.36	8.912	110.21	7.863	.34
Mid-term exam, Grade 10	78.57	15.199	86.45	12.712	.56
China Education Panel Survey, 2013 (representative of the population of middle-schoolers)					
Final exam, Grades 7 and 9	72.86	31.39	86.20	27.53	.45
Huang, 2011					
CET-4, undergraduates	365.22	41.52	387.23	42.99	.52
Wei, Zuo, Chen, and Yang, 2015					

CET-4, undergraduates	353.43	58.12	387.79	64.82	.56
Du and Wang, 2017					
Model test of CET-4, undergraduates	39.05	9.22	45.24	9.52	.66

*Notes.* CET-4 refers to College English Test (Band 4), which is the national test of English proficiency for undergraduates in mainland China. Positive Cohen's *ds* here mean that girls outperform boys in a particular test.

When the effect size indicators (Cohen's *ds*) found from Study 1 are translated into common-language effect sizes (as shown in Table 9.5), the exaggeration of the achievement component becomes even more apparent. The largest gender difference in Table 9.4, reported by Du and Wang (2017), means that in reality, 74.1% of males' and females' achievement distributions overlap, and there is only a 68.0% chance that a random girl may outperform a random boy. Yet, participants in Study 1 stereotypically thought of boys' and girls' achievement as much less similar (with less than 70% overlapping coefficients, see the third column in Table 9.5). They also amplified the likelihood that a girl beats a boy in English.

**Table 9.5: Practical Interpretations of Achievement Component Measured in Study 1**

Participant	Cohen's <i>d</i> , 95% CI	Overlapping coefficient	Probability of superiority
Guardians	.840, [.790, .891]	67.4%	72.4%
Students	.817, [.741, .892]	68.3%	71.8%
Teachers	.968, [.697, 1.238]	62.8%	75.3%

Some audiences might refuse to take the achievement component as too much of a misrepresentation of genuine gender difference. After all, the overlapping coefficients and probabilities of superiority are not substantially different between FALS achievement and the gender performance gap identified by Du and Wang (2017). Their assertion is reasonable, but it does not warrant any disregard of the achievement aspect of FALS. A careful examination of Table 9.4 discloses an upward trend of gender achievement gap as students move from early adolescence to young adulthood. Among the middle-schoolers (aged between 12 and 15) investigated by Gan (2006) and in China Education Panel Survey (CEPS, 2013), the gender disparities are Cohen's *ds* = .34 and .45. But in the remaining studies, whose participants are either in senior high schools or universities, the effect sizes inflate. In fact, the same tendency for initially small gender differences in school to develop into increasingly apparent ones as boys and girls move along their educational trajectories also exists in other cultures (e.g., Evans, Schweingruber, and Stevenson, 2002; Retelsdorf, Schwartz, and Asbrock, 2015). As Kollmayer et al. (2018) have convincingly argued, the widening gender

performance gap can be traced to the gender-stereotypical expectations from significant others, including teachers and parents. When such GSs are internalised by students, they become self-fulfilling prophecies that lead students to behave stereotype-congruently, enlarging the originally marginal gender disparities in learning outcomes (Jussim, Eccles, and Mardon, 1996).

### **9.1.2.3 Affect Component as Over-Simplification**

Empirical studies attest to girls' stronger motivation (e.g., Williams, Burden, and Lanvers, 2002) and more pleasure (e.g., Dewaele, MacIntyre, Boudreau, and Dewaele, 2016) regarding foreign languages. Thus, there seems to be a kernel of truth underlying the affect component of FALS, which characterises girls' affective relationship with English more positively than that of boys. Yet, such findings might not necessarily apply to Chinese students: in a study examining positive emotions among English learners in a Chinese university, no statistically significant gender differences were detected (Jiang and Dewaele, 2019). Likewise, Shen investigated the motivations for learning English among Taiwanese secondary school students (2005), which also yielded non-statistically significant results. Nevertheless, female learners in both studies did rate themselves higher than their male counterparts with regards to enjoyment and motivation. The failure to reject the null hypothesis that no gender differences existed, therefore, can be rooted in lack of statistical power: the female learners' stronger motivation prevailing in China might be too trivial to capture in neither studies.

The duality of learners' affective experiences with language courses further complicates the picture: the abovementioned positive emotions constitute only one facet the issue; in fact, a much larger body of research is devoted to negative emotions, with anxiety at the centre stage (e.g., Bailey, Daley, and Onwuegbuzie, 1999; Hu, 2003). Girls, according to this line of inquiry, tend to worry more over their mistakes (Dewaele et al., 2016), fear more of negative feedbacks (Abu-Rabia, 2010), and suffer from more severe communication apprehension (Liu, 2006) than boys. This gendered pattern, unlike the case with positive emotions earlier, is also discernible in Chinese societies (e.g., Liu, 2006; Jiang and Dewaele, 2019). Additionally, female participants in Study 4 also stated their mixed feelings towards English (Section 8.2.1): while they claimed to be fond to English, they also recounted instances where English brought upon stresses and distresses among them. Thus, not only does the affect component of FALS dramatise the amplitude of an essentially imperceptible gender difference in FL

enjoyment and motivation, it also masks the fact that girls suffer from more negative emotions in language classrooms.

#### **9.1.2.4 Summary: Female-Advantage-in-Languages Stereotype and Real Gender Differences**

To understand the extent to which FALS mirrors reality, each FALS component was compared against authentic gender differences in language aptitude, performance, and affect, respectively. According to research assessing individuals' verbal ability, females generally tend to have a stronger linguistic competence than males, but the female edge is minor in practical terms. Since the effects size of the aptitude component of FALS as endorsed by students, their guardians, and teachers of English are larger than the genuine gender difference in language aptitude, the aptitude aspect seem to be an exaggeration. The achievement dimension of FALS, likewise, distorts the real-world gender performance gap to a more extreme version. In addition, its potential as self-fulfilling prophecies is also worth pondering. Finally, the affect component of FALS oversimplifies learners' two-fold feelings towards languages by only conveying the positive side of the story. All in all, FALS appears to contain some element of truth, but the way it overestimates some gender difference while ignoring others still calls for attention from language educators and policy-makers.

#### **9.1.3 Three Convergences among Guardians, Students, and Teachers**

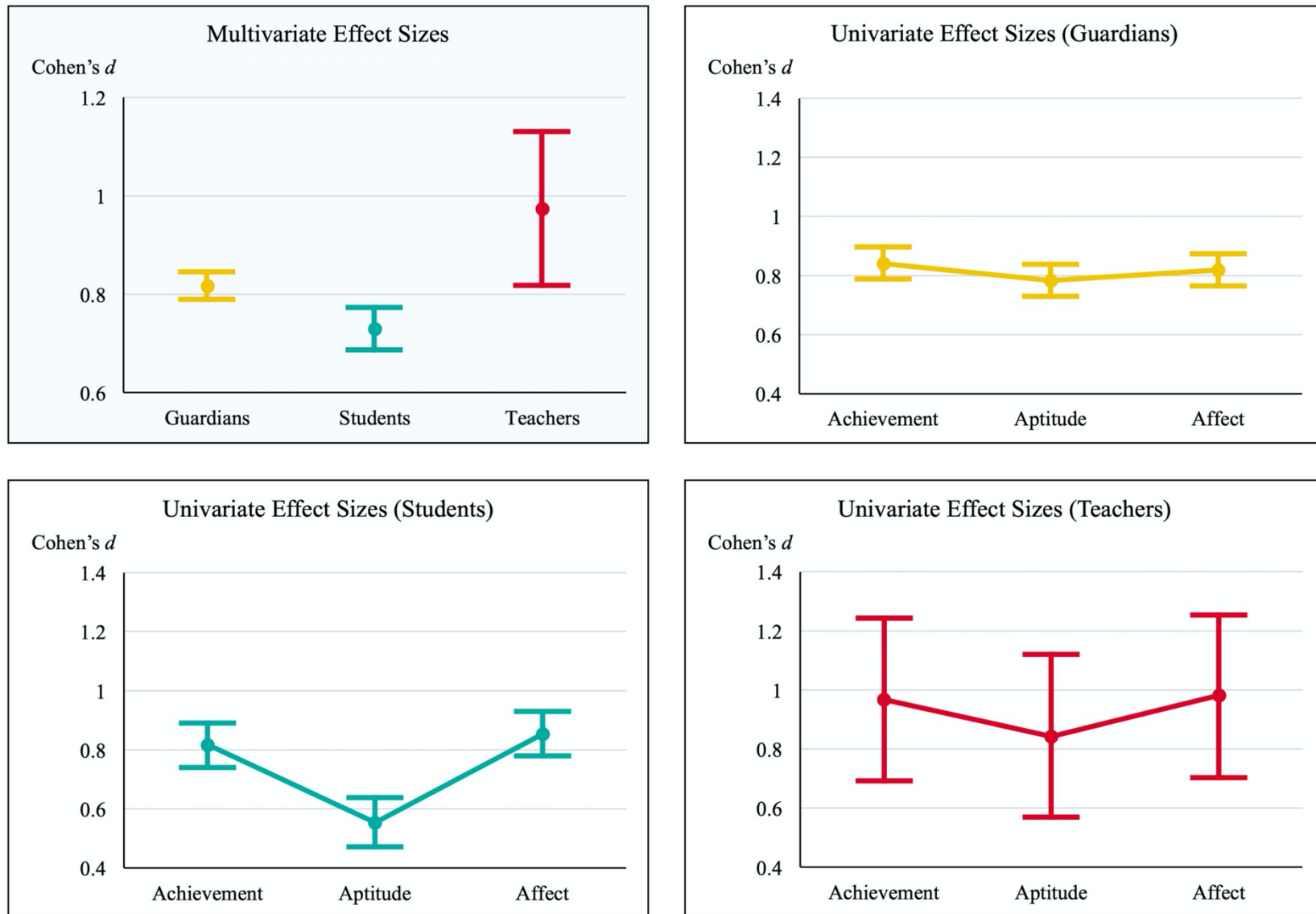
In Studies 1, 2 and 4, agreement among groups of participants was discovered in three aspects. The first consensus is that out of three FALS components, aptitude was the weakest and least widespread one. Another agreement concerns variabilities in FALS endorsement, as manifested by the large standard deviations in all three groups. Finally, teachers and students both perceived a more intense stereotypical additive connection between affect and achievement than that between aptitude and achievement. These convergences will be clarified respectively from Sections 9.1.3.1-3.

##### **9.1.3.1 Aptitude as Least Endorsed Component in All Three Groups**

Figure 9.1 compares the FALS components across datasets in Study 2, which reveals the first agreement among all three groups of participants—the aptitude seems to be the least endorsed dimension by all. In addition, it was held by the least number of teachers (11 out of 20), as well as students (8 out of 24), in Studies 2 and 4. Such a finding has theoretical and practical implications for scholarly examination of academic GSs, which over-emphasises competence-based stereotypes, as analysed in Section 3.2.2. Because the overlooked affect and achievement components, as found in Study 1, had larger amplitudes, the role of gender



**Figure 9.1: Comparison of FALS Magnitudes across Guardians, Students (Macro Level), and Teachers**



stereotyping in language classrooms can easily be underestimated if the foreground is taken by competence-based stereotypes.

Cautious audiences might doubt whether this finding can be generalised to other cultures and societies. Indeed, under Confucian influence, learners were more likely to attribute performance to effort, instead of talent, which contrasts with those in Western societies (Shen, 2005). Furthermore, as Section 2.1.2 has outlined, the traditional Chinese construction of gender stresses not the essential different natures between men and women, but their hierarchical social statuses and relative roles in public and private lives. This propensity to downplay aptitude, therefore, might have been culture-specific. However, such tendency can transcend cultural borders for two reasons. First, language domains are typically thought to rely heavily on effort and sustained attention, unlike math or science subjects which may require certain cognitive ability (e.g., Lummis and Stevenson, 1990; Swinton, 2012). In addition, because the cross-cultural male-competence stereotype emerged from Study 2, it is also likely to co-exist with FALS in cultures other than Chinese. Its prevalence, as a result, might also reduce the strength of FALS beyond the Chinese society. To sum up, there are two reasons suggesting that the small magnitude of aptitude component is specific to Chinese population, but two other causes point out that the finding might potentially be generalisable to other societies. Thus, transnational studies are needed in future to settle the issue here.

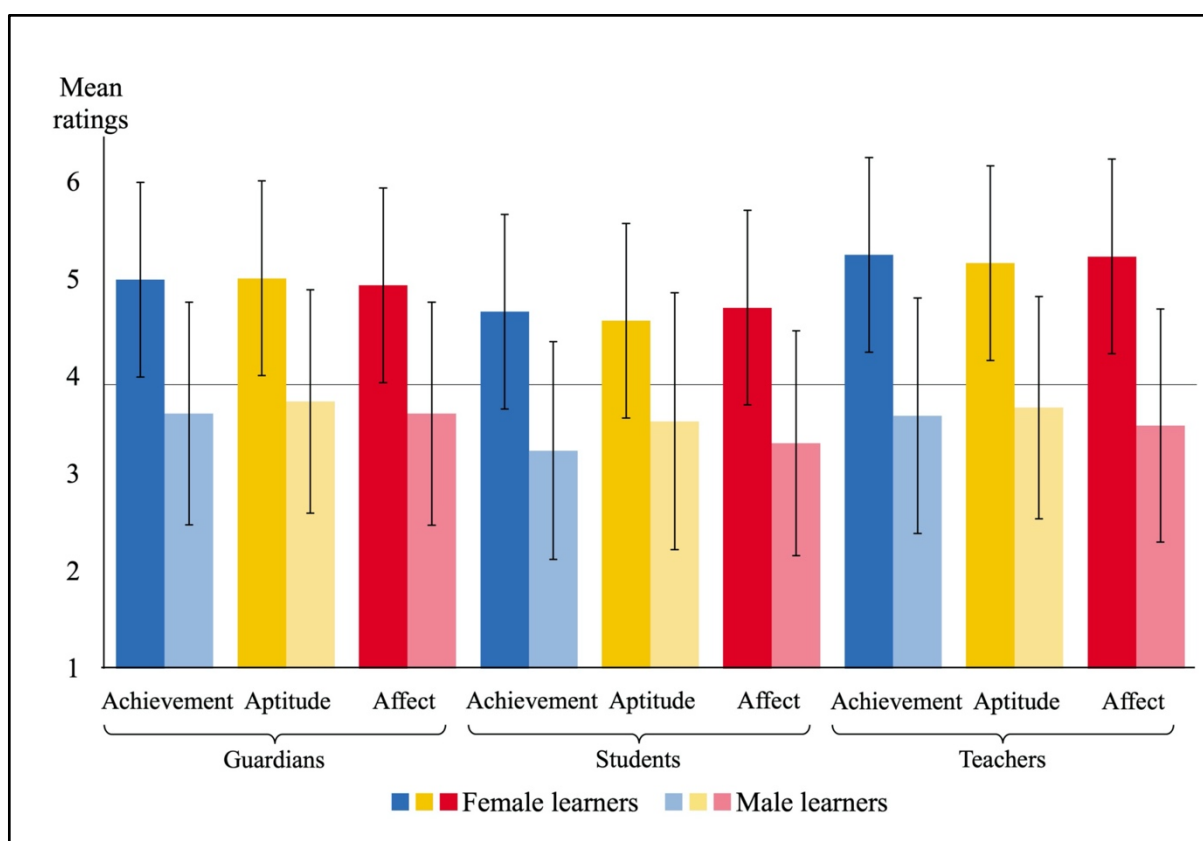
#### **9.1.3.2 Variabilities within Each Group**

The second similarity among three groups of participants in Study 1 is the variabilities in their gender-stereotypical perspectives. In Study 1, participants indicated their stereotypical images of boys and girls on scales from one to seven, where one, four and seven each represented extreme negative, neutral, and extreme positive attitudes. As illustrated in Figure 9.2, students, guardians, and teachers all rated boys unfavourably (below four) and girls favourably (above 4). Yet, all ratings have standard deviations larger than one (specific statistics can be found in Table 5.1 for guardian dataset, Tables 5.3 and 5.5 for student dataset, and Table 6.2 for teacher dataset), suggesting considerable heterogeneity in stereotypical beliefs within each dataset. Due to lack of space in the thesis, the variation in stereotype endorsement is not examined directly using questionnaire data. The research intends to explore this phenomenon when preparing the thesis findings for journal publication in future.

Nevertheless, Study 2 did manage to uncover some idiosyncrasies in gender-stereotypical attitudes, including those rejecting FALS (e.g., Jian and Yuehan, see Section

6.2.4) or reluctant to overgeneralise FALS to all learners (see Section 7.4). In addition, Studies 2 and 4 provide evidence that sometimes, FALS-believers might only endorse one or two FALS components. All these findings from the qualitative strand support the observation in Study 1 that FALS endorsement can be heterogeneous.

**Figure 9.2: Mean Ratings for Male and Female Learners by Guardians, Students, and Teachers**



### 9.1.3.3 Stereotypical Causal Connections between Components among Teachers and Students

In both the teacher and student datasets, a stronger stereotypically additive relationship between affect and achievement components was found, compared to that between aptitude and achievement components. This converging finding further highlights the difficulty of challenging FALS, as its components are causally linked by its believers, making the tripartite FALS logically self-consistent. To tackle FALS, therefore, a synergy of different interventions targeting each component is in need. For example, informing teachers and students the triviality of the real sex difference in linguistic capacity can strategically minimise or even eliminate their endorsement of the aptitude component. Additionally, to combat the achievement and affects components, teachers and student will benefit from

understanding the self-fulfilling powers of GSs, such as how GSs lower students' interest, engagement, and performance in counter-stereotypical domains (Kollmayer et al., 2008).

This consensus might arise from the tendency to downplay the role of aptitude regarding a) language domains and b) Chinese society (see the last paragraph in Section 9.1.3.1). Besides, it might reflect the influence of teachers' gender attitudes on students, as Section 9.1.4.3 will later entertain.

#### **9.1.4 Five Divergences among Guardians, Students, and Teachers**

In Studies 1, 2 and 4, five pieces of disagreement among groups of participants were discovered. The first difference is the magnitude of FALS endorsement, which will be pondered in Section 9.1.4.1. Then, Section 9.1.4.2 will continue to another source of disagreement regarding FALS: which component is perceived strongest by each group of participants. Furthermore, disparities between teachers and students regarding their gender attitudes will be delineated in Section 9.1.4.3. Finally, disparities in stereotype endorsement regarding participant gender and region will be considered in Sections 9.1.4.4-5.

##### **9.1.4.1 Disparities in Magnitudes of Female-Advantage-in-Languages Stereotype across Three Groups**

In Study 1, teachers had the strongest stereotype, compared to guardians or students, which can be understood through the lens of the reality principle (see Section 3.1.1). This principle proposes that individuals construct stereotypical images of social groups via their contact and experiences with the social world (Oakes, Haslam, and Turner, 1994). That is, stereotypes reflect, instead of distort, reality. Given that consistent and reliable gender disparities favouring girls have been witnessed regarding achievement and participation rate in language domains in China (e.g., Liu and Li, 2011; Ma, 2001; Wen, 2005; Zhang, 2016), it seems reasonable that teachers, the devoted practitioners with extended engagement with learners, would endorse FALS to a larger extent than the other two groups. A similar case could be argued, of course, for students, who, curiously, had the weakest FALS. This seeming oddity can be explained by the generation gap in stereotype endorsement (see the following paragraph) and differences between teachers' and students' gender attitudes (see Section 9.1.4.3 later).

The disparities in FALS strengths indicates a generation gap in endorsement. In Section 5.3, it was discovered through MANOVA procedures that guardians, the older generation, had a stronger gender stereotype about aptitude than students, the younger generation. This gap might as well have existed between students and teachers, too, since students had a weaker

commitment to FALS than teachers, as illustrated by the multivariate effect sizes exhibited in Table 9.1. When the same gap has been identified by some studies inspecting gender-stereotypical beliefs about math (e.g. Hyde, Fennema, Ryan, Frost, and Hopp, 1990; Plante, Theoret, and Favreau, 2009), it is interpreted as a result of efforts to promote gender-egalitarian thinking and increase gender fairness in education (e.g., Kollmayer et al., 2018). Section 9.1.4.3, however, will offer an alternative understanding.

This gap, as it happens, manifests itself only in magnitude, not in scope: in Study 4, FALS was endorsed by 23 out of 24 interviewees (95.8%), that same percentage uncovered from teachers is 90% in Study 2. The reason for students' weaker yet widespread FALS also lies with the divergence between teachers' and students' gender attitudes, as Section 9.1.4.3 will unfold.

#### **9.1.4.2 Disagreement over Strongest Component across Three Groups**

Table 9.1 shows that among students and teachers, affect is the strongest component with the largest effect size. Guardians, on the contrary, believed in the achievement component most adamantly. Comparing results from Studies 2 and 4, another disagreement between teachers and students also appears: in Study 2, it was achievement that was most popular (18 teachers), but in Study 4, affect was the most widespread component among students (23 believers). This divergence among three groups of participants features the need to study stereotype endorsement of different groups simultaneously: understanding the plausible disagreement can help devise more effective, strategic interventions to combat gender-stereotypical beliefs among various groups. Given students' inclination towards the affect component, promoting the educational and professional affordances of multilingual competence might refute the stereotypical portrayal of languages as a feminine pursuit. Considering that guardians might have lower expectations of their sons due to their more rigid endorsement of the achievement component, awareness-raising initiatives should be devised to sensitise guardians of their own gender-stereotypical beliefs and the effects they might exert on students.

#### **9.1.4.3 Discrepancies in Gender Attitude between Teachers and Students**

In Study 2, teachers displayed relatively conservative gender attitudes: except for two non-believers of FALS, the teachers overwhelmingly subscribed to FALS. Even with three male veteran who postulated that they would not overgeneralise the gendered patterns they personally witness to all language learners, such critical perspectives are rare in frequency (one by each) and proportion (three out of 20 interviewees, a mere 15%). Students in Study 4,

on the other hand, demonstrated higher levels of critical thinking towards FALS by challenging its validity.

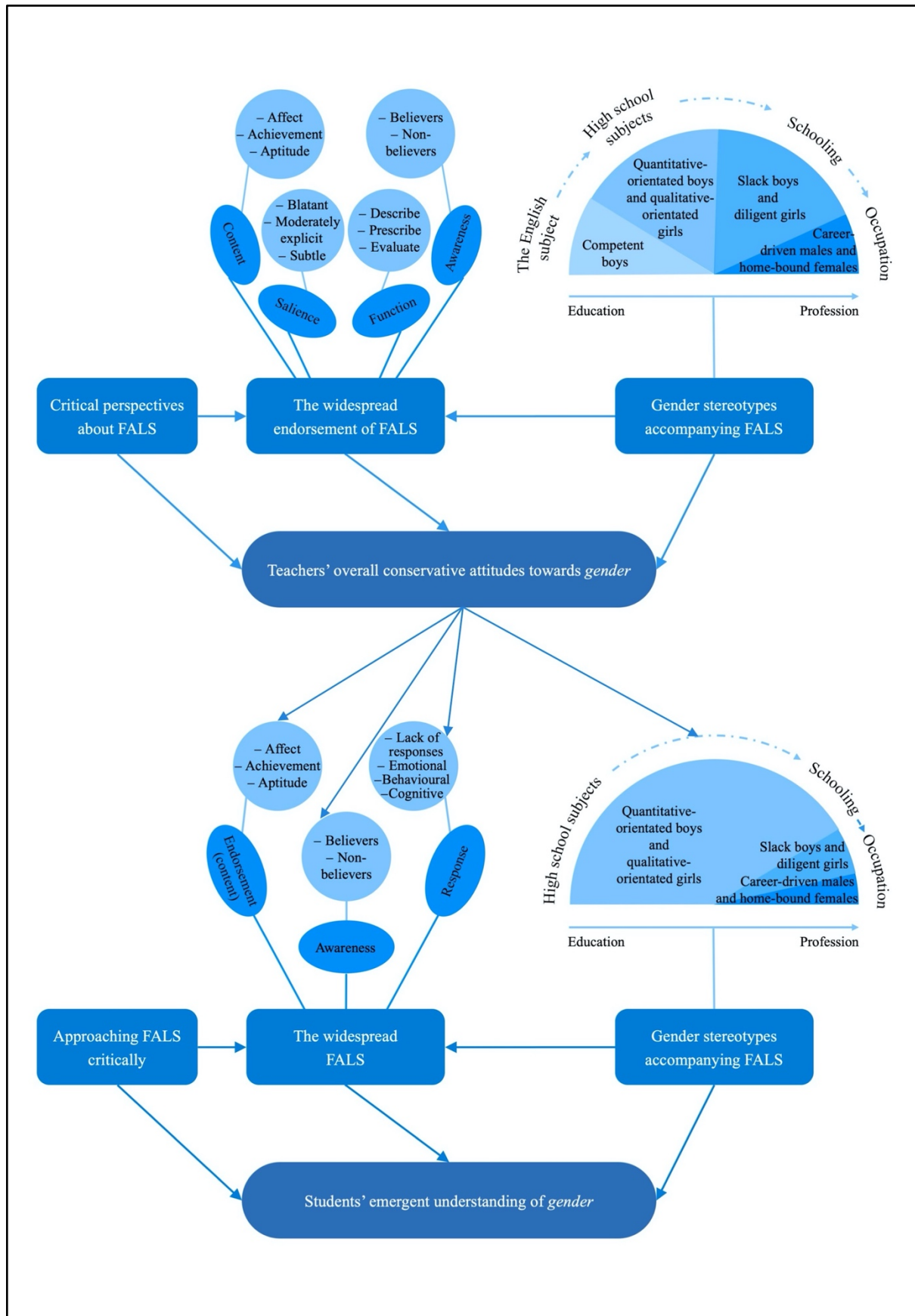
However, this does not imply that students were in anyway more gender-egalitarian than their teachers. As mentioned in Section 9.1.4.1, a higher percentage of students than teachers endorse FALS; unlike the teachers, there were no non-believers among students. Moreover, both teachers and students held a range of educational and professional GSs. In Study 2, these GSs were described by teachers as commonplace and well-accepted; in Study 4, students conformed to the gender norms prescribed by GSs, though one girl was confused by the career-driven males and home-bound females stereotypes. These critical tendencies and confusions among students point to the emergent nature, not necessarily a more gender-egalitarian one, of their gender attitudes.

Thus, Study 2 identifies that teachers' gender attitudes are readily established and generally conservative, with few exceptions. In comparison, Study 4 points out that students are still developing their understanding of *gender*, which explains the presence of critical voices and confusions. Combining the pertinent themes emerged from both studies, there is an obvious connection between teachers' gender attitudes and students' interpretation of *gender*, as exposed in Figure 9.3. Teachers convey their GSs, explicitly and implicitly, to students (see relevant quotes in Section 7.1); then, the majority of students conform to or passively accept the gender norms prescribed by these stereotypes (see Sections 8.2.4-5). A few are perplexed, and some developed critical understanding of some stereotypes but not others (see Section 8.2.6). Therefore, the discrepancies here signal both the possibility of stereotype maintenance *and* opportunities for stereotype change. On the one hand, traces of teachers' conservative gender attitudes seep into children's gender-related beliefs. On the other, the fact that students are still exploring their gender identities and developing their understanding of *gender* means that a more gender-egalitarian generation is still in the making, if educators work on cultivating critical thinking capacities among students.

#### **9.1.4.4 Gender Difference or Similarity in Stereotype Endorsement**

Empirical research presents inconsistent findings in respect of gender differences in stereotype endorsement (Section 3.2.3). Some report gender differences in a self-serving manner: male participants endorsed male-math stereotypes more strongly than female ones (Tomasetto et al., 2015), or female students are more convinced of female-language stereotypes (Plante et al., 2009). However, studies documenting a lack of gender difference (e.g. Whitehead, 1996; Muntoni and Retelsdorf, 2019), or even gender difference in a self-

**Figure 9.3: Influence of Teachers' Gender-Stereotypical Beliefs on Students' Gender Attitudes**



handicapping manner (i.e. members of a gender group expressing a more stubborn belief in negative academic GSs about themselves; e.g. Yee and Eccles, 1992; Liu, et al., 2010).

In Study 1, no gender differences were found in guardian or teacher datasets. The exception is students' FALS endorsement on the micro level: boys were revealed to harbour stronger stereotypes regarding grammar achievement, writing aptitude, and grammar aptitude; girls, instead, perceived a larger stereotypical female advantage in listening and speaking affect. The fact that gender differences were detected at the micro level but not the macro level among students actually provides compelling evidence that there might be similar trends among teachers on the micro level. But because it was not feasible to recruit enough teachers to answer questions on the micro level while preserving enough statistical power from the eight participating schools, this possibility was not examined. Future studies utilising online questionnaire might be able to locate a larger sample size and answer this question.

With regards to guardians, given the size of valid responses (1,904), it is unlikely that Type I error would have occurred. The finding of this a lack of gender difference, therefore, might be valid. Such homogeneity of guardians' FALS further underscores the necessity to increase scholarly exploration into guardians' gender stereotyping, an angle previously overlooked by research into academic GSs.

#### **9.1.4.5 Regional Variation or Consistency in Stereotype Endorsement**

According to Section 3.1.4.3, subcultural variations and regional specificity in stereotype endorsement may exist. Indeed, there was evidence that guardians from the south held a stronger stereotype regarding language aptitude. This disparity might have arisen because of the more task-orientated sub-culture in the south (Huo and Randall, 1991). As Hofstede (2011) has pointed out, people in such societies tend to subscribe to more traditional gender role beliefs than those from people-orientated contexts. Nevertheless, for both northerners and southerners, the stereotypical belief that 'English is not a masculine domain' is still strong, because both gave higher ratings for female learners in terms of English aptitude and affect.

Interestingly, regional variation was missing from the student dataset (see Section 5.2.3.3) and teacher dataset (see Section 6.5). It would seem inadvisable to decide that no regional variations existed with only the current evidence for three reasons: 1) there is collaborating evidence from previous literature suggesting regional differences in stereotype



endorsement (Section 3.1.4.3); 2) the effect sizes concerning regional difference in three datasets are relatively small ( $\eta_p^2 < .01$  in all three datasets), implying a need for larger sample sizes; and 3) the observed powers were only .152 and .32 in student and teacher datasets, respectively. Of course, this is also a possibility that the regional variation among guardians is indeed a chance finding, given the relatively small number of schools in the sample. That is, the guardian samples from the participating schools may not be representative of the guardian population in the southern and northern regions. Thus, to avoid committing Type II error, i.e. missing genuine regional differences in FALS endorsement, larger and more representative samples are needed for future studies.

### **9.1.5 Gender Stereotypes Accompanying Female-Advantage-in-Languages Stereotype**

Together, Sections 3.1.4.4 and 3.2.1 review complementary academic GSs in the literature. Male students are stereotypically characterised more positively in STEM, besides being troublesome, less academically motivated and easily bored with schoolwork, while girls are perceived more favourably in the Humanities, as well as being compliant, averagely more devoted and enjoying school more (e.g., Bonnot and Jost, 2014; Carr and Pauwels, 2006; Jones and Myhill, 2004). Studies 2 and 4 also uncover evidence for these GSs accompanying FALS, as Figure 9.3 reveals. Moreover, the career-driven men and home-bound women stereotypes emerging from the two interview studies approximate the agentic men and communal women stereotypes extensively analysed in Section 3.1.3. The findings here indicate that *gender* is considered as a differentiating in English classrooms, in schools, and in Chinese society.

To challenge these circulating GSs in China, senior high schools have to play a more active role in refuting the gender dichotomy primarily prescribed by the quantitative-males and qualitative-females stereotypes: researchers have shown that in environments where gender is emphasised and/or salient functionally, gender stereotyping tendencies tend to increase (Liben and Bigler, 2002). In Chinese senior high schools, gender is especially notable due to the policy of making students to choose between Sciences subjects and the Humanities ones when they proceed to the second year. So long as the quantitative-males and qualitative-females stereotypes promote and perpetuate the gendered pathways in subject selection, other relevant academic and professional GSs will work in concert to maintain current divide between boys and girls.

## 9.2 Effects of Female-Advantage-in-Languages Stereotype

The second goal of the current project was to unearth whether FALS could have influenced learners' test performance, in addition to self-concept, emotions and behaviours concerning language learning. To this end, two studies were conducted sequentially in Phase 2: Study 3, a field experiment, and Study 4, group interviews. Section 9.2.1 will interpret and contextualise the gender-differential impacts on test performance. Then, Section 9.2.2 will turn to ponder learners' cognitive, emotional, and behavioural responses to FALS.

### 9.2.1 Gender-Specific Effect on Test Performance

Study 3 discovered that boys and girls were affected differently by FALS: boys, after being primed of FALS, showed a substantial decrease in performance (4.54 marks) compared to their peers in the control group. Girls' performance, on the other hand, was somewhat boosted (.80 mark) after they had been reminded of FALS.

To interpret the amplitude of boys' performance decrement, two frameworks describing language proficiencies at different learning levels need to be consulted—China's Standards of English Language Ability (CSE<sup>48</sup>) and the Common European Framework of Reference (CEFR<sup>49</sup>, as explained on <https://www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions>). A typical English learner in Chinese high schools is expected to reach at least Level 4 of CSE, which is equivalent to Level B2 or higher in CEFR. According to the designers of OOPT, a score range of 60-80 equals B2, a level achieved by all but one group of participants: boys in the stereotype threat condition. They, instead, got an average of 59.963, a score corresponding to B1, a level lower in CEFR. In other words, being reminded of FALS caused boys to drop one level down when being judged against CEFT, a widely accepted assessment framework of English proficiency. This result has significant real-work applications: boys' performance in high-stake language qualification tests could be constantly and consistently threatened by FALS, a prevalent academic GS as shown in Chapters 5-8.

In comparison, the boost in girls' performance is trivial in practical terms. This minor 'benefit' can be easily dismissed, if the effects of FALS on their self-concept, emotions

---

<sup>48</sup> CSE, the first full-range English proficiency scale in China, is established in 2017 by the National Education Examinations Authority under the instruction of The People's Republic of China's Ministry of Education. It is intended to provide a set of reliable and valid standards of assessing English proficiency. The English version can be retrieved from <http://cse.neea.edu.cn/html1/report/18112/9627-1.htm>.

<sup>49</sup> CEFR, developed by the Council of Europe, consists of a series of descriptions of language abilities at different learning levels. Similar to CSE, it is also used as a criterion to interpret and compare different language qualifications.

towards languages, and learning behaviours are taken into account, as the next section will unfold.

### **9.2.2 Effects on Self-Cognition, Emotions and Behaviours Concerning English Learning**

In Study 4, after having acknowledged FALS's presence in their lives, students were encouraged to comment on the effects of FALS on them and how they had responded to it. In general, negative effects on learners' self-cognition, emotions and behaviours concerning English learning are found among both boys and girls. For boys, in order to protect their self-esteem, some decided that they had grown too accustomed to FALS to bother reacting to it. Some others internalised FALS into their self-concepts, believing that they were not endowed with strong linguistic competence. As a result, they accepted FALS's prescription of failure, reducing effort in English in an attempt to deflect any personal responsibilities for the bad grades, and relying on FALS to explain their lack of interest in English. Perhaps the only comfort from Study 4 is that, although one boy learnt English well, he did not recall being bullied or teased because of his counter-stereotypical status. Instead, he resorted to 'Wen wu shuang quan' ('be adept with both the pen and the sword'), a type of ideal masculinity in traditional Chinese culture, to expound his achievement. Yet, his readiness to cite this expression might still indicate that he was so familiar with FALS-fuelled narratives that he had to find a suitable model of masculinity to protect his self-regard.

To girls, surprisingly, no positive effects happened. Regarding self-cognition, some girls reported that they felt obliged to improve their English because of FALS. Curiously, no girls mentioned any real effort they made in order to live up to the expectations laid out by FALS. In fact, the affect component of FALS actually caused a girl to feel rejected by her own gender group: as a girl who was not attracted to English, she internalised the male aloof attitude and literally called herself a boy. In addition, a few girls experienced feelings of inadequacy or hurt due to FALS: those with unsatisfactory English grades feared that they were letting their FALS-subscribing parents down, while those attaining high scores were harmed by their parents' taking their success as granted.

As commented in Section 3.3, there is a dearth of research into the effects of language-related academic GS on learners, especially those utilising qualitative methods. Among the few exceptions, Carr and Pauwels's book (2006) is instrumental in understanding how boys interpret and react to the female stereotyping of language fields in Australian schools.

however, in their study, the focus was almost exclusively on boys—there is even a chapter where girls talked about boys. Thus, the current study contributed to GS literature by giving a voice to girls, who appear not to be encouraged or comforted by FALS, sadly.

### **9.3 Educational Implications**

Gender stereotypes feminising language-related domains have largely prevailed across the world, despite decades-long efforts to challenge them (Carr & Pauwels, 2006; Schmenk, 2004; Zhao, Li, & Yin, 2014). This project provides the following implications for educators and policy-makers concerned with rigid and restrictive gender stereotypes in schools and families.

First, it pinpoints problems with current attempts to combat gender stereotypes circulating around language classrooms. Much previous research primarily approached ability-related stereotypes (“girls are more gifted with words”, for example). Only a limited few acknowledged that stereotyping could extend beyond views on biology to prescribe gender-typed feelings and attainment patterns in languages. But combined together, Studies 1 and 2 has shown that FALS consists of three interconnected elements—gender-based linguistic aptitude, gender-typical affect, and gender-differential achievement, with the former two fuelling the third component. The aptitude component, in fact, is the least endorsed one among guardians, students, and teachers. These findings underscore the fact that FALS has essentially been operating as a logically coherent, potentially self-perpetuating belief system in schools and households, and a focus on ability-based gender stereotyping is surely not be enough. Therefore, to tackle FALS effectively, an integrative intervention programme contains respective strategies targeting each component is needed.

Furthermore, this project sheds light on how to specifically dismantle the aptitude, affect, and achievements aspects of FALS. For example, in Section 9.1.2.1, the aptitude component has been revealed as an exaggeration of genuine sex difference in verbal abilities. Thus, educating language teachers about the triviality of actual sex difference while juxtaposing the polarised portrayals of boys and girls by the aptitude component of FALS might reverse their stereotypical underestimation of boys’ linguistic competence. With regards to the achievement element, Sections 9.1.2.2 and 9.2.1 have worked in concert to illustrate that not only do stereotypical beliefs in gender-differential accomplishment inflate the size of real gender achievement gap in languages, they *are* simultaneously suppressing boys’ performance and slightly boosting that of girls. Such empirical evidence of the self-fulfilling nature of the achievement aspect should be highlighted to language teachers and

school leaders, as they tend to be invested in improving students' academic performance. Additionally, if the affect component is disclosed as an over-simplification of boys' and girls' affective responses, this should serve as a reality-check for both teachers and parents/guardians, sensitising them to the emotional toll of their gender-related beliefs on learners.

Additionally, this research offers three caveats for educators and policy-makers with the hope of establishing gender-equal and -inclusive language classrooms. The foremost concern is the relatively large magnitude of teachers' FALS, as argued in Section 9.1.4.1 earlier. Given that ample research has accentuated the influence of teachers' beliefs on learners' ability self-beliefs, academic choices, and achievement-related behaviours, on informing teachers' pedagogical practice and on shaping learners' learning experiences (e.g., Gunderson, Ramirez, Levine, & Beilock, 2012; Muntoni & Retelsdorf, 2018), it is essential that awareness-raising programmes are implemented among teachers. A related caution surfaces from the two interview-based studies (Studies 2 and 4), where a range of substantially similar educational and professional GSs were uncovered among teachers and students from nine different schools. Clearly, challenging gender stereotypes is made harder when gender binary thinking permeates in academic, professional and social aspects of teachers' and students' lives. Synergised efforts, therefore, are called for in order to encourage boys and girls to freely and fully develop their potentials and explore their interest in different domains.

The final caveat pertains to the cultural specificity of the project. All four studies were conducted in the context of English curriculum in high schools across mainland China. Admittedly, the idea of *gender* (see Section 2.1.2) and the value of English proficiency (see Section 2.3) in this particular context can have an impact on the socio-cultural relevance of the findings and their implications. For example, Section 9.1.3.1 has pointed out that attributing success to effort, rather than exceptional talent, is central to Confucian philosophy of learning. As a result, the finding of downplaying aptitude in Studies 1 and 2 might not be applicable to other culture and societies outside the reach of Confucianism. Therefore, a consideration of the cultural embeddedness is required if findings from this project are to be relied on by educators and school administrators worldwide.

## 9.4 Limitations

Limitations within each study and across studies have been identified. In Study 1, the Teacher/Guardian Questionnaire measured each of the three stereotype component (aptitude,

affect, and achievement) with a single self-report item. Had a more complex measure incorporating multiple items for each component been utilised, a more comprehensive understanding of FALS might have been achieved. Furthermore, Study 1 did not account for school-level effects in stereotype endorsement because such multi-level analysis would have required 20 or more schools to be included.

Study 2's limitation lies with the interview schedule. Although neutral prompts were employed, and the topic of *gender* arose naturally when teachers described characteristics of ideal language learners, it was likely that gender issues were centralised during interviews. Consequently, some participants might have provided answers they had thought that the researchers had wanted to hear, i.e. accounts and even explanations of gender differences.

Study 3's intrinsic limitation arises from its context. In order to increase ecological validity, a total of five cues were used to activate negative gender stereotypes targeting males in languages under the experimental condition (See p. 78 in Section 4.3.1.1). But Study 3 could only provide evidence that being reminded of such stereotypes would have lowered boys' performance in an immediately subsequent English test of their vocabulary and reading achievement. The study's reveal of ST effect, therefore, could be considered restricted to this particular test context. The adverse effects of stereotypes feminising language domains, in fact, have to be evaluated and interpreted with help from other methodologies and wider contexts.

The major concern with Study 4 is the possibility that group thinking might have led to certain consensuses, consciously or subconsciously hiding away plausible dissenting voices. Of course, the decision to conduct group interviews, instead of one-to-one sessions, was to address power asymmetry between the researcher and student interviewees (see p.83-4 in Section 4.3.2), and the likely group thinking could arguably be inevitable at least to some extent. This was why the interviewer, being fully aware of this risk, paid attention to interviewees who appeared less outspoken or articulate in each group, encouraged them to contribute to the discussion (verbally and non-verbally, such as via nodding and smiling), and waited for them to speak when pauses arose. Still, it could still be the case that groups constructed ideas together in their conversation, and some interviewees might dissent. Yet, acknowledging this limitation does not undermine that findings of adverse effects of FALS on learners, as there were between-group consensus, too.

Two major limitations have been identified across studies: imbalanced samples and the large amount of qualitative data. Although eight schools with varying geological locations, socio-economic statuses, and education sectors participated in Study 1, the sample was still unbalanced: there were more schools from southern cities than ones from northern cities; more public schools from the north joined the survey than those from the south. Thus, there is issue with representativeness in Study 1's sample. In Study 2, a related problem is the scarcity of male novice teachers. I did strive to find more teachers fitting this demographic profile, but the second candidate never appeared. This mere fact suggests how feminised the language teaching profession has become in China. Actually, the teacher and students interviewees in Studies 2 and 4 recognised this phenomenon on multiple instances: teachers acknowledged that boys and men commonly regarded English learning and teaching, as a feminine pursuit; students themselves even called male English teachers 'unacceptable'. Thus, in a way, the incapacity to convey more voices from male novice teachers in this thesis encapsulates the gendered nature of English education in China.

The last plausible constraint of this project lies with the analysis of relatively large quantities of qualitative data from Studies 2 and 4. As Pope and Denicolo (1986) have acknowledged, reporting findings from qualitative data inevitably takes the form of data reduction, because researchers typically aim for the recognition of themes and commonalties in the originally descriptive, multi-layered dataset. In this work, a total of 28 interviews were conducted, which turned out to be over 24 hours of audio recordings and around 290,000 words in transcriptions (Chinese characters). Therefore, to combat this issue of data reduction, I was committed to familiarising myself with the interview data (both recordings and transcripts), coding, re-coding, and cross-case synthesising, and citing rich and detailed descriptions from interviews in the thesis. Furthermore, in order to protect the credibility and authenticity of findings against my assumptions, I adopted the technique of member-checking in Studies 2 and 4. I also discussed my analytic framework, coding schemes, and plausible presumptions interfering with data analysis with my supervisor in email correspondence and face-to-face meetings. Finally, to avoid misrepresenting interviewees' intentions or opinions in the quotes in this thesis during my translating them from Chinese to English, all interview excerpts were reviewed by a professional translator.

## Chapter 10: Conclusions

This concluding chapter consists of six sections. In Sections 10.1-2, a summary of major findings from the two-phase mixed methods project will be provided. Then, the limitations of the project will be considered in Section 10.3. Subsequently, Section 10.4 will delineate the theoretical and practical contributions of the project, which is followed by suggestions of future research in Section 10.5. Finally, in Section 10.6, I will reflect on how I have developed personally and professionally during the PhD. journey.

### 10.1 Summary of Findings in Phase 1

Phase 1 investigated whether students in Chinese senior secondary schools, their guardians, and their teachers of English stereotypically believed girls to be better language learners than boys, adopting a questionnaire survey (Study 1) and an interview-based study (Study 2). In general, all three groups endorsed the ‘female-advantage-in-languages stereotype’ (FALS), regarding girls as more gifted, more enthusiastic, and higher-achieving learners than boys. The students, in addition, also harboured FALS on a micro level, deeming a female advantage in language skills (listening, speaking, reading, and writing) and knowledge (grammar and vocabulary). Moreover, as student proceeded to higher grades, they stereotypically perceived a widening gender difference in various aspects of language achievement and affect, as well as in grammar aptitude.

There is some evidence of variations in FALS endorsement due to gender and region in some groups of participants. Gender difference emerged among students on the micro level: boys tended to harbour stronger stereotypes regarding grammar achievement, writing aptitude, and grammar aptitude; girls, instead, perceived a larger stereotypical female edge in affect for listening and speaking. Furthermore, region specificity arose among guardians (southern, task-orientated regions displaying a stronger stereotype concerning aptitude), but not among the other two groups.

Among all three FALS components, aptitude was the weakest: not only did it have the smallest magnitude in all three groups of questionnaire respondents, it was also endorsed by the fewest number of teachers in Study 2 (and also so among student interviewees in Study 4, Phase 2). However, the participants in Study 1 disagreed on which component was the strongest: for students and teachers, it was affect; but for guardians, it was achievement. Another discrepancy among participants is that older generations, i.e., guardian and teachers, subscribed to FALS to a greater extent. Nevertheless, a consensus among teacher and student



interviewees (from Studies 2 and 4) appeared: a stronger stereotypically additive relationship between affect and achievement components was assumed, compared to that between aptitude and achievement components.

The teacher interviewees, in general, endorsed at least one component of FALS, except for two non-believers (10%). All believers expressed FALS moderately explicitly in various instances, but some also did so either blatantly or occasionally, subtly. Additionally, among the believers, FALS was most frequently used descriptively, but also sometimes in prescriptive or evaluative ways. Meanwhile, less than half FALS believers simultaneously held the ‘competent males stereotype’, characterising boys as more intellectually endowed and academically accomplished irrespective of domains. Three sets of complementary GSs along the education-profession continuum were also held by some interviewees, stereotypes that depicting males as a quantitative-orientated, slack, yet career-driven group, while females as a qualitative-orientated, diligent, nevertheless home-bound crowd. Finally, there were traces of critical voices regarding FALS, where teachers either rejected FALS or refused to overgeneralise FALS to all learners. However, as these critical perspectives are small in number, the teachers’ attitudes towards *gender* were still overall conservative.

## **10.2 Summary of Findings in Phase 2**

Phase 2 was primarily designed to understand the effects of FALS on learners. In Study 3, the field experiment, gender-divergent performance patterns were detected: while girls’ performance was marginally boosted after having being reminded of FALS, boys under the same condition experienced a substantial decrease in performance compared to their peers in the control group (stereotype boost among girls and stereotype threat among boys).

The boost in performance, however, cannot be taken as evidence of girls benefiting from FALS, despite four girls in Study 4, the group interview, reporting that they felt obliged to learn English better because FALS prescribed so. Because during the interviews, FALS was found to negatively influence girls’ self-concept and emotional wellbeing. Because of FALS, some girls questioned their gender identity, felt uneasy or even hurt. In the case of boys, the negative effects encompassed cognitive, emotional, and behavioural aspects: boys internalised FALS’s portrayal of them as remedial English learners, felt upset due to FALS, and reduced their effort to learn. These findings from the interviews also explained how FALS could have led to performance decrements in the experiment: during the test, boys had to exert extra effort into suppressing troubling thoughts and regulating negative emotions while keeping on track of the test at hand.

### 10.3 Theoretical and Practical Contributions of Doctoral Project

This project contributes to research into gender stereotypes (GSs) and language education. In previous works, the male stereotyping of STEM subjects has been receiving much more attention than the female stereotyping of language and literacy domains. This negligence is unfortunate and unfair, considering the cognitive gains and psychological benefits that can be enjoyed by boys and men should they be free to pursue their interests and potentials in languages (e.g., Cambridge Public Policy SRI, 2015; Croft, Schmader, Block, 2015). The relatively limited investigation into GSs targeting language learners, nevertheless, tends to over-emphasise competence-based stereotypes, even if the literature indicates that affect- and achievement-related stereotypes are also circulating. Therefore, the first and foremost original contribution of this project is the conceptualisation and measurement of FALS, a tripartite construct encompassing aptitude, achievement, and affect components. Furthermore, this project also advances scholarly understanding of academic GSs, through its revelation that aptitude, notwithstanding the spotlight it has been attracting, was actually the weakest component.

With regards to the field of language education, a longstanding tradition is to locate gender differences regarding motivation, attainment, and etc, a custom with an inbuilt gender-stereotypical bent. Yet, this thesis, by its comprehensive and in-depth analysis of gender stereotyping in English classrooms using a mixed-methods design, leads to a more critical and nuanced appreciation of gender's role in language domains. FALS, in addition to other academic and professional GSs, are witnessed among teachers and students alike. Moreover, the current project expands stereotype threat (ST) and stereotype boost research by expanding it to EFL contexts. Therefore, this project was instrumental in understanding the gendered patterns in engagement, devotion, and interest in language subjects.

From a practical perspective, this project helps to promote a gender-equitable and -inclusive environment for language learners. Two findings are of specific significance here: a) students, their guardians, and teachers endorsed one or more FALS components to varying degrees and b) there was a stereotypically stronger additive relationship between affect and achievement components than that between aptitude and achievement. Together, they highlight the difficulty of tackling FALS, a multifaceted, logically self-consistent stereotype. Therefore, to combat FALS and establish a gender-fair learning environment, awareness-raising programmes are called for to sensitise parents, teachers, and teacher educators to their gender-related beliefs. Furthermore, a synergy of assorted interventions targeting each

component, and reaching students, guardians, and teachers is in need. Finally, the detection of ST effect among male learners and other negative impact on both male and female learners should alert policy-makers to the poisoning influences of the female stereotyping of language domains across cultures and countries.

#### **10.4 Future Directions**

Building on this doctoral thesis, future studies can take several routes to forward the study of gender stereotyping in language classrooms. First, teachers' perspectives can be explored more extensively and intensively using a range of methods and methodologies. For example, online questionnaires can potentially reach larger samples, enabling the examination of FALS endorsement on the micro level by teachers. In addition, due to the plausible lack of male novice teachers, future researchers can adopt a phenomenological approach to understanding the barriers preventing men from becoming language teachers. Meanwhile, some researchers might inspect how academic GSs accompanying FALS might have been incorporated by teachers in classroom interactions via methods such as diary, observation, and case study.

Taking a cross-cultural approach, researchers can also carry out transnational studies to identify similarities and differences in FALS endorsement in diverse language-learning contexts. For example, the discovery of the aptitude component's small amplitude has been attributed to four plausible reasons: a) the traditionally marginal position of gender essentialism in Chinese culture; b) the Confucian tendency to downplay intellectual prowess when accounting for achievement; c) the widespread presumption that language learning requires sustained attention and effort, rather than great aptitude; and d) the prevalent male-agency stereotype potentially mitigating the strength of FALS. Since the first two causes are specific to the Chinese society, if the magnitude of aptitude is more substantial in other cultures, these two culture-sensitive explanations will probably seem more convincing. Otherwise, the latter two cross-cultural rationales shall be thought to possess more explanatory power.

Alternatively, the field experiment (Study 3) can be applied to different age groups on various language tasks across cultures. In this project, for instance, ST and stereotype boost has emerged among, respectively, boys and girls sitting an online test of English in a Chinese senior high school. Replication studies on elementary school students and adult learners might be conducted to gain insights into the robustness and reliability of the abovementioned effects. Furthermore, because the test employed in Study 3 does not assess writing and

speaking proficiency, the influences of FALS on performance in these two productive skills remain elusive, a topic that future research can explore. Another suggestion is to qualitatively inspect the experience of high-achieving males, particularly how they manage to succeed in the presence of FALS.

### **10.5 Personal reflections**

Growing up, I have always been haunted by unrelenting senses of self-doubt: ‘am I doing well enough?’ ‘Is this accomplishment due to a stroke of good luck?’ ‘When will fortune take its favour away?’ In middle school, I stumbled upon readings on *gender*, GSs in particular, revealing to me that girls and women all over the world, at least at some point in their lives, share similar apprehensions. Gradually, I started to appreciate how the social environment shaped my understanding of *gender* and the development of my own gender identity. Yet, I still thought that conforming to prescribed gender norms would make life happier and easier. This is why my ideal future in high school was to ‘be successful in a womanly manner’: I did not enjoy being called a tough girl. During undergraduate years, an added worry loomed large—‘as an English major, what should I do after graduation?’ English teachers and translators were the obvious choices then, but both required creativity, patience, attention to detail, and tenacity. I was hardly any of those things. Sitting in one of the best universities in China, I was overwhelmed with confusion and insecurity.

Fortunately, such perplexity and lack of self-assurance propelled me to work harder and read more, setting me onto the path of becoming an educational researcher. I feel especially grateful for this opportunity to focus on GSs in my doctoral project, which has developed into a journey of professional and personal empowerment. With guidance and encouragement from my supervisor, advisor, and colleagues, I have honed my ability to manage a multi-phase research project, conduct questionnaire and interview surveys, and experiments, and analyse multiple strands of data. Pondering the psychosocial, socio-cultural, and environmental factors behind gender issues, I also practised skills essential to researchers: synthesising and critically engaging with the literature, practicing data analysis techniques, utilising software packages to manage references and data files, writing journal papers clearly and concisely, and presenting at academic conferences. I joined the PhD programme as an insecure and bewildered young woman, but the project has shown me that I can be innovative, perseverant, prudent and resilient. I feel confident, ambitious, and ready to commit myself to addressing gender relations and equality issues in China’s sprawling education labyrinth.

## References

- Altan, M. Z. (2012). Pre-service EFL teachers' beliefs about foreign language learning. *European Journal of Teacher Education*, 35(4), 481-493. doi:10.1080/02619768.2011.643399
- Andre, T., Whigham, M., Hendrickson, A., & Chambers, S. (1999). Competency beliefs, positive affect, and gender stereotypes of elementary students and their parents about science versus other school subjects. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 36(6), 719-747. doi: 10.1002/(SICI)1098-2736(199908)36:6%3C719::AID-TEA8%3E3.0.CO;2-R
- Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: Stereotype threat prior to test taking. *Educational Psychology Review*, 24(4), 609-635. doi: 10.1007/s10648-012-9200-4
- Archer, J., & Lloyd, B. (2002). *Sex and gender*. Cambridge: Cambridge University Press.
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology*, 16(1), 94–98. doi: 10.1037/a0017564
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of experimental social psychology*, 35(1), 29-46. doi: 10.1006/jesp.1998.1371
- Abu-Rabia, S. (1995). Attitudes and cultural background and their relationship to English in a multicultural social context: The case of male and female Arab immigrants. *Educational Psychology*, 15(3), 323-336.
- Asassfeh, S. M. (2015). Prospective EFL teachers: What language learning beliefs do they hold? *The Asia - Pacific Education Researcher*, 24(1), 13-26. doi: 10.1007/s40299-013-0154-1
- Bagnoli, A., Demey, D., & Scott, J. (2014). Young people, gender, and science. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment* (pp. 321–345). Cambridge University Press. doi: 10.1017/CBO9781139128933.019
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of personality and social psychology*, 68(2), 181-198. doi: 10.1037/0022-3514.68.2.181

- Barnard, S., Hassan, T., Bagilhole, B., & Dainty, A. (2012). 'They're not girly girls': An exploration of quantitative and qualitative data on engineering and gender in higher education. *European Journal of Engineering Education*, 37(2), 193-204. doi: 10.1080/03043797.2012.661702
- Baron, A. S., Schmader, T., Cvencek, D., & Meltzoff, A. N. (2014). The gendered self-concept: How implicit gender stereotypes and attitudes shape self-definition. *Gender and Development*, 109-132. doi: 10.4324/9781315870878
- Barton, A. (1997). Boys' under-achievement in GCSE modern languages: Reviewing the reasons. *The Language Learning Journal*, 16(1), 11-16. doi: 10.1080/09571739785200211
- Bastian, B., & Haslam, N. (2007). Psychological essentialism and attention allocation: Preferences for stereotype-consistent versus stereotype-inconsistent information. *The Journal of Social Psychology*, 147(5), 531-541. doi: 10.3200/SOCP.147.5.531-542
- Beck, V., Fuller, A., & Unwin, L. (2006). Safety in stereotypes? The impact of gender and 'race' on young people's perceptions of their post-compulsory education and labour market opportunities. *British Educational Research Journal*, 32(5), 667-686. doi: 10.1080/01411920600895718
- Becker, J. C., & Sibley, C. G. (2016). Sexism. In T. d. Nelson (ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (2<sup>nd</sup> ed, pp. 315-336). New York NY: Psychology Press.
- Berger, P., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Doubleday.
- Bem, S. L. (1983). Gender Schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs: Journal of Women in Culture and Society*, 8, 598-616.
- Bem, S. (1993). *The lenses of gender: transforming the debate on sexual inequality*. New Haven, Connecticut: Yale University Press.
- Bhanot, R., & Jovanovic, J. (2005). Do parents' academic gender stereotypes influence whether they intrude on their children's homework? *Sex Roles*, 52, 597-607. doi:10.1007/s11199-005-3728-4.
- Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355, 389-391. doi:10.1126/science.aah6524

- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality And Social Psychology*, 66(1), 5-20. doi:10.1037/0022-3514.66.1.5
- Biernat, M. (1991). Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis. *Journal of Personality and Social Psychology*, 61(3), 351-365. doi: 0.1037/0022-3514.61.3.351
- Bigler, R. S., & Liben, L. S. (1999). Cognitive mechanism in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, 63, 1351–1363. doi: 10.2307/1131561
- Bonnot, V., & Jost, J. T. (2014) Divergent effects of system justification salience on the academic self-assessments of men and women. *Group processes and intergroup relations*, 17(4), 453-464. doi: 10.1177/1368430213512008
- Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative research*, 8(1), 137-152.
- Breckler, S. J. (1984). Empirical validation of affect, behaviour, and cognition as distinct components of attitude. *Journal of personality and social psychology*, 47(6), 1191-1205. doi: 10.1037/0022-3514.47.6.1191
- Briggs, C. L. (2007). Anthropology, interviewing, and communicability in contemporary society. *Current Anthropology*, 48(4), 551-580.
- Brinkmann, S. (2013). *Qualitative interviewing: Understanding qualitative research*. New York, NY: Oxford University Press.
- Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research*. Los Angeles, California: SAGE Publications Ltd.
- Broaded, C. M., & Liu, C. (1996). Family background, gender and educational attainment in urban China. *The China Quarterly*, 145, 53-86.
- Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues*, 28(2), 59–78. doi: 10.1111/j.1540-4560.1972.tb00018.x
- Brown, R. (2010). *Prejudice: Its social psychology* (2nd ed.). Chichester: Wiley-Blackwell
- Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, 39(6), 626-633. doi: 10.1016/S0022-1031(03)00039-8
- Bryman, A. (2004). *Social research methods* (2nd Edition ed.). Oxford: Oxford University Press.

- Butler, Y. G. (2015). English language education among young learners in East Asia: A review of current research (2004–2014). *Language teaching*, 48(03), 303-342.
- Cambridge Public Policy SRI. (Oct. 2015). *The value of languages*. Retrieved from <http://www.languagesciences.cam.ac.uk/national-languages-policy/report-the-value-of-languages/view>
- Carr, J., & Pauwels, A. (2006). *Boys and foreign language learning: Real boys don't do languages*. Hampshire: Palgrave Macmillan Ltd.
- Carter, D. B., & Patterson, C. J. (1982). Sex roles as social conventions: The development of children's concepts of sex role stereotypes. *Developmental Psychology*, 18, 812–824. doi: 10.1037/0012-1649.18.6.812
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and social psychology bulletin*, 25(4), 413-423. doi: 10.1177/0146167299025004002
- Chalabaev, A., Stone, J., Sarrazin, P., & Croizet, J. C. (2008). Investigating physiological and self-reported mediators of stereotype lift effects on a motor task. *Basic and Applied Social Psychology*, 30(1), 18-26. doi: 10.1080/01973530701665256
- Chatard, A., Selimbegović, L., Konan, P., & Mugny, G. (2008). Performance boosts in the classroom: Stereotype endorsement and prejudice moderate stereotype lift. *Journal of Experimental Social Psychology*, 44(5), 1421-1424. doi: 10.1016/j.jesp.2008.05.004
- Chen, X [陈霞]. (2007). Investigating teachers' gender attitudes [教师性别意识现状调查分析]. *Journal of Heilongjiang College of Education [黑龙江教育学院学报]*, 26(9), 54-56. doi: 1001-7836(2007)09-0054-03
- Cheng, X. (2011). The 'English Curriculum Standards' in China: Rationales and issues. In A. Feng (Ed.), *English language education across greater China* (pp. 133-150). Bristol: Multilingual Matters.
- China Education Panel Survey (2013). *Student Data* [Data file]. Retrieved from <http://www.cnsda.org/index.php?r=projects/view&id=72810330>
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th Edition ed.). Abingdon, Oxon: Routledge.
- Collier, A. (1994). *Critical realism: an introduction to Roy Bhaskar's philosophy*. London: Verso.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications, Inc.



- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th Edition ed.). Thousand Oaks, CA: Sage publications.
- Croft, A., Schmader, T., & Block, K. (2015). An under-examined inequality: Cultural and psychological barriers to men's engagement with communal roles. *Personality and Social Psychology Review*, 19(4), 343-370. doi: 10.1177/1088868314564789
- Cuddy, A. J., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of personality and social psychology*, 92(4), 631-648. doi: 10.1037/0022-3514.92.4.631
- Dasgupta, N., & Stout, J. G. (2014). Girls and Women in Science, Technology, Engineering, and Mathematics: STEMing the Tide and Broadening Participation in STEM Careers. *Policy Insights from the Behavioural and Brain Sciences*, 1(1), 21–29. doi:10.1177/2372732214549471
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28(12), 1615-1628. doi: 10.1177/014616702237644
- DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. C., & Johnson, R. C. (1974). Near identity of cognitive structure in two ethnic groups. *Science*, 183(4122), 338-339. doi: 10.1126/science.183.4122.338
- Department of Education, the United Kingdom. (2018) *Subject tables: SFR01/2018. Revised GCSE and equivalent results in England: 2016 to 2017* [Data file]. Retrieved from <https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2016-to-2017>
- Dewaele, J. M. (2007). Predicting language learners' grades in the L1, L2, L3 and L4: The effect of some psychological and sociocognitive variables. *International Journal of Multilingualism*, 4, 169–197. doi: 10.2167/ijm080.0
- Dewaele, J. M., MacIntyre, P., Boudreau, C., & Dewaele, L. (2016). Do girls have all the fun? Anxiety and enjoyment in the foreign language classroom. *Theory and Practice of Second Language Acquisition*, 1(2), 41-63.
- Diekmann, A. B., & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, 26, 1171–1188. doi:10.1177/0146167200262001
- Eagly, A. H. (1987). *Sex differences in social behaviour: A social-role interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (p. 269–322). McGraw-Hill.
- Eagly, A. H., & Steffen, V. J. (2000). Gender stereotypes stem from the distribution of women and men into social roles. In C. Stangor (ed.), *Stereotypes and prejudice: Essential readings* (pp. 142-160). Ann Arbor, MI: Edwards Brothers.
- Eagly, A. H., Wood, W., & Diekmann, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In E. Thomas (ed.), *The Developmental Social Psychology of Gender* (pp.123-174).
- Eccles, J. S. (2014). Gender and achievement choices. In E. T. Gershoff, R. Mistry, & D. A. Crosby (Eds.), *Societal contexts of child development: Pathways of influence and implications for practice and policy* (pp. 19-34). Oxford University Press.
- Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs: Journal of women in culture and society*, 11(2), 367-380. doi:10.1086/494229.
- Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, 46, 183–201. doi:10.1111/j.1540-4560.1990.tb01929.x.
- Education First. (2011-2019). *EF English Proficiency Index*. Retrieved from <https://www.ef.co.uk/epi/>
- Educational Testing Service. (2017). Test and score data summary for TOEFL iBT [PDF document]. Retrieved from [https://www.ets.org/s/toefl/pdf/94227\\_unlweb.pdf](https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf)
- Egan, S.K., & Perry, D.G. (2001). Gender identity: a multidimensional analysis with implications for psychosocial adjustment. *Developmental Psychology*, 37, 451–463. doi: 10.1037//0012-1649.37.4.451
- Ellemers, N. (2018). Gender stereotypes. *Annual Review of Psychology*, 69, 275-298. doi: 10.1146/annurev-psych-122216-011719
- Epting, L. K., Rand, L. A., & D'Antuono, H. L. (2014). Literacy in the College Classroom: Gendered Perceptions of Reading, Writing, Revision, and Grammar. *North American Journal of Psychology*, 16(3), 537-555.
- E-Stat. (2012). *Basic Survey on Education* [Data file]. Retrieved from <https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00400001&tstat=000001011528>
- Erikson, E. H. (1995). *Child and society*. London: Vintage Books.

- Evans, E. M., Schweingruber, H., & Stevenson, H. W. (2002). Gender differences in interest and knowledge acquisition: The United States, Taiwan and Japan. *Sex Roles*, 47, 153–167. doi:10.1023/A:1021047122532
- Fabes, R. A., Hayford, S., Pahlke, E., Santos, C., Zosuls, K., Martin, C. L., & Hanish, L. D. (2014). Peer influences on gender differences in educational aspiration and attainment. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment* (pp. 29–52). Cambridge University Press. doi: 10.1017/CBO9781139128933.004
- Fagot, B. I. (1977). Consequences of moderate cross-gender behaviours in pre-school children. *Child Development*, 48, 902–907. doi: 10.2307/1128339
- Fan, C. C. (1996). Language, gender, and Chinese culture. *International Journal of Politics, Culture, and Society*, 10(1), 95-114.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95–103. doi: 10.1037/0003-066X.43.2.95
- Feng, A. (2012). Spread of English across Greater China. *Journal of Multilingual and Multicultural Development*, 33(4), 363-377.
- Feng, X., & Xiao, J. [风笑天, 肖洁]. (2014). Differences in gender role attitudes between urban and rural regions [中国女性性别角色意识的城乡差异研究]. *The Journal of Humanities [人文杂志]*, 11, 107-116. doi: 10.15895/j.cnki.rwzz.2014.11.014
- Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. *Educational Studies in Mathematics*, 21, 55–69. doi: 10.1007/ BF00311015.
- Fine, M., & Gordon, S. M. (1989). Feminist transformations of/despite psychology. In M. Crawford & M. Gentry (Eds.), *Gender and thought* (pp. 146–174). New York, NY: Springer- Verlag.
- Fisher, L. (2001). Modern foreign languages recruitment post-16: The pupils' perspective. *The Language Learning Journal*, 23, 33–40. doi: 10.1080/09571730185200071
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6), 878-902. doi: 10.1037/0022-3514.82.6.878
- Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of school psychology*, 53(1), 25-44. doi: 10.1016/j.jsp.2014.10.002

- Francis, B. (2000). The gendered subject: students' subject preferences and discussions of gender and subject ability. *Oxford Review of Education*, 26(1), 35-48. doi: 10.1080/030549800103845
- Francis, B., & Paechter, C. (2015). The problem of gender categorisation: Addressing dilemmas past and present in gender and education research. *Gender and Education*, 27(7), 776–790. doi: 10.1080/09540253.2015.1092503
- Francisca del Rio, M., & Strasser, K. (2013). Preschool children's beliefs about gender differences in academic skills. *Sex Roles*, 68(3–4), 231–238. doi:10.1007/s11199-012-0195-6
- Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of personality and social psychology*, 74(2), 435-452. doi: 10.1037//0022-3514.74.2.435
- Fusch, P. I., & Ness, L. R. (2015). Are we there yet? Data saturation in qualitative research. *The qualitative report*, 20(9), 1408-1416.
- Gaither, S. E., Remedios, J. D., Schultz, J. R., & Sommers, S. R. (2015). Priming White identity elicits stereotype boost for biracial Black-White individuals. *Group Processes & Intergroup Relations*, 18(6), 778-787.
- Gao, X [高雪玉]. (2004). A comparative study of the changed conception of women's social gender in Zhejiang [浙江妇女社会性别观念状况的比较分析]. *Journal of Ningbo University (Liberal arts edition)* [宁波大学学报(人文科学版)], 17(5), 133-137.
- Gelman, S. A., Heyman, G. D., & Legare, C. H. (2007). Developmental changes in the coherence of essentialist beliefs about psychological characteristics. *Child Development*, 78, 757–774. doi: 10.1111/j.1467-8624.2007.01031.x
- Gelman, S. A., & Taylor, M. G. (2000). Gender essentialism in cognitive development. In P. H. Miller & E. Kofsky Scholnick (Eds.), *Toward a feminist developmental psychology* (pp. 169–190). Taylor & Frances/Routledge.
- Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs. *Monographs of the Society for Research in Child Development*, 69(1), vii, 116–127. doi: 10.1111/j.1540-5834.2004.06901001.x
- Georgiou, S. N. (2008). Beliefs of experienced and novice teachers about achievement. *Educational Psychology*, 28(2), 119–131. doi: 10.1080/01443410701468716

- Gil, J., & Adamson, B. (2011). The English language in mainland China: A sociolinguistic profile. In A. Feng (Ed.), *English language education across greater China* (pp. 23-45). Bristol: Multilingual Matters.
- Gillham, B. (2005). *Research interviewing: The range of techniques*. Maidenhead: Open University Press.
- Goetz, T., Frenzel, A. C., Hall, N. C., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, 33, 9–33.  
doi:10.1016/j.cedpsych.2006.12.002
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. doi: 10.1037//0022-3514.74.6.1464
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197–216. doi: 10.1037/0022-3514.85.2.197
- Guba, E. G., & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 191–215). Thousand Oaks, CA: Sage.
- Guimond, S. (2008). Psychological similarities and differences between women and men across cultures. *Social and Personality Psychology Compass*, 2(1), 494–510. doi: 10.1111/j.1751-9004.2007.00036.x
- Guimond, S., & Roussel, L. (2001). Bragging about one's school grades: Gender stereotyping and students' perception of their abilities in science, mathematics, and language. *Social Psychology of Education*, 4(3-4), 275-293. doi: 10.1023/A:1011332704215
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66(3-4), 153-166. doi: 10.1007/s11199-011-9996-2
- Guo, C., Tsang, M., & Ding, X [郭丛斌, 曾满超, 丁小浩]. (2007). Gender difference in the education and employment of science & engineering students in HEIs [中国高校理工类学生教育及就业状况的性别差异]. *Journal of Higher Education [高等教育研究]*, 28(11), 89-101.

- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The Times They Are a-Changing... or Are They Not? A Comparison of Gender Stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3), 353-363. doi: 10.1177/0361684316634081
- Hakstian, A. R., & Cattell, R. B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. *Journal of Educational Psychology*, 70(5), 657-669. doi: 10.1037/0022-0663.70.5.657
- Halim, M. L., & Ruble, D. (2010). Gender identity and stereotyping in early and middle childhood. In J.C. Chrisler & D. R. McCreary, *Handbook of Gender Research in Psychology* (pp. 495-525). NY: Springer.
- Halpern, D. F. (2012) *Sex Differences in Cognitive Abilities* (4<sup>th</sup> ed.). New York: Psychology Press.
- Haslam, S. A., Turner, J. C., Oakes, P. J., Reynolds, K. J., & Doosje, B. (2002). From personal pictures in the head to collective tools in the world: How shared stereotypes allow groups to represent and change social reality. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (pp. 157-185). Cambridge, UK: Cambridge University Press.
- Hedges, L. V., & Nowell, A. (1995, July 7). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41–45. doi: 10.1126/science.7604277
- Herbert, J., Stipek, D. T. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, 26, 276-295. doi: 10.1016/j.appdev.2005.02.007
- Heyman, G. D., & Legare, C. H. (2004). Children's beliefs about gender differences in the academic and social domains. *Sex Roles*, 50(3-4), 227-239. doi: 10.1023/B:SERS.00000015554.12336.30
- Hofstede, G. (1980), *Culture's consequences, international differences in work-related values*. London, England: SAGE Ltd.
- Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture, 2(1). <https://doi.org/10.9707/2307-0919.1014>
- Hu, G., & McKay, S. L. (2012). English language education in East Asia: Some recent developments. *Journal of Multilingual and Multicultural Development*, 33(4), 345-362.

- Huang, Y [黄燕青]. (2011). Gender differences in undergraduates' performance in CET-4 [大学英语四级考试成绩的性别差异]. *Journal of Guangxi College of Education* [广西教育学院学报], 114, 140-143.
- Huo, P. Y., & Randall, D. M. (1991). Exploring subcultural differences in Hofstede's value survey: The case of the Chinese. *Asia Pacific Journal of Management*, 8(2), 159–173. doi: 10.1007/BF01731938
- Hyde, J. S. (2014). Gender similarities and differences. *Annual review of psychology*, 65, 373-398. doi: 10.1146/annurev-psych-010213-115057
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53–69. doi: 10.1037/0033-2909.104.1.53
- Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion how being the target of prejudice affects self-control. *Psychological Science*, 17(3), 262-269. doi: 10.1111/j.1467-9280.2006.01695.x
- Inzlicht, M., & Schmader, T. (2012). Introduction. In M. Inzlicht & T. Schmader, (Eds.) *Stereotype threat: Theory, process, and application* (pp. 3-14). New York; Oxford: Oxford University Press.
- Jon, J., & Chung, H. (2013). *STEM report—Republic of Korea*. Retrieved from <https://acola.org/wp-content/uploads/2018/12/Consultant-Report-Korea.pdf>
- Johns, M. J., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, 137, 691–705. doi: 10.1037/a0013834
- Johnson, H. J., Barnard-Brak, L., Saxon, T. F., & Johnson, M. K. (2012). An experimental study of the effects of stereotype threat and stereotype lift on men and women's performance in mathematics. *The Journal of Experimental Education*, 80(2), 137-149. doi: 10.1080/00220973.2011.567312
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Johnson, W., & Bouchard Jr, T. J. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence*, 35(1), 23-39. doi: 10.1016/j.intell.2006.03.012

- Jones, S., & Myhill, D. (2004). 'Troublesome boys' and 'compliant girls': Gender identity and perceptions of achievement and underachievement. *British Journal of Sociology of Education*, 25(5), 547-561. doi: 10.1080/0142569042000252044
- Jost, J. T., & Kay, A. C. (2005). Exposure to Benevolent Sexism and Complementary Gender Stereotypes: Consequences for Specific and Diffuse Forms of System Justification. *Journal of Personality and Social Psychology*, 88(3), 498–509. doi: 10.1037/0022-3514.88.3.498
- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, 29(3), 371–381. doi: 10.1177/0146167202250218
- Kessels, U. 2005. Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education* 20 (3), 309–323. doi: 10.1007/BF03173559
- Kessels, U., Heyder, A., Latsch, M., & Hannover, B. (2014). How gender differences in academic engagement relate to students' gender identity. *Educational Research*, 56(2), 220-229. doi: 10.1080/00131881.2014.898916
- Kite, M. E., Deaux, K., & Haines, E. L. (2008). Gender stereotypes. In F. L. Denmark & M. A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (2<sup>nd</sup> ed., pp. 205-236). Westport, CT: Praeger Publishers.
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, 9, 1086. doi: 10.3389/fpsyg.2018.01086
- Kohlberg, L.A., 1966. A cognitive-developmental analysis of children's sex role concepts and attitudes. In: Maccoby, E.E. (Ed.), *The Development of Sex Differences* (pp. 82–173). CA, Stanford: Stanford Univ. Press.
- Kollmayer, M., Schober, B., & Spiel, C. (2018). Gender stereotypes in education: Development, consequences, and interventions. *European Journal of Developmental Psychology*, 15(4), 361–377. doi:10.1080/17405629.2016.1193483
- Krendl, A. C., Gainsburg, I., & Ambady, N. (2012). The effects of stereotypes and observer pressure on athletic performance. *Journal of Sport & Exercise Psychology*, 34(1), 3–15. doi: /10.1123/jsep.34.1.3
- Kriesi, I., & Buchmann, M. (2014). Beginning school transition and academic achievement in mid-elementary school. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in*



- aspirations and attainment* (pp. 53–78). Cambridge University Press. doi: 10.1017/CBO9781139128933.005
- Kurtz-Costes, B., Copping, K. E., Rowley, S. J., & Kinlaw, C. R. (2014). Gender and age differences in awareness and endorsement of gender stereotypes about academic abilities. *European Journal of Psychology of Education*, 29(4), 603–618. doi:10.1007/s10212-014-0216-7
- Lamont, R. A., Swift H. J., Abrams D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology of Aging*, 30(1):180–193. doi: 10.1037/a0038586
- Leaper, C., Farkas, T., & Brown, C. S. (2012). Adolescent girls' experiences and gender related beliefs in relation to their motivation in maths/science and English. *Journal of Youth and Adolescence*, 41(3), 268–282. doi:10.1007/s10964-011-9693-z
- Levy, B. R. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, 71(6), 1092–1107.
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, 41, 63–76. doi: 10.1080/0013188990410106
- Liben, L. S., & Bigler, R. S. (2002). The developmental course of gender differentiation: conceptualising, measuring, and evaluating constructs and pathways. *Monographs of the Society for Research in Child Development*, 67, vii-147. doi: 10.1111/1540-5834.t01-1-00187
- Liben, L. S., Bigler, R. S., & Hilliard, L. J. (2014). Gender development: From universality to individuality. In E. T. Gershoff, R. Mistry, & D. A. Crosby (Eds.), *Societal contexts of child development: Pathways of influence and implications for practice and policy* (pp. 3-18). Oxford University Press.
- Lim, L. (2002). Work-related values of Malaysians and Japanese: A Re-Examination of Hofstede's Propositions. *Journal of Transnational Management Development*, 6(3-4), 39-56.
- Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2011). Paradigmatic controversies, contradictions, and emerging confluences, revisited. In N. K. Denzin, & Y. S. Lincoln, *The Sage handbook of qualitative research* (4th Edition ed., pp. 97-128). Thousand Oaks, CA: SAGE Publications, Inc.
- Liu, R. (2018). Gender-math stereotype, biased self-assessment, and aspiration in STEM careers: The gender gap among early adolescents in China. *Comparative Education Review*, 62(4), 522-541. doi: 10.1086/699565

- Liu, W [刘玮玮]. (2008). On gender attitudes among Daoists and in Daoism [论道家和道教的性别观]. *Journal of Hunan University of Technology (Social Science Edition)* [湖南工业大学学报(社会科学版)], 13(4), 82-84.
- Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education*, 22(2), 164-184. doi: 10.1080/08957340902754635
- Liu, M., Hu, W., Jiannong, S., & Adey, P. (2010). Gender stereotyping and affective attitudes towards science in Chinese secondary school students. *International Journal of Science Education*, 32(3), 379-395. doi: 10.1080/09500690802595847
- Liu, D. Z., Huang, H. X., Jia, F. Q., Gong, Q., Huang, Q., & Li, X. [刘电芝, 黄会欣, 贾凤芹, 龚茜, 黄颀, 李霞] (2011). A new sex-role inventory (CSRI-50) indicates changes of sex role among Chinese college students [新编大学生性别角色量表揭示性别角色变迁]. *Acta Psychologica Sinica* [心理学报], 43(6), 639-649. doi: 10.3724/SP.J.1041.2011.00639
- Lummis, M., & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Developmental Psychology*, 26(2), 254-263. doi:10.1037/0012-1649.26.2.254
- Ma L., You, Y., Xiong, Y., Dong, L., Wang, M., & Kou, K. (2016). Gender gap in college major choice: An empirical study on 85 Universities. *Journal of Higher Education*, 37(5), 36-42.
- MacNell, L., Driscoll, A., Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40, 291-303. doi: 10.1007/s10755-014-9313-4
- Maqsud, M. (1993). Relationships of some personality variables to academic attainment of secondary school pupils. *Educational Psychology*, 13, 11-18. doi:10.1080/0144341930130102
- Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, 25, 80-88. doi: 10.1037/0012-1649.25.1.80
- Martinot, D., Bages, C., & Desert, M. (2012). French children's awareness of gender stereotypes about mathematics and reading: when girls improve their reputation in math. *Sex Roles*, 66(3-4), 210-219. doi:10.1007/s11199-011-0032-3

- Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91(2), 243–254. doi: 10.1037/0022-3514.91.2.243
- Mascret, N., & Cury, F. (2015). “I’m not scientifically gifted, I’m a girl”: Implicit measures of gender-science stereotypes – preliminary evidence. *Educational Studies*, 41(4), 462–465. doi: 10.1080/03055698.2015.1043979
- Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum: Qualitative Social Research*, 11(3).
- Maykut, P., & Morehouse, R. (2005). *Beginning qualitative research: A philosophical and practical guide*. London: The Falmer Press.
- Maylor, E. A., Reimers, S., Choi, J., Collaer, M. L., Peters, M., & Silverman, I. (2007). Gender and sexual orientation differences in cognition across adulthood: Age is kinder to women than to men regardless of sexual orientation. *Archives of sexual behavior*, 36(2), 235–249. doi: 10.1007/s10508-006-9155-y
- McGarty, C. (2002). Stereotype formation as category formation. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (pp. 16–37). Cambridge, UK: Cambridge University Press.
- McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). Social, cultural and cognitive factors in stereotype formation. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes as explanations: The formation of meaningful beliefs about social groups* (pp. 1–15). Cambridge, UK: Cambridge University Press.
- McGeown, S. P., & Warhurst, A. (2019). Sex differences in education: Exploring children’s gender identity. *Educational Psychology*, 1–17. doi: 10.1080/01443410.2019.1640349
- McGlone, M. S., & Aronson, J. (2007). Forewarning and forearming stereotype-threatened students. *Communication Education*, 56(2), 119–133. doi: 10.1080/03634520601158681
- Meece, J. L. (1987). The influence of school experiences on the development of gender schemata. In L. S. Liben & M. L. Signorella (Eds.), *Children’s gender schemata* (pp. 57–73). San Francisco: Jossey-Bass.
- Miller, C. F., Trautner, H., & Ruble, D. N. (2006). The role of gender stereotypes in children’s preferences and behavior. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 293–323). New York: Psychology Press.

- Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107*(3), 631-644.
- Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18*(1), 37-45. doi: 10.1016/j.tics.2013.10.011
- Miller, L. R., Bilimoria, R. N., & Pattni, N. (2000). Do women want new men'? Cultural influences on sex-role stereotypes. *Psychology, Evolution & Gender, 2*(2), 127-150. doi: 10.1080/14616660050200913
- Mischel, W. (1970). Sex typing and socialization. In P. H. Mussen (Ed.), *Carmichael's handbook of child psychology* (Vol. 2, pp. 3-72). New York: Wiley.
- Morgan, D. L. (1998). *Planning focus groups*. Thousand Oaks, CA: Sage.
- Morgan, D. (2014). Pragmatism as a paradigm for mixed methods research. In Morgan, D. *Integrating qualitative and quantitative methods* (pp. 25-44). London: SAGE Publications, Inc. doi: 10.4135/9781544304533
- Morse, J. (2010). Procedures and practice of mixed method design: Maintaining control, rigour, and complexity. . In A. Tashakkori, & C. & Teddlie, *SAGE handbook of mixed methods in social & behavioural research* (2nd Edition ed., pp. 339-352).
- Moss-Racusin, C. A., Phelan, J. E., & Rudman, L. A. (2010). When men break the gender rules: Status incongruity and backlash against modest men. *Psychology of Men & Masculinity, 11*(2), 140–151. doi: 10.1037/a0018093
- Muntoni, F., & Retelsdorf, J. (2019). At their children's expense: How parents' gender stereotypes affect their children's reading outcomes. *Learning and Instruction, 60*, 95–103. doi: 10.1016/j.learninstruc.2018.12.002
- Murphy, M. C. & Taylor, V. J. (2012) The role of situational cues in signalling and maintaining stereotype threat. In M. Inzlicht & T. Schmader, (Eds.) *Stereotype threat: Theory, process, and application* (pp. 17-33). New York; Oxford: Oxford University Press.
- Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*(3), 747–759. doi: 10.1037/0012-1649.43.3.747
- National Assessment of Educational Progress. (2011). *Writing Assessment* [Data file]. Retrieved from <https://www.nationsreportcard.gov/ndecore/xplore/NDE>
- National Assessment of Educational Progress. (2015). *Vocabulary Assessment* [Data file]. Retrieved from <https://www.nationsreportcard.gov/ndecore/xplore/NDE>

- National Assessment of Educational Progress. (2019). *Reading Assessment* [Data file]. Retrieved from <https://www.nationsreportcard.gov/ndecore/xplore/NDE>
- National Science Board, National Science Foundation. (2019). Higher Education in Science and Engineering. *Science and Engineering Indicators 2020*. NSB-2019-7. Alexandria, VA. Available at <https://nces.nsf.gov/pubs/nsb20197/>.
- Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics: Theory, Research, and Practice*, 6(1), 101-115.
- Nguyen, H. H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314-1334. doi: 10.1037/a0012702
- Torgimson, B. N., & Minson, C. T. (2005). Sex and gender: What is the difference? *Journal of Applied Physiology*, 99(3), 785-787. doi: 10.1152/japplphysiol.00376.2005
- Operario, D., & Fiske, S. T. (2004). Stereotypes: Content, structures, processes, and context. In M. B. Brewer, M. Hewstone, M. B. Brewer, M. Hewstone (Eds.), *Social cognition* (pp. 120-141). Malden: Blackwell Publishing.
- Organization for Economic Co-operation and Development. (2017). *Education at a glance/ Students, access to education and participation/ Enrolment by gender, program orientation and mode of study/ Enrolment by field* [Data file]. Retrieved from <https://stats.oecd.org>
- Organization for Economic Cooperation and Development (OECD), Programme for International Student Assessment (PISA). (2018). *Reading, Mathematics and Science Assessments* [Data file]. Retrieved from <https://pisadataexplorer.oecd.org/ide/idepisa/>
- Organization for Economic Cooperation and Development (OECD) (2019), *Skills matter: Additional results from the survey of adult skills*. Paris: OECD Publishing. doi: 10.1787/1f029d8f-en.
- Peacock, M. (2001). Pre-service ESL teachers' beliefs about second language learning: A longitudinal study. *System*, 29(2), 177-195. doi: 10.1016/S0346-251X(01)00010-0
- Petersen, J., & Hyde, J. S. (2014). Gender-related academic and occupational interests and goals. In L. S. Liben & R. S. Bigler (eds), *Advances in child development and behaviour* (Vol. 47, pp. 43-76). CA, San Diego: Academic Press.
- Plaks, J. E., Stroessner, S. J., Dweck, C. S., & Sherman, J. W. (2001). Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic

- information. *Journal of Personality and Social Psychology*, 80(6), 876–893. doi: 10.1037/0022-3514.80.6.876
- Plante, I., Theoret, M., & Favreau, O. E. (2009). Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language. *Educational Psychology*, 29(4), 385–405. doi: 10.1080/01443410902971500
- Pomerleau, Q., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first 2 years of life. *Sex Roles*, 22, 359–367. doi: 10.1007/BF00288339
- Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly*, 26(4), 269–281. doi: 10.1111/1471-6402.t01-1-00066
- Prentice, D. A., & Miller, D. T. (2006). Essentialising differences between women and men. *Psychological Science*, 17, 129–135. doi: 10.1111/j.1467-9280.2006.01675.x
- Priess, H. A., & Hyde, J. S. (2010). Gender and academic abilities and preferences. In J.C. Chrisler & D. R. McCreary (eds), *Handbook of Gender Research in Psychology* (pp. 297–316). NY: Springer.
- Pring, R. (2000). The 'false dualism' of educational research. *Journal of Philosophy of Education*, 34(2), 247–260.
- Pritchard, R. M. O. (1987). Boys' and girls' attitudes towards French and German. *Educational Research*, 29(1), 65–72. doi: 10.1080/0013188870290108
- Ramsey, L. R. (2017). Agentic traits are associated with success in science more than communal traits. *Personality and Individual Differences*, 106, 6–9. doi: 10.1016/j.paid.2016.10.017
- Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok — Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, 48(3), 731–737. doi: 10.1016/j.jesp.2011.12.012
- Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PloS one*, 7(7), e39904. doi: 10.1371/journal.pone.0039904
- Retelsdorf, J., Schwartz, K. & Asbrock, F. (2015) "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, 107(1), 186–194.

- Rhodes, M., & Brickman, D. (2008). Pre-schoolers' responses to social comparisons involving relative failure. *Psychological Science*, 19, 968–972. doi: 10.1111/j.1467-9280.2008.02184.x
- Rhodes, M., & Gelman, S. A. (2008, January). *Cultural and developmental influences on the conceptual structure of social categories*. Poster presented at the annual meeting of the Society for Personality and Social Psychology, Albuquerque, NM.
- Rhodes, M., & Gelman, S.A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, 59, 244–274. doi: 10.1016/j.cogpsych.2009.05.001
- Ruble, D. N., Alvarez, J., Bachman, M., Cameron, J., Fuligni, A., Coll, C. G., & Rhee, E. (2004). The development of a sense of “we”: The emergence and implications of children's collective identity. In M. Bennett & F. Sani (Eds.), *The development of the social self* (pp.29-75). Sussex, UK: Psychology Press.
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender Development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (pp. 858–932). John Wiley & Sons Inc.
- Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in organizational behaviour*, 28, 61-79. doi: 10.1016/j.riob.2008.04.003
- Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. Retrieved from <http://www.oecd.org/pisa/publications/pisa-2018-results.htm>
- Schmader, T., & Beilock, S. L. (2012). An integration of processes that underlie stereotype threat. In M. Inzlicht & T. Schmader, (Eds.) *Stereotype threat: Theory, process, and application* (pp. 34-50). New York; Oxford: Oxford University Press.
- Schmader, T., Whitehead, J., & Wysocki, V. H. (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex roles*, 57(7-8), 509-514. doi: 10.1007/s11199-007-9291-4
- Schmenk, B. (2004). Language learning: a feminine domain? The role of stereotyping in constructing gendered learner identities. *TESOL Quarterly*, 38(3), 514-538. doi: 10.2307/3588352
- Schoon, I., & Eccles, J. S. (2014). Introduction: Conceptualizing gender differences in aspirations and attainment. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment: A life course perspective* (pp.3-26). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781139128933.002

- Seale, C. (1999). Quality in qualitative research. *Qualitative inquiry*, 5(4), 465-478.
- Seginer, R., & Vermulst, A. (2002). Family environment, educational aspirations, and academic achievement in two cultural settings. *Journal of Cross-Cultural Psychology*, 33, 540–558. doi:10.1177/ 00220022102238268
- Seliger, H. W., & Shohamy, E. (1989). *Second language research methods*. Oxford, UK: Oxford University Press.
- Shannon-Baker, P. (2016). Making Paradigms Meaningful in Mixed Methods Research. *Journal of Mixed Methods Research*, 10(4), 319–334. doi: 10.1177/1558689815575861
- Shields, S. A., & Dicicco, E. C. (2011). The Social Psychology of Sex and Gender: From Gender Differences to Doing Gender. *Psychology of Women Quarterly*, 35(3), 491–499. doi: 10.1177/0361684311414823
- Shih, M. J., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80–83. doi: doi.org/10.1111/1467-9280.00111
- Shteivi, M. (2003). Gender role stereotypes in primary school textbooks in Jordan. *Dirasat: Human and Social Sciences*, 30(1), 89–104.
- Shu, X., & Zhu, Y. (2012). Uneven transitions: Period-and cohort-related changes in gender attitudes in China, 1995–2007. *Social science research*, 41(5), 1100-1115. doi: 10.1016/j.ssresearch.2012.05.004
- Siebert, L. L. (2003). Student and teacher beliefs about language learning. *The ORTESOL Journal*, 21, 7–39.
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, 13, 147–183. doi:10.1006/drev.1993.1007
- Silverman, D. (2013). *Interpreting Qualitative Data* (4th Edition ed.). London: SAGE publications.
- Siyanova-Chanturia, A., Warren, P., Pesciarelli, F., & Cacciari, C. (2015). Gender stereotypes across the ages: On-line processing in school-age children, young and older adults. *Frontiers in psychology*, 6, 1388.
- Skelton, C., Francis, B., & Smulyan, L. (2006). *The SAGE handbook of gender and education*. London: SAGE. doi: 10.4135/9781848607996
- Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, 67, 617-629. doi: 10.1007/s11199-012-0209-4



- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415-437. doi: 10.1146/annurev-psych-073115-103235
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology*, 35(1), 4-28.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Cambridge, MA: Cambridge.
- Stangor, C., & Schaller, M. (2000). Stereotypes as individual and collective representations. In C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings* (pp. 64-82). Ann Arbor, MI: Edwards Brothers.
- Steensma, T. D., Kreukels, B. P., de Vries, A. L., & Cohen-Kettenis, P. T. (2013). Gender identity development in adolescence. *Hormones and behaviour*, 64(2), 288-297. doi: 10.1016/j.yhbeh.2013.02.020
- Steffens, M. C., & Jelenec, P. (2011). Separating implicit gender stereotypes regarding math and language: Implicit ability stereotypes are self-serving for boys and men, but not for girls and women. *Sex Roles*, 64, 324-335. doi: 10.1007/s11199-010-9924-x
- Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: What are the roles of personality and achievement motivation?. *European Journal of Personality: Published for the European Association of Personality Psychology*, 22(3), 185-209. doi: 10.1002/per.676
- Steinmetz, J., Bosak, J., Szczesny, S., & Eagly, A. H. (2014). Social role effects on gender stereotyping in Germany and Japan. *Asian Journal of Social Psychology*, 17(1), 52-60. doi: 10.1111/ajsp.12044
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of personality and social psychology*, 69(5), 797-811. doi: 10.1037//0022-3514.69.5.797
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379-440. doi: 10.1016/S0065-2601(02)80009-0
- Stone, J. (2002). Battling Doubt by Avoiding Practice: The Effects of Stereotype Threat on Self-Handicapping in White Athletes. *Personality and Social Psychology Bulletin*, 28(12), 1667-1678. doi: 10.1177/014616702237648
- Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching*, 33(04), 203-223. doi: 10.1017/S0261444800015688

- Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, 66, 21-36. doi: 10.1037/0022-3514.66.1.21
- Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin*, 22(5), 507-519. doi: 10.1177/0146167296225008
- Tajfel, H., & Forgas, J. P. (2000). Social categorisation: Cognitions, values and groups. In C. Stangor (Ed.), *Stereotypes and prejudice: Essential readings* (pp. 49-63). Ann Arbor, MI: Edwards Brothers.
- Taylor, F., & Marsden, E. J. (2014). Perceptions, attitudes, and choosing to study foreign languages in England: An experimental intervention. *The Modern Language Journal*, 98(4), 902-920. doi: 10.1111/modl.12146
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioural sciences*. Thousand Oaks, CA: SAGE Publications, Inc.
- Tenenbaum, H. R. (2009). 'You'd be good at that': Gender patterns in parent-child talk about courses. *Social Development*, 18(2), 447-463. doi: 10.1111/j.1467-9507.2008.00487.x
- Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities? *Developmental psychology*, 39(1), 34-47. doi: 10.1037/0012-1649.39.1.34
- The International English Language Testing System (IELTS). (2018). [Graph illustration of mean IELTS performance by gender]. *Test taker performance 2018*. Retrieved from <https://www.ielts.org/teaching-and-research/test-taker-performance>
- The State Council. (2014). *Implement Opinions of the State Council on Deepening the Reform of the enrolment and entrance examination system* (TSC Publication No. [2014]35). Beijing, BJ: The State Council of the People's Republic of China.
- Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, 92(1), 144-151. doi: 10.1037/0022-0663.92.1.144
- Tomasetto, C., Mirisola, A., Galdi, S., & Cadinu, M. (2015). Parents' math-gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemporary Educational Psychology*, 42, 186-198. doi: 10.1016/j.cedpsych.2015.06.007

- Tsui, M. [崔明]., & Venator, E. (2008) Stereotype threat and the academic performance of Chinese students [传统固有观念的心理暗示与中国学生的测试表现]. *Society: Chinese Journal of Sociology [社会]*, 28,191-202.
- Upadyaya, K., & Eccles, J. S. (2014). Gender differences in teachers' perceptions and children's ability self-concepts. In I. Schoon & J. S. Eccles (Eds.), *Gender differences in aspirations and attainment* (pp. 79–100). Cambridge University Press. doi: 10.1017/CBO9781139128933.006
- Van de Gaer, E., Pustjens, H., Van Damme, J., & De Munter, A. (2009). School engagement and language achievement: A longitudinal study of gender differences across secondary school. *Merrill-Palmer Quarterly (55)*4, 373-405.
- van de Vijver, F. J. R. (2007) Cultural and gender differences in gender-role beliefs, sharing household task and child-care responsibilities, and well-being among immigrants and majority members in the Netherlands. *Sex Roles*, 57, 813-824. doi: 10.1007/s11199-007-9316-z
- van der Slik, F. W., van Hout, R. W., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PloS one*, 10(11), e0142056. doi: 0.1371/journal.pone.0142056
- Vera, H. (2016). Rebuilding a classic: 'The Social Construction of Reality' at 50. *Cultural Sociology*, 10(1), 3-20.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. doi: 10.1037/a0036620
- Wall, K., & Statistics Canada. (2019). *Persistence and representation of women in STEM programmes. insights on Canadian society*. Statistics Canada. Statistics Canada. Retrieved from <https://search-ebshost-com.ezp.lib.cam.ac.uk/login.aspx?direct=true&db=eric&AN=ED594933&site=ehost-live&scope=site>
- Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, 39(5), 456– 467. doi: 10.1016/S0022-1031(03)00019-2
- Wang, R. R. (2003) *Images of women in Chinese thought and culture: Writings from the pre-Qin period through the Song Dynasty*. Indianapolis, IN: Hackett Publishing.

- Waterman, A. S. (1982). Identity development from adolescence to adulthood: An extension of theory and a review of research. *Developmental psychology*, 18(3), 341-358. doi: 10.1037/0012-1649.18.3.341
- Watson, A., Kehler, M., & Martino, W. (2010). The problem of boys' literacy underachievement: Raising some questions. *Journal of Adolescent & Adult Literacy*, 53(5), 356-361. DOI: <https://doi.org/10.1598/JAAL.53.5.1>
- Wengraf, T. (2001). *Qualitative research interviewing: Biographic narrative and semi-structured methods*. London: SAGE Publications, Inc.
- Wei, R., & Su, J. (2012). Statistics of English in China. *English today*, 28(3), 10-14.
- Weschler, D. (1995). Weschler adult intelligence scale. *San Antonio, TX: The Psychological Corporation*.
- Whitehead, J. M. (1996). Sex stereotypes, gender identity and subject choice at A-level. *Educational Research*, 38(2), 147-160. doi: 10.1080/0013188960380203
- Williams, J. E., Satterwhite, R. C., & Best, D. L. (1999). Pancultural gender stereotypes revisited: The five factor model. *Sex roles*, 40(7-8), 513-525. doi: 10.1023/A:1018831928829
- Williams, M., Burden, R., & Lanvers, U. (2002). 'French is the language of love and stuff': Student perceptions of issues related to motivation in learning a foreign language. *British educational research journal*, 28(4), 503-528. doi: 10.1080/0141192022000005805
- Wood, W., & Eagly, A. H. (2010). Gender. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., vol. 1, pp. 629-668). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behaviour. In J. M. Olson, & M. P. Zanna (Eds.), *Advances in experimental social psychology* (vol. 46, pp. 55-123). Burlington: Academic Press.
- Xiao, W [肖巍]. (2017, Oct. 17). Analysis of traditional Confucian thinking and modern, gender-egalitarian attitudes [儒家传统思想与现代男女平等观念之辩]. Retrieved from <http://paper.cnwomen.com.cn/content/2017-10/17/042901.html>
- Xu, Q. [许琪]. (2016). Trend, source and heterogeneity of the change of gender-role attitude in China: A case of two indicators [中国人性别观念的变迁趋势、来源和异质性——以“男主外，女主内”和“干得好不如嫁得好”两个指标为例]. *Collection of Women's Studies [妇女研究论丛]*, 3, 33-43.

- Yang, J., Li, H., & Zhu, G. [杨菊华, 李红娟, 朱格]. (2014). Changing trends in gender outlook in China from 1990 to 2010 [近 20 年中国人性别观念的变动趋势与特点分析]. *Collection of Women's Studies [妇女研究论丛]*, 6, 28-36.
- Yao, Y. [姚艳梅]. (2011) Reasons for male students' underachievement in foreign language studies in higher education institutions and proposed strategies [高校外语专业男生学习弱势的原因及对策]. *Journal of Hebei University (Philosophy and Social Sciences)* [河北大学学报(哲学社会科学版)], 36(1), 138-140.
- Yin, R. (2014). *Case study research: Design and methods*. Thousand Oaks, CA: SAGE publications.
- Zhang, H [张红萍]. (2015, Sept. 15). Rethinking the gender attitudes of Confucius and the Confucian tradition [孔子及儒家女性观再辨析]. Retrieved from <http://paper.cnwomen.com.cn/content/2015-09/15/020508.html>
- Zhang, X [张祥龙]. (2002). The status and consequences of 'gender difference' in Chinese and Western Philosophy [“性别”在中西哲学里的地位及其思想后果]. *Jiangsu Social Sciences [江苏社会科学]*, 6, 1-9.
- Zhao, H. (2011). *Gender construction and negotiation in the Chinese EFL classroom*. Cambridge Scholars Publishing.
- Zhao, J., Li, P., & Yin, L [赵建芬, 李蓓, 尹丽娟]. (2014). The influence of gender stereotypes on students' language learning [性别刻板印象对学生语言表达能力的影响]. *Journal of Shijiazhuang University [石家庄学院学报]*, 16(4), 87-91. doi: 10.13573/j.cnki.sjzxyxb.2014.04.040

## Appendix A: Online Questionnaire for Guardians and Teachers in Pilot Study 1

*Dear Guardian,*

*You are invited to take part in a survey about how parents attribute academic success in senior secondary education. This survey is conducted by Miss Jing LI, a PhD. student from the Faculty of Education at University of Cambridge. It is part of her PhD. research project.*

*This survey has two parts. Your participation will require approximately 5-10 minutes and is completed online at your computer. There are no known risks or discomforts associated with this survey. Taking part in this study is completely voluntary. If you choose to be in the study you can withdraw at any time. Your responses will be kept strictly confidential, and digital data will be stored in secure computer files. Any report of this research that is made available to the public will not include your name or any other individual information by which you could be identified. If you have questions or want a copy or summary of this study's results, you can contact the researcher at the following email address: jl806@cam.ac.uk.*

*Click ☐ **Start**: This means that you have agreed to participate in the survey. and will answer the questions based on your opinion.*

*Click ☐ **Quit**: This means that you have decided to not to participate in the survey. Thank you for your support.*

尊敬的家长，

您好！本研究项目由李婧（现为英国剑桥大学教育学系在读博士生）所设计和领导，目的在于发掘高中生家长对于高中各个学科的学业成败因素的态度和观点。

本问卷共有两部分，填写大约需要5-

10分钟。在回答本问卷时，您的所有回答将会被严格保密，研究者承诺这些回答仅供用于研究目的。问卷回答完毕之后，如果您想对本项目做进一步了解，可通过以下电子邮箱联系研究者：jl806@cam.ac.uk

点击☐ **开始**：表示您同意参加本次研究，并保证根据个人的真实想法填写问卷。

点击☐ **退出**：表示您不同意参加本次研究。再次感谢您的支持。

### Part 1

*Please read each question and select the number that best corresponds to your attitude or belief. Among the 11 numbers, 0 means 'Not at all', 5 means 'Cannot decide', and 10 means 'Totally'.*

请您阅读每一道题后，选出最符合您观点的数字。其中，0表示“完全反对”，5表示“无法确定”，10表示“完全同意”。

NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定						TOTALLY 完全 同意			
1	Do you believe that girls are gifted in English? 你是否认为女生普遍擅长学习英语？	0	1	2	3	4	5	6	7	8	9	10
2	Do you believe that girls are willing to learn English? 你是否认为女生普遍喜欢学习英语？	0	1	2	3	4	5	6	7	8	9	10
3	Do you believe that girls do well in English? 你是否认为女生普遍英语成绩好？	0	1	2	3	4	5	6	7	8	9	10

## Part 2

Please read each question and select the number that best corresponds to your attitude or belief. Among the 11 numbers, 0 means 'Not at all', 5 means 'Cannot decide', and 10 means 'Totally'.

请您阅读每一道题后，选出最符合您观点的数字。其中，0表示“完全反对”，5表示“无法确定”，10表示“完全同意”。

NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定						TOTALLY 完全 同意			
1	Do you believe that boys are gifted in English? 你是否认为男生普遍擅长学习英语？	0	1	2	3	4	5	6	7	8	9	10
2	Do you believe that boys are willing to learn English? 你是否认为男生普遍喜欢学习英语？	0	1	2	3	4	5	6	7	8	9	10
3	Do you believe that boys do well in English? 你是否认为男生普遍英语成绩好？	0	1	2	3	4	5	6	7	8	9	10

➤ Please indicate your relationship to your child in XX High School:

请选择您的性别：

☐ Mother 女

☐ Father 男

\*\* This is the end of the questionnaire. Thank you for your participation!

问卷题目到此结束。感谢您的参与！

## Appendix B: Online Questionnaire for Students in Pilot Study 1

*Dear Student,*

*You are invited to take part in a survey about how students attribute academic success in the English subject. This survey is conducted by Miss Jing LI, a PhD. student from the Faculty of Education at University of Cambridge. It is part of her PhD. research project.*

*This survey has two parts. Your participation will require approximately 8-10 minutes and is completed online at your computer. There are no known risks or discomforts associated with this survey. Taking part in this study is completely voluntary. If you choose to be in the study you can withdraw at any time. Your responses will be kept strictly confidential, and digital data will be stored in secure computer files. Any report of this research that is made available to the public will not include your name or any other individual information by which you could be identified. If you have questions or want a copy or summary of this study's results, you can contact the researcher at the following email address: jl806@cam.ac.uk.*

*Click ☐ Start: This means that you have agreed to participate in the survey. and will answer the questions based on your opinion.*

*Click ☐ Quit: This means that you have decided to not to participate in the survey. Thank you for your support.*

亲爱的同学，

你好！本研究项目由李婧（现为英国剑桥大学教育学系在读博士生）所设计和领导，目的在于发掘高中生对于高中英语课的学业成败因素的态度和观点。

本问卷共有两部分，填写大约需要8-

10分钟。在回答本问卷时，你的所有回答将会被严格保密，研究者承诺这些回答仅供用于研究目的。问卷回答完毕之后，如果你想对本项目做进一步了解，可通过以下电子邮箱联系研究者：jl806@cam.ac.uk

点击☐开始：表示你同意参加本次研究，并保证根据个人的真实想法填写问卷。

点击☐退出：表示你不同意参加本次研究。再次感谢您的支持。









### Part 1





*Please read each question and select the number that best corresponds to your attitude or belief. Among the 11 numbers, 0 means 'Not at all', 5 means 'Cannot decide', and 10 means 'Totally'.*



请你阅读每一道题后，选出最符合你观点的数字。其中，0表示“完全反对”，5表示“无法确定”，10表示“完全同意”。

NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定						TOTALLY 完全 同意			
1	Do you believe that girls are gifted in understanding English utterances? 你是否认为女生普遍擅长理解英文录音？	0	1	2	3	4	5	6	7	8	9	10
2	Do you believe that girls are gifted in communicating in English? 你是否认为女生普遍擅长用英文对话？	0	1	2	3	4	5	6	7	8	9	10
3	Do you believe that girls are gifted in understanding English texts? 你是否认为女生普遍擅长读懂英文文章？	0	1	2	3	4	5	6	7	8	9	10
4	Do you believe that girls are gifted in writing English essays? 你是否认为女生普遍擅长用英文写作？	0	1	2	3	4	5	6	7	8	9	10
5	Do you believe that girls are gifted in learning English grammar? 你是否认为女生普遍擅长学习英文语法？	0	1	2	3	4	5	6	7	8	9	10
6	Do you believe that girls are gifted in memorising English vocabulary? 你是否认为女生普遍擅长记英文单词？	0	1	2	3	4	5	6	7	8	9	10
7	Do you believe that girls are willing to develop their English listening skills? 你是否认为女生普遍愿意培养自己的英文听力？	0	1	2	3	4	5	6	7	8	9	10
8	Do you believe that girls are willing to practice English speaking skills? 你是否认为女生普遍喜欢练习英文口语？	0	1	2	3	4	5	6	7	8	9	10
9	Do you believe that girls are willing to read in English? 你是否认为女生普遍喜欢阅读英文？	0	1	2	3	4	5	6	7	8	9	10




10	Do you believe that girls are willing to learn skills in writing English essays? 你是否认为女生普遍喜欢学习英文写作技巧？	0 1 2 3 4 5 6 7 8 9 10 		
NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定	TOTALLY 完全 同意
11	Do you believe that girls are willing to learn English grammar? 你是否认为女生普遍喜欢学习英文语法知识？	0 1 2 3 4 5 6 7 8 9 10 		
12	Do you believe that girls are willing to learn English vocabulary? 你是否认为女生普遍喜欢学习英文词汇知识？	0 1 2 3 4 5 6 7 8 9 10 		
13	Do you believe that girls do well in English listening? 你是否认为女生普遍英文听力的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
14	Do you believe that girls do well in English speaking? 你是否认为女生普遍英文口语的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
15	Do you believe that girls do well in English reading? 你是否认为女生普遍英文阅读的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
16	Do you believe that girls do well in English writing? 你是否认为女生普遍英文作文写得好？	0 1 2 3 4 5 6 7 8 9 10 		
17	Do you believe that girls have a solid grasp of English grammar? 你是否认为女生普遍英文语法知识巩固？	0 1 2 3 4 5 6 7 8 9 10 		









18	Do you believe that girls have a large English vocabulary? 你是否认为女生普遍英文词汇量大?	0 1 2 3 4 5 6 7 8 9 10 
19	Do you believe that girls are gifted in English? 你是否认为女生普遍擅长学习英语?	0 1 2 3 4 5 6 7 8 9 10 
NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对 CANNOT DECIDE 无法 确定 TOTALLY 完全 同意
20	Do you believe that girls are willing to learn English? 你是否认为女生普遍喜欢学习英语?	0 1 2 3 4 5 6 7 8 9 10 
21	Do you believe that girls do well in English? 你是否认为女生普遍英语成绩好?	0 1 2 3 4 5 6 7 8 9 10 









## Part 2



Please read each question and select the number that best corresponds to your attitude or belief. Among the 11 numbers, 0 means 'Not at all', 5 means 'Cannot decide', and 10 means 'Totally'.

请你阅读每一道题后，选出最符合你观点的数字。其中，0表示“完全反对”，5表示“无法确定”，10表示“完全同意”。

NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对 CANNOT DECIDE 无法 确定 TOTALLY 完全 同意
1	Do you believe that boys are gifted in understanding English utterances? 你是否认为男生普遍擅长理解英文录音?	0 1 2 3 4 5 6 7 8 9 10 
2	Do you believe that boys are gifted in communicating in English? 你是否认为男生普遍擅长用英文对话?	0 1 2 3 4 5 6 7 8 9 10 
3	Do you believe that boys are gifted in understanding English texts? 你是否认为男生普遍擅长读懂英文文章?	0 1 2 3 4 5 6 7 8 9 10 

4	Do you believe that boys are gifted in writing English essays? 你是否认为男生普遍擅长用英文写作？	0 1 2 3 4 5 6 7 8 9 10 		
5	Do you believe that boys are gifted in learning English grammar? 你是否认为男生普遍擅长学习英文语法？	0 1 2 3 4 5 6 7 8 9 10 		
NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定	TOTALLY 完全 同意
6	Do you believe that boys are gifted in memorising English vocabulary? 你是否认为男生普遍擅长记英文单词？	0 1 2 3 4 5 6 7 8 9 10 		
7	Do you believe that boys are willing to develop their English listening skills? 你是否认为男生普遍愿意培养自己的英文听力？	0 1 2 3 4 5 6 7 8 9 10 		
8	Do you believe that boys are willing to practice English speaking skills? 你是否认为男生普遍喜欢练习英文口语？	0 1 2 3 4 5 6 7 8 9 10 		
9	Do you believe that boys are willing to read in English? 你是否认为男生普遍喜欢阅读英文？	0 1 2 3 4 5 6 7 8 9 10 		
10	Do you believe that boys are willing to learn skills in writing English essays? 你是否认为男生普遍喜欢学习英文写作技巧？	0 1 2 3 4 5 6 7 8 9 10 		
11	Do you believe that boys are willing to learn English grammar? 你是否认为男生普遍喜欢学习英文语法知识？	0 1 2 3 4 5 6 7 8 9 10 		

12	Do you believe that boys are willing to learn English vocabulary? 你是否认为男生普遍喜欢学习英文词汇知识？	0 1 2 3 4 5 6 7 8 9 10 		
13	Do you believe that boys do well in English listening? 你是否认为男生普遍英文听力的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
NO. 题号	QUESTION 题目	NOT AT ALL 完全 反对	CANNOT DECIDE 无法 确定	TOTALLY 完全 同意
14	Do you believe that boys do well in English speaking? 你是否认为男生普遍英文口语的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
15	Do you believe that boys do well in English reading? 你是否认为男生普遍英文阅读的成绩好？	0 1 2 3 4 5 6 7 8 9 10 		
16	Do you believe that boys do well in English writing? 你是否认为男生普遍英文作文写得好？	0 1 2 3 4 5 6 7 8 9 10 		
17	Do you believe that boys have a solid grasp of English grammar? 你是否认为男生普遍英文语法知识巩固？	0 1 2 3 4 5 6 7 8 9 10 		
18	Do you believe that boys have a large English vocabulary? 你是否认为男生普遍英文词汇量大？	0 1 2 3 4 5 6 7 8 9 10 		
19	Do you believe that boys are gifted in English? 你是否认为男生普遍擅长学习英语？	0 1 2 3 4 5 6 7 8 9 10 		

20	Do you believe that boys are willing to learn English? 你是否认为男生普遍喜欢学习英语？	<div>0 1 2 3 4 5 6 7 8 9 10</div> <div></div>
21	Do you believe that boys do well in English? 你是否认为男生普遍英语成绩好？	<div>0 1 2 3 4 5 6 7 8 9 10</div> <div></div>

### DEMOGRAPHICS 个人信息

- Please choose your gender: 请选择你的性别：
  - ☐ Female 女
  - ☐ Male 男
  
- Please choose your year of study: 请选择你的年级：
  - ☐ Year 1 高一
  - ☐ Year 2 高二
  - ☐ Year 3 (including students repeating their third year) 高三 (含复读学生)
  
- Please choose your academic branch: 请选择你所在的班级类型：
  - ☐ The Liberal Arts Branch 文科班
  - ☐ Haven't been allocated to either branch 尚未分文理班
  - ☐ The Science Branch 理科班
  - ☐ My school has abolished the branching system 所在学校/地区取消文理分科制

\*\* This is the end of the questionnaire. Thank you for your participation! 问卷题目到此结束。感谢你的参与！

## Appendix C: Paper-and-Pencil Questionnaire for Guardians and Teachers in Study 1

Dear guardian,

Thank you for consenting to participate in the survey. This questionnaire is designed for guardians of high school students. It has three parts, and your participation will require approximately 5 minutes.

尊敬的监护人，

您好！感谢您参加本研究。《问卷一》是为中学生的监护人所设计的，共有三部分，完成时间约 5 分钟。

### Part 1 第一部分

Below are three statements describing learning activities and abilities of Chinese high school students. Please read each statement and circle the number that best reflects YOUR OWN attitude or belief. Among the 7 numbers, 1 means 'Totally Disagree', 4 means 'Cannot decide', and 7 means 'Totally Agree'.

下表包含三个针对中国高中学生的学习行为和学习能力的陈述。请您阅读每个陈述后，圈出右侧箭头中最符合您个人态度或看法的数字。其中，1 表示“完全反对”，4 表示“无法确定”，7 表示“完全同意”。

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
1	Do you believe that girls are gifted in English. 在我看来，一般来说，女生有学习英语的天赋。	1	2	3	4	5	6	7
2	I believe that generally speaking, girls do well in English. 我认为女生往往英语成绩好。	1	2	3	4	5	6	7
3	I suppose that girls usually are willing to learn English. 我相信通常情况下，女生喜欢学习英语。	1	2	3	4	5	6	7

## Part 2 第二部分

Below are three statements describing learning activities and abilities of Chinese high school students. Please read each statement and circle the number that best reflects YOUR OWN attitude or belief. Among the 7 numbers, 1 means 'Totally Disagree', 4 means 'Cannot decide', and 7 means 'Totally Agree'.

下表包含三个针对中国高中学生的学习行为和学习能力的陈述。请您阅读每个陈述后，圈出右侧箭头中最符合您个人态度或看法的数字。其中，1 表示“完全反对”，4 表示“无法确定”，7 表示“完全同意”。

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
1	Do you believe that boys are gifted in English. 在我看来，一般来说，男生有学习英语的天赋。	1	2	3	4	5	6	7
2	I suppose that boys usually are willing to learn English. 我相信通常情况下，男生喜欢学习英语。	1	2	3	4	5	6	7
3	I believe that generally speaking, boys do well in English. 我认为男生往往英语成绩好。	1	2	3	4	5	6	7



### Part 3 第三部分

Please fill out the table about your background by circling the options that apply to you.

请你根据个人情况填写以下表格：圈出符合你个人情况的选项即可。

Age 年龄				Education 最高学历			Ethnicity 民族	
36-40	41-45	46-50	Other: <u>    </u> 其他: <u>    </u>	No Degree 本科以下	Degree or Equivalent 大专及本科	Post-graduate Degree 研究生及以上	Chinese 汉族	Others: <u>            </u> 其他: <u>            </u>
Occupation 职业 <sup>50</sup>								
Administrative staff in public institutions 机关和事业单位管理人员		Administrative staff in large and medium-sized enterprises 大中型企业高中层管理人员		Private entrepreneur 私营企业主	Professional and technical personnel 专业技术人员	Service personnel in business 商业服务业人员	Mobile worker 进城务工人员	
Industrial worker 产业工人		Agricultural labourer 农业劳动者 (农林牧渔)		Military 军人	Pensioner 离退休人员	Unemployed 无业	Others: <u>            </u> 其他: <u>            </u>	

Lastly, it is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise months of effort (the researchers') and the time of other participants could be wasted. So please answer the following three questions by circling the choice that applies to you.

为了确保研究研究结果的真实性，我们希望能尽可能多地收集到真实、有效的数据；否则，研究者的精力和所有参与研究人员的时间都可能会被浪费。因此，希望你能如实回答以下问题，帮助研究者判断你的回答的可靠性。

I gave this survey 我在答问卷时的专心程度为：	<input type="radio"/> almost none of <input type="radio"/> 几乎为零	<input type="radio"/> very little of <input type="radio"/> 不太专心	<input type="radio"/> some of <input type="radio"/> 有些专心	<input type="radio"/> quite a bit of <input type="radio"/> 比较专心	my attention. 。
-------------------------------------	--	--	---	--	--------------------

\*\* This is the end of the questionnaire. Thank you for your participation! 问卷到此结束。感谢您的参与！

<sup>50</sup> The occupations are divided based on the demographic sections found in China Household Finance Survey (<http://www.chfsdata.org/intro-14.html>).

## Appendix D: Paper-and-Pencil Questionnaire for Students in Study 1

Dear Student,

Thank you for consenting to participate in the survey. This questionnaire is designed for high school students. It has three parts, and your participation will require approximately 5 minutes.

亲爱的同学，

你好！感谢你参加本研究。《问卷三》是为高中学生所设计的，共有三部分，完成时间约 8 分钟。

### Part 1 第一部分

Below are a number of statements describing learning activities and abilities of Chinese high school students. Please read each statement and circle the number that best reflects YOUR OWN attitude or belief. Among the 7 numbers, 1 means 'Totally Disagree', 4 means 'Cannot decide', and 7 means 'Totally Agree'.

下表包含一系列针对中国高中学生的学习行为和学习能力的陈述。请您阅读每个陈述后，圈出右侧箭头中最符合您个人态度或看法的数字。其中，1 表示“完全反对”，4 表示“无法确定”，7 表示“完全同意”。

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
1	Do you believe that girls do well in English writing. 我相信女生往往英语作文写得好。	1	2	3	4	5	6	7
2	I guess that typically, girls have a solid grasp of English grammar. 我觉得女生往往英文语法知识巩固。	1	2	3	4	5	6	7
3	I guess girls are usually willing to develop English listening skills. 在我看来，女生通常愿意培养自己的英文听力。	1	2	3	4	5	6	7

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
4	I believe that generally, girls are gifted in communicating in English. 我觉得女生普遍擅长用英文对话。	1	2	3	4	5	6	7
5	I believe that usually, girls are gifted in understanding English texts. 在我看来，女生通常擅长读懂英文文章。	1	2	3	4	5	6	7
6	I suppose girls are usually willing to practice English speaking skills. 我认为通常情况下，女生喜欢练习英文口语。	1	2	3	4	5	6	7
7	I suppose girls are usually willing to learn English grammar.我觉得 通常情况下，女生喜欢学习英文语法知识。	1	2	3	4	5	6	7
8	I suppose girls are usually gifted in understanding English utterances. 我相信通常情况下，女生擅长听懂英语语音材料。	1	2	3	4	5	6	7
9	My high school does not offer English classes. 我所在的高中没有开设英语课。	1	2	3	4	5	6	7
10	I suppose that girls are generally willing to learn English vocabulary. 我认为一般来说，女生喜欢学习英文词汇知识。	1	2	3	4	5	6	7
11	Do you believe that girls do well in English reading. 我认为女生往往英文阅读的成绩好。	1	2	3	4	5	6	7
12	I believe that typically, girls are gifted in writing English essays. 我觉得女生普遍擅长用英文写作。	1	2	3	4	5	6	7
13	I suppose girls are usually gifted in memorising English vocabulary. 我认为通常情况下，女生擅长记英文单词。	1	2	3	4	5	6	7

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
14	I am convinced that girls do well in English listening. 我相信女生普遍英文听力的成绩好。	1	2	3	4	5	6	7
15	I believe that typically, girls do well in English speaking. 我相信女生普遍英语口语好。	1	2	3	4	5	6	7
16	Do you believe thatgenerally speaking, girls are willing to read in English. 在我看来，女生一般喜欢阅读英文。	1	2	3	4	5	6	7
17	I suppose typically, girls are willing to learn English writing skills. 我觉得一般情况下，女生喜欢学习英文写作技巧。	1	2	3	4	5	6	7
18	I believe that typically, girls are gifted in learning English grammar. 我相信一般来说，女生擅长学习英文语法。	1	2	3	4	5	6	7
19	I am convinced that girls have a large English vocabulary. 在我看来，一般来说，女生英语词汇量大。	1	2	3	4	5	6	7
20	I am convinced that girls typically are willing to learn English. 我认为一般来说，女生喜欢学习英语。	1	2	3	4	5	6	7
21	I believe that typically, girls are gifted in English. 我觉得女生普遍有学英语的天赋。	1	2	3	4	5	6	7
22	Do you believe thatgirls girls do well in English. 我觉得女生通常英语成绩好。	1	2	3	4	5	6	7

## Part 2 第二部分

Below are a number of statements describing learning activities and abilities of Chinese high school students. Please read each statement and circle the number that best reflects YOUR OWN attitude or belief. Among the 7 numbers, 1 means 'Totally Disagree', 4 means 'Cannot decide', and 7 means 'Totally Agree'.

下表包含一系列针对中国高中学生的学习行为和学习能力的陈述。请您阅读每个陈述后，圈出右侧箭头中最符合您个人态度或看法的数字。其中，1 表示“完全反对”，4 表示“无法确定”，7 表示“完全同意”。

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
1	I suppose boys are usually gifted in understanding English utterances. 我相信通常情况下，男生擅长听懂英语语音材料。	1	2	3	4	5	6	7
2	I believe that typically, boys are gifted in writing English essays. 我觉得男生普遍擅长用英文写作。	1	2	3	4	5	6	7
3	I suppose that boys are generally willing to learn English vocabulary. 我认为一般来说，男生喜欢学习英文词汇知识。	1	2	3	4	5	6	7
4	I guess that typically, boys have a solid grasp of English grammar. 我觉得男生往往英文语法知识巩固。	1	2	3	4	5	6	7
5	Do you believe that generally speaking, boys are willing to read in English. 在我看来，男生一般喜欢阅读英文。	1	2	3	4	5	6	7
6	I suppose boys are usually willing to learn English grammar. 我觉得通常情况下，男生喜欢学习英文语法知识。	1	2	3	4	5	6	7

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
7	I believe that typically, boys are gifted in learning English grammar. 我相信一般来说，男生擅长学习英文语法。	1	2	3	4	5	6	7
8	I suppose boys are usually gifted in memorising English vocabulary. 我认为通常情况下，男生擅长记英文单词。	1	2	3	4	5	6	7
9	I believe that generally, boys are gifted in communicating in English. 我觉得男生普遍擅长用英文对话。	1	2	3	4	5	6	7
10	Do you believe that boys do well in English reading. 我认为男生往往英文阅读的成绩好。	1	2	3	4	5	6	7
11	I guess boys are usually willing to develop English listening skills. 在我看来，男生通常愿意培养自己的英文听力。	1	2	3	4	5	6	7
12	I am convinced that boys do well in English listening. 我相信男生普遍英文听力的成绩好。	1	2	3	4	5	6	7
13	I suppose typically, boys are willing to learn English writing skills. 我觉得一般情况下，男生喜欢学习英文写作技巧。	1	2	3	4	5	6	7
14	In my high school, we use at least one textbook in English lessons. 在我所就读的高中，英语课至少用一本课本。	1	2	3	4	5	6	7
15	I believe that usually, boys are gifted in understanding English texts. 在我看来，男生通常擅长读懂英文文章。	1	2	3	4	5	6	7
16	I suppose boys are usually willing to practice English speaking skills. 我认为通常情况下，男生喜欢练习英文口语。	1	2	3	4	5	6	7

No. 编号	Statements 陈述	Totally Disagree 完全 反对	Strongly Disagree 强烈 反对	Slightly Disagree 有些 反对	Cannot Decide 无法 确定	Slightly Agree 有些 同意	Strongly Agree 强烈 同意	Totally Agree 完全 同意
17	Do you believe that boys do well in English writing. 我相信男生往往英语作文写得很好。	1	2	3	4	5	6	7
18	I am convinced that boys have a large English vocabulary. 在我看来，一般来说，男生英语词汇量大。	1	2	3	4	5	6	7
19	I believe that typically, boys do well in English speaking. 我相信男生普遍英语口语好。	1	2	3	4	5	6	7
20	Do you believe that boys do well in English. 我觉得男生通常英语成绩好。	1	2	3	4	5	6	7
21	I am convinced that boys typically are willing to learn English. 我认为一般来说，男生喜欢学习英语。	1	2	3	4	5	6	7
22	I believe that typically, boys are gifted in English. 我觉得男生普遍有学英语的天赋。	1	2	3	4	5	6	7

### Part 3 第三部分

Please fill out the table about your background by circling the options that apply to you.

请你根据个人情况填写以下表格：圈出符合你个人情况的选项即可。

Gender 性别		Age 年龄	Year of study 年级				Branch 班级属性		
Male 男	Female 女	___ Years old ___ 周岁	Year 1 高一	Year 2 高二	Year 3 高三	Others: _____ 其他: _____	LAB 文科	SB 理科	Others: _____ 其他: _____

Lastly, it is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise months of effort (the researchers') and the time of other participants could be wasted. So please answer the following three questions by circling the choice that applies to you.

为了确保研究研究结果的真实性，我们希望能尽可能多地收集到真实、有效的数据；否则，研究者的精力和所有参与研究人员的时间都可能会被浪费。因此，希望你能如实回答以下两个问题，帮助研究者判断你的回答的可靠性。

I gave this survey 我在答问卷时的专心程度为：	<input type="radio"/> almost none of <input type="radio"/> 几乎为零	<input type="radio"/> very little of <input type="radio"/> 不太专心	<input type="radio"/> some of <input type="radio"/> 有些专心	<input type="radio"/> quite a bit of <input type="radio"/> 比较专心	my attention. 。
-------------------------------------	--	--	---	--	--------------------

\*\* This is the end of the questionnaire. Thank you for your participation!

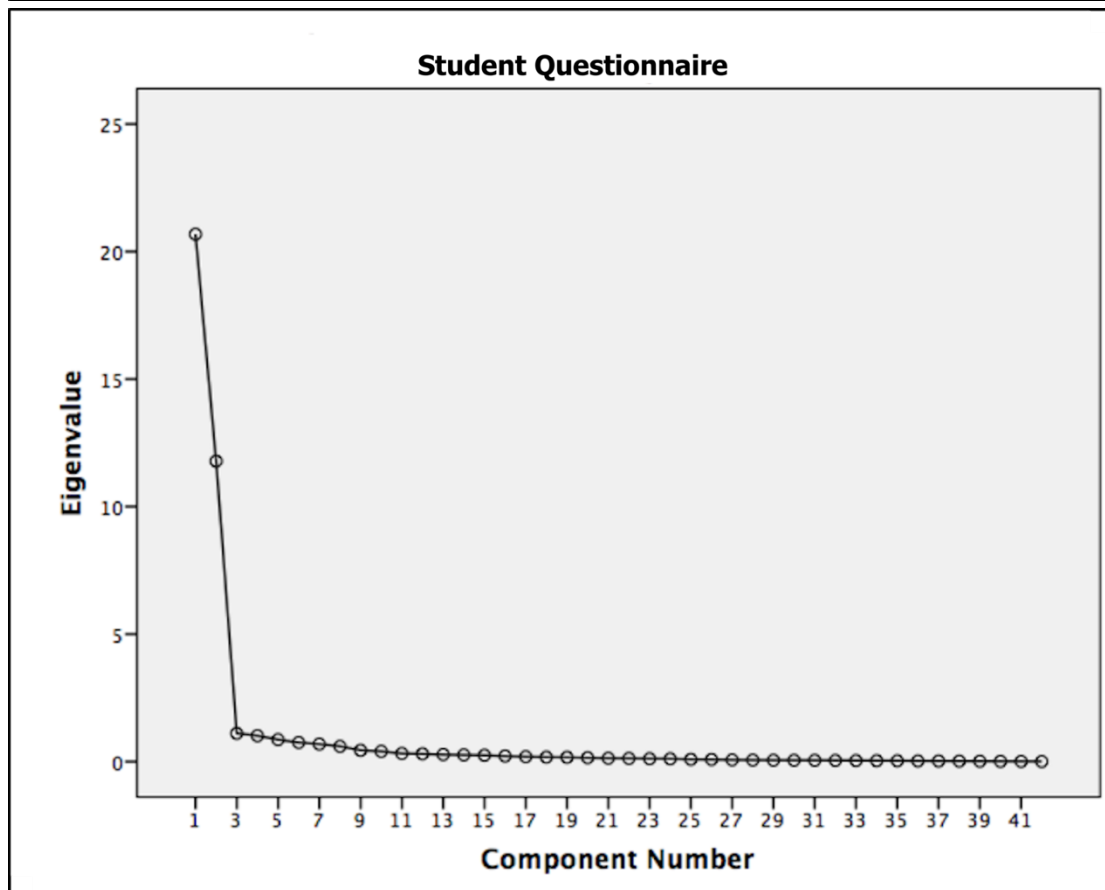
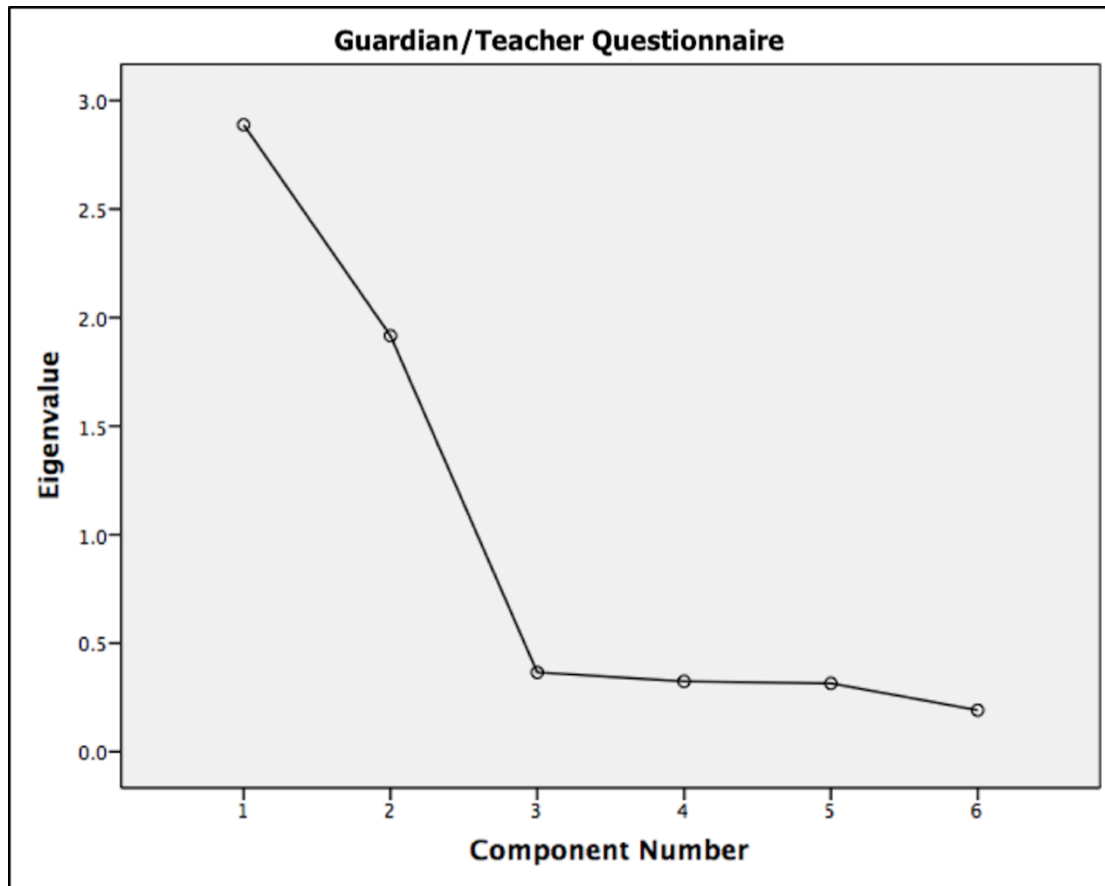
问卷到此结束。感谢你的参与！



**Appendix E: Component Matrix of Two-Component Model of  
Guardian/ Teacher Questionnaire in Pilot Study 1**

	<b>Component</b>	
	<b>1</b>	<b>2</b>
Do you believe that girls are gifted in English?	.932	
Do you believe that girls genrally do well in English?	.898	
Do you believe that typically are willing to learn English?	.872	
Do you believe that boys are gifted in English?		.881
Do you believe that boys genrally do well in English?		.878
Do you believe that boys typically are willing to learn English?		.876
<b>Eigenvalues</b>	<b>2.89</b>	<b>1.92</b>

## Appendix F: Scree Plots of Questionnaires in Pilot Study 1



**Appendix G: Component Matrix of Two-Component Model of  
Student Questionnaire in Pilot Study 1**

	<b>Component</b>	
	<b>1</b>	<b>2</b>
Do you believe that boys do well in English reading?	.931	
Do you believe that boys have a large English vocabulary?	.922	
Do you believe that boys are willing to learn English?	.921	
Do you believe that boys are gifted in learning English grammar?	.918	
Do you believe that boys are gifted in understanding English texts?	.918	
Do you believe that boys do well in English listening?	.915	
Do you believe that boys have a solid grasp of English grammar?	.914	
Do you believe that boys are gifted in writing English essays?	.911	
Do you believe that boys are willing to learn English grammar?	.904	
Do you believe that boys are gifted in memorising English vocabulary?	.903	
Do you believe that boys are gifted in English?	.900	
Do you believe that boys do well in English writing?	.898	
Do you believe that boys do well in English?	.896	
Do you believe that boys do well in English speaking?	.895	
Do you believe that boys are willing to read in English?	.890	
Do you believe that boys are gifted in communicating in English?	.888	
Do you believe that boys are gifted in understanding English utterances?	.888	
Do you believe that boys are willing to learn English vocabulary?	.864	
Do you believe that boys are willing to develop English listening skills?	.862	
Do you believe that boys willing to practice English speaking skills?	.858	
Do you believe that boys are willing to learn writing English essays?	.794	
Do you believe that girls are gifted in understanding English texts?		.932
Do you believe that girls do well in English?		.914
Do you believe that girls are gifted in learning English grammar?		.910
Do you believe that girls do well in English speaking?		.892
Do you believe that girls do well in English reading?		.891
Do you believe that girls are gifted in understanding English utterances?		.886

Do you believe that girls do well in English writing?	.881	
Do you believe that girls are willing to develop English listening skills?	.880	
Do you believe that girls do well in English listening?	.877	
Do you believe that girls are gifted in English?	.877	
Do you believe that girls are gifted in communicating in English?	.871	
Do you believe that girls are willing to learn writing English essays?	.852	
Do you believe that girls have a solid grasp of English grammar?	.839	
Do you believe that girls gifted in memorising English vocabulary?	.836	
Do you believe that girls are willing to learn English vocabulary?	.829	
Do you believe that girls are gifted in writing English essays?	.826	
Do you believe that girls are willing to read in English?	.825	
Do you believe that girls are willing to learn English?	.821	
Do you believe that girls are willing to practice English speaking skills?	.816	
Do you believe that girls have a large English vocabulary?	.806	
Do you believe that girls are willing to learn English grammar?	.711	
<b>Eigenvalues</b>	<b>20.68</b>	<b>11.78</b>

## Appendix H: Consent Form in Study 1

Dear Guardians:

You and your child are invited to participate in a survey conducted by Miss Jing LI, a PhD. student from the Faculty of Education at University of Cambridge. This survey aims to investigate how parents and high school students attribute academic successes in school subjects.

Please read the information below and sign your name and date in designated places to indicate your consent.

- **Procedures**

Inside the brochure you will find 1) two copies of this consent form (one to return after signed, one for parents to keep); 2) Questionnaire 1 (for fathers' use only); 3) Questionnaire 2 (for mothers' use only); and 4) Questionnaire 3 (for students' use only).

If you are willing to participate yourself, please sign your name and date below and fill out the questionnaire for your use.

If you (minimum one parent) are willing to allow your child to participate, please sign your name and date below. You can then ask your child to fill out Questionnaire 3 independently.

Your and child's participation will require 5-8 minutes. When returning the brochure, please tear off ONE copy of consent form and return THE REST of the brochure.

- **Risks and Benefits**

There are no known risks or discomforts associated with this survey.

Your and your child's participation will benefit your child in two ways. First, the researcher will produce a report summarising the major findings and suggesting ways to help parents and students form positive attitudes towards academic results.

In addition, the researcher has agreed to deliver lectures to participating students, topics of lectures including how to prepare for TOFEL/IELTS exams, how to improve academic writing skills, and how to apply for academic programmes abroad. These lectures have been delivered to several participating schools, and they were received with high remarks.

- **Miscellaneous**

Your and your child's participation will be entirely voluntary.

If you or your child choose to participate, you can withdraw at any time without adversely affecting your relationship with anyone in the school or with the researcher.

Your responses will be kept strictly confidential, and digital data will be stored in secure computer files. Any report of this research that is made available to the public will not include your name or any other individual information by which you could be identified.

If you have questions or want a copy or summary of this study's results, you can contact the researcher at the email address on the top right corner of this page.

---

**I understand the procedures described above, and I agree to participate in this study. I have been given a copy of this form.**

Signature of guardian (female): _____	Date: _____
Signature of guardian (male): _____	Date: _____
Signature of child: _____	Date: _____

**I agree to allow my child to participate in this study.**

Signature of guardian (either): _____	Date: _____
---------------------------------------	-------------

---

**Appendix I: Sequences of Items for Male and Female Learners among  
Eight Versions of Questionnaire Brochures for Participating Families**

<b>Version</b>	<b>Questionnaire for Male Guardians</b>		<b>Questionnaire for Female Guardians</b>		<b>Questionnaire for Students</b>	
	<b>Part 1</b>	<b>Part 2</b>	<b>Part 1</b>	<b>Part 2</b>	<b>Part 1</b>	<b>Part 2</b>
<b>1</b>	Female	Male	Female	Male	Female	Male
<b>2</b>	Female	Male	Female	Male	Male	Female
<b>3</b>	Female	Male	Male	Female	Female	Male
<b>4</b>	Female	Male	Male	Female	Male	Female
<b>5</b>	Male	Female	Female	Male	Female	Male
<b>6</b>	Male	Female	Female	Male	Male	Female
<b>7</b>	Male	Female	Male	Female	Female	Male
<b>8</b>	Male	Female	Male	Female	Male	Female

### Appendix J: Interview Schedule in Pilot Study 2

Section 1: Demographic Information												
Name	Gender		Age	Head-teacher		Length of teaching	Current class size	Current students			Current branch	
_____	M	F	_____	Y	N	_____ year(s)	_____ students	Y1	Y2	Y3	LAB	SB
Email:						Mobile:			Social media:			

Section 2: Interview Schedule								
Date:			Time:			Location:		
Introduction	<p>Thank you for being willing to take part in this research.</p> <p>First of all, I want to assure you that you will remain absolutely anonymous in this research. I guarantee that no records of the interview will be labelled with your name on them. When presenting the data, I will refer you to the alias XXX (chosen by each interviewee from a pool).</p> <p>Secondly, I would like to ask for your permission to audio record this interview. The reason is to ensure the accuracy and authenticity of your responses and opinions. Besides, this will also enable a more comprehensive analysis of the data.</p> <p>Now I will briefly introduce you to the purpose of this research. The research aims to investigate English teaching and learning activities in Chinese high schools, so there will be questions about your teaching, your students' learning behaviours, their performances in classes and in exams, and their motivations. The names and places you mention will be faithfully transcribed, but when presenting results, these details will be substituted with the letters N (for names) and P (for places) so that no one can use them to identify you. If you feel like the questions are inappropriate, or you feel uncomfortable during the interview, you are completely free to withdraw at any point.</p> <p>During the interview, I will be asking the questions in Mandarin Chinese, but you can answer in English, Mandarin, or your dialect so long as you can convey your ideas freely and expressively.</p> <p>OK. Now, do you have any questions about the interview? (Answer if the interviewee has any questions.)</p> <p>Great. Now I would first like to know a bit more about your experience as an English teacher. ...</p>							

Topic	Questions & Prompts	Aim
I. Working experience	1. First, can you tell me for how long have you worked as an English teacher? – Have you always taught in the same school? – How many students do you usually get? – Have you worked as a head-teacher before? – How many colleagues do you have in the school?	Demographic information  Gender-ratio (teaching staff)
II. Current class	2. Good. Now can you describe the class(es) that you are currently teaching? – What grade do you teach now? <i><b>If Year 2 and above:</b></i> – Do you teach the LAB or the SB? – How many students do you have? How many boys and how many girls? <i><b>If Year 2 and above:</b></i> – Is this gender ratio common for that branch? – Are you now a head-teacher of any class?	  Gender-ratio (class)  Gender-ratio (branch)
III. Students' performance	3. Let's talk about your teaching routine then. How many lessons do you teach a week? – How do you usually organise a lesson?	Teaching context
	4. How much time do you spend on introducing vocabulary in class? – Have you noticed any students who can remember new words fast? – Have you noticed any students who can remember new words accurately? – What kind of students do you think can be considered gifted in learning vocabulary?	Knowledge about vocabulary  Gendered patterns?
	5. How much time do you spend on explaining grammar in class? – Have you noticed any students who learn grammar well? – What kind of students do you think can be considered gifted in learning grammar?	Knowledge of grammar  Gendered patterns?
	6. How much time do you spend on training listening skills in class? – Have you noticed any students who perform well in listening tasks? – What kind of students do you think can perform better in listening tasks?	Proficiency in listening  Gendered patterns?
	7. How much time do you spend on training speaking skills in class? – Have you noticed any students who perform well in speaking tasks? – What kind of students do you think can perform better in speaking tasks?	Proficiency in speaking  Gendered patterns?
	8. How much time do you spend on training reading skills in class?	Proficiency in reading



	<ul style="list-style-type: none"> <li>– Have you noticed any students who perform well in reading tasks?</li> <li>– What kind of students do you think can perform better in reading tasks?</li> </ul>	Gendered patterns?
	9. How much time do you spend on training writing skills in class? <ul style="list-style-type: none"> <li>– Have you noticed any students who perform well in writing tasks?</li> <li>– What kind of students do you think can perform better in writing tasks?</li> </ul>	Proficiency in writing Gendered patterns?
	10. What about study sessions in the evenings? <ul style="list-style-type: none"> <li>– Do you usually ask your students to study by themselves or do you make these in-put sessions?</li> </ul> <p><b><i>If ask the students to study by themselves:</i></b></p> <ul style="list-style-type: none"> <li>– Do you stay in the classroom and offer assistance to those ask for it or do you leave students alone?</li> </ul> <p><b><i>If stay to offer assistance:</i></b></p> <ul style="list-style-type: none"> <li>– Have you noticed anyone in your class who is always willing to ask for your help?</li> <li>– What kind of students do you think are more willing to ask for your help?</li> </ul>	Motivation for English (student)  Gendered patterns?
	11. What kind of homework do you leave? <ul style="list-style-type: none"> <li>– Which skills or type of knowledge do you wish to develop by administering the homework?</li> </ul> (Link to Questions 11-16)	Knowledge or proficiency
IV. Students' exam results & aptitude	12. Who typically rank top in English exams? <ul style="list-style-type: none"> <li>– Can you describe what reasons have contributed to his/her success?</li> <li>– What role do you think gift plays in this success?</li> <li>– What kind of students do you think may be more gifted in English?</li> </ul>	Test performances Attribution of success Aptitude for English Gendered pattern?
V. Students' affection/ motivation	13. Have you noticed anyone in your class who finds English important? <ul style="list-style-type: none"> <li>– Can you describe what behaviours or remarks by him/her make you think that way?</li> <li>– Have you noticed anyone in your class who you assume dismisses English and also describe his/her indicative actions and words?</li> <li>– Do you think your students in general find English important? Are there any patterns?</li> </ul>	Value attached to English (student)  Gendered patterns?
	14. Have you noticed anyone in your class who likes English? <ul style="list-style-type: none"> <li>– Can you describe what behaviours or remarks by him/her make you think that way?</li> <li>– Have you noticed anyone in your class who is not a fan of English and also describe his/her indicative actions and words?</li> </ul>	Affection for English (student)

	– Do you think your students in general like English? Are there any patterns?	Gendered patterns?
	15. Have you noticed anyone in your class who enjoys learning English? – Can you describe what behaviours or remarks by him/her make you think that way? – Have you noticed anyone in your class who finds learning English a chore and also describe his/her indicative actions and words? – Do you think your students in general enjoy learning English? Are there any patterns?	Affection for English  Gendered patterns?
	16. Have you noticed anyone in your class who is highly motivated to learn English? – Can you describe what behaviours or remarks by him/her make you think that way? – Have you noticed anyone in your class who requires a bit more pushing in English learning and also describe his/her indicative actions and words? – Do you think your students in general enjoy learning English? Are there any patterns?	Motivation for English (student)  Gendered patterns?
	17. Do you call or meet with parents to discuss their children's English results with them? – Do parents meet with you and discuss their children's English scores? – What do parents say about the English subject?	Value attached to English (parent)
VI. Interactions with parents	18. Have you noticed any parent who is proud of his/her child's English achievements? – What does he/she tell you about the reasons for the child's success? <i>If head-teacher:</i> – Do you meet with parents and discuss their children's performances in other subjects? – What subjects are they commonly concerned with? – Have you noticed any parent who is proud of his/her child's academic results? – What does he/she tell you about the reasons for the child's success? Are there any differences among subjects?	Attribution of success (parent) Gendered patterns? Value attached to English (parent) Attribution of success (parent) Gendered patterns?
Closure	It seems that we have discussed quite a lot about your teaching and your students. Thank you very much for your patience and honesty. Do you think there is anything we might have missed out? Or do you have anything to add about what we have covered, or about the research as a whole? Great, so this is the end of our interview. Is it okay that I send you the transcript of your interview in January so that you can review it for accuracy? Thank you again so much for your cooperation.	

### Appendix K: Interview Schedule in Study 2

Part 1: Demographic Information												
Name	Gender		Age	Class-teacher		Length of teaching	Current class size	Current students			Current branch	
_____	M	F	_____	Y	N	_____ year(s)	_____ students	Y1	Y2	Y3	LAB	SB
Email:						Mobile:			Social media:			

Part 2: Interview Agenda		
Date:		Time:
<b>Introduction</b>	<p>Thank you for being willing to take part in this research.</p> <p>First of all, I want to assure you that you will remain absolutely anonymous in this research. I guarantee that no records of the interview will be labelled with your name on them. When presenting the data, I will refer you to the alias XXX (chosen by each interviewee from a pool).</p> <p>Secondly, I would like to ask for your permission to audio record this interview. The reason is to ensure the accuracy and authenticity of your responses and opinions. Besides, this will also enable a more comprehensive analysis of the data.</p> <p>Now I will briefly introduce you to the purpose of this research. The research aims to investigate English teaching and learning activities in Chinese high schools, so there will be questions about your teaching, your students' learning behaviours, their performances in classes and in exams, and their motivations. The names and places you mention will be faithfully transcribed, but when presenting results, these details will be substituted with the letters N (for names) and P (for places) so that no one can use them to identify you. If you feel like the questions are inappropriate, or you feel uncomfortable during the interview, you are completely free to withdraw at any point.</p> <p>During the interview, I will be asking the questions in Mandarin Chinese, but you can answer in English, Mandarin, or your dialect so long as you can convey your ideas freely and expressively.</p> <p>OK. Now, do you have any questions about the interview? (Answer if the interviewee has any questions.)</p> <p>Great. Now I would first like to know a bit more about your experience as an English teacher. ...</p>	
<b>Topic</b>	<b>Questions &amp; Prompts</b>	
<b>I.</b>	1. First, can you tell me for how long have you worked as an English teacher?	
	<b>Aims</b>	
	Demographic information	

<b>Working experience</b>	<ul style="list-style-type: none"> <li>– Have you always taught in the same school?</li> <li>– How many students do you usually get?</li> <li>– Have you worked as a class-teacher before?</li> <li>– How many colleagues do you have in the school?</li> </ul>	Gender-ratio (teaching staff)
<b>II. Current class</b>	<p>2. Good. Now can you describe the class(es) that you are currently teaching?</p> <ul style="list-style-type: none"> <li>– What grade do you teach now?</li> </ul> <p><b><i>If Year 2 and above:</i></b></p> <ul style="list-style-type: none"> <li>– Do you teach the LAB or the SB?</li> <li>– How many students do you have? How many boys and how many girls?</li> </ul> <p><b><i>If Year 2 and above:</i></b></p> <ul style="list-style-type: none"> <li>– Is this gender ratio common for that branch?</li> <li>– Are you now a class-teacher of any class?</li> </ul>	<p>Gender-ratio (class)</p> <p>Gender-ratio (branch)</p>
<b>III. Proficiency in English skills</b>	<p>3. Let's talk about your teaching routine then. How many lessons do you teach a week?</p> <ul style="list-style-type: none"> <li>– How do you usually organise a lesson?</li> </ul>	Teaching context
	<p>4. How do you assess students' mastery of the four English skills, listening, speaking, reading, and writing?</p> <p><b><i>Listening skills:</i></b></p> <ul style="list-style-type: none"> <li>– Have you noticed any students who usually perform well in listening tasks?</li> <li>– Can you name a few?</li> <li>– What do these students have in common?</li> </ul> <p><b><i>Speaking skills:</i></b></p> <ul style="list-style-type: none"> <li>– Have you noticed any students who perform well in speaking tasks?</li> <li>– Can you name a few?</li> <li>– What do these students have in common?</li> </ul> <p><b><i>Reading skills:</i></b></p> <ul style="list-style-type: none"> <li>– Have you noticed any students who usually perform well in reading tasks?</li> <li>– Can you name a few?</li> <li>– What do these students have in common?</li> </ul> <p><b><i>Writing skills:</i></b></p> <ul style="list-style-type: none"> <li>– Have you noticed any students who usually perform well in writing tasks?</li> <li>– Can you name a few?</li> <li>– What do these students have in common?</li> </ul>	<p>Proficiency in four skills</p> <p>Gendered patterns?</p>
	<p>5. Do you think there are gender differences in the mastery of English skills?</p>	FES endorsement

	– Do you think others share your opinion?	FES awareness
<b>IV. Mastery of English knowledge</b>	6. How do you assess students' mastery of vocabulary and grammar knowledge? <b><i>Vocabulary knowledge:</i></b> – Have you noticed any students who can remember new words fast? – Have you noticed any students who can remember new words accurately? – Can you name a few? – What do these students have in common? <b><i>Grammar knowledge:</i></b> – Have you noticed any students who learns grammar well? – Can you name a few? – What do these students have in common?	Mastery of English knowledge  Gendered patterns?
	7. Do you think there are gender differences in the mastery of English skills? – Do you think others share your opinion?	FES endorsement FES awareness
<b>V. Students' exam performance &amp; aptitude</b>	8. Who typically rank top in English exams? – Can you describe what reasons have contributed to his/her success? – What role do you think gift plays in this success? – Who do you think may be more gifted in English?	Test performances Attribution of success Aptitude for English Gendered pattern?
<b>VI. Students' affect/ motivation</b>	9. Do you have self-study sessions in the evenings? – Do you usually ask your students to study by themselves or do you make these in-put sessions? <b><i>If ask the students to study by themselves:</i></b> – Do you stay in the classroom and offer assistance to those ask for it or do you leave students alone? <b><i>If stay to offer assistance:</i></b> – Have you noticed anyone in your class who is always willing to ask for your help? – Who are more willing to ask for your help when they are in trouble?	Motivation for English (student)  Gendered patterns?
	10. Have you noticed anyone in your class who finds English important? – Can you describe what behaviours or remarks by him/her make you think that way? – Have you noticed anyone in your class who you assume dismisses English and also describe his/her indicative actions and words? – Do you think your students in general find English important? Are there any patterns?	Value attached to English (student)  Gendered patterns?

	11. Have you noticed anyone in your class who likes English? – Can you describe what behaviours or remarks by him/her make you think that way? – Have you noticed anyone in your class who is not a fan of English and also describe his/her indicative actions and words? – Do you think your students in general like English? Are there any patterns?	Affection for English (student)  Gendered patterns?
	12. Have you noticed anyone in your class who is highly motivated to learn English? – Can you describe what behaviours or remarks by him/her make you think that way? – Have you noticed anyone in your class who requires a bit more pushing in English learning and also describe his/her indicative actions and words? – Do you think your students in general enjoy learning English? Are there any patterns?	Motivation for English (student)  Gendered patterns?
<b>Closure</b>	It seems that we have discussed quite a lot about your teaching and your students. Thank you very much for your patience and honesty. Do you think there is anything we might have missed out? Or do you have anything to add about what we have covered, or about the research as a whole? Great, so this is the end of our interview. Is it okay that I send you the transcript of your interview in January so that you can review it for accuracy? Thank you again so much for your cooperation.	

## Appendix K: Consent Form in Study 2



Jing LI  
Faculty of Education  
University of Cambridge  
Mobile (China): +86 152 0143 5754  
Email: jl806@cam.ac.uk

### CONSENT TO PARTICIPATE IN RESEARCH

#### Investigating Teaching and Learning of English: From the Perspectives of High School English Teachers in China

#### 关于参与

#### 《英语的教与学之调查：基于中国高中英语教师的态度与看法》

#### 研究项目的知情同意书

Dear Mr./Ms. \_\_\_\_\_,

You are invited to participate in an interview-based study conducted by Miss Jing LI, a doctoral student from the Faculty of Education at University of Cambridge. This research is part of her doctoral thesis. Your participation in this study is entirely voluntary. Please read the information below and ask questions about anything you do not understand and/or need clarification, before deciding whether or not to participate.

亲爱的 \_\_\_\_\_ 老师，

您好！本人李婧是剑桥大学教育学系的博士研究生，很荣幸能邀请您参加我的研究项目。本项目是我博士毕业论文的一部分。您的参与全凭自愿。在研究开始之前，请您阅读并签署这份知情同意书；在决定参与与否之前，如果您对本研究有任何疑问或不解，均可以向我提出。

#### PURPOSE OF THE STUDY • 研究目的

This interview-based survey aims to investigate how English teachers in Chinese high schools describe and feel about their teaching activities and their interactions with students.

本研究旨在通过访谈的方式，调查中国高中的英语老师们如何描述和看待他们的教育活动和与学生的沟通互动。

#### PROCEDURES • 研究步骤

In this study, you will be asked to join a 45-60-minute-long one-to-one interview with the researcher. The interview questions will be focusing on your teaching experiences, your students' learning activities, their performances in classes and exams, and their motivations. All the interview questions will be asked in Mandarin Chinese. Your responses can be in Chinese and/or in English partly or completely. Later, you will be invited to review the transcript of your interview and be allowed to make revisions or additions to your responses. You will also be invited to review parts of the analysis to guarantee that the analyses accurately reflect your opinion.

在本研究过程中，您会接受一个 45-60 分钟的一对一采访。采访内容包括您的教学经历、您的学生的学习情况、课堂表现、考试成绩以及学习动机等。所有问题均以汉语普通话提问；回答时，您可自由使用全部或部分的汉语（普通话或方言）及英语。日后，您会被邀请复审访谈的文字记录，并做出相应的改动或增补。您还会被邀请审核我对访谈数据的分析，以保证我的分析能如实反映您的观点。

#### CONFIDENTIALITY AND ANONYMITY • 保密性和匿名处理

Unless you object, this interview will be audio-recorded digitally. Should you choose not to be recorded, the researcher will take notes instead. You can request that the recording be stopped at any time during the interview, either permanently or temporarily. Both the audiotape and the transcript of your interview will only be accessed by the researcher. The recording and the transcript will be password protected when digitally stored.

经过您的同意之后，采访过程将会被录音。如果您不同意采访过程被录音，那么我会在采访过程中使用记笔记的方式。您可以在采访中的任何时候要求中止录音。采访录音和文字记录不会向其他任何人公开，并会被加密储存。

Any information that is obtained in connection with this study and that can be identified with you will remain confidential. Anonymity will be maintained by means of using an alias in transcriptions and any written report of the research.

任何在采访过程中获得的能够指明您身份的信息都不会被公开。在采访的文字记录和日后的研究结果报告中，我会使用化名来指代您，保证您的匿名参与。

### **RISKS AND BENEFITS • 风险与利益**

There are no known risks or discomforts associated with this survey.

参与本研究不会给您带来任何的风险或不适。

When reviewing the analyses, you may read about other teachers' experiences and opinions, which may be similar to or different from your own. Anyway, their insights may benefit your teaching by generating new ways of understanding your own experiences.

当您参与审核我对访谈数据的分析时，您也许会看到我所引用的其他老师的观点或经历。这些可能与您不谋而合，也可能与您大相径庭，但来自别人的视角也许会让您对自身经历产生新的理解，从而为您的教学助力。

You will receive a gift from the researcher as compensation for participating in the research as your interview ends.

此外，您还会在采访结束时收到一份来自研究者的小礼物，用来感谢您的参与。

### **RECORD OF CONSENT • 知情同意**

Your signature below indicates that you have understood the information about this research and consent to your participation. Your participation is entirely voluntary and you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer. If you have further questions about this study, please contact the researcher.

如果您在下方签名，这表示您已经知晓了本研究的信息并同意成为受访者。接受采访为完全自愿行为，您可以拒绝回答采访中的任何问题，也可以随时要求中止采访。如果您还有任何疑问，欢迎随时联系研究者。

---

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

我已经阅读了这份知情同意书。我对本研究的所有疑惑已被详尽解答。我自愿参加这项研究，并保有一份《同意书》留作个人参考。

---

(Signature of Participant) ( 受访者签名 )

---

(Date) ( 日期 )

---

(Signature of Researcher) ( 研究者签名 )

---

(Date) ( 日期 )

---



# ENGLISH TEST

2017.JUN

---

## test instructions

English skills are crucial in higher education and for future employment. Yet currently, the mental processes of achieving proficiency in foreign language is unclear. This research seeks to establish better understanding about what makes some people better at English than others.



This test has been piloted in several middle schools. Analyses of previous results have shown that females perform better on this test. In other words, this English test shows that females have higher levels of English proficiency than males.

# ENGLISH TEST

2017.JUN

---

## test instructions

English skills are crucial in higher education and for future employment. Yet currently, the mental processes of achieving proficiency in foreign language is unclear. This research seeks to establish better understanding about what makes some people better at English than others.



This test has been piloted in several middle schools. Analyses of previous results have shown that females and females perform equally well on this test. In other words, this English test shows that females and females have similar levels of English proficiency.

## Appendix O: Instrument for Manipulation Check in Study 3

### Feedback on *Dragon and Phoenix*

***Dragon and Phoenix* is a sitcom about high school students. Please watch a clip from it and finish the five items below to indicate your feedback. (Items marked with \* are mandatory)**

1. Both *Act I: Shirts* and *Act III: After School* show some popular assumptions about boys' and girls' differential attitudes towards clothing and diet. Please tick off the statements you believe are mentioned in the video clip. (Select all that apply) \*

- ☐ Boys don't wear pink clothes, which are usually considered not masculine.
- ☐ Generally speaking, the pants-shirt combination is more suitable for boys than for girls.
- ☐ If a boy displays feminine behaviours (such as wearing pink or having a sweet tooth), he is likely to be negatively judged (for example, being ridiculed by his peers).
- ☐ Having a sweet tooth is usually considered a feminine trait.

2. Both *Act II: An English Lesson* and *Act III: After School* show some popular assumptions about boys' and girls' differential attitudes towards schooling and careers. Please tick off the statements you believe are mentioned in the video clip. (Select all that apply) \*

- ☐ Boys usually do not enjoy English lessons.
- ☐ Typically, males do not choose careers as a school teacher.
- ☐ English teachers are more likely to be females.
- ☐ Boys are not that willing to spend time on tasks that require extra attention and patience, such as practising English pronunciation.
- ☐ Compared to boys, girls are more gifted in tasks involving reading, writing, or memorising English words.
- ☐ Boys are not generally gifted in subjects dealing with textual materials, such as English.
- ☐ Compared to boys, girls are more willing to spend time studying.
- ☐ Compared to boys, girls pay more attention to details when studying.

3. You are from Class \_\_\_\_\_. \* You are ☐ Male ☐ Female. \* Your name is \_\_\_\_\_. \*

### Answers (withheld from participants)

(This item aims to remind participants of popular GSs. Participants are supposed to choose at least one from Options 1, 3, and 4.)

✓ 1

☐ 2

✓ 3

✓ 4

(This item aims to remind participants of GS about English. Although all options here are mentioned in the video, participants are supposed to choose at least one from Options 1, 3, 4, 5, and 6.)

✓ 1

✓ 2

✓ 3

✓ 4

✓ 5

✓ 6

✓ 7

✓ 8

### Appendix P: Interview Schedule in Pilot Study 4

Part 1: Demographic Information							
Name	Gender	Age	Years of English learning	Mobile	Wechat	QQ	Email
	F M						
	F M						
	F M						
Current Year		Y1	Y2	Y3	Current branch	LAB	SB

Part 2: Interview Agenda			
Date:		Time:	Location:
<b>Introduction</b>	<p>Thank you for being willing to take part in this research.</p> <p>First of all, I want to assure you that you will remain absolutely anonymous in this research. I guarantee that no records of the interview will be labelled with your names on them. When presenting the data, I will refer you to aliases.</p> <p>Secondly, I would like to ask for your permissions to audio record this interview. The reason is to ensure the accuracy and authenticity of your responses and opinions. Besides, this will also enable a more comprehensive analysis of the data.</p> <p>Now I will briefly introduce you to the purpose of this research. The research aims to investigate high school students' experiences with and opinions about English learning, so there will be questions about your learning activities, your views about other English learners, and your feelings about English. The names and places you mention will be faithfully transcribed, but when presenting results, these details will be substituted with the letters N (for names) and P (for places) so that no one can use them to identify you. If you feel like the questions are inappropriate, or you feel uncomfortable during the interview, you are completely free to withdraw at any point.</p> <p>During the interview, I will be asking the questions in Mandarin Chinese, but you can answer in English, Mandarin, or your dialect so long as you can convey your ideas freely and expressively.</p> <p>OK. Now, do you have any questions about the interview? (Answer if the interviewee has any questions.)</p> <p>Great. Now I would first like to know a bit more about your experience as an English learner. ...</p>		
<b>Topic</b>	Questions & Prompts		Aims

<b>I. Learning English in general</b>	1. First, can you tell me for how long you have learnt English? – What kind of activities do you do in English classes? (four abilities and two types of knowledge) – Do you also attend English lessons outside school? – How many hours do you learn English in an average week?	Exposure to English
<b>II. Views about English learners</b>	2. Given that you've all learnt English for so many years, you must have some ideas about what characteristics good English learners have. – Can you name someone you know that is a good English learner and describe what makes him/her such a good learner? (Affection? Aptitude? Effort? Parental influence? SES?) – Why do you think that these characteristics are essential for a good English learner? – Are these characteristics specifically beneficial to English learning, or common to all academic subjects?	Characteristics of good English learners?  Gendered patterns?
	○ <i>If affection in Q2</i> – In your class, who typically express that they like learning English? What speeches or actions you observe from them make you think that they do? – What characteristics do they have in common?	
	○ <i>If aptitude in Q2</i> – In your class, who looks like gifted English learners? What speeches or actions you observe from them make you think that they are? – What characteristics do they have in common?	
<b>III. Comparing boys with girls</b>	3. In your class, who typically perform better in English exams, boys or girls? Why do you think this gender difference has happened?	Performance
	4. In your class, who typically like English more, boys or girls? Why do you think this gender difference has happened? (link to Q2.1)	Affection
	5. In your class, who typically are more talented in English, boys or girls? Why do you think this gender difference has happened? (link to Q2.2)	Aptitude
	6. Do you think male and female English learners are different, or are they similar?	GS

	<ul style="list-style-type: none"> <li>– What are the advantages girls have in English learning? What about their disadvantages?</li> <li>– What are the advantages boys have in English learning? What about their disadvantages?</li> </ul>	
<b>IV. Circulating stereotypes and their influences</b>	<p>7. Are you familiar with the saying that “Girls learn English better than boys”?</p> <ul style="list-style-type: none"> <li>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</li> <li>– What do you think of such sayings? Do you agree or disagree with them?</li> <li>– Do you think about such saying when you learn English?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <ul style="list-style-type: none"> <li>– What possible influences do such sayings have on your English learning?</li> </ul>	Opinions and experiences with GS relating to performance
	<p>8. Are you familiar with the saying that “Girls are more talented English learners compared to boys”?</p> <ul style="list-style-type: none"> <li>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</li> <li>– What do you think of such sayings? Do you agree or disagree with them?</li> <li>– Do you think about such saying when you learn English?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <ul style="list-style-type: none"> <li>– What possible influences do such sayings have on your English learning?</li> </ul>	Opinions and experiences with GS relating to aptitude



	<p>9. Are you familiar with the saying that “Girls like LA subjects more than boys do”?</p> <ul style="list-style-type: none"> <li>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</li> <li>– What do you think of such sayings? Do you agree or disagree with them?</li> <li>– Do you think about such saying when you learn English?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <p>– What possible influences do such sayings have on your English learning?</p>	<p>Opinions and experiences with GS relating to affect</p>
	<p><b><i>If participants say that they are not familiar with sayings above:</i></b></p> <p>10. Typically, which kind of subjects do you think English is more similar to, LA ones or Science ones?</p> <p>– Why?</p> <p>11. Are you familiar with the saying that “Girls learn LA subjects better than boys”?</p> <ul style="list-style-type: none"> <li>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</li> <li>– What do you think of such sayings? Do you agree or disagree with them?</li> <li>– Do you think about such saying when you learn English?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <p>– What possible influences do such sayings have on your English learning?</p> <p>12. Are you familiar with the saying that “Girls are more talented LA subjects learners compared to boys”?</p>	<p>In case of denial:</p> <p>Opinions and experiences with GS relating to performance</p> <p>Opinions and experiences with GS relating to aptitude</p>

	<p>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</p> <p>– What do you think of such sayings? Do you agree or disagree with them?</p> <p>– Do you think about such saying when you learn English?</p> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <p>– What possible influences do such sayings have on your English learning?</p> <p>13. Are you familiar with the saying that “Girls like LA subjects more than boys do”?</p> <p>– From what sources have you heard about sayings like this? In what occasions have you heard people saying such things? How did you feel about such remarks?</p> <p>– What do you think of such sayings? Do you agree or disagree with them?</p> <p>– Do you think about such saying when you learn English?</p> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulties or get bad grades in English, will you think of such sayings and use them to explain what you are going through?</li> </ul> <p><b>If girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you get good grades in English, will you think of such sayings and use them to explain your success?</li> </ul> <p>– What possible influences do such sayings have on your English learning?</p>	Opinions and experiences with GS relating to affect
Closure	<p>It seems that we have discussed quite a lot about English learning. Thank you very much for your patience and honesty. Do you think there is anything we might have missed out? Or do you have anything to add about what we have covered, or about the research as a whole?</p> <p>Great, so this is the end of our interview. Is it okay that I send you the transcript of your interview in July so that you can review it for accuracy?</p> <p>Thank you again so much for your cooperation.</p>	



### Appendix Q: Interview Schedule in Study 4

Part 1: Demographic Information								
Position	Name	Gender	Age	Years of English learning	Mobile	Wechat	QQ	Email
Left		F    M						
Middle		F    M						
Right		F    M						
Current Year		Y1		Y2	Y3	Current branch	LAB	SB

Part 2: Interview Agenda				
Date		Time		Location
<b>Introduction</b>	<p>Thank you for being willing to take part in this research.</p> <p>First of all, I want to assure you that you will remain absolutely anonymous in this research. I guarantee that no records of your discussion will be labelled with your names on them. When presenting the data, I will refer you to aliases. Secondly, I would like to ask for your permissions to audio record this group discussion. The reason is to ensure the accuracy and authenticity of your responses and opinions. Besides, this will also enable a more comprehensive analysis of the data.</p> <p>Now I will briefly introduce you to the purpose of this research. The research aims to investigate high school students' experiences with and opinions about English learning, so the discussion will involve topics like your learning activities, your views about other English learners, and your feelings about English. During the discussion, if you agree with someone else, you can give examples and reasons to support your shared opinion; but if you do not have a shared opinion about a certain topic, it is okay to have a debate among yourselves. Of course, if sometimes you feel uncomfortable talking about something, you are completely free to remain silent or withdraw at any point.</p> <p>During the group discussion, I will be introducing the topics in Mandarin Chinese, but you can talk in English, Mandarin, or your dialect so long as you can convey your ideas freely and expressively.</p> <p>OK. Now, do you have any questions about the group discussion? (Answer any questions the participants raise.)</p> <p>Great. Now I would first like to know a bit more about your experience as English learners. ...</p>			

Topic	Questions & Prompts	Aims	Notes		
			Left (     )	Mid (     )	Right (     )
<b>I. Learning English in general</b>	1. Can each of you share with the rest of us your experience with English learning? – Such as when did you start to learn English, what textbooks you have used, and what is your general opinion about English?	Exposure to English and general attitude			
<b>II. Views about English learners</b>	2. It seems that you all have learnt English for quite some time. – Can you discuss among yourselves what it takes to become a good English learner? – (Affection? Aptitude? Effort? Parental influence? SES?) – Why do you think that these characteristics are essential for a good English learner? – Are these characteristics specifically beneficial to English learning, or common to all academic subjects?	Characteristics of good English learners?  Gendered patterns?			
<b>III. Comparing boys and girls</b>	<b>For girls:</b> 3.1 As girls, can you discuss among yourselves: compared to boys, what advantages, if any, do girls have when learning English? – What about disadvantages? 4.1 Among yourselves, can you discuss how you feel that boys differ from girls when they learn English?	GS			
	<b>For boys:</b> 3.2 As boys, can you discuss among yourselves: compared to girls, what advantages, if any, do boys have when learning English? – What about disadvantages? 4.2 Among yourselves, can you discuss how you feel that girls differ from boys when they learn English?				
	5. Judging from your own experiences, can you discuss among yourselves who typically perform better in English exams, boys or girls? – Why or why not do you think there is a gender difference?	Achievement			
		Disagreement?			

	6. Judging from your own experiences, can you discuss among yourselves who typically like English more, boys or girls? – Why or why not do you think there is a gender difference?	Affect Disagreement?			
	7. Judging from your own experiences, can you discuss among yourselves who typically are more talented in English, boys or girls? – Why or why not do you think there is a gender difference?	Aptitude Disagreement?			
	8. There seems to be a wide spread opinion that girls typically learn English better than boys. – Can you discuss among yourselves: from what sources or whom have you heard about opinions like this? In what occasions? – Can you share with the rest of us how such remarks make you feel? – Can you share your opinions about such remarks? Do you agree or disagree with them? 1. What possible influences do such remarks have on your learning English?	Opinions and experiences with GS relating to achievement			
IV. Circulating gender stereotypes and their influences	<b>For girls:</b> • For example, when you memorise English words or understand grammar well, will you think of such remarks and use them to explain your success like telling yourselves “I am a girl, so naturally I am talented in English”? • Do other people, such as your parents, teachers, or peers, use such remarks to explain your success, like telling you “Girls just have a talent for English”?				
	<b>For boys:</b> • For example, when you have difficulty in memorising English vocabulary or understanding English grammar, will you think of such remarks and use them to explain what you are going through, like telling yourselves “I am a boy, so it is no surprise that I am not that talented in English”? • Do other people, such as your parents, teachers, or peers, use such remarks to comfort you when you get unsatisfactory results, like telling you “Boys just don't have a talent for English”? – Are there any other influences?				

	<p>9. There seems to be a wide spread opinion that girls typically are more talented English learners than boys.</p> <ul style="list-style-type: none"> <li>– Can you discuss among yourselves: from what sources or whom have you heard about opinions like this? In what occasions?</li> <li>– Can you share with the rest of us how such remarks make you feel?</li> <li>– Can you share your opinions about such remarks? Do you agree or disagree with them?</li> </ul> <p>2. What possible influences do such remarks have on your learning English?</p> <p><b>For girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you memorise English words or understand grammar well, will you think of such remarks and use them to explain your success like telling yourselves “I am a girl, so naturally I am talented in English”?</li> <li>• Do other people, such as your parents, teachers, or peers, use such remarks to explain your success, like telling you “Girls just have a talent for English”?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you have difficulty in memorising English vocabulary or understanding English grammar, will you think of such remarks and use them to explain what you are going through, like telling yourselves “I am a boy, so it is no surprise that I am not that talented in English”?</li> <li>• Do other people, such as your parents, teachers, or peers, use such remarks to comfort you when you get unsatisfactory results, like telling you “Boys just don't have a talent for English”?</li> </ul> <ul style="list-style-type: none"> <li>– Are there any other influences?</li> </ul>	Opinions and experiences with GS relating to aptitude			
	<p>10. There seems to be a wide spread opinion that girls typically enjoy learning English more than boys do.</p> <ul style="list-style-type: none"> <li>– Can you discuss among yourselves: from what sources or whom have you heard about opinions like this? In what occasions?</li> <li>– Can you share with the rest of us how such remarks make you feel?</li> </ul>	Opinions and experiences with GS relating to affect			

	<p>– Can you share your opinions about such remarks? Do you agree or disagree with them?</p> <p>3. What possible influences do such remarks have on your learning English?</p>				
	<p><b>For girls:</b></p> <ul style="list-style-type: none"> <li>• For example, when you find English learning enjoyable or feel motivated to learn English, will you think of such remarks and use them to explain your feelings, like telling yourselves “English is what girls enjoy”?</li> <li>• Do other people, such as your parents, teachers, or peers, use such remarks to explain your interest in English, like telling you “English is just what girls like”?</li> </ul> <p><b>For boys:</b></p> <ul style="list-style-type: none"> <li>• For example, when you find English learning boring or do not feel like learning English, will you think of such remarks and use them to explain what you are going through, like telling yourselves “boys are just not interested in English”?</li> <li>• Do other people, such as your parents, teachers, or peers, use such remarks to comfort you when you get unsatisfactory results, like telling you “Boys just are not interested in English”?</li> </ul> <p>– Are there any other influences?</p>				
<b>Closure</b>	<p>It seems that we have discussed quite a lot about English learning. Thank you very much for your patience and honesty.</p> <p>Do you think there is anything we might have missed out? Or do you have anything to add about what we have covered, or about the research as a whole?</p> <p>Great, so this is the end of our interview. Is it okay that I send you the transcript of your interview in July so that you can review it for accuracy?</p> <p>Thank you again so much for your cooperation.</p>				

## Appendix R: Data Screening Process of Guardian Dataset in Study 1

Category		Notes
1 Self- Reported low amount of attention	Explanation	There were 148 cases in the guardian sample where participants reported that they exerted little or no attention in the survey, among which 12 cases (marked in bold) also appeared in the following two categories.
	Cases identified	#34, #72, #80, #81, #82, #91, #99, #102, #124, #135, #152, #186, #198, #199, #233, #237, #270, #271, #294, #313, #383, #384, #396, #400, #404, #415, #431, #441, #451, #458, #466, #472, #488, #495, #517, #539, #595, #620, #652, #661, #670, #680, #741, #745, #747, #774, #785, #803, #823, #826, #854, #866, #871, #891, #901, #983, #984, #1010, #1019, #1033, #1037, #1047, #1070, #1101, #1102, #1113, #1127, #1130, #1131, #1135, #1137, #1195, #1197, #1226, #1227, #1235, #1299, #1304, #1312, #1328, #1343, #1348, #1370, #1396, #1474, #1500, #1518, #1519, #1520, #1543, #1564, #1580, #1618, #1645, #1649, #1690, #1695, #1708, #1721, #1723, #1724, #1739, #1766, #1772, #1788, #1804, #1813, #1843, #1844, #1863, #1890, #1894, #1901, #1911, #1915, #1927, #1933, #1938, #1949, #1971, #1978, #1999, #2001, #2029, #2060, #2072, #2075, #2076, #2088, #2103, #2107, #2114, #2115, #2150, #2225, #2233, #2282, #2286, #2296, #2329, #2351, #2362, #2378, #2386, #2442, #2449, #2456, #2484;
2 Uniform response pattern	Explanation	Participants answering all items with one answer were likely to be careless. There were 421 cases in the guardian sample where uniform response patterns emerged, 13 of which were duplicate cases (marked in bold).
	Cases identified	#12, #24, #25, #29, #32, #33, #42, #46, #63, #67, #74, #77, #83, #97, #109, #122, #150, #151, #170, #184, #202, #216, #218, #230, #254, #268, #270, #279, #284, #296, #305, #314, #317, #345, #387, #388, #449, #460, #469, #470, #474, #491, #511, #513, #516, #534, #538, #547, #565, #573, #580, #582, #586, #602, #606, #611, #612, #616, #624, #629, #631, #637, #647, #648, #650, #668, #669, #673, #681, #686, #690, #694, #700, #703, #725, #726, #751, #764, #766, #767, #781, #786, #792, #793, #794, #796, #798, #799, #805, #813, #822, #823, #829, #856, #857, #862, #871, #887, #904, #905, #906, #907, #908, #909, #910, #911, #912, #913, #914, #915, #916, #917, #918, #919, #920, #921, #923, #924, #925, #926, #927, #928, #929, #930, #931, #932, #933, #934, #935, #936, #937, #938, #939, #940, #941, #942, #943, #944, #945, #946, #947, #948, #949, #950, #951, #952, #953, #954, #955, #956, #957, #958, #959, #960, #961, #962, #963, #964, #965, #966, #967, #968, #969, #970, #976, #977, #978, #979, #980, #981, #982, #987, #990, #997, #1021, #1030, #1031, #1035, #1038, #1039, #1042, #1045, #1048, #1057, #1063, #1070, #1081, #1084, #1086, #1090, #1099, #1112, #1116, #1127, #1134, #1156, #1157, #1159, #1164, #1170, #1174, #1180, #1181, #1184, #1186, #1189, #1190, #1200, #1205, #1210, #1212, #1215, #1216, #1218, #1221, #1225, #1228, #1229, #1232, #1233, #1237, #1239, #1241, #1244, #1260, #1270,

3 Random response pattern		#1273, #1282, #1314, #1346, #1371, #1385, #1394, #1417, #1419, #1428, #1451, #1459, #1466, #1479, #1490, #1496, #1516, #1517, <b>#1519</b> , #1558, #1560, #1566, #1571, #1582, #1592, #1631, #1662, #1682, #1688, #1694, #1703, #1704, <b>#1708</b> , #1719, #1722, #1725, #1743, #1762, #1781, #1782, #1784, #1785, #1821, #1827, #1829, #1835, #1875, #1892, #1898, #1908, #1934, #1945, <b>#1949</b> , #1955, #1961, #1993, #2011, #2013, #2016, #2021, #2022, <b>#2030</b> , #2043, #2048, #2050, #2059, <b>#2072</b> , #2086, <b>#2088</b> , #2091, #2100, #2153, #2154, #2155, #2156, #2157, #2158, #2159, #2160, #2161, #2162, #2163, #2164, #2165, #2166, #2167, #2168, #2169, #2170, #2171, #2172, #2173, #2174, #2175, #2176, #2177, #2178, #2179, #2180, #2181, #2182, #2183, #2184, #2185, #2186, #2187, #2188, #2189, #2190, #2191, #2192, #2193, #2194, #2195, #2196, #2197, #2198, #2199, #2200, #2201, #2202, #2203, #2204, #2205, #2206, #2207, #2208, #2209, #2210, #2211, #2212, #2213, #2214, #2215, #2216, #2217, #2218, #2219, #2220, #2221, #2222, #2223, #2224, #2235, #2246, #2265, #2269, #2270, #2276, #2279, #2281, #2284, #2304, #2307, #2308, #2318, #2319, #2330, #2357, #2359, #2360, #2361, #2364, #2366, #2380, #2383, #2387, #2390, #2396, #2405, #2408, #2409, #2416, #2420, #2421, #2422, #2423, #2426, #2429, <b>#2430</b> , #2436, #2444, #2445, #2446, #2454, #2460, #2465, #2468, #2471, #2475, #2487, #2489, #2495, #2497, #2498;
	Explanation	Random responses can also signal in laxity. As assessed by Mahalanobis distance ( $p > .001$ ), there were 40 multivariate outliers in the current sample, among which 5 cases (in bold) already appeared once in the previous categories.
	Cases identified	#23, #70, #112, #126, #144, #302, #363, #444, #522, #662, #699, #720, #721, #765, <b>#781</b> , #791, <b>#854</b> , <b>#1019</b> , #1357, #1515, #1551, #1573, #1641, #1642, #1653, #1700, #1771, #1777, #1778, #1812, #1825, #1834, #1967, #2014, <b>#2030</b> , #2227, #2290, #2339, <b>#2430</b> , #2483.

### Appendix S: Data Screening Process of Student Dataset in Study 1

Category		Notes
<b>1</b> <b>Self-Reported low amount of attention</b>	Explanation	There were 88 cases in the parent sample where participants reported that they exerted little or no attention in the survey, among which 37 cases also appeared in the following three categories. In addition, Cases 908 and 920 appeared simultaneously in Categories 1 and 3.
	Cases identified	#3, #6, #53, #81, #83, #94, #99, #131, #135, #147, #248, #251, <b>#272, #294</b> , #315, #384, <b>#396</b> , #400, #413, #431, <b>#441</b> , #453, #459, #475, #500, <b>#528</b> , #529, <b>#539</b> , #540, <b>#567, #595</b> , #628, <b>#652, #654</b> , #662, #674, <b>#699</b> , #700, #734, #741, #744, #745, #747, <b>#762</b> , #766, #798, #810, #821, #823, #828, <b>#852, #858</b> , #865, #889, <b>#898, #908</b> , #916, <b>#917, #920, #922</b> , #923, <b>#927, #943, #948, #952</b> , #953, <b>#954</b> , #1004, <b>#1009</b> , #1013, #1033, #1034, #1041, <b>#1047</b> , #1056, #1072, #1075, <b>#1080, #1101, #1102</b> , #1136, #1137, #1161, #1166, <b>#1197</b> , #1200, #1203, <b>#1238</b> ;
<b>2</b> <b>Uniform response pattern</b>	Explanation	Participants answering all items with one answer were likely to be careless. There were 22 cases in the parent sample where uniform response patterns emerged, among which 11 also appeared either in Category 1 or Category 4. In addition, Cases 908 and 920 appeared simultaneously in Categories 1, 2 and 4.
	Cases identified	<b>#122</b> , #322, #713, #794, <b>#888</b> , #894, #903, <b>#908</b> , #915, <b>#918, #920</b> , #925, <b>#926</b> , #928, <b>#958, #960</b> , #1038, <b>#1081, #1159, #1180</b> , #1205, #1275;
<b>3</b> <b>Random response pattern</b>	Explanation	Participants who randomly chose answers may have rushed through the questionnaire, thus giving unengaged responses (Johnson, 2005). As assessed by Mahalanobis distance ( $p > .001$ ), there were 98 multivariate outliers in the current sample, among which 15 also appeared in Categories 1 and 2.
	Cases identified	#4, #19, #23, #27, #32, #34, #38, #43, #50, #53, #55, #57, #59, #61, #70, #71, #82, #88, #94, #102, #126, #136, #140, #143, #156, <b>#175</b> , #181, #193, <b>#279</b> , #284, #295, #326, <b>#329</b> , #330, #331, #332, #334, #340, #345, #354, #356, #376, <b>#392, #393</b> , #394, #401, <b>#404</b> , #413, #444, #447, #457, #464, #487, #489, #490, #576, #579, <b>#581</b> , #583, #602, #627, #650, #666, #684, #702, #707, #720, #727, #728, #737, #744, #756, <b>#779</b> , #789, #795, <b>#909, #930</b> , #1019, #1069, #1150, #1153, #1167, <b>#1177</b> , #1250, #1251, #1252, #1258, #1260, #1261, #1265, #1268, #1289, #1290, #1291, #1294, #1295, #1296, #1297.
<b>4</b> <b>Answering the bogus</b>	Explanation	There were 153 cases in the student dataset where participants answered the bogus item wrongly in the survey, among which 51 cases (in bold) also appeared in the previous three categories.



<b>item wrongly</b>	Cases identified	#109, #122, #154, #175, #213, #229, #230, #270, #272, #279, #294, #329, #349, #352, #353, #370, #386, #392, #393, #396, #404, #420, #422, #433, #440, #441, #462, #463, #466, #477, #497, #502, #504, #505, #514, #528, #538, #539, #552, #567, #571, #572, #574, #577, #581, #595, #605, #614, #616, #641, #643, #647, #652, #653, #654, #663, #692, #697 698, #699, #762, #764, #773, #779, #785, #808, #814, #822, #834, #841, #843, #845, #846, #852, #858, #861, #869, #872, #874, #880, #885, #888, #898, #904, #908, #909, #917, #918, #919, #920, #921, #922, #924, #926, #927, #929, #930, #933, #934, #936, #937, #943, #947, #948, #952, #954, #957, #958, #960, #961, #963, #966, #968, #969, #970, #1003, #1008, #1009, #1027, #1047, #1062, #1064, #1067, #1080, #1081, #1082, #1101, #1102, #1106, #1116, #1123, #1124, #1125, #1128, #1130, #1148, #1156, #1159, #1176, #1177, #1180, #1183, #1192, #1194, #1197, #1202, #1236, #1238, #1240, #1242, #1243, #1259, #1273;
-------------------------	---------------------	---

## Appendix T: Sub-Codes Categorising Teachers' Endorsement of the Female-Advantage-in-Languages Stereotype in Study 2

Categorisation	Code	Notes
Content	Aptitude	<b>Description</b> Gender-stereotypical accounts, opinions, or attitudes that assume female learners generally have a stronger aptitude in language learning than males.
		<b>Inclusion criteria</b> Words or phrases that refer to female's biological superiority, cognitive advantage, and other 'gift'-related notions (for example, 'the language faculty') in language learning.
		<b>Typical exemplar</b> <i>I do tell my students [that girls tend to do better in English]. But((a short pause)) not for the purpose of hitting them over the head, just to let ((a short pause)) the students know that this is the status quo. Probably when it comes to language, girls, in this area of language, may be just slightly more gifted. So boys, they must learn from girls more often. They need to open their mouths more often, and read and recite more often. Boys probably just are not that into reciting. [Ann, Paragraph 126]</i>
		<b>Atypical exemplar</b> <i>Girls just understand [English texts] faster. In terms of language, girls probably have a relatively stronger faculty of understanding. [Carolyn, Paragraph 122]</i>
	Affect	<b>Description</b> Gender-stereotypical accounts, opinions, or attitudes that assume female learners typically have stronger and/or more favourable attitudes towards language subjects compared to their male counterparts.
		<b>Inclusion criteria</b> Words or phrases that describe affective, emotive, motivational or attitudinal factors contributing to the success of language learning.
		<b>Typical exemplar</b> <i>Still, it's interest that contributes a lot [to how successful a student can be in English]. Interest contributes more [to learning success]. Because when I introduced many great apps to students, those who showed interest, like those who joined my extra-curriculum group, are mainly girls. Yes. I looked at the group where more than twenty students joined, there was no boys. [Bernie, Paragraph 178]</i>
		<b>Atypical exemplar</b> <i>Among the low-achieving students, there are definitely more boys. Like, in classes I teach, students whose scores are below 50, perhaps in both classes, there are on average 10 students or so of this kind. And about seven or eight would be boys. Yeah. Seven or eight of them are</i>

			<p><i>boys. ... These students, in fact, the obstacles to their learning success, should actually have been there when they first started learning English, when they were in junior high. Because when I talk to them, I definitely have to talk with them [about their learning experiences], and they would all say, [the reason that they are not doing well enough] would be that in junior high, they did not build a solid foundation. [They would say that] it probably was because they did not form good learning habits and others. [Gerrie, Paragraph 92-96]</i></p>
Achievement		<b>Description</b>	Gender-stereotypical accounts, opinions, or attitudes that assume female learners tend to outperform the male ones in language skills (listening, speaking, reading, and writing), language knowledge (vocabulary and grammar), tests and exams, and other language-related tasks or activities.
		<b>Inclusion criteria</b>	Words or phrases that are concerned with performance or achievement in language, instead of language learning behaviours or habits.
		<b>Typical exemplar</b>	<i>Regarding performance in oral English, I would have to say it's girls whose performance, whose expression, is better than boys. ... Because for the same question, the exact same question, if you ask boys and girls to answer, girls' answers will be more fluent, and more logical. But boys, when they answer the question, their oral expression will be less fluent, [and their performance] will be slightly ((a short pause)) worse. [Charlie, Paragraph 70-72]</i>
		<b>Atypical exemplar</b>	<i>Now [what I examine in dictations is] just words. So if you are willing to work on it in your own time, you will remember words well, and you will do well in dictations. This is where I feel that girls can be patient and exert enough efforts. So probably it is girls who typically do well [in dictations]. [Kylie, Paragraph 112]</i>
Salience	Blatant claim	<b>Description</b>	Gender-stereotypical accounts that demonstrates a participant's tendency to highlight gender differences and issues without any prompt from the researcher. Some may also reflect the participant's effort to incorporate them into teaching strategies.
		<b>Inclusion criteria</b>	Gender-stereotypical accounts that are voluntarily expressed by participants, i.e., without being prompted to compare learners on the basis of gender, participants themselves see gender as a differentiating variable in language learners and bring up assumed gender differences on their own.

		<b>Typical exemplar</b>	<i>I:: think that, for so long since I started teaching, typically girls do better [in English]. It might be related to the aptitude for language learning. ... I think girls learn, they have this gift for learning English. ((She giggles)) Boys' [aptitude] may be, <u>relatively speaking</u>, worse. But there are boys who do quite well [in English], relatively speaking, still not as well as girls can do. [Daphnie, Paragraph 114-116]</i>
		<b>Atypical exemplar</b>	<i>I tell the students, "if you want to study English well, you try to get close, as close as possible to me<sup>51</sup>." That's what I tell them. You must get as close to me as possible, "if you want to learn English well<sup>52</sup>", right? ... Like today, I was in Class 2, and I said, "you boys, the four Yangs, why do you <u>never</u> want to get close to me? Is it because we are both males and like charges repel? Girls all like me <u>so much</u>. They are all willing to get close to me, and they always ask me questions after class." I said, "boys <u>never</u> come forward to me. Is it because you think males, like charges repel?" They responded, "No::, sir. We are just timid." So I said, "Never mind your timidity, you have to get close to me to learn English well. If you keep me at arm's length, how can you learn English well? Right?" [Donnie, Paragraph 52]</i>
	<b>Moderately explicit claim</b>	<b>Description</b>	Gender-stereotypical accounts that demonstrates a participant's tendency to highlight gender differences and issues with some prompt from the researcher. Some may also reflect the participant's effort to incorporate them into teaching strategies.
		<b>Inclusion criteria</b>	Gender-stereotypical accounts that are elicited, i.e., when being prompted to compare learners on the basis of gender, participants are inclined to discuss how male and female learners differ.
		<b>Typical exemplar</b>	<i>((Interviewer asks, "so, as we just talked about vocabulary and grammar, do you think that boys and girls differ in their mastery of the two, or is there no gender difference?")) I think we should leave the top [students] out of our discussion. The remaining students, I think girls are still:: I'm not sure if it is because girls have this natural aptitude or whatever, they just master [vocabulary and grammar] faster. Some boys, even if you've explained to him, he still won't quite understand. ... But in fact, sometimes I think for these boys, the problems they have with grammar are a result from their vocabulary range. Because they have no idea what that word</i>

<sup>51</sup> Originally in English.

<sup>52</sup> Originally in English.

Subtle expression		<i>means at all. They don't know about parts of speech at all. So when you talk to them about grammar, oh, it can make you really annoyed. [Eileen, Paragraph 173-176]</i>
	Atypical exemplar	<i>((Interviewer asks, "so in terms of listening performance, do you think there is a difference between boys and girls?")) You mean a difference based on gender? ((Interviewer confirms, "Yes. Is it like, boys do better, or do girls do better? Or are they actually similar?")) Well, ((a short pause)) How do I put it? ... I think typically:: Typically speaking, speaking of listening, girls would still be, as a whole, they would be better than boys. This is the case. I am talking about the difference between the Liberal Arts Branch and the Sciences Branch, actually. Because in Liberal Arts class, in fact, it's also better [than the Sciences Branch] ... ((Interviewer asks, "are you saying that because there are more girls in the Liberal Arts Branch, it does better in English listening than the Sciences Branch?")) Yes, yes, exactly. [Frankie, Paragraph 81-86]</i>
	Description	Gender-stereotypical accounts that indicates a participant's potential to be gender-aware, but the accounts themselves are neither blatant nor moderately explicit.
	Inclusion criteria	Accounts that imply participants' endorsement of the female-language-advantage stereotype, sometimes without explicitly comparing male and female learners.
	Typical exemplar	<i>((Interviewer asks, "you've just talked about many students, some of them are doing well, but some need to improve. So generally speaking, do you think that boys and girls differ in their English achievement? Or do boys and girls have similar results?")) I think, if a boy is willing to learn, he will learn better than girls, faster than girls. ((Interviewer asks a follow-up question, "can you elaborate on that?")) I mean:: generally, don't boys put their minds into the science subjects, or in playing games? ((She giggles)) Cell phones are a big problem among students. Our school does not allow cell phones, but where there is a policy, there is a strategy against that policy. You know, parents would always prioritise their children's needs, [so they buy cell phones for their children.] Yet, they don't know that the students want cell phones for games. ((She pauses)) So you know, if a boy can put his mind into English, then he will learn it rather well. [Gwen, Paragraph 91-94]</i>
	Atypical exemplar	<i>Basically there are always boys who have uneven performances across subjects in any class in the Sciences Branch, boys who have gone overboard with the sciences subjects. ... [Girls of this kind also] exists. There is one in my class. She's also gone overboard with the sciences</i>

		<i>subjects. She is actually the opposite of that boy. This girl's English result is not good, but her results in sciences subjects are still not, relatively speaking, <u>very</u> good. But perhaps with regards to her English results, she's gone overboard with the sciences subjects. [Carolyn, Paragraph 210-212]</i>
Function	Description	Gender-stereotypical beliefs, opinions or attitudes portraying what male and female learners typically do.
	Inclusion criteria	Accounts that are merely used to sketch the respective attributes of male or female learners, instead of setting gender-differential standards or norms for learners of a given gender.
	Typical exemplar	<i>Speaking of this, yes, there are [gender differences in writing]. In this area, I suppose that girls are, I suppose their performance tend to be ((a short pause)) slightly be stronger. The first reason would be in handwriting. Girls' handwriting can be better than boys', tidier and better-looking. And then, also in expression, [girls'] expression, compared to that of boys', can be more, relatively, I mean in sentence structures and everything, they form better [writing products]. [Hollie, Paragraph 124]</i>
	Atypical exemplar	<i>I mean students who get good exam results in English tend to be interested in the subject. They memorise words fast, and they learn fast. And when they learn, there seems to be not much strain on their shoulders. They seem quite relaxed. ... These students ((a short pause)), I suppose, there should be more girls like this. Girls are definitely the majority. [Archie, Paragraph 106-108]</i>
	Description	Gender-stereotypical beliefs, opinions or attitudes defining what male and female learners should do.
	Inclusion criteria	Accounts that are used to prescribe gender-differential labels to male or female learners.
	Typical exemplar	<i>((Interviewer asks, "you've mentioned a lot of students. As you see it, generally speaking, do boys and girls differ in the four English skills? Or is there no difference between them?")) If we are talking about these skills, listening, speaking, reading, and writing, as far as I see it, probably ((a short pause)) female students perhaps should just do better. I mean as a <u>whole</u>, as a whole. Because I feel, I mean if you, especially if you are talking about language learning ((she giggles)). I mean it's like, and just like we just talked about, choosing between the Liberal Arts and Sciences Branches. Probably the most important thing would be, for boys,</i>

Evaluation		<p><i>they just don't think that their English is very good, so they will go for the Science branch. And the same goes for Chinese. Because these [subjects] require students to memorise stuff, and perhaps they, if they are not keen on that, they just won't make an effort to memorise anything, and neither will they choose the Liberal Arts Branch. In other words, perhaps boys, relatively speaking, are unwilling to memorise stuff. I think that's what happens. [Hollie, Paragraph 126]</i></p>
	Atypical exemplar	<p><i>Boys, in fact, the extent to which they are interested in English is not as big as girls. So probably they are just more interested in Physics, Chemistry, Biology, or maybe Math, the sort of subjects that involve computation, or calculation. ... Um:: I mean, if you give his time to study on his own, no boys will, because some self-study session are not monitored by designated teachers, they are just there for students to study on their own. And boys will probably just choose to do exercise books in Sciences subjects, but not English, or Chinese. When doing exercise books, they may just prefer those tasks, computation and calculation stuff. Because, Chinese and English, they should belong to the area of reading and understanding texts, and more girls [tend to be interested in them]. ((Interviewer asks, "is this what you've witnessed among your students?")) In my classes, if I am monitoring a self-study session, I won't allow them to study other subjects. ((She laughs)) So this, this is, because I teach many classes in the Science Branch, this is what I feel must be like. [Brie, Paragraph 104-108]</i></p>
	Description	Gender-stereotypical beliefs, opinions or attitudes causing gender-differential expectations and/or treatments.
	Inclusion criteria	Accounts that imply a participant's use of gender stereotypes as standards against which individual learners of a given gender are compared against.
	Typical exemplar	<p><i>And in my classes, students who are relatively good at English:: tend to be mostly girls. Boys are rare. Boys, but there is one exception, a boy who is not quite good at any other subjects, but his English is very outstanding. ... Because his English can get as high as, I mean last term, the highest score in our school was more than 120, and he got as high as 117. ... But in terms of other subjects. ((She laughs)) Because that was the final exam, I took a peek at his academic report, and he got less than 20 for the remaining subjects, out of 100. The full mark was 100, and he got less than 20. So it was a dramatic contrast. ... But if you look at, only look at his English results, his score was top-notch. ... He has gone overboard with just English. And this is a boy. Isn't this astonishing? ... I mean, boys who are relatively bad at English are</i></p>

*pretty large in number, pretty common. So this, this boy, is very peculiar. I often want to, every time I, ((She clicks her tongue)) I just want to tell him that he needs to work on other subjects, too. It is such a pity, you know, to have a boy whose English is so good is something once in a blue moon [but at the same time his performance in other subjects is dragging him down].*  
[Carolyn, Paragraph 80-88]

**Atypical  
exemplar**

*We just started [“Wordsman Test”] last week, and all prizes went to girls. ((She giggles)) Yeah. Diamond, Gold, and Silver [prizes], all girls. Yeah. Just take a look when I teach, you’ll see that those cautiously and conscientiously taking notes are all girls. Boys won’t write a word down, and you just don’t know if they will prepare for the test after the class. ... As for this week’s Wordsman Test, I’ve been telling them since the beginning of this week, I said, “We’ll have another Wordsman Test this week. You’ll need to review the words we learn during the class. I won’t throw you a curve ball or anything, but the words I taught will definitely be tested.” I am laying rules down here. ... And we’ll see. In the end, tonight, we’ll definitely see girls winning again. Yeah. Boys are just like that. Um:: they would always think, “the words I learnt from the class, I can just sit back and listen, right?” And that’s just what they do. They just don’t have that, that kind of deadly serious attitude to learning stuff like this.*  
[Laureen, Paragraph 152]