

**THE BIG FIVE PERSONALITY TRAITS AND
FOREIGN LANGUAGE SPEAKING CONFIDENCE
AMONG JAPANESE EFL STUDENTS**

A Dissertation
Submitted
to the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF EDUCATION

by
Matthew Thomas Apple
May, 2011

Examining Committee Members:

David Beglar, Advisory Chair, CITE/TESOL

Mark Sawyer, External Member, Kwansei Gakuin University

Yoko Kozaki, External Member, Mukogawa Women's University

Peter Gobel, External Member, Kyoto Sangyo University

Jim Sick, External Member, J. F. Oberlin University

©
by
Matthew Thomas Apple
2011
All Rights Reserved

ABSTRACT

This research examined the relationships between the Big Five human personality traits, favorable social conditions, and foreign language classroom speaking confidence. Four research questions were investigated concerning the validity of the Big Five for a Japanese university sample, the composition of Foreign Language Classroom Speaking Confidence, the degree to which the Big Five influenced Foreign Language Classroom Speaking Confidence, and the degree to which perceptions of classroom climate affect Foreign Language Classroom Speaking Confidence.

The first stage of the research involved three pilot studies that led to the revision of the Big Five Factor Marker questionnaire and the creation of a new instrument for measuring foreign language classroom speaking confidence that included both cognitive and social factors as theorized in mainstream social anxiety research. The second stage of the research involved the collection and analysis of data from 1,081 participants studying English in 12 universities throughout Japan. Data were analyzed using a triangulation of Rasch analysis, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA) in order to verify the construct validity of the eleven hypothesized constructs. Following validation of the measurement model, the latent variables were placed into a structural regression

model, which was tested by using half of the data set as a calibration sample and confirmed by using the second half of the data set as a validation sample.

The results of the study indicated the following: (a) four of the five hypothesized Big Five personality traits were valid for the Japanese sample; (b) Foreign Language Classroom Speaking Confidence comprised three measurement variables, Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English; (c) Emotional Stability and Imagination directly influenced Foreign Language Classroom Speaking Confidence, and; (d) Current English Classroom Perception and Perceived Social Value of Speaking English directly influenced Foreign Language Classroom Speaking Confidence. The findings thus demonstrated a link between personality, positive classroom atmosphere, and foreign language classroom speaking confidence. The implications of the findings included the possibility that foreign language anxiety is not situation-specific as theorized, and that improved social relations within the foreign language classroom might help reduce speaking anxiety.

ACKNOWLEDGMENTS

First and foremost, I must acknowledge the assistance of my dissertation advisor, Dr. David Beglar, who not only introduced me to Rasch measurement analysis, but who also helped me to improve my overall knowledge of the field of Second Language Acquisition as a whole as I worked my way through both master's level and doctoral level courses. His "superhuman" efforts to assist and advise my research, through draft upon draft of the proposal, data analysis using new techniques, up to and including of course the dissertation defense and the revisions, made it possible for me to explore research ideas, analysis techniques, and theoretical implications of this study. I cannot express how fortunate I feel to have been by taught and advised by him. He is, quite frankly, one of the best teachers I have had at any level of education.

I would also express my appreciation for the other members of my dissertation defense committee: Drs. Mark Sawyer, Yoko Kozaki, Peter Gobel, and Jim Sick. I'm sure it was not easy to read through a 400-page dissertation on the eve of the New Year, and I am grateful for all their suggestions and comments. I also extend a special thank you to Dr. Mark Sawyer. The first course session I ever attended at Temple University was taught by Dr. Sawyer, and the last course session was also taught by him. It may have been merely a coincidence, but it did seem entirely appropriate.

Appreciation is also due to Peter Neff, who started the doctoral program with me in 2005 during our first year as teaching colleagues. Even though we went our separate ways during the course of the program, we continued to exchange ideas about teaching and research. It was thanks to Peter that I learned about the Big Five model and the IPIP home page; his eye for detail was instrumental in the presentation of the pilot study at the Pacific Rim Objective Measurement Symposium 2008 conference in Tokyo and Rasch 2010 conference in Denmark.

I was extremely grateful, and flattered, by the large number of my Osaka doctoral cohort classmates who took the time to join the audience for my dissertation defense. Throughout my doctoral coursework, I was very fortunate to be surrounded by such an intelligent, helpful, and supportive group of fellow language teachers. I would also like to express thanks to those instructors throughout Japan who helped me collect questionnaire data. They must remain anonymous, but they know who they are, and I am grateful.

Finally, I cannot even begin to express the appreciation felt for my family—my lifelong partner Etsuko Shimo, who has given me endless encouragement throughout the doctoral course program and dissertation writing stages, and my daughter Erina, whose toddler antics provided a bit of anxiety but much more relief and laughter. I simply couldn't have done it without them.

Dedicated to Jared Phillip Apple,
and the wonderful teacher he would have become,
and to Erina Shimo Apple,
for the inspiring teacher she may yet grow up to be.

TABLE OF CONTENTS

	PAGE
ABSTRACT	iii
ACKNOWLEDGMENTS.....	v
DEDICATION	vii
LIST OF TABLES.....	xvii
LIST OF FIGURES	xxvi
CHAPTER	
1. INTRODUCTION	1
Background of the Issue	1
Statement of the Problem	3
Purposes of the Study.....	6
The Audience for the Study.....	8
Delimitations.....	9
The Organization of this Study	10
2. REVIEW OF THE LITERATURE	11
Big Five Personality Traits: A Summary	12
History of the Big Five: Lexical and Questionnaire Traditions....	12

Personality Trait Measurements: Factor Markers vs.	
NEO-PI-R	15
Competing Models to the Big Five	17
Big Five Studies in East Asia.....	21
Chinese Big Five Studies	23
Korean Big Five Studies	28
Japanese Big Five Studies.....	29
Personality Traits and Second Language Acquisition.....	34
Anxiety and Confidence	46
Origins of Social Anxiety	46
The Japanese Context.....	49
L1 Influences	49
Educational Influences	50
The L1 Context: Communication Apprehension.....	55
The L2 Context: The Foreign Language Class Anxiety Scale	57
Foreign Language Anxiety Studies in Japan	59
Testing Foreign Language Anxiety.....	62
Constructing Confidence	65
Theorized Enduring Influences of the Heuristic	
Pyramid	66

Theorized Situational Influences of the Heuristic Pyramid	68
L2 Confidence in the Heuristic Pyramid: Some Questions	70
Redefining the Construct of Foreign Language Confidence	72
Confidence and Social Situations	75
Gaps in the Existing Literature	81
Statement of Purpose	88
Research Questions	88
3. METHODS	89
Participants	89
School 1 Participants	90
School 2 Participants	90
School 3 Participants	91
School 4 Participants	91
School 5 Participants	92
School 6 Participants	93
School 7 Participants	93
School 8 Participants	94
School 9 Participants	94

School 10 Participants	95
School 11 Participants	95
School 12 Participants	96
Summary of Participants in this Study	96
Instrumentation	98
Revised Big Five Factor Markers.....	98
Foreign Language Classroom Speaking Confidence	100
Foreign Language Classroom Speaking Anxiety	101
Perceived Foreign Language Speaking	
Self-Competence.....	101
Desire to Speak English	102
Past English Classroom Experiences	102
Current English Classroom Perception	103
Perceived Social Value of Speaking English.....	104
Research Design.....	106
Analytical Procedures.....	107
Overview of Analysis Techniques.....	107
Rasch Analysis.....	107
Confirmatory Factor Analysis and Structural	
Education Modeling.....	113

Testing the Proposed Models	117
Justification for Analysis Techniques.....	118
4. PRELIMINARY ANALYSIS	121
Pilot Study Analysis	121
Instrument Validation.....	129
Background of the Main Study Participants	130
The Revised Big Five Factor Marker Questionnaire.....	133
Rasch PCA of Item Residuals	133
Rasch Analysis of the Five Constructs Comprising the	
R-B5FM	149
Summary of Rasch Results for the R-B5FM	
Constructs.....	167
Exploratory Factor Analysis.....	171
Comparison of the Rasch Analysis and EFA Results.....	179
Summary of Instrument Validation for the R-B5FM	
Questionnaire.....	182
The Foreign Language Classroom Speaking Confidence	
Questionnaire.....	183
Rasch PCA of Item Residuals	183

Rasch Analysis of the Six Constructs Comprising the FLCSC.....	199
Summary of Rasch Results for the FLCSC Constructs.....	227
Exploratory Factor Analysis.....	232
Comparison of the Rasch Analysis and EFA Results.....	242
Summary of Preliminary Analysis Findings.....	246
5. STRUCTURAL EQUATION MODEL RESULTS	248
Confirmatory Factor Analysis Procedures	249
Research Question One: Validity of the Big Five Personality Traits.....	253
Research Question Two: Foreign Language Classroom Speaking Confidence	260
Confirmatory Factor Analysis: Favorable Social Environment.....	266
Summary of the Confirmatory Factor Analysis Findings.....	271
Structural Equation Modeling.....	272
Justification for the Big Five Personality Trait Variable Paths.....	274

Justification for the Favorable Social Conditions	
Variable Paths.....	276
Justification for the Foreign Language Classroom	
Speaking Confidence Variable Paths.....	278
Research Question Three: Personality and Speaking	
Confidence	279
Research Question Four: Classroom Climate and	
Speaking Confidence	283
Equivalent Models	284
Equivalent Model 2.....	285
Equivalent Model 3.....	287
Choosing a Model.....	289
Summary of the Structural Regression Models.....	292
6. DISCUSSION.....	293
Research Question One: Validity of the Big Five.....	293
Summary of the Results for Research Question One.....	294
Interpretation of the Results for the Validity of the	
Big Five	295
Research Question Two: Foreign Language Classroom Speaking	
Confidence	309

Summary of the Results for Research Question Two	310
Interpretation of the Results for Foreign Language	
Classroom Speaking Confidence	310
Research Question Three: Personality and Speaking Confidence	314
Summary of the Results for Research Question Three	315
Interpretation of the Results for the Relation between	
Personality and Speaking Confidence.....	316
Research Question Four: Classroom Climate and Speaking	
Confidence	319
Summary of the Results for Research Question Four.....	319
Interpretation of the Results for the Influence of	
Classroom Climate on Foreign Language Classroom	
Speaking Confidence.....	320
Theoretical Implications.....	328
Pedagogical Implications.....	335
7. CONCLUSION.....	340
Summary of the Findings	340
Limitations.....	344
Suggestions for Future Research.....	347
Final Conclusions.....	349

REFERENCES CITED.....	351
APPENDICES	
A. ORAL SCRIPT OF INFORMED CONSENT AND RIGHT NOT TO PARTICIPATE IN THE STUDY	381
B. INFORMED CONSENT AND RIGHT NOT TO PARTICIPATE IN THE STUDY PROVIDED IN WRITING TO PARTICIPANTS.....	382
C. REVISED BIG FIVE FACTOR MARKERS (ENGLISH VERSION).....	384
D. REVISED BIG FIVE FACTOR MARKERS (JAPANESE VERSION)	386
E. FOREIGN LANGUAGE CLASSROOM SPEAKING CONFIDENCE (ENGLISH VERSION)	388
F. FOREIGN LANGUAGE CLASSROOM SPEAKING CONFIDENCE (JAPANESE VERSION).....	392
G. LIST OF ABBREVIATIONS AND ACRONYMS	395

LIST OF TABLES

Table	Page
1. Percentage of Activities Experienced by Study Participants in the Foreign Language Classroom	97
2. Background of the Participants	131
3. Major Fields of Study of the Participants	132
4. Rasch Principal Components Analysis for the Revised Big Five Factor Markers Questionnaire (37 Items).....	135
5. Rasch Principal Components Analysis for the Items Measuring Extraversion-Introversion and Agreeableness	137
6. Rasch Principal Components Analysis for the Items Measuring Extraversion-Introversion	138
7. Rasch Principal Components Analysis for the Items Measuring Extraversion-Introversion Including Item AGR 5.....	140
8. Rasch Principal Components Analysis for the Items Measuring Agreeableness.....	142
9. Rasch Principal Components Analysis for the Items Measuring Agreeableness Excluding Item AGR 5.....	143
10. Rasch Principal Components Analysis for the Items Measuring Emotional Stability, Conscientiousness, and Intellect/Imagination.....	145

11. Rasch Principal Components Analysis for the Items Measuring Emotional Stability.....	145
12. Rasch Principal Components Analysis for the Items Measuring Intellect/Imagination and Conscientiousness.....	146
13. Rasch Principal Components Analysis for the Items Measuring Intellect/Imagination.....	148
14. Rasch Principal Components Analysis for the Items Measuring and Conscientiousness	149
15. Category Structure Functioning for Extraversion-Introversion	151
16. Rasch Statistics for the Items Measuring Extraversion-Introversion.....	151
17. Rasch Statistics for the Items Measuring Extraversion-Introversion Excluding Item EI 7.....	152
18. Category Structure Functioning for Agreeableness	155
19. Rasch Statistics for the Items Measuring Agreeableness	156
20. Category Structure Functioning for Emotional Stability.....	159
21. Rasch Statistics for the Items Measuring Emotional Stability.....	159
22. Category Structure Functioning for Intellect/Imagination.....	162
23. Rasch Statistics for the Items Measuring Intellect/Imagination	163
24. Rasch Statistics for the Items Measuring Intellect/Imagination Excluding Item INT 5.....	163

25. Category Structure Functioning for Conscientiousness	165
26. Rasch Statistics for the Items Measuring Conscientiousness	166
27. Explained and Unexplained Variance for the Constructs Including Variance Explained by the First Contrast.....	169
28. Item and Person Separations and Reliabilities for the Constructs Comprising the Revised Big Five Factor Marker Questionnaire.....	170
29. Correlational Matrix of Rasch Person Measures for Five Constructs Comprising the Revised Big Five Factor Markers Questionnaire	171
30. Descriptive Statistics for the Revised Big Five Factor Marker Questionnaire Items.....	173
31. Exploratory Factor Analysis Results of the Revised Big Five Factor Marker Questionnaire (Five-Factor Solution).....	175
32. Exploratory Factor Analysis Results of the Revised Big Five Factor Marker Questionnaire (Four-Factor Solution)	177
33. Factor Correlation Matrix for the Four Factors Comprising the Revised Big Five Factor Marker Questionnaire.....	179
34. A Comparison of the Big Five Constructs Identified by Rasch PCA of Item Residuals and EFA.....	181
35. Problematic Items Identified by Rasch PCA of Item Residuals and EFA.....	182

36. Rasch Principal Components Analysis for the Items Measuring the Six Foreign Language Classroom Speaking Confidence Constructs (52 Items).....	185
37. Rasch Principal Components Analysis for the Items Measuring Foreign Language Classroom Speaking Anxiety.....	186
38. Rasch Principal Components Analysis of Item Residuals for Items Measuring Desire to Speak English, Past English Classroom Experiences, Perceived Foreign Language Speaking Self-Competence, Current English Classroom Perception, and Perceived Social Value of Speaking English (41 Items).....	188
39. Rasch Principal Components Analysis for the Items Measuring Past English Classroom Experiences.....	189
40. Rasch Principal Components Analysis for the Items Measuring Perceived Foreign Language Speaking Self-Competence, Desire to Speak English, Current English Classroom Perception, and Perceived Social Value of Speaking English (33 Items).....	191
41. Rasch Principal Components Analysis for the Items Measuring Desire to Speak English.....	193
42. Rasch Principal Components Analysis for the Items Measuring Perceived Foreign Language Speaking Self-Competence.....	194

43. Rasch Principal Components Analysis of Item Residuals for the Items Measuring Current English Classroom Perception and Perceived Social Value of Speaking English (16 Items)	195
44. Rasch Principal Components Analysis for the Items Measuring Current English Classroom Perception.....	196
45. Rasch Principal Components Analysis for the Items Measuring Perceived Social Value of Speaking English.....	198
46. Category Structure Functioning for Foreign Language Classroom Speaking Anxiety	200
47. Rasch Analysis for the Items Measuring Foreign Language Classroom Speaking Anxiety.....	201
48. Rasch Statistics for the Items Measuring Foreign Language Classroom Speaking Anxiety Excluding Item FLCSA 3	202
49. Rasch Principal Components Analysis for Foreign Language Classroom Speaking Anxiety Excluding Item FLCSA 3	203
50. Category Structure Functioning for Past English Classroom Experiences (6-point Scale)	206
51. Category Structure Functioning for Past English Classroom Experiences (5-point Scale)	206

52. Category Structure Functioning for Past English Classroom Experiences (4-point Scale)	206
53. Rasch Statistics for the Items Measuring Past English Classroom Experiences	207
54. Rasch Statistics for the Items Measuring Past English Classroom Experiences Excluding Item PECE 8	209
55. Rasch Principal Components Analysis for the Items Measuring the Past English Classroom Experiences Excluding Item PECE 8	209
56. Category Structure Functioning for Desire to Speak English.....	211
57. Rasch Statistics for the Items Measuring Desire to Speak English.....	212
58. Category Structure Functioning for Perceived Foreign Language Speaking Self-Competence	214
59. Rasch Statistics for the Items Measuring Perceived Foreign Language Speaking Self-Competence.....	215
60. Category Structure Functioning for Current English Classroom Perception	218
61. Rasch Statistics for the Items Measuring Current English Classroom Perception	218
62. Rasch Statistics for the Items Measuring Current English Classroom Perception Excluding Item CECP 1	219

63. Rasch Principal Components Analysis for the Items Measuring Current English Classroom Perception Excluding Item CECP 1	220
64. Category Structure Functioning for Perceived Social Value of Speaking English (6-point Scale).....	224
65. Category Structure Functioning for Perceived Social Value of Speaking English (5-point Scale).....	224
66. Category Structure Functioning for Perceived Social Value of Speaking English (4-point Scale).....	225
67. Rasch Statistics for the Items Measuring Perceived Social Value of Speaking English.....	225
68. Explained and Unexplained Variance for the Constructs Including Variance Explained by the First Contrast.....	228
69. Item and Person Separations and Reliabilities for the Constructs Comprising the Foreign Language Classroom Speaking Confidence Questionnaire.....	230
70. Correlational Matrix of Rasch Person Measures for the Six Constructs Comprising the Foreign Language Classroom Speaking Confidence Questionnaire	231
71. Descriptive Statistics for the Foreign Language Classroom Speaking Confidence Questionnaire	233

72. Six-Factor Solution for Exploratory Factor Analysis of the Foreign Language Classroom Speaking Confidence Questionnaire	238
73. Factor Correlation Matrix of the Six Factors Comprising the Foreign Language Classroom Speaking Confidence Questionnaire	242
74. A Comparison of the Foreign Language Classroom Speaking Confidence Constructs Identified by Rasch PCA and EFA.....	244
75. Problematic Items Identified by Rasch PCA and EFA.....	245
76. Comparison of Fit Indices for the Calibration Sample and Validation Sample for the Big Five Personality Traits.....	259
77. Cross-Validation Statistics for Latent Variable Paths in the Confirmatory Factor Analysis of the Big Five Personality Traits	259
78. Comparison of Fit Indices for the Calibration Sample and Validation Sample for Foreign Language Classroom Speaking Confidence	265
79. Cross-Validation Statistics for Latent Variables Paths in the Confirmatory Factor Analysis of Foreign Language Classroom Speaking Confidence.....	265
80. Comparison of Fit Indices for the Calibration Sample and Validation Sample for Favorable Social Conditions	271

81. Cross-Validation Statistics for Latent Variable Paths in the Confirmatory Factor Analysis of Favorable Social Conditions.....	271
82. Comparison of Fit Indices for the Calibration Sample and Validation Sample for Structural Regression Model 1.....	282
83. Cross-Validation Statistics for Latent Variable Paths in the Structural Regression Model 1.....	283
84. Path Strengths among the Favorable Social Conditions and Foreign Language Classroom Speaking Confidence Variables of the Three Equivalent Regression Models.....	290

LIST OF FIGURES

Figure	Page
1. Path analysis of L2 Communication Frequency from MacIntyre and Charos (1996), including Perceived Competence, L2 Anxiety, Integrativeness, Attitudes toward the Learning Situation, L2 Willingness to Communicate, and Motivation.....	41
2. The Heuristic Model of Variables Influencing WTC, from MacIntyre et al. (1998).....	67
3. Model of Second Language (L2) Communication among Anglophones from Clément et al. (2003).....	77
4. Model of Second Language (L2) Communication among Francophones from Clément et al. (2003).....	77
5. The structural path portion of the structural regression model.....	117
6. The item-person map of the Extraversion-Introversion construct.....	153
7. The item-person map of the Agreeableness construct	157
8. The item-person map of the Emotional Stability construct.....	160
9. The item-person map of the Intellect/Imagination construct	164
10. The item-person map of the Intellect/Imagination construct	168
11. The item-person map of the Foreign Language Classroom Speaking Anxiety construct.....	204

12.	The item-person map of the Past English Classroom Experiences construct	210
13.	The item-person map of the Desire to Speak English construct.....	213
14.	The item-person map of the Perceived Foreign Language Speaking Self-Competence construct.....	216
15.	The item-person map of the Current English Classroom Perception construct	222
16.	The item-person map of the Perceived Social Value of Speaking English construct.....	226
17.	Hypothesized model of a causal structure for foreign language classroom speaking anxiety.....	252
18.	Initial confirmatory factor analysis model of the Big Five Personality Traits, consisting of 34 items.	254
19.	Final confirmatory factor analysis model of the Big Five Personality Traits, consisting of 14 items.	258
20.	Initial confirmatory factor analysis model of Foreign Language Classroom Speaking Confidence latent variables, consisting of 28 items.....	262

21.	Final confirmatory factor analysis model of Foreign Language Classroom Speaking Confidence latent variables, consisting of 16 items.....	264
22.	Initial confirmatory factor analysis model of Favorable Social Conditions latent variables, consisting of 20 items.	267
23.	Final confirmatory factor analysis model of Favorable Social Conditions latent variables, consisting of 12 items.	269
24.	Revised structural regression model of a causal structure for foreign language classroom speaking anxiety.....	273
25.	Structural regression model 1 of a causal structure for foreign language classroom speaking anxiety after model path respecification.....	281
26.	Structural regression equivalent model 2 of a causal structure for foreign language classroom speaking anxiety, in which DSE influences PFLSS.....	286
27.	Structural regression equivalent model 3 of a causal structure for foreign language classroom speaking anxiety, in which FLCSA influences DSE.	288

CHAPTER 1

INTRODUCTION

The Background of the Issue

The issues discussed in this study stem from a question that I was asked in the spring of 2005. I was participating in a teacher-training seminar held in Vietnam. After one workshop, a Vietnamese teacher approached me and, upon finding out that I was living in Japan, asked me, “Why don’t Japanese people speak to me?” It seemed that the teacher had occasionally encountered Japanese tourists in Vietnam, and was confused by the tourists’ apparent reluctance to strike up a conversation.

I realized that I had often asked myself the same question: Why were many Japanese reluctant to speak? However, I also was aware that the question itself may have been a bit stereotypical; from personal experience, I knew that some Japanese were not at all reluctant to speak, depending on the situation and depending on the extent to which they personally knew and felt comfortable with their interlocutors. I was also aware, and constantly reminded on a regular basis by my Japanese teaching colleagues, that Japanese students are often considered “shy,” and that this shyness is said to be a personality trait that was directly responsible for university students’ apparent lack of communicative confidence in English.

On the other hand, I was also reminded of my own experiences speaking in public, particularly in the classroom. Personally, I have always found it challenging in even a semi-formal setting to speak in front of a group. One especially embarrassing high school history class presentation stands out in my memory: during the course of a five-minute individual presentation, I gradually turned pale and became drenched with sweat. After I asked in a squeaky voice to be excused from the class, I fled to the nearest men's room and nearly fainted. My history teacher thought I had an attack of influenza, but I later realized that it had been the result of extreme speaking anxiety. Since that incident I have not experienced speaking anxiety to that extreme when using my first language (L1), although I continue to struggle with anxiousness when speaking English in front of a large number of people. As a university student I often experienced anxiety at the mere thought of speaking in front of classmates in my French and German classes. Even today after living in Japan for over a decade, I still often feel an acute lack of confidence speaking Japanese in various public situations, ranging from faculty meetings to ordering food in a restaurant. My L1 speaking anxiety seems somehow related to my experience of speaking anxiety in other languages, but the degree to which this anxiety seems to affect my speaking confidence seems to depend on the social situation and the familiarity I have with the interlocutors.

As I began to consider topics for a doctoral dissertation, the question asked to me by the English teacher in Vietnam came back to me, and I started to make connections with my own language learning and speaking experiences. I couldn't help feeling that there was some relationship between aspects of my personality and anxiety or lack of confidence in speaking. My ongoing difficulty with speaking confidence in Japanese, despite having spoken it for almost ten years, led me to wonder whether my university students in Japan were experiencing similar feelings. The questions began to form in my mind: What were the factors involved in speaking confidence? How could I accurately measure these factors with my current university students? What relationships exist between my students' personalities and speaking confidence or lack of confidence? These questions, fueled by personal curiosity and experience, led me to choose the complexly intertwined topics of personality and speaking confidence in this study.

Statement of the Problem

While studies concerning both personality traits and anxiety have a long history in first language contexts, studies about personality and anxiety in the field of second language acquisition have failed to demonstrate overall robust findings. The frustrating lack of significant findings have convinced many researchers that psychological constructs such as personality and anxiety are best measured on an

individual basis using qualitative techniques, such as case studies or ethnographic research. However, researchers in the field of second language acquisition (SLA) have made many errors by failing to validate their measurement instruments, by using small numbers of participants, and by relying too much on simple correlations rather than more complex statistical analyses to demonstrate relationships among the psychological constructs in their studies.

The first major problem underlying flaws in previous SLA studies of personality and anxiety is the lack of instrument validity. This problem has occurred due to a misunderstanding of the nature of instrument creation. A prime example of this problem is a reliance on the relatively unsophisticated statistical procedures of bivariate correlational analysis and item reliability estimates as measured by Cronbach's alpha purportedly to demonstrate the "construct validity" of the construct being measured. SLA researchers have often assumed that their instruments, frequently borrowed from previous researchers, were indeed measuring the intended constructs. However, without the crucial step of instrument validation, further data analyses cannot be considered meaningful or trustworthy, because it is not possible to know whether the data actually represent the latent constructs that the researcher wants to investigate. One must examine the data "noise," or the standard residuals, to verify the unidimensionality of the construct measurement instrument.

The second problem is the misuse of Likert-scale data. Many researchers add raw scores obtained from the factor analysis of Likert-scale items in order to create “composite scores” for various constructs as a preliminary step for correlational or other analyses. Not only do such composite indexes tend to ignore the existence of a different number of items per factor, the addition of Likert-scale data is based on the unwarranted assumption that an answer of “strongly disagree” is the same interval distance from “somewhat agree” on any given item. Measurement studies have repeatedly demonstrated that Likert-scale data is not true interval data; by consistently treating it as such, SLA researchers run the serious risk of compromising internal validity. Even assuming that the variable was measured as intended, using combined raw scores leads to inaccurate estimates of the location of each individual participant on the latent variable, because they are ordinal, not interval measures of the construct.

In addition to a lack of instrument validation and the misuse of Likert-scale data, SLA studies of personality and anxiety have often suffered from a number of methodological flaws. The most crucial flaw in many studies has been small sample sizes in proportion to the number of variables being measured. Several studies of foreign language anxiety in particular have used questionnaires with an almost equal number of items to that of participants in the study. Furthermore, several researchers showed a lack of understanding of the nature of statistical

factors, claiming to find “factors” that were defined by three items or fewer. Although statistical software will often run analyses using sample sizes of under 100 participants or under the recommended case-to-variable ratio of 10:1 (Stevens, 2009; Tabachnick & Fidell, 2007), results from studies that consist of 60 to 80 participants taking a questionnaire with 60 to 80 items are questionable at best. It is unsurprising that many such personality and anxiety studies have remained difficult to duplicate, because the findings differ depending on the participant sample.

Finally, an overemphasis on high internal reliability demonstrates a lack of understanding of what reliability indices represent. For example, a small number of “bipolar” items (i.e., “I like English” and “I do not like English”) will naturally be correlated highly because the items are measuring the same concept. However, high reliability does not indicate that the intended construct was measured or that individual participant fit on the construct was verified (Green, Lissitz, & Mulait, 1977; Sijtsma, 2009). A reliability estimate only shows that the respondents’ answers were consistent; it cannot show whether the respondents have been measured precisely.

Purposes of the Study

The first purpose of this study is to address validity problems in existing SLA personality and anxiety research through the use of Rasch analysis to validate

questionnaires measuring the constructs of personality, speaking anxiety, and perceived self-competence. The Rasch model directly addresses the misuse of Likert-scale data present in many previous studies. The methodological problems that have plagued many studies are also addressed in this study: The *N*-size is in excess of 1,000 participants and relatively large numbers of good performing items are used to measure each latent construct. In addition, the Rasch model can confirm the unidimensionality of the constructs without relying on Cronbach's alpha estimates or bivariate correlational analysis. The second purpose of this study concerns the personality theories used in SLA research. By using the Big Five model of personality traits, the study will help encourage SLA individual differences research to fall more in line with current trends in mainstream psychology. Through instrument validation, the study can additionally confirm whether all five of the hypothesized Big Five personality traits exist for a Japanese sample.

The third purpose of this study concerns the concept of foreign language confidence. The study is focused on only one aspect of confidence—that of foreign language classroom speaking confidence—in order to measure this complex psychological construct with a high degree of unidimensionality. The relationship between genetic, or inherited personality traits, and social variables is also examined to bring theories of foreign language speaking confidence into line with

theories from mainstream social anxiety. Finally, structural equation modeling is used to confirm the theory that trait personality is the key to understanding how foreign language speaking confidence affects foreign language users in a classroom setting. The misuse of data-driven models will be addressed through the use of a calibration sample to test the structural equation model and a validation sample to confirm the model after path respecification. Alternate models will also be examined to avoid confirmation bias.

The Audience for the Study

This study should benefit researchers interested in personality and affective constructs in two ways. First, researchers can use two validated questionnaires designed to measure personality traits and foreign language classroom speaking confidence. Second, some researchers will discover new statistical methods of inquiry that can be used to examine the interaction between genetic and social aspects of speaking anxiety in the foreign language classroom.

Foreign language classroom teachers should also benefit by learning more about the psychological aspects at play among learners in the foreign language classroom. Particularly in a foreign language, as opposed to second language, learning environment, students' perceptions of self- and peer-competence as well as

perceptions of peer attitudes towards foreign language classroom speaking can strongly influence speaking anxiety and confidence.

In Japan, university English classrooms that focus on oral communication often have at least 25 students, and classes consisting of 40 to 45 students or more are not uncommon at larger schools. Because class sizes and schedules are usually decided by university language program administrators, it is also hoped that the results will inform such administrative decisions as determining how many students there should be in one class, as well as including students with mixed major fields of study in a single classroom.

Delimitations

The first delimitation of this study is the nature of the participants: All are Japanese students who speak the same native language and are approximately the same age (average age 18 to 19 years old). Caution must be made when attempting to generalize the results to a larger Japanese population or to populations in Asia or other parts of the world.

A second delimitation is the nature of the method used in the study to gather data. Only one data collection method was used: two Likert-scale questionnaires implemented once during class time. Thus, the study is based entirely on self-reported information consisting of participants' perceptions. Additionally, the

study can give only a “snap-shot,” cross-sectional perspective of participants at a specific point in time. Because data were collected from 49 intact classes in twelve separate institutions spread across the main island of Honshu, other methods of data collection such as interview and classroom observation were impractical.

The Organization of this Study

A review of the literature concerning the history and development of the Big Five personality traits, social anxiety, and foreign language confidence in language education is presented in Chapter 2. Chapter 3 details the study participants, instruments to be used, research design, analytical techniques, and statistical models to be tested. Chapter 4 presents the results of instrument validation using a combination of Rasch measurement model analysis and exploratory factor analysis. Chapter 5 presents the results from confirmatory factor analysis and the structural regression modeling. Chapter 6 follows with a discussion of the results, and Chapter 7 concludes the study and offers possible directions for future studies.

CHAPTER 2

REVIEW OF THE LITERATURE

In this chapter, I discuss literature concerning the concept of the Big Five personality traits, the current state of second or foreign language speaking confidence research, and the relationships among personality and foreign language speaking confidence. In the first section of the chapter, I summarize the history behind the development of the Big Five model for personality traits, with particular attention paid to the differences between the Big Five and the Five Factor Model. I then follow the historical summary with a brief examination of competing models to the Big Five, including Eysenck's Big Three "PEN" model and non-English language studies of the Big Six and Big Seven. Finally, I describe personality trait studies that have been carried out in East Asia, specifically in China, Korea, and Japan.

In the second section of the chapter, I begin my discussion of the concept of confidence by reviewing the concepts of trait anxiety and social anxiety from mainstream psychology. Next, I discuss the concepts of Communicative Apprehension and Perceived Speaking Self-Competence in the L1 context of North America. After that, I summarize the development of the Foreign Language Classroom Anxiety Scale and its relation to the ever-changing construct of L2

Confidence in the field of SLA. This is followed by an examination of recent foreign language speaking confidence questionnaire studies, with a particular focus on studies conducted in East Asia. Finally, I conclude with a brief summary of methodological and analytic flaws in existing literature that the present study is intended to address.

Big Five Personality Traits: A Summary

History of the Big Five: Lexical and Questionnaire Traditions

The concept of using language to reduce basic human personality traits to a finite number of factors has its origins in the late 1880s with the adjective studies of Galton (1884), who believed that all human languages encoded individual distinctions of personality by means of the adjectives used. Galton postulated that from 1000 adjectives whose meanings were slightly different, a smaller number of adjectives could be found whose meanings overlapped or shared enough in common to express human character.

However, the first indication of the existence of the five-factor model came in the mid-1930s when Thurstone, a pioneer in factor analysis development, chose 60 “common” adjectives and asked 1300 raters to underline which adjective they would use in conversation with a person whom they knew well (Thurstone, 1934). The factor analysis of the data resulted in five independent factors. Strangely

enough, Thurstone chose not follow up on his discovery and instead focused on developing the Thurstone Temperamental Schedule. He did not even bother to name the five factors (Digman, 1996). Cattell (1943), along with many other psychologists, could not believe that personality could be accounted for using as few as five factors. Instead, Cattell used 35 variables and found 12 factors using oblique rotational factoring; other researchers have been unable to duplicate more than five of these factors using orthogonal rotational methods (see Digman, 1990; Goldberg, 1990).

Interest in the search for universals in personality traits intensified in the first few decades following Thurstone's and Cattell's pioneering research using factor analysis. From the 1940s to the 1980s, this interest was exemplified by a veritable "alphabet soup" of numerous personality trait instruments and their acronyms, such as the Minnesota Multiphasic Personality Inventory, or MMPI (1939), the Yawate-Guilford Personality Inventory, or Y/GPI (1953); the Norman Peer Rating Scales, or NPRS (Norman, 1963); the Eysenck Personality Questionnaire, or EPQ (Eysenck & Eysenck, 1975), and the Hogan Personality Inventory, or HPI (Hogan, 1986). Despite the numerous, and sometimes antagonistic, attempts to confirm personality traits, a lack of reproducibility of factors from various studies and questionnaires led to doubts that the search for personality universals would reach a consensus.

However, in the early 1980s a movement to consolidate the various personality inventories began. Goldberg had previously worked on the Eysenck PEN model (Peabody & Goldberg, 1989; see the Competing Models to the Big Five section), but in a review of personality trait studies, Goldberg termed the five-factor model “robust” and noted that “*any* model structuring individual difference will have to encompass...something like these ‘big five’ dimensions” (Goldberg, 1981, p. 159). Goldberg thus is said to be the first researcher to use the phrase “Big Five” to describe the five-factor lexical model. He went on to develop what he termed Big Five “factor markers,” and eventually produced two self-rating scales of 50 items and 100 items (Goldberg, 1992, 1993). The Big Five Factor Markers dimensions of personality are called Extraversion-Introversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect-Imagination.

Meanwhile, Costa and McCrae (1985) had independently developed a three-factor “personality inventory” model based on neuroticism, extraversion, and openness to experience, which they called the NEO-PI (Costa & McCrae, 1985). They later added additional items to conform to the emerging five-factor consensus, making a Revised NEO-PI with 240 items (McCrae & Costa, 1985; 1987), and further produced a shorter version of 60 items called the NEO Five Factor Inventory (NEO-FFI) (Costa & McCrae, 1992). In studies of personality traits, the terms Big Five and Five-factor Model (FFM) have thus come to be used

interchangeably, and the Big Five moniker is now regarded as a “template” or “taxonomy” for personality trait researchers (De Fruyt, McCrae, Szirmak, & Nagy, 2004; De Raad & Perugini, 2002; John & Srivastava, 1999). In practice, personality trait researchers who use their own instruments typically validate their findings by correlating their instrument with Goldberg’s Factor Markers questionnaire or Costa & McCrae’s NEO-PI-R, or both (e.g., Gow, Whiteman, Pattie, & Deary, 2005).

Personality Trait Measurements: Factor Markers vs. NEO-PI-R

While the two main instruments for measuring personality traits seem to be in close agreement, the factors measured by each instrument may differ. Each instrument uses the terms Factor I, Factor II, Factor III, Factor IV, and Factor V; however, only Factor I, Factor III, and Factor IV have consistently remained the same during the past twenty years of research. Factor I was originally known as “Surgency,” but now is generally called “Extraversion.” Factor II is known variously as “Agreeableness,” “Friendliness,” and “Socialization”; the overall meaning seems to involve individual perceptions of other people or relationships with other people to some extent. Factor III was once called “Will to Achieve” but is now known as “Conscientiousness.” Factor IV is called “Emotional Stability” in the Big Five construct and “Neuroticism” in the FFM construct. This labeling of neuroticism seems to reflect the large number of negative items in this factor as

well as the essentially negative viewpoint of human personality problems from past researchers in clinical and developmental psychology (see Digman, 1990, p. 423, for a more comprehensive listing of past factor labels).

Factor V remains the most troubling factor, and indeed, many researchers argue that this factor may differ depending on the language or culture of the participants in the studies. Although this factor has come to be known as “Intellect/Imagination” in the Big Five model and “Openness to New Experiences” in the Five Factor model, it was, somewhat ironically, initially called “Culture” (Goldberg, 1993; Norman, 1963). Many researchers working outside the North American, English language-only context have argued strenuously that Factor V ought to be separated into two related factors based on cultural or linguistic coding differences in participants. Indeed, recent studies in East Asian countries have pointed to the possibility of six or seven factor model versions based on lexical differences in Asian languages (see the Big Five Studies in East Asia section). The nature of the lexical items included within this factor may point to the existence of two factors, which may be termed “Imagination” or “Abstract Thought” and “Intellect” or “Linguistic Awareness.”

Just before the turn of the 21st century, a number of studies testing the cross-cultural validity of the Big Five structure were published. Saucier, Hampton, and Goldberg (2000) reviewed different language versions of the Big Five

instrument and reported good replicability in several languages, including Indo-European languages, such as German and Russian, and non-Indo-European languages, such as Hebrew and Filipino (Saucier, Hampton, & Goldberg, 2000). McCrae (2000) additionally claimed that, due to recent consensus on the five-factor model of personality traits, "...the demonstration that the same structure of traits can be found in many quite different languages and cultures opens the way to a new approach to personality and culture" (McCrae, 2000, p. 11). By this statement, McCrae perhaps meant that the search for universal personality traits was over, and that researchers could focus on how cultural settings influence the development of the Big Five personality traits in localized contexts.

However, the debate on whether there truly are five personality traits shared by all human beings seems far from over. As I discuss in the following section, alternate models to the Big Five existing prior to the Big Five still garner international support, while new models of personality traits have been proposed by several researchers in Asian countries.

Competing Models to the Big Five

In the mid to late 1990s the Big Five, a. k. a., Five Factor personality trait model rapidly rose to prominence in mainstream psychology and is currently the favorite model in the field; the term Big Five has even entered popular culture, and

one can now find freely available Big Five personality quizzes on the Internet. However, even as the five-factor model was becoming the dominant personality trait paradigm, the debate raged in psychological journals between the main proponents of the five-factor model, Costa and McCrae, and the main proponent of the three-factor Psychoticism, Extraversion, and Neuroticism (PEN) model, Eysenck (Costa & McCrae, 1992; Eysenck, 1992).

The PEN model, which is measured with the Eysenck Personality Questionnaire, or EPQ (Eysenck, 1992; Eysenck & Eysenck, 1975), was proposed originally as a primarily “nervous” explanation of personality. Eysenck later amended the term nervous as “genetic factors,” including the human body’s physical tolerance to external stimuli (Eysenck, 2000). Eysenck criticized the lexical Big Five tradition from social psychology as a mere “description” of personality behaviors, rather than following the scientific approach used in clinical psychology of first theorizing, developing a model, and then testing the model based on observable phenomena (Eysenck, 1991).

Thus, when several personality trait studies based on the lexical model in non-English speaking cultures failed to replicate one or more of the Big Five factors, researchers quickly turned to the PEN model and hailed it as a “Big Three.” For example, researchers in Italy claimed that only three of the five factors were stable across language groups due to cultural differences (Di Blas & Forzi, 1999).

Even working in Dutch and German, which are related to English lexically, researchers have found it difficult to reproduce all Big Five factors and suggested a return to the Big Three (Peabody & De Raad, 2002).

In fact, many recent researchers working in non-English language contexts have been unable to duplicate all five factors from the Five Factor Model. The Factor V of Intellect/Imagination or Openness to New Experiences has proved the most difficult to duplicate, but other factors, such as Factor I (Extraversion) and Factor II (Agreeableness) have also proved difficult to analyze as separate, uncorrelated factors. As a result, many personality researchers in non-English language contexts have proposed alternative fifth factors, while others have proposed six- or even seven-factor models.

A commonly agreed upon alternative fifth factor is Humility or Honesty, which has been discovered in several non-Germanic language-based lexical personality trait studies. For example, in Mexico, Ortiz et al. (2007) found that factors from a Mexican inventory of personality traits supported four of the factors of the five-factor model. However, in the place of an Openness to New Experiences or Intellect/Imagination factor, the researchers found a Humility-Honesty factor (Ortiz, Church, de Jesus Vargas-Flores, Ibanez-Reyes, Flores-Galaz, Iuit-Briceno, & Escamilla, 2007). Other European language studies agreed with the results of the Mexican study; many personality trait researchers in European contexts reported

the existence of between five to seven factors and proposed a new factor variously called Humility, Honesty, or Integrity in studies carried out in French, Hungarian, Italian, Dutch, Turkish, Czech, and Polish (Ashton, Lee, de Vries, Perugini, Gnisci, & Sergi, 2006; Ashton, Perugini, de Vries, Boies, Lee, Szarota, Blas, & Raad, 2004; Mlačić & Goldberg, 2007).

In addition to Humility or Honesty, other researchers have proposed separating factors into positive and negative components. In studies conducted in Hebrew, Spanish, and Filipino, researchers reported seven factors. These seven factors included negative emotionality, positive emotionality, negative valence, and positive valence (Ashton & Lee, 2001; Peabody & De Raad, 2002). The same researchers additionally suggested that *judgment* could be used instead of *valence*, perhaps due to the influence (consciously or subconsciously) of the Myers-Briggs Type Indicator (MBTI) model, which includes a *judging-feeling* bipolar scale. Interestingly, six- and seven-factor studies by Ashton and colleagues were conducted using the same languages and in the same countries as recent Five Factor Model studies reported by Saucier, Hampton, and Goldberg (2000).

Thus, even in European language-based studies support both for and against the Big Five Factor Model is extent. Next, I discuss a few studies from East Asia, which is a similar context to that in which the present study was conducted.

Big Five Studies in East Asia

Studies carried out in the so-called collectivist societies, such as China, Korea, and Japan, have both supported and discounted the five-factor model. The concept of the individual personality in these countries is frequently described as a mutual understanding of societal roles termed the *interdependent self*, compared to the Western concept of the individual as an *independent self* who is separate from others in society. The interdependent self is thought to be particularly strong in China, Korea, and Japan, which are often considered Confucian-influenced cultures (Kim, 2005). In these societies, emphasis is placed upon the “importance of understanding one’s roles within a hierarchy and of fulfilling obligations to others that are associated with these roles” (Heine et al., 2001, p. 600).

The importance of group roles and interpersonal obligations in Japanese society in particular is highlighted by a reliance on external evaluations by others to determine the sense of self (Kitayama, Snibbe, Markus, & Suzuki, 2004; Kitayama & Uchida, 2003) and the need to preserve “face,” or the sense of how others evaluate the self (Heine, Takemoto, Moskalkenko, Lasaleta, & Henrich, 2008). Speakers of Japanese are thus often likely to hesitate to speak their opinions openly due to concern about using words appropriate to their perceived role in the social group as *senpai* (senior, older, more experienced) or *kohai* (junior, younger, less experienced) (Heine, Kitayama, Lehman, Takata, Ide, Leung, & Matsumoto, 2001;

Heine, Lehman, Peng, & Greenholtz, 2002; Kitayama & Markus, 1999; Markus & Kitayama, 1991).

On the other hand, some researchers have argued that the interdependent-independent/collective-individual dichotomy is a false stereotype of Eastern and Western cultures, and that Western cultures can at times be more collective than so-called collectivist cultures. For example, a recent study comparing Japanese and American university students found that, although Japanese students were more likely to take other people's opinions into account before speaking, American students were more likely to help others who were in need (Noguchi, 2007). This could be partly explained by a "moralist aspect" of Americans helping others due to religion (Noguchi, 2007, p. 140) or by a tendency of people to compare themselves to others within their own cultural framework rather than to people of other cultures (Heine et al., 2002).

An alternative explanation to the mixed findings of the Big Five in East Asian societies could be, as Hofstede (2007) argued, that personality inventories simply do not take into account the complicated nature of human relationships and are based on North American paradigms of individualism and collectivism. Thus, Big Five personality studies conducted in China, Korea, and Japan find confusion between the expected factors of Extraversion and Agreeableness, and Conscientiousness and Intellect/Imagination. Further, studies in East Asia often

hint at the existence of at least a sixth factor, which Hofstede suggested could be labeled Dependence on Others (Hofstede, 2007, p. 19). Other researchers, noting the successful translation of the Chinese Personality Assessment Inventory into English—and its subsequent renaming as the Cross-Cultural Personality Assessment Inventory—have noted that Asian personality researchers emphasize the social nature rather than the independent nature of the human psyche, and that studies conducted in Asia could actually “inform the blind spot in Western trait measures of personality” (Cheung, Cheung, Wada, & Zhang, 2003, p. 287). I describe a few of these recent East Asian personality studies in the next three sections.

Chinese Big Five Studies. In the early 1970s, anthropologist Francis Hsu argued that Chinese culture did not consider individual personality separate from society, noting that the Chinese word for “man” (*ren* or *jen*) encompassed the individual person’s societal and cultural context necessary for the construction of the individual personality (Hsu, 1971). Recently, Chinese personality researchers have attempted to demonstrate this “uniqueness” of Chinese personality through the creation of several personality measurement instruments unique to the Chinese lexicon. Using confirmatory factor analysis and comparing results from the NEO-FFI to a Chinese questionnaire called the Chinese Personality Assessment

Inventory (CPAI), Chinese researchers reported that the NEO-FFI factors of Extraversion and Agreeableness were intermixed among Chinese students and that the Openness to New Experiences factor did not appear at all. Based on the results, the researchers argued that personality inventories such as the NEO-FFI that were based on Western traditions did not measure a new factor that they termed *Interpersonal Relatedness*. According to the researchers, Interpersonal Relatedness exists in Asian cultures because Asians value interpersonal relationships more highly than those in Western cultures due to the historical influence of Confucian beliefs and ethics (Cheung, Leung, Zhang, Sun, Gan, Song, & Xie, 2001).

Working in the lexical tradition of the Factor Marker questionnaire, Wang, Cui, and Zhou (2005) conducted a comparative study of the NEO-PI-R to their Chinese Personality Scale (QZPS). A total of 2540 participants in mainland China took both the Chinese version of the NEO-PI-R and the QZPS, including university students' friends and family members, representing age groups between 15 and 60 years old. The first five factors from the NEO-PI-R data explained 52.7% of the variance and had correlations between $r = -.31$ and $r = .66$ with the seven factors of the QZPS, whose total variance was not reported. When the 420 items from both questionnaires were combined, the seven factors explained 53.9% of total variance. These seven factors were Extraversion, Kindness, Behavior Styles, Talents, Emotionality, Human Relations, and Ways of Life. The researchers concluded that,

because the five factors of the NEO-PI-R were dispersed and did not have a complete correspondence to the seven factors of the QZPS, the Big Five structure did not adequately explain all facets of Chinese personality. The researchers also noted that, similar to other studies in Eastern Asia, the Openness to New Experiences factor did not emerge at all. They speculated that this was due to an emphasis in Eastern cultures on evaluation and judgment of personal character, rather than a Western emphasis on objective description and analysis of personal characteristics. Thus, the NEO-PI-R had an “emic-component,” meaning that items were culturally-biased in favor of Western expectations of personality.

Problems with the study included the deletion of over half of the items on the NEO-PI-R and almost 20% of the QZPS items due to lack of factor loading, as well as a lack of unidimensionality tests for either questionnaire. Moreover, the researchers simply assumed that the NEO-PI-R had five factors, and while they presented four-, six-, and seven-factor solutions for the QZPS, they concluded that the QZPS had seven factors based on a combination of scree plots and variance accounted for by the solution. The total variance accounted for by the QZPS was between 52.7% (four factor solution) and 53.9% (seven factor solution). Put another way, the results indicated that between 46.1% and 47.3% percent of the variance was not accounted for by the questionnaire. Even with an impressive *N*-size of close to 2500 participants, the missing variance may have contained a

sizeable number of additional factors resulting from correlated residual errors after the factor extraction.

The researchers in this study supported their claim that the QZPS was more suited to measuring Chinese personality on the basis of having to eliminate fewer non-loading items as the NEO-PI-R. However, unidimensionality cannot be confirmed through the elimination of questionnaire items. Without a measurement of the residuals after factor extraction, the researchers could not be sure how accurately the constructs were measured. A final caveat is the comparison of the non-lexical tradition NEO-PI-R to the lexically based QZPS by means of correlational analysis in order to confirm construct validity. The NEO-PI-R factors correlated between $r = -.40$ and $r = .87$ with various QZPS factors; four of the five factors correlated at a level of $r = .75$ or higher with at least one QZPS factor, but Openness to New Experiences correlated at between $r = .34$ and $r = .48$ with three QZPS factors. After correlating the two questionnaires, the researchers stated that they had kept only those items in the NEO-PI-R that “reflected Chinese personality characteristics,” while eliminating other “inappropriate items” (Wang et al., 2005, p. 111). The researchers thus claimed that their results indicated the inappropriateness of the NEO-PI-R for measuring a Chinese population, and “confirmed the construct validity of the QZPS” (p. 111). Again, the problem is that unidimensionality and construct validity cannot be confirmed through correlational analysis; such analysis

can confirm to what degree the participants consistently respond to questionnaire items, but it cannot confirm that the response patterns reflected unidimensional constructs in the participants.

In Hong Kong, Lun, and Bond (2006) conducted a study of 96 undergraduate students using the NEO-FFI 60-item short version questionnaire. Students first took the questionnaire, and then performed three group tasks during a period of three months. At the end of the three-month period, two of the student group tasks were assessed by a teaching assistant. Students were asked to list the names of people in the group and were given an “Interpersonal relationship harmony inventory” in order to evaluate the level of “harmony” between the group members (Lun & Bond, 2006). Contrary to the researchers’ expectations, the Extraversion and Agreeableness factors did not correlate significantly with the group harmony factors, whereas the Openness to Experience and Conscientiousness factors showed significant medium negative correlations to group harmony. None of the personality factors correlated significantly with group performance, but group harmony was statistically significant with a strong correlation ($r = .63$). Flaws in the study were a short time span (three months), a required course assignment, intact classes and lack of randomization, and a total of 95 variables for 96 participants. Further, both the group harmony and group performance factors were essentially self-rater, peer-rater, and teacher-rater scores, all of which might have

contained rater bias. Without a Rasch or other logit analysis to account for rating bias, particularly because the group performance data relied on the opinions of only one rater, the data results that were based on these factors must be considered tentative.

Korean Big Five Studies. Yoon et al. (2002) examined 717 Korean company employees using the NEO-PI-R 240-item questionnaire and found five factors in agreement with the FFM model. However, the researchers also noted higher levels of Neuroticism and lower levels of Openness to New Experiences and Extraversion when compared to results from a North American sample. The Agreeableness and Conscientiousness factors were moderately correlated ($r = .38$), leading the researchers to speculate that the Korean respondents had a “modesty bias” and tended to perceive themselves as less capable than peers due to “the Confucian ethic” (Yoon, Schmidt, & Ilies, 2002, p. 231).

In another study of 435 Korean university students, researchers presented three-, four-, five-, six-, and seven-factor solutions (Hahn, Lee, & Ashton, 1999). The researchers described the five-factor solution as showing strong similarity to the Big Five model. However, the researchers also noted that the Conscientiousness and Intellect factors were not clearly separated, and speculated that, perhaps due to the nature of the Korean lexicon, “a large number of ‘controlled’ intellect terms,

which correlate positively with Conscientiousness...may tend to move the Conscientiousness factor in the direction of Intellect” (Hahn et al., 1999, p. 278). The researchers suggested that a six- or seven-factor model of personality traits might be more appropriate for a Korean population.

In both Yoon et al. and Hahn et al., no mention was made of construct validity checks to confirm the unidimensionality of hypothesized constructs. The number of factors was determined based solely on correlations with other questionnaires, and item validity and reliability was checked with Cronbach’s alpha reliability estimates.

Japanese Big Five Studies. There have been comparatively few studies of personality conducted in Japanese using the Big Five model. In one of the first Big Five-related personality studies, Bond et al. (1975) found evidence of a five-factor structure among 91 Japanese university students at Kwansai Gakuin University, Kobe. In the study, the researchers used a questionnaire based on the Norman Peer Rating Scale (Norman, 1963), which consisted of 20 polar scales of four items each. The researchers then performed correlational analyses of the data and factor analyzed the correlation matrix. After assuming the existence of the same five factors as studies carried out in the United States (Norman, 1963) and the Republic of the Philippines (Guthrie & Bennett, 1971), the researchers compared the results

and discovered that although the factors for Extroversion, Agreeableness, and Conscientiousness seemed similar in the three studies, the final two factors of Emotional Stability and Culture were more problematic both in terms of the number of items with complex loadings and the low factor loadings of items in general. The researchers interpreted the problematic factors as deriving from cultural differences: “It is probable that the behaviors which constitute these dimensions are specific to each culture, assuming that these two dimensions are used in those cultures” (Bond, Nakazato, & Shiraishi, 1975, p. 354). A major flaw of the study was the small sample size ($N = 91$) and relatively large number of questionnaire items (80), leading to a lack of power and generalizability in the statistical findings.

The first sign of the Big Five in Japan was Isaka (1990), who conducted three successive studies in the psycho-lexical tradition and ultimately concluded that there were ten factors that roughly corresponded to the same five factors as the Bond et al study. Several Japanese language studies then appeared in Japanese psychology journals shortly thereafter (Kashiwagi, 1999; Kashiwagi & Wada, 1996; Kashiwagi, Wada, & Aoki, 1993; Kashiwagi & Yamada, 1995; Wada, 1996). An English-language review of these studies appeared in Kashiwagi (2002), who additionally conducted two studies of his own and reported the results. In the study, 218 Japanese university students were given a list of 200 adjectives and were asked

to respond *Yes*, *Questionable*, or *No* as to whether the adjectives best described themselves. The researcher conducted a number of varimax and incomplete orthogonal procrustes rotations, eliminating several items step by step before finally arriving at a final list of 105 Japanese adjectives that corresponded to the Big Five structure. Although he found high reliability for most factors (Extraversion, 28 items, $\alpha = .93$; Agreeableness, 19 items, $\alpha = .87$; Conscientiousness, 15 items, $\alpha = .83$; Neuroticism, 21 items, $\alpha = .90$; Openness, 22 items, $\alpha = .90$), the Conscientiousness factor correlated moderately with Agreeableness ($r = .34$) and the Openness factor correlated moderately with Extraversion ($r = .46$).

Similar to other Big Five studies, the Kashiwagi study committed four statistical flaws. First, the number of cases (218) was not in a ratio of least five to one with the number of variables (105). Second, the existence of five factors was assumed without examining the possibility of other factor solutions. Third, correlation among items on a factor cannot demonstrate construct validity. Finally, the use of a non-Likert-scale based answering key with two response options and a non-response option (*Yes*, *Questionable*, and *No*) did not place person responses on an interval scale according to item endorsement difficulty level, thus leading to results of questionable psychometric quality.

As part of a larger studying combining data from several smaller, unpublished studies to compare the Big Five traits in five countries, Yik et al., examined 450 students at Doshisha University, in Kyoto, Japan, using a circumplex model with the Big Five traits as predictor variables of affective variables (Yik, Russell, Ahn, Dols, & Suzuki, 2002). Personality was measured by the NEO-FFI 60-item questionnaire, and affect was measured by an unspecified number of items from five separate scales of bipolar measures of various related types of affect (e.g., pleasure, tiredness, and calmness). Data from the Japan-based study were correlated with studies in the United States, Spain, China, and Korea and input into a CIRCUM structure modeling program, which placed the correlations into a “two-dimensional space” that “captures the core of affect” (p. 90).

For data analysis “40 structural models in all” (p. 91) were computed, each personality dimension being assigned three separate figures: angle, zeta, and VAF (Variance Accounted For). These figures indicated the “location” of the personality variable in the circumplex model, the correlation between the personality variable and the affective variable, and the model fit of the personality variable in the circumplex model, respectively. Results indicated that neuroticism (.43) and extraversion (.24) showed the greatest mean zeta (i.e., correlation between personality and affect).

However, there were several flaws in the study. First, the number of participants per study varied from 233 (Spanish) to 535 (English), meaning that correlations were performed between studies consisting of an unequal number of participants. Also, while the researchers listed the Cronbach's alpha estimates for each personality factor per study, ranging widely from a low of .62 to a high of .87, no reliability estimates were provided for the various affect scales. The number of items was also not given for the affect scales, perhaps on the assumption that readers would be familiar with the five scales used. Further, the measurement instruments used were not examined for construct unidimensionality; thus, the researchers had no way of knowing whether the constructs were measured as intended.

Other personality trait studies conducted in Japan have taken place within the context of SLA research. In the next section, I begin by discussing a few key SLA studies concerning the relationship between personality variables and language learning, particularly with reference to studies conducted in Japan and other Asian contexts. However, I pay special attention to two studies: One of the most influential empirical studies regarding personality and confidence, and one of the quite possibly most influential, and also controversial, theoretical model of affective factors and foreign language production in the field of SLA.

Personality Traits and Second Language Acquisition

The potential influence of personality traits on non-linguistic, affective factors related to second language acquisition was noted early in the history of SLA. Even in the field of psychology, most researchers agreed that Extraversion and Neuroticism were most strongly linked to affect (e.g., Eysenck, 1992; Eysenck & Eysenck, 1985; McCrae & Costa, 1991), and a few researchers reported on the relationship among other factors, such as Conscientiousness, Agreeableness, and Openness to experience and affective factors, such as confidence and self-esteem (e.g., Watson & Clark, 1992). Since affect greatly influences second or foreign language speaking, SLA researchers had some reason to believe that personality would thus be key to understanding the roots of foreign language speaking anxiety and confidence. However, early SLA researchers found mixed results, leading to Lalonde and Gardner's (1984) pessimistic statement that "[b]ased on such research, there is little reason to conclude that personality variables are directly implicated to any great extent in second language acquisition" (p. 225).

On the other hand, early SLA researchers might have been looking in the wrong places. Many personality traits studies in research during the 1970s and 1980s included empathy, anxiety, creativity, field dependence/independence, deliberateness, analytic orientation, and self-esteem. Of these so-called traits, only anxiety was ever accepted by mainstream psychology as an aspect of personality

(see the Origins of Social Anxiety section). In the numerous past investigations of the relationship of personality variables to second language acquisition (SLA), extraversion quickly attracted the most attention, so much so that the success of extraverted students at learning and communicating in a foreign language became a learning belief. In fact, the majority of SLA personality researchers investigating the extraversion-introversion aspect of personality paid particular interest to whether extraverts or introverts would be more successful at language achievement as measured either by course grades or by standardized test scores; the results were somewhat mixed.

One of the most widely cited SLA studies concerning personality was conducted by Ehrman, who searched for traits and habits of the good language learner in the 1990s. Ehrman (1990) used results from several Likert-scale questionnaires, such as the Myers-Briggs Type Indicator (M/BTI), to test the theory that extraverts were better learners and communicators of a foreign language than introverts. However, despite the large sample size of over 1,000 students, the researcher failed to find any statistically significant differences between extraverts and introverts. The results led her to speculate that “most language use is dyadic,” or occurring in pairs, “which is often as comfortable for introverts as for extraverts” (p. 418). The researcher further hypothesized that extraverts were more likely to have an advantage in large groups, such as full-classroom participation activities.

In a more recent study, Ehrman (2008) compiled a total N -size of 3,145 participants culled from U.S. Foreign Service Institute databases of native English speaker students in intensive foreign language programs across the United States. She reported that the largest number of students ($n = 272$) were from the introversion-sensing-thinking-judging (ISTJ) personality type. Using a small sub-sample of students of undetermined n -size who achieved Level Four of the Foreign Language Institute Oral Proficiency Interview, the researchers performed a series of correlations and chi-squared tests based on frequency statistics from the 16 possible M/BTI personality types. Her results found one significant group (introverted-intuitive-thinking-judging, or INTJ), which consisted of 16.5% of the sample.

The biggest drawback of the study was the small n -size reported, which, based on the frequency statistics and rounded-up participant numbers seemed between 60 and 70. However, the instrument used as well as the statistical methods may have been questionable. First, the M/BTI is generally perceived by mainstream psychologists as demonstrating poor psychometric qualities because its use of a dichotomous answering key system (*Yes*, *No*, and occasionally *Maybe*), rather than a Likert scale, results in ordinal data rather than interval data. Further, chi-squared tests based on nonparametric, categorical or rank-ordered data are sample-specific and generally lack statistical power. Finally, correlational analyses based on results

of chi-squared tests lack external validity and cannot be generalized across sample populations.

While researchers such as Ehrman focused on extraversion in the English as a Second Language (ESL) context of North America, several researchers in Asian English as a Foreign Language (EFL) contexts similarly conducted extraversion-introversion trait-based SLA studies. Influenced by the Big Three P-E-N model, Iwawaki, Eysenck, and Eysenck (1980) used the 86-item Eysenck Personality Questionnaire (EPQ) developed by Eysenck and Eysenck (1975) to determine correlations among extraversion-introversion and English achievement tests for a sample of 96 female students at a junior college in western Japan. The researchers failed to find significant differences on overall test scores between extraverts and introverts identified by the EPQ (Iwawaki, Eysenck, & Eysenck, 1980). The results were possibly due to a low *N*-size, a high number of questionnaire items, or the use of raw questionnaire item and test scores rather than standardized *z*-score variables.

Similarly, Busch also attempted to discern whether extraverts had an advantage over introverts in a study of Japanese university students using course grades as the outcome variable (Busch, 1982). To justify the reasons for choosing the extraversion-introversion variable as a possible predictor of English language learning, Busch (1982) cited results from the good learner study of Naiman,

Frohlich, and Stern (1975) that 31% of students felt that extraversion was helpful for communication skills. As Griffiths (1991) pointed out, however, “[Busch] might have observed...that 69% of the students did not” feel that extraversion was helpful (Griffiths, 1991, p. 106). Indeed, Busch’s results indicated no significant differences for course grades between extraverts and introverts. However, the largest drawback may have been the use of English course grades; grades given by a single teacher to a group of students might not provide a sufficiently objective measure of overall English proficiency. The definition of what constituted proficiency might also have affected the results. For example, if written quizzes or examination comprised a substantial percentage of course grades, introverted students seem as likely as, or more than, extraverted students to receive high grades. If personality influences nonlinguistic factors such as self-competence beliefs or speaking confidence, the completion of writing or speaking tasks that are judged by at least two or three raters is necessary for a reliable and valid assessment of English productive skills prior to conducting correlational analysis with personality traits.

In another EFL-context study, Carrell, Prince, and Astika (1996) investigated the relationship between extraversion-introversion and achievement tests among 75 Indonesian university students. Following the lead of Ehrman and Oxford (Ehrman & Oxford, 1990; Oxford & Ehrman, 1993), Carrell et al. used the Myers-Briggs

Type Indicator (MBTI) personality inventory rather than a psychometric personality questionnaire such as the EPQ, NEO-PI, or Factor Markers (Carrell, Prince, & Astika, 1996). As with the Iwawaki et al. study, Carrell et al. found only low correlations between personality types and language learning, leading them to state, “[W]e must be circumspect in drawing implications for EFL classrooms” (p. 96). Again, similar to the Iwawaki et al. study and the Busch study, Carrell et al. might have failed to find significant results due to an extremely small *N*-size and a proportionally high number of questionnaire items, as well as the use of the generally poor psychometric characteristics of the M/BTI instrument.

One of the first studies in SLA to use the Big Five personality traits, and also one of the first to use the concept of Willingness to Communicate (WTC) in SLA, was MacIntyre and Charos (1996). Based on a model of WTC as a personality trait among native language (L1) speaker adults (McCroskey & Richmond, 1987), MacIntyre and Charos (1996) tested their theory that each of the Big Five personality traits contributed to language learning motivation and/or WTC in one’s second language (L2) only in an indirect fashion. The researchers surveyed 92 participants learning French in a language immersion setting in Ontario, Canada. The participants took a series of measurements: (a) a 35-item version of Greenberg’s (1993) Factor Markers, (b) three items for Integrativeness, (c) three items for Motivation comprising Desire to Learn French, Motivational Intensity,

and Attitude toward Learning French, (d) two items for Attitudes toward the Learning, (e) two items for Language Anxiety, (f) twelve items for Perceived Competence in French, (g) twelve items for Frequency of Communication in French, (i) twenty items for Willingness to Communicate, based on the L1 model (McCroskey & Richmond, 1987), and (j) two items for Social Context, to measure how much French was used at work and at home. The questionnaire thus contained a total of 99 items.

Before conducting the analyses, the researchers created a path model of personality, attitudes, and *WTC*, based on prior research by Gardner (1985) and MacIntyre (1994) (Figure 1). They proposed a path from Extraversion to L2 Anxiety, supposing that socially oriented people would feel less anxiety speaking with members of the L2 group. However, the researchers hypothesized that Emotional Stability would not lead directly to L2 Anxiety, stating that prior research had not demonstrated a strong connection between foreign language anxiety and general trait anxiety. The research to which was being referred was not noted. Instead, Emotional Stability was connected to Integrativeness, based on the assumption that people who do not have positive attitudes toward the L2 language group feel uncomfortable talking with a member of that group. The researchers performed a path analysis (Figure 1) and concluded that personality exerted only an indirect influence on *WTC* after being filtered through the latent variables of

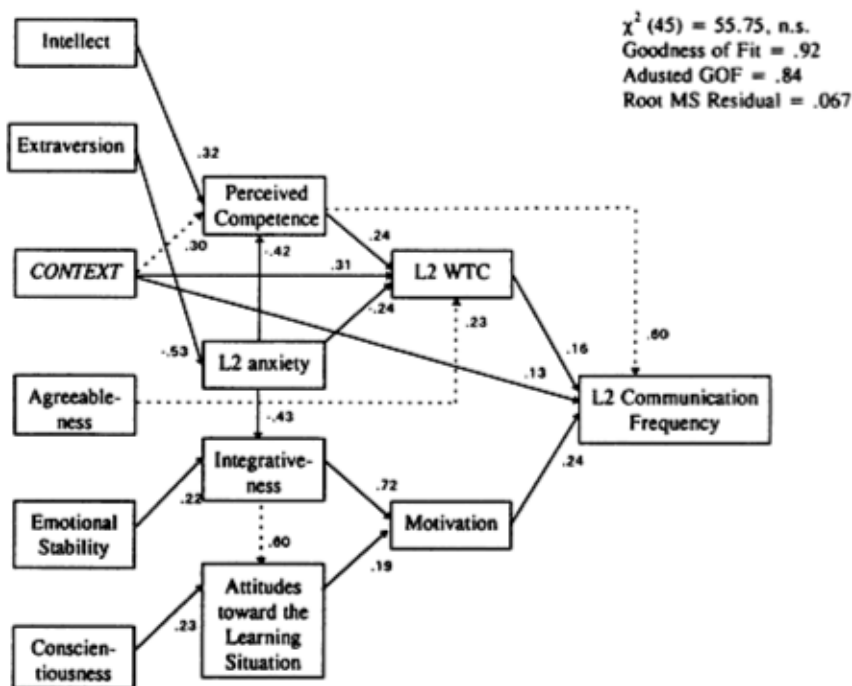


Figure 1. Path analysis of L2 Communication Frequency from MacIntyre and Charos (1996), including Perceived Competence, L2 Anxiety, Integrativeness, Attitudes toward the Learning Situation, L2 Willingness to Communicate, and Motivation.

Attitudes toward the Learning Situation, Integrativeness, L2 Anxiety, and Perceived Competence (MacIntyre & Charos, 1996).

However, similar to previous personality studies in SLA, the methodology was flawed by too few participants and too many variables. Although the researchers began the analysis with the original WTC model from McCroskey and Richmond (1987) as a template, the researchers were forced to add four new post-hoc paths for the data to fit the model because the number of participants ($N =$

96) was outnumbered by the number of variables (99). The model had adequate fit, with a non-significant *chi*-squared of 55.75, a root mean square residual (RMSR) of .067, and a goodness of fit (GFI) of .92; however, the adjusted GFI was .84 due to the small sample size and cross-sample confirmation with a new set of data was not conducted. Moreover, several of the factors, such as Desire to Learn French and L2 Anxiety, consisted of only two or three items and no tests of unidimensionality were conducted for any of the factors.

As far as anxiety is concerned, there was no mention of the “prior research” upon which the researchers’ assumptions concerning links between trait anxiety and L2 anxiety were based. The lack of a link between the personality trait of Emotional Stability and L2 Anxiety was unsupported by evidence; instead, the researchers theorized that Extraversion alone negatively predicted L2 Anxiety, and further posited that Emotional Stability influenced Integrativeness based on the assumption that people who do not have the desire to become like members of the L2 community feel uncomfortable talking with them. Logically, this seems to be setting the cart before the horse, as personality traits are expected to be inherited genetic factors rather than dependent upon external social situations; the researchers’ reasoning should have been that people with unstable emotional traits find it difficult to imagine themselves a member of another language community. MacIntyre and Charos’ claim that “language anxiety is not strongly related to

general trait anxiety” was unsupported in their literature review. Mainstream psychology theory and research indicates that social anxiety is a combination of genetic and environmental influences; thus, both introverted tendencies (i.e., the lack of tolerance of prolonged social exposure) and lack of emotional stability (i.e., the inability to control one’s emotions) should logically have lead the researchers to hypothesize strong negative relationships among Extraversion, Emotional Stability, and L2 Anxiety. Finally, because L2 Anxiety was measured by only two items, generalizing the results to different participant samples is difficult at best.

Supported partly by the results from the MacIntyre and Charos (1996) study, MacIntyre, Clément, Dörnyei, and Noels (1998) proposed a “heuristic pyramid model” of L2 WTC, in which personality was placed at the bottom level of “Layer VI: Social and Individual Context.” According to the pyramid model, instead of directly influencing language learning outcomes, “personality helps to set the context in which the language learning occurs” (MacIntyre, Clement, Dörnyei, & Noels, 1998b, p. 558). This social context is further influenced by other variables, such as L1-L2 intergroup climate, disposition towards foreigners, and linguistic self-confidence. The authors’ concept of social context in L2 WTC has been challenged in recent years; the heuristic model is discussed further in the Anxiety and Confidence section.

A more recent Japanese study carried out at Temple University Japan in Tokyo surveyed 320 university students using the Yawate-Guilford Personality Inventory (Y/GPI), an early pre-PEN and pre-Big Five personality trait instrument with 12 subscales consisting of ten items per scale (Brown, Robson, & Rosenkjar, 2001). Six instruments were used in the study to measure personality, motivation, anxiety, strategies usage, English grammar knowledge, and overall English proficiency. By counting the number of sub-scales of the instruments, the authors claimed that a “total of 34 variables” were analyzed (p. 361); however, the authors used the term “variable” in the sense of “factor.” Even considering only the Likert-scale items used in the factor analyses, the number of combined measurement instrument variables (408) was far larger than the participant size ($N = 320$). Thus, the statistics from the study lacked predictive power and the results of the correlational analyses between the factor analysis results of the Likert-scale data and the proficiency data might have been of limited value.

In one of the most recently published Big Five and SLA studies to date, Verhoeven and Vermeer (2002) examined the relationship between the Big Five personality traits and communicative competence among 144 native Dutch-speaking children and 69 non-native Dutch-speaking children. The results indicated significant correlations between individual Big Five traits such as Conscientiousness and Openness and various aspects of communicative

competence; however, the study contains several flaws. First, the personality trait measurement instrument was “30 pairs of judgments,” the nature of which was not noted, but which the researchers indicated were based on the concept of the Big Five rather than any specific existing instrument (Verhoeven & Vermeer, 2002, p. 366). Second, the teachers of the children rather than the children themselves evaluated the personality items; given that there were 213 children in the study, it is doubtful that the teachers could accurately judge each student’s individual personality. Finally, the small and unevenly distributed participant size between the two groups of L1 and L2 Dutch speakers might have led to nonparametric data distributions, and thus reduced the generalizability of the results to other samples.

Some of the problems encountered by SLA researchers might be a lack of understanding of the nature of personality trait and foreign language confidence constructs. Early sections discussed the various personality trait models available, and the Big Five Factor model in particular, knowledge of which might assist SLA researchers in future studies. The track record of confidence studies in SLA has less depth and is shorter than in other fields; therefore, I start the next section with a brief exploration of the origins of the concept of anxiety in the fields of social and cognitive psychology and its complicated relationship to the constructs of foreign language anxiety and foreign language confidence.

Anxiety and Confidence

Origins of Social Anxiety

Although general anxiety was originally proposed as a personality trait (Eysenck, 1970), anxiety seems to be a complicated combination of both genetic, or hereditary (i.e., trait-like anxiety), and social-oriented (i.e., situation-specific) origins. On the hereditary side, Cattell (1973) proposed that people who exhibit overall anxiety tendencies tend to have sympathetic nervous systems that are easily “activated” through external stimuli (Cattell, 1973). There is also some evidence indicating that infants and children who are temperamental, fearful of strangers, and inhibited in general tend to become anxiety-prone in later life (Berberian & Snyder, 1982; Kagan, Reznick, & Snidman, 1988; Scarr & Salapatek, 1970). It has also been reported that people labeled as “shy,” “anxious,” or “introverted” have a stronger tendency than non-anxious, extraverted people to develop health problems, such as eczema, asthma, hay fever, and stomach cramps (Bell, Jasnoski, Kagan, & King, 1990). Psychiatrists have theorized that neurotransmitters and neurochemicals in the brain related to emotional control might be responsible for the connection between anxiety and physical ailments, leading to the conclusion that anxiety can be at least partially attributable to genetic causes (Kagan, Snidman, Julia-Sellers, & Johnson, 1991).

Certain personality types are also theorized to have strong connections to general social anxiety. Leary and Kowalski (1995) listed five traits that predispose people to experience a sense of anxiety in social situations: general trait anxiety, self-consciousness, approval motivation, self-esteem, and interpersonal confidence (Leary & Kowalski, 1995). In this view, general trait anxiety is perceived as a personality trait as measured by the Trait Anxiety Scale (Taylor, 1953), a 50-item, true-false questionnaire based on the Minnesota Multiphasic Personality Inventory and frequently used in clinical psychology even today. However, the construct validity of the Trait Anxiety Scale (or, the Biographical Inventory, as Taylor called it) was challenged almost as soon as it was created (Jessor & Hammond, 1957), and although the scale became popular as a short version questionnaire for trait anxiety, the items themselves were never statistically validated nor the unidimensionality of the construct confirmed.

In addition, the remaining four “traits” of social anxiety cannot clearly be said to be part of one’s personality. Instead, over-sensitivity to social evaluation, the need for approval from others, the desire to avoid social rejection, and the inability to respond appropriately in social situations are all conditioned from a young age by parental figures and peers. Researchers have pointed out that self-consciousness and reticence in children is often the result of parental figures who are too demanding or critical of their children due to concerns over the perceptions by

people outside the children's family or group. An overemphasis on achievement either academically or socially can also lead to an acute awareness of one's shortcomings and others' expectations of personal behavior (Klonsky, Dutton, & Libel, 1990).

On the other hand, parental figures who exhibit a certain amount of disinterest, disapproval, or even outright rejection of children can create the need for approval and the need to avoid disapproval, thus leading to an overawareness of the evaluative perceptions by others of the individual (Allaman, Joyce, & Crandall, 1972). Similarly, excessively harsh or forbidding parental figures tend to cause a sense of low self-worth and the need for social approval. This need leads to the existence of anxiety and a low level of or lack of confidence as a response to any perceived threat to a sense of inclusion and membership in a social group (Baumeister & Tice, 1990; Leary, 1990; Leary, Tambor, Terdal, & Downs, 1995). This sense of needing group membership is one of the defining characteristics of Japanese culture; the following sections will examine the L1 and educational influences in Japan on foreign language anxiety and confidence in the foreign language classroom.

The Japanese Context

Hofstede (1988) famously found that Japan as a “culture” scored lower than all but two countries on his Risk Avoidance scale. His finding was not surprising; one of the most-well known sayings in Japan is, “The fence post that stands up will be hammered down” (in Japanese, *Deru kui ha utareru*), an admonition to avoid standing out and to conform to social norms for the good of the group. The avoidance of risk, and fear of mistakes, can be attributed at least partly to a sense of interdependence that makes Japanese extremely reluctant to openly voice opinions contrary to perceived opinions of those around them. Indeed, the “interdependent self” of Japanese society is well attested in personality psychology literature (e.g., Heine, Kitayama, & Lehman, 2001; Heine, Kitayama, Lehman, Takata, Ide, Leung, et al., 2001; Heine, Lehman, Peng, & Greenholtz, 2002; Kitayama, 2000; Kitayama & Markus, 1999; Kitayama, Markus, & Lieberman, 1995).

L1 Influences. Part of Japanese tendencies to avoid conflict in favor of smooth personal relations might be embedded in the Japanese language itself: before a Japanese person says the words *I* or *you* in conversation, he or she is expected to consider several social factors, including gender, age, social status, personal familiarity, and the present disposition of the interlocutor. In fact, there have historically been over 50 different words meaning *I* and over 80 words

meaning *you* (Sugawara, 1989), each of which at least half a dozen are used in modern Japanese. It is no wonder that a polite speaker of Japanese generally avoids the usage of *I* and *you* whenever possible, so as to avoid potential conflict and loss of face from making a social faux pas.

The general reluctance even to express words in one's own native language can easily be extended to low second or foreign language speaking confidence. Japanese students can find it particularly anxiety-provoking to speak English to people from different age groups and social rank groups. For example, when talking with an older student or a teacher, a Japanese student might feel reluctant to start a conversation, to ask questions, or to volunteer information without first being asked directly to do so by the older person in the conversation. English does not have the same sense of humble or honorific expressions that Japanese language and society would expect students to use when addressing social "superiors" or "inferiors"—this lack of honorific expressions can leave Japanese students literally at a loss for words.

Educational Influences. Another factor to consider is the effect of schooling on social anxiety and thus a corresponding lack of confidence among Japanese students of English. With a traditional focus on university entrance examinations, high school education in particular in Japan has a strong tendency toward

anxiety-creating contexts. Mainstream secondary schools traditionally have favored a teacher-centered classroom; in fact, the classroom teacher is not only seen as the authority figure from whom all knowledge flows, but is even physically raised above the level of the students on a platform in front of the classroom. In a typical Japanese high school English classroom, only the teacher speaks, usually while diagramming complicated grammar patterns on the blackboard, while the students silently copy what the teacher writes (or fall asleep). This style of teaching, known as *yakudoku*, or grammar-translation, has remained the dominant form of English language education in Japan for over a century (Hino, 1988; Watanabe, 1996). Despite the best attempts of the Japanese Ministry of Education (MEXT) to introduce other methods of language learning, *yakudoku* still thrives (Nishino, 2009).

Recent changes in educational policy by MEXT were exemplified by the “Action Plan to Cultivate ‘Japanese with English Abilities,’” whose aim was to encourage the majority of English classes in junior and senior high school be taught through the medium of the English language (MEXT, 2003). However, the so-called Action Plan created only a limited number of Super English Language High Schools (SELHis), in which regular content classes are taught through the medium of English language by both Japanese and foreign national instructors, in a limited time period from 2002 to 2009. In public junior and senior high schools

throughout Japan, students often only experience communicative language activities during classes with native or near-native English language speakers called Assistant Language Teachers. Public senior high schools in particular were encouraged by MEXT, starting in 1994, to offer a course called Oral Communication for first year students; however, this course only meets once or twice per week for 45 minutes, and other English courses typically are taught by a Japanese instructor who feels obligated to prepare students for the perceived discrete grammar points test on university entrance examinations.

Students themselves generally perceive the native English speaker-taught course as a student-friendly, communicative, environment that uses pair and group work. On the other hand, students generally perceive the Japanese speaker-taught course as a traditional teacher-fronted, studious environment in which individual student, instructor-led grammar-translation activities dominate (Shimizu, 1999). Due to the fact that the overwhelming majority of high school English teachers in Japan are Japanese speakers, it is fair to assume that English education at the vast majority of secondary schools in Japan thus affords students few opportunities in the English class to communicate openly either in Japanese or in English prior to entering university.

Indeed, this assumption has recently been supported in two wide-ranging surveys of public high school English oral communication teaching practices,

conducted by MEXT in 2007 and 2010. Both surveys included over 3,600 public high schools throughout Japan, including over 25,000 Japanese teachers of English. Approximately 70% of the schools in both surveys offered the English Oral Communication course; teacher self-reported results indicated that, in 2007, 20.7% of Oral Communication courses were conducted “mostly” in English and 33.9% were conducted “more than half” in English (MEXT, 2007). These numbers fell to 19.6% for “mostly” and 32.8% for “more than half” in the 2010 survey (“Concern over use of Japanese language,” 2010). These figures were widely reported in English-language newspapers in Japan in early 2011 as “cause for concern” for MEXT, which has stated that all English Oral Communication classes will be mandatory starting in 2013 (“English classes mostly in Japanese,” 2011; “Only 20% of English conversation teachers,” 2011). Given that the surveys reported percentages of English language instruction by teachers for courses intended to promote English conversation, it seems reasonable to expect that the percentage of English language use by teachers is much lower in the English reading and writing courses that dominate the Japanese high school English curriculum; if teachers are using Japanese most of the time in English classes, it is also reasonable to assume that the students likewise mostly use Japanese, if they speak at all.

In addition to a lack of speaking opportunities, the number of students per class in a typical Japanese high school is 40; in some private high schools and

universities this number can reach well over 100 or even 150 depending on the type of English class (e.g., reading, listening, and “repeaters only”). While MEXT might be sincere in its desire to increase the number of Japanese students who can actively communicate their ideas in written or spoken English, individual institutions are allowed complete freedom to ignore the “recommendations” for using communicative teaching methods in the English classroom. Although the Ministry of Education’s goal that, “The majority of an English class will be conducted in English and many activities where students can communicate in English will be introduced” (MEXT, 2003), was elucidated as early as the 1992 Course of Study (Gorsuch, 2000), Japanese high school teachers resist the communicate teaching of English out of the belief that *yakudoku* and the teaching of discrete grammar points for university entrance examinations is more important (Gorsuch, 1998; 1999; Nishino, 2009).

It has been known for almost two decades that the number of college-bound students has been decreasing and the number of students that enter university through the recommendation system, thus bypassing regular university entrance examinations, has been increasing (Mori, 2002). Despite these facts, English high school classes continue to rely heavily on grammar-translation methodology through the medium of Japanese with virtually no interaction between the students. In such a classroom context, the possibility for social anxiety increases and for

foreign language speaking confidence decreases, leading to greater fear of speaking and a decrease in communication. In short, the educational context of the average Japanese high school or university English classroom might be at least partly responsible for the lack of foreign language speaking confidence among Japanese students of English.

In order to further understand the nature of speaking confidence, in the next section I briefly examine the concept of communication apprehension from the field of communications in a North American L1 context. Following that, I summarize early SLA research into foreign language confidence and continue with a description of the Foreign Language Classroom Anxiety Scale. I also examine the relationship between foreign language anxiety and the concept of linguistic self-confidence, as well as examine flaws in the research in which the foreign language confidence “construct” has been used. Finally, I offer suggestions on improving the concept of foreign language speaking confidence through the inclusion of both desire to speak English and measurements of the social situation in which communication occurs.

The L1 Context: Communication Apprehension

Known best for pioneering the concept of Willingness to Communicate, McCroskey (1970, 1977) also pioneered the concept of Communication

Apprehension (CA), which he defined as “an individual’s level of fear or anxiety associated with either real or anticipated communication” (McCroskey, 1977, p. 78). McCroskey developed several Personal Report Communication Apprehension scales for measuring self-reported communication apprehension across age groups: the PRCA-College, for U.S. university students; the PRCA-Ten, for U.S. senior high school 10th graders; the PRCA-Seven, for American junior high school 7th graders; and the PRPSA, which was designed specifically to measure public speaking apprehension (McCroskey, 1970). Other PRCA versions followed: the PRCA-24, consisting of 24 items; the PRCA-Long Form consisting of 25 items; and the PRCA-Short Form, consisting of 10 items (McCroskey, 1978).

In contrast to general anxiety scales of the time, McCroskey’s PRCA-24 was focused specifically on oral communication in four social contexts and with three types of interlocutors. The social contexts involved speaking with only one interlocutor, speaking in a small group, speaking at a large meeting, and speaking in public. Interlocutors included friends, passing acquaintances, and strangers. Although McCroskey’s work concerned L1 speakers in an L1 context (e.g., native speakers of English in the United States), his work strongly influenced speaking anxiety and confidence research in the field of SLA, particularly that of Gardner, Clément, and MacIntyre (see the Testing Foreign Language Anxiety and Constructing Self-Confidence sections).

The L2 Context: The Foreign Language Class Anxiety Scale

Although early SLA researchers and teachers instinctively felt that anxiety had some part to play in the second language learning process, some studies relating anxiety to language achievement or performance reported inconsistent results (Chastain, 1975; Scovel, 1978) while other studies conducted primarily for the purposes of motivation research claimed to find consistent negative correlations between anxiety and language performance (Clément, Gardner, & Smythe, 1977; Gardner, Smythe, Clément, & Glikzman, 1976). In a review of anxiety research at the end of the 20th century, MacIntyre (1999) summarized difficulties in early anxiety research as primarily stemming from the use of test and trait anxiety measurement instruments, rather than the use of an instrument specifically designed to measure foreign language anxiety.

Although Gardner (1985) was one of the first SLA researchers to develop an instrument designed to measure foreign language anxiety—a short five-item scale called “French classroom anxiety”—it was the study by Horwitz, Horwitz, and Cope (1986) that popularized the term *foreign language anxiety* through the creation of the Foreign Language Classroom Anxiety Scale (FLCAS). The FLCAS consisted of 33 positively and negatively worded items, which Horwitz (1986) claimed to validate through correlation with existing anxiety scales. Horwitz et al., argued that foreign language anxiety was composed of three primary components:

communication apprehension (CA), fear of negative evaluation, and test anxiety. Foreign language anxiety, however, differed from L1 CA in that L2 learners had already acquired their L1, and the gap between linguistic knowledge in their L1 and L2 would lead L2 learners to experience frustration and apprehension at not being able to express themselves or to comprehend the foreign language. Negative evaluation was similarly enhanced by the students' awareness that they might not be understood by classmates due to lack of L2 knowledge, thus leading to poor social impressions. The FLCAS subsequently subsumed Gardner's five items as well as items from scales measuring test anxiety, speech anxiety, and communication apprehension.

Unfortunately, the items of the FLCAS were never validated and the unidimensionality of the FLCAS was never examined, even though the originators admitted to deliberately including items from what they believed were three separate constructs. Additionally, the original study in which the FLCAS was tested had an *N*-size of only 78; although Horwitz (1986) reported that over 300 undergraduate students at the University of Texas had taken the FLCAS at the time of her validation, the largest single sample size was $N = 108$ and the smallest was $N = 32$. The FLCAS was correlated with various anxiety-related scales (State-Trait Anxiety Inventory, PRCA, Fear of Negative Evaluation, Test Anxiety Scale) to varying degrees ranging from $r = .29$ to $r = .53$, and was found statistically

significant or close to significant in all comparisons. Despite the reliance entirely on correlational analyses, Horwitz concluded, “These results suggest that foreign language anxiety can be discriminated from these related constructs” (p. 561).

Foreign Language Anxiety Studies in Japan

Overall, surprisingly few foreign language anxiety studies of a statistical nature have been conducted in Japan. One study of note was conducted by Matsuda and Gobel (2004), who examined anxiety among 252 students at a large Japanese university in several first year four-skills, second year reading and listening, and third year reading courses. The researchers used two instruments to measure anxiety: the FLCAS and the Foreign Language Reading Anxiety Scale (FLRAS) (Saito, Garza, & Horwitz, 1999). The Test of English for International Communication (TOEIC) was used as a measurement of L2 proficiency. The FLCAS and the FLRAS were first factor analyzed into two and three factors, respectively, and factor scores were obtained to transform the Likert-scale data into standardized scores. A MANOVA was then conducted using gender and overseas experience as independent variables and the two FLCAS factors, the three FLRAS factors, and total TOEIC scores as the dependent variables. No statistically significant differences were found for gender on either anxiety questionnaire,

although a follow-up MANOVA using a smaller sample of only first-year students ($n = 89$) revealed significant differences between gender and TOEIC scores.

As with other anxiety studies, the construct validity of the FLCAS and the FLRAS was determined through the use of Cronbach's alpha reliability estimates and principal components factor analysis (Matsuda & Gobel, 2004, p. 26). The researchers suggested that relatively low alpha estimates of the FLCAS ($\alpha = .78$) and FLRAS ($\alpha = .71$) might have affected the factor analysis results; however, it is more likely that the multidimensional nature of both anxiety questionnaires was at least partly responsible for the failure of several items to load onto the hypothesized factors. Finally, the second factor of the FLCAS in this study—called Low Self-confidence in Speaking English by the researchers—was the only factor of either questionnaire that significantly predicted language proficiency as measured by the TOEIC. It thus stands to reason that using a speaking anxiety measurement instrument rather than a multidimensional general anxiety instrument would logically lead to more trustworthy results.

One of the most recent studies of anxiety in Japan was conducted by Kimura (2008), who examined 452 students at a Japanese university for L2 listening anxiety. The study used a Japanese translation of a 33-item questionnaire called the Foreign Language Listening Anxiety Scale (FLLAS), originally designed for Korean university students (Kim, 2000). A principal components analysis was

conducted, and three factors of listening anxiety were found: Emotionality ($\alpha = .85$), Worry ($\alpha = .85$), and Anticipatory Fear ($\alpha = .80$). Total variance accounted for was 38.55%. Three follow-up two-way ANOVAs were also conducted to examine differences according to gender and academic major, and significant differences for academic major were discovered on the Emotionality factor, with an extremely small effect size ($\eta^2 = .03$).

Similar to other anxiety studies, a major flaw of the study was the reliance on Cronbach's alpha estimates to demonstrate construct unidimensionality. This flaw was evident from a close examination of the wording of the items in each factor. The Emotionality factor "reflected not just lack of confidence but other emotional reactions as well: discomfort, dislike, annoyance, alienation, and intimidation" (Kimura, 2008, p. 185); however, the actual wording of the items on the factor included phrases such as "It is difficult to...", "It is easy to...", and "It often happens that..." Further, one item asked participants to identify their familiarity with English intonational and word stress patterns, while another asked whether participants translated word by word. Thus, the factor seemed more about participant beliefs or behaviors rather than emotional states. The Anticipatory Fear factor additionally was measured by only two items (factor loadings of .77 and .67) that began with the phrases "I am nervous when listening to..." and "I feel tense when listening to..." Words such as "nervous" and "tense" seem to indicate worry,

rather than fear, and highlight the danger of relying on Cronbach's alpha estimates and factor analysis to determine construct unidimensionality. A closer examination of the specific wording of the questionnaire items using Rasch analysis to aid in item revision in a smaller preliminary pilot study might have led to clearer factor analysis solution and better ANOVA results in the main study.

Testing Foreign Language Anxiety

Gardner, Day, and MacIntyre (1992) conducted a study to evaluate the theoretical claims concerning foreign language anxiety by Horwitz et al., using a combination of questionnaires and experimental design methods. The researchers tested 52 men and 52 women individually for one hour each. At the start of the session, the participants completed a questionnaire consisting of 70 items from six separate anxiety-related scales. The participants first completed a "test phase" consisting of a computerized multiple-choice vocabulary test of 38 French words. The participants were then shown a series of 38 paired associates of English and French words on a computer screen for 2.5 seconds as a "learning phase." The test phase and learning phase cycle recurred five times, and a State Anxiety measurement of 20 items was given after the first, third, and fifth cycle. At the conclusion of the fifth test, the participants were given six oral prompts by one of the researchers. Three of the prompts had to be answered orally and three had to be

answered in writing with only two minutes allowed for each prompt. At this point, the participants were randomly assigned to High Pressure or Low Pressure conditions; the participants in the High Pressure conditions were given a new item if they failed to answer any item within fifteen seconds. Following the series of prompts, the participants were then given a Free Recall test, in which they were asked to remember orally as many pairs of words as possible within a span of four minutes.

Analysis consisted of first factor analyzing the questionnaire data into two factors, which the researchers named General Anxiety and Communicative Anxiety, then performing a series of analyses using analysis of variance (ANOVA), multiple analysis of variance (MANOVA), and analysis of covariance (ANCOVA). The results indicated that, although Horwitz et al. (1986) included test anxiety as an aspect of foreign language anxiety, test anxiety was also related to general anxiety and was not specific to foreign language learning. On other hand, items concerning French class anxiety, French use anxiety, English class anxiety, and Audience sensitivity formed a distinct factor of Communicative Anxiety, which in turn predicted poorer language recall performances. French class anxiety in particular was the only factor significantly correlated to oral performance. Although state anxiety was somewhat inconsistent, the researchers interpreted the results as indicating that state anxiety was a product of rather than a predictor of language

performance. The overall results led the researchers to postulate that learners who are poor at language performance might experience repeated episodes of state anxiety, which would in turn lead into a more situation-specific anxiety such as classroom anxiety.

Unfortunately, as with many other SLA anxiety studies, the Gardner et al, 1992, study suffered from a dearth of participants compared to a plethora of variables. The *N*-size was 104, while the total number of questionnaire variables, including multiple implementations of the State anxiety scale, was 130; Tabachnick and Fidell (2007) recommended at least 150 participants and preferably 300 for principal components analysis, and Stevens (2009) recommended a case to variable ratio of at least 5:1, preferably 10:1. In the Gardner et al. study, this ratio is less than 1:1, leading to possibly unreliable statistical results. The use of vocabulary item pairs also is debatable as a means of evaluating “learning” and “producing” a foreign language. Given the time constraints and manner of learning, the tests in the study seemed to favor participants with better working memory (Atkins & Baddeley, 1998) rather than those who were more proficient in either speaking or writing French.

MacIntyre, one of the researchers in the Gardner, et al., 1992 study, continued to research foreign language communication, delving deeply into how anxiety and perceived competence influence WTC (MacIntyre & Charos, 1996; see the

Personality and SLA section). In later studies, anxiety as a separate construct became incorporated into the concept of L2 Self-Confidence. To further examine the relationship of foreign language anxiety and the construct of foreign language speaking confidence in the following section I briefly examine how Clément, his associates, and his students originally conceived the construct and how the construct has evolved, or not, during the past three decades of SLA research.

Constructing Confidence

Researchers studying Willingness to Communicate (WTC) in the socio-educational tradition have focused on a construct variously termed L2 self-confidence or L2 communication confidence. In fact, many terms, such as apprehension, perceived self-confidence, and perceived competence, are used in the literature either as interchangeable terms or as parts of each other.

The term *linguistic self-confidence* was first defined as self-rated language proficiency and the lack of anxiety (Clément, 1980). Since that time, the construct has been renamed and redefined so many times that one can hardly keep track of it. In 1985, it was renamed L2 confidence (Clément & Kruidenier, 1985), redefined as low L2 anxiety and high perceived self-competence. In 1994, Clément et al. renamed the construct again as L2 self-confidence and defined it as a combination of two factors: an affective factor, namely lack of communication apprehension,

and a cognitive factor, namely high self-evaluation or perceived competence (Clément, Dörnyei, & Noels, 1994). The name once again changed in 1996 to linguistic self-confidence (Noels, Pon, & Clément, 1996).

The research tradition of L2 confidence was eventually combined with research theories about WTC and motivation in the now well-known WTC heuristic pyramid (Figure 2) of MacIntyre et al. (1998). The pyramid model consists of six layers, from Layer VI at the bottom to Layer I at the top. The bottom three layers (Layer VI, V, and IV) are considered “enduring influences” that remain relatively unchanged over time, while the top three layers (Layer III, II, and I) are perceived as “situational influences” that can change depending on the circumstances of L2 interaction (MacIntyre, Clément, Dörnyei, & Noels, 1998a, p. 546).

Theorized enduring influences of the heuristic pyramid. The bottom three layers of the pyramid comprise the enduring, or more or less permanent, influences on L2 communication. Layer VI, is termed the Social and Individual Context. This layer consists of Intergroup Climate, which is the social power relationship between the L1 and L2 communities, and Personality, which is perceived as not only as individual personality types and traits, but also as a predisposition to act

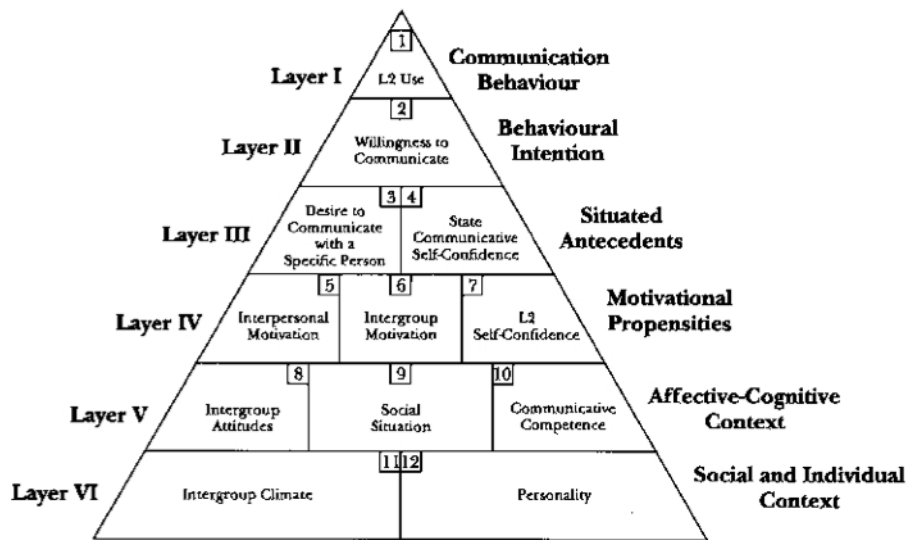


Figure 2. The Heuristic Model of Variables Influencing WTC from MacIntyre et al. L2 = second language; WTC = Willingness to Communicate (1998).

positively or negatively to members of a social group (p. 558). Just above the bottom layer is Layer V, which is termed the Affective-Cognitive Context. This layer consists of three components: Intergroup Attitudes, Social Situation, and Communicative Competence. Intergroup Attitudes includes the desire or the fear of integrating or assimilating with the L2 community, as well as positive or negative experiences in the language classroom (p. 552). Social Situation refers to the participants, setting, purpose, topic, and means of communication (p. 553). Communicative Competence comprises five communicative language abilities: linguistic competence, discourse competence, “actional” or pragmatic competence, sociocultural competence, and strategic competence (p. 554).

Layer IV, the final layer of the enduring influences, is termed Motivational Propensities and comprises Interpersonal Motivation, Intergroup Motivation, and L2 Self-Confidence. Interpersonal Motivation is described as primarily one of two orientations between individual interlocutors: “control” (i.e., hierarchical, task-based relations involving interaction between specific social roles such as doctor-patient) or “affiliation” (i.e., friendship relationships formed on the basis of physical characteristics such as beauty, similarity, closeness, and repeated contact) (p. 550). Intergroup Motivation similarly consists of control and affiliation orientations, but on the social level of groups rather than individuals.

The final component of this layer, L2 Self-Confidence, is defined first as “a lack of anxiety combined with a sufficient level of communicative competence, arising from a series of reasonably pleasant L2 experiences” (p. 548), then as corresponding to “the overall belief in being able to communicate in the L2,” and finally as “judgments of proficiency and feelings of apprehension” (p. 551). The two types of motivation are described as “affective and social aspects of the motivation to communicate,” while L2 Self-Confidence somehow combines the affective and social in a “relationship between the individual and the L2” (p. 551).

Theorized situational influences of the heuristic pyramid. Situational influences consist of the top three layers, including the top layer of L2

communication itself. Layer III is termed Situated Antecedents and consists of Desire to Communicate with a Specific Person (whose name speaks for itself) and State Communicative Self-Confidence, which is defined as separate from L2 Self-Confidence because it can be affected not only by current situations and persons, but also by both positive and negative previous experiences using the L2. Thus, although comprised of the same two components of anxiety and perceived competence, “trait-like self-confidence” is not the same as a “momentary feeling of confidence” (p. 549).

Layer II, termed Behavioral Intention, consists of only one component, the Willingness to Communicate. WTC is defined as a “readiness to enter into discourse at a particular time with a specific person or persons, using a L2” (p. 547). Unlike the McCroskey and Baer (1985) definition of L1 WTC as a trait-like variable, L2 WTC is a “situation-based” variable that, according to the authors, is primarily influenced by self-confidence and social context (p. 558), but also is a type of planned behavior (p. 548) with desire to speak and state communicative confidence as the most immediate causes (p. 549). The final layer, Layer I, is L2 Use, or the act of L2 communication itself. L2 communication is broadly defined as including various activities such as speaking L2 in class, reading L2 newspapers, watching TV in L2, or using L2 at work (p. 547).

L2 confidence in the heuristic pyramid: Some questions. In the heuristic pyramid, L2 Self-Confidence is created through the interaction of communicative competence and social situation, influenced by personality as expressed through attitudes toward the L2 that are formed through contact and exposure both to the L2 and to users of the L2. The text description of L2 Self-Confidence follows the Clément et al., 1980, analysis of the construct as comprised a lack of anxiety and the existence of perceived self-competence in using the L2. However, there are several problems with the construct of L2 Self-Confidence in the heuristic pyramid model.

First, considering L2 Self-Confidence as a “motivational propensity” on Layer IV seems somewhat misplaced. While the other two components of Layer IV were described as relationships among social roles within a group or among individuals in different groups, L2 Self-Confidence was described as a combination of lack of foreign language anxiety and the presence of perceived self-competence in the foreign language, both of which are said to arise from previous positive foreign language experiences. This description of L2 confidence as an individual foreign language ability belief seems at odds with motivations related to interactions with other people.

Second, the division of foreign language confidence into two constructs of L2 Self-Confidence and State Communicative Self-Confidence seems unwarranted,

given the explanations of the constructs themselves. If foreign language confidence arises from one's previous language experiences, then it can be changed in the future. Thus, all L2 confidence is a kind of state variable by nature. Furthermore, the argument that L2 confidence is a cause of state communicative confidence is evidently based on the research in Gardner et al. (1992), which speculated that state anxiety was a result of rather than a cause of poor language performance (see the Testing Foreign Language Anxiety section). If L2 confidence is both a cause and a result of foreign language use, it stands to reason that a more complicated model than a unidirectional pyramid model is necessary.

Finally, although L2 use was broadly defined as practically any language activity using the L2, it seems evident that the confidence required to speak up in front of one's peers in a foreign language classroom is much different than the level of confidence necessary to watch a foreign language TV program in the privacy of one's room. The use of L2 Use in Layer I of the model seems rather odd, given the use of Desire to Speak with a Specific Person in Layer III. This "desire" implies speaking, as opposed to reading or listening, as the communicative method in question. Thus, the social situations, roles, group memberships, and personal relationships concerning the L2 interlocutor become crucial, as the model suggests; likewise, foreign language confidence also must be primarily concerned with speaking in the heuristic model. Unfortunately, this is not made clear. Particularly

considering the importance of social situations and perceived communicative competence necessary to develop L2 confidence in the first place, I believe a new definition or a new model of L2 confidence is needed both from a theoretical as well as a statistical perspective.

Redefining the Construct of Foreign Language Confidence

The construct of foreign language confidence has seen a dizzying array of permutations, including self-confidence, L2 confidence, L2 self-confidence, state self-confidence, state communicative self-confidence, communication confidence in a L2, confidence in L2 communication, self-confidence in L2 communication, and L2 communication confidence (MacIntyre et al., 1998; Yashima, 2002; Yashima, Zenuk-Nishide, & Shimizu, 2004). Regardless of the construct name, however, there seems to be broad consensus among SLA researchers that foreign language anxiety and perceived self-competence in the target language are related concepts.

However, there are two problems with considering the concepts of foreign language anxiety and perceived foreign language self-competence as comprising a “high order construct” called foreign language confidence. The first problem is theoretical, while the second problem is statistical.

First, both low foreign language anxiety and high perceived foreign language self-competence could be exhibited by those who do not care about communicating effectively in their L2. Many students are required to learn a foreign language—for example, students in Japan are now required to learn English for at least six years throughout junior and senior high school, and most university students are additionally required to take between two and four English courses during their first and second years—with little chance of daily English communicative need. It stands to reason that there may be many students in Japan who experience little anxiety speaking English, or who may have high self-competence, because of a lack of experience or a lack of desire to speak. If these students were measured using existing foreign language anxiety and perceived foreign language self-competence instruments with no additional variables, according to the existing “higher order construct” used by Clément, MacIntyre, and others, such students would be labeled as having high foreign language confidence even though they may have no intentions whatsoever of initiating L2 communication.

The second problem is the designation of the concept of L2 self-confidence as a higher order construct. Originally used by Clément (1980), the term *higher order construct* was also used in MacIntyre and Charos (1996), and has continued to be used in recent WTC studies by Yashima (2002) and Yashima et al. (2004). However, the correct label in statistically terms ought to be *latent variable*, because

a construct only consists of one concept, not two, that has demonstrable unidimensional properties (Thurstone, 1931; Wright, 1999). A latent variable such as foreign language confidence ought to consist of at least three related constructs, or observed variables. If only two observed variables are used to indicate the latent variable, and both of the observed variables are highly correlated (i.e., $r > .90$), then the observed variables are essentially the same variable. On the other hand, if the two observed variables are not correlated highly enough, then the latent variable is on shaky statistical grounds. Either case leads to problems of multicollinearity and empirical underidentification of the model in question (Chen, Bollen, Paxton, Curran, & Kirby, 2001; Kenny, 1979; Kenny, Kashy, & Bolger, 1998).

With these facts in mind, many of the models used in previous SLA studies concerning WTC, foreign language confidence, and foreign language anxiety have contained a crucial flaw, namely, the division of items designed to measure a single construct, such as WTC or Anxiety, into two or three separate components in order to measure what has been incorrectly termed a latent construct (e.g., MacIntyre, Babin, & Clément, 1999; MacIntyre, MacMaster, & Baker, 2001; Yashima, Zenk-Nishide, & Shimizu, 2004). In essence, this approach results in a single measurement scale defining itself rather than a latent variable. Moreover, the use of a single measurement instrument divided into three observed variables leads to strongly correlated (e.g., $r > .90$) variables that violate the assumption of lack of

multicollinearity and lead to empirical underidentification of the model (Kenny, 1978).

The solution to both of these problems is to use instruments that measure at least three separate, related constructs to identify the latent variables. In order for foreign language confidence to be considered a latent variable for SLA modeling research, it would be prudent to identify at least one more factor in addition to foreign language anxiety and perceived foreign language self-competence.

Confidence and Social Situations

The heuristic model by MacIntyre et al. (1998) proposed that the variables Intergroup Climate, Intergroup Attitudes, and Social Situation might play crucial roles in the amount and type of contact with foreign language users. Exposure to and attitudes towards users of the foreign language would in turn influence perceptions of one's foreign language competence and thus increase or decrease foreign language confidence. The question remained, To what degree do social situations influence foreign language confidence? In one study designed partly to examine this question, Clément, Baker, and MacIntyre (2003) attempted to test two social models of second language that encompassed the concepts of subjective L2 norms as well as foreign language confidence to use or to avoid the target language.

The researchers administered a questionnaire consisting of 70 items to two separate groups of self-reported Anglophones ($n = 248$) and Francophones ($n = 130$) studying at the University of Ottawa, Ontario, Canada. More than two-thirds of the participants were female, and 80% were under the age of 20. A major research objective was to test the influences of L2 confidence and L2 Norms in predicting identity and frequency of communication in L2. A path model was created using six variables (frequency of L2 contact, quality of L2 contact, L2 norms, L2 confidence, Francophone/Anglophone identity L2 WTC) to predict frequency of L2 communication. The model was tested twice using Anglophones (Figure 3) and Francophones (Figure 4) separately. In each test, the LaGrange Multiplier test was used to modify the model four times.

Although the researchers had theorized that L2 norms would predict L2 WTC, in both revised models, the results indicated that L2 norms predicted ethnic identity and additionally predicted L2 confidence ($r = .18$) for Anglophones. In the theoretical model, L2 confidence predicted both ethnic identity and L2 WTC, both of which then predicted frequency of L2 communication. However, in the revised models, L2 confidence directly predicted frequency of L2 communication ($r = .49$ and $r = .42$ for Anglophones and Francophones, respectively). Multiple regression analyses were performed separately for both ethnic groups using L2 confidence and

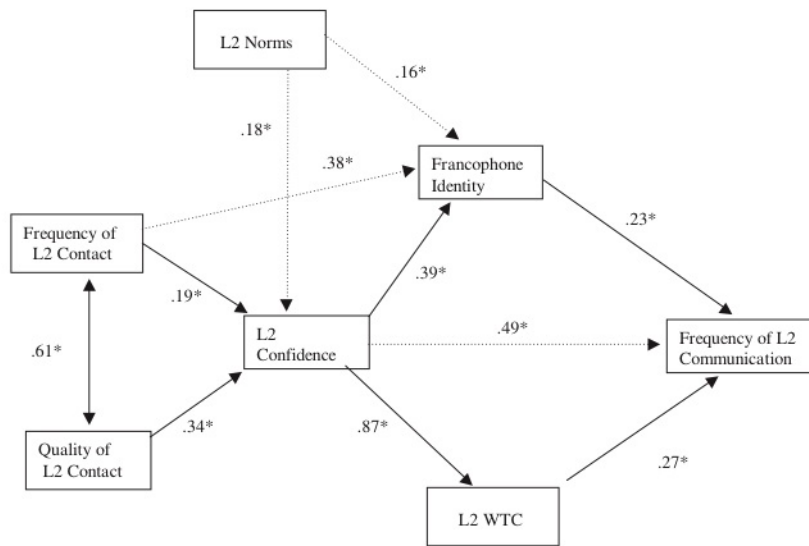


Figure 3. Model of Second Language (L2) Communication among Anglophones from Clément et al. (2003). WTC = Willingness to Communicate.

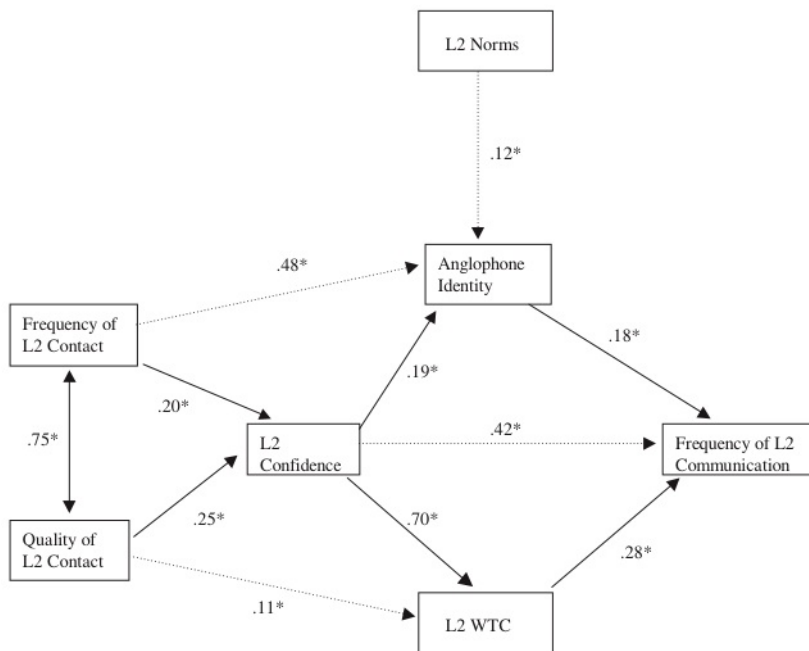


Figure 4. Model of Second Language (L2) Communication among Francophones from Clément et al. (2003). WTC = Willingness to Communicate.

L2 norms as predictor variables of L2 identity. For both ethnic groups, there was a significant correlation between L2 confidence and L2 identity. There was an additional significant coefficient for the influence of L2 norms on L2 confidence for the Francophone group, but not for the Anglophone group.

The researchers noted that French is a minority language in Canada, with only 6% of Ontario residents speaking French natively, and even with a policy of bilingual education, the University of Ottawa was comprised of only 34% native French speakers (p. 195). These statistics make the findings that Francophones have more L2 confidence and frequency of L2 use due to more L2 communicative opportunities than Anglophones somewhat of a given. However, the researchers also stated that the results indicated that for Anglophones, “having the support of significant others decreases anxiety and increases perception of competence” (p. 205). In other words, where there is a the lack of opportunity for frequent L2 contact with members of an L2 community, peer pressure and support in the form of positive-reinforcement social norms can help Anglophones become more confident using their L2. Such a finding could have implications for Japanese learners of English in an EFL context, where there is little opportunity for L2 use outside the immediate classroom.

Some flaws in the Clément et. al, study included a relatively small *N*-size for the Francophone group ($n = 130$), the use of path analysis as an exploratory

technique rather than a confirmatory technique, and the use of items from the WTC scale to create separate, unvalidated scales for L2 norms, frequency of L2 communication, and L2 confidence. As with many SLA studies, the small *N*-size was the most crucial flaw, as small sample sizes can lead to measurement error as well as lack of generalizability. The total number of items across the seven instruments used was 70; according to Stevens (2009), a minimum sample should contain at least five cases per item. In this study, that would mean a minimum of 350 cases. Anglophones and Francophones combined would have been a total *N*-size of 378, which would have exceeded the required minimum sample size; however, the model was not tested using the combined sample size. The combined sample size would also have allowed the researchers first to test their model and respecify parameters before confirming the model using the two ethnic group samples separately. Instead, the researchers allowed the results to determine the model, resulting in a data-driven model, rather than a theory-driven model.

Further, the use of WTC items to measure L2 confidence is the likely cause of the high correlations between L2 confidence and L2 WTC ($r = .87$ and $r = .70$ for Anglophones and Francophones, respectively). Particularly for Anglophones, these high correlations may indicate that L2 confidence and L2 WTC are the same construct, leading to a violation of the assumption of a lack of multicollinearity for the multiple regression analyses.

The study in Clement et al. was conducted in Ottawa, Canada, which can be considered an ESL context for the Francophone participants. However, there is a crucial difference between foreign language confidence in ESL and EFL contexts. In the heuristic pyramid model of MacIntyre et al. (1998), the theoretical variables Intergroup Climate, Intergroup Attitudes, and Social Situation were described in terms of the ways in which members of an L1 group react or relate to the members of an L2 group. However, in a largely monolingual EFL situation such as Japan, there is no L1 community with which to interact. This begs the question whether the L2 social group has any bearing on foreign language confidence at all for EFL learners; because Japanese students do not have direct contact with English users on a daily basis, they might find it difficult even to determine their attitudes toward the English-speaking social group, whomever that may be (e.g., American, British, Singaporean, or Indian). If social situations and contact between users of the foreign language influence Japanese students' perceived foreign language self-competence, foreign language confidence, and the desire to communicate in a foreign language, then one must assume that communication between members of the same foreign language community (i.e., the immediate foreign language classroom context) plays an important role.

In fact, recent studies in Japan have questioned the validity of the social situations in the theoretical model, which seemed to assume regular and frequent

contact between members of the L1 and L2-speaking communities. In a Japanese EFL context, Yashima (2002, 2004) proposed a construct that she termed International Posture, defined as positive attitudes towards what English represents (i.e., foreign countries and foreigners in general) without wanting to integrate or to be in close contact with a specific L2 culture, in the place of the intergroup variables of the L2 WTC heuristic model. However, the International Posture construct seems to confuse the original concepts of Intergroup Attitudes and Intergroup climate, without taking into account the reality of the EFL learning context; students in an EFL classroom necessarily must interact with other members of the classroom, all of whom speak the same native language and none of whom speak the target language as native speakers. Yashima et al. (2004) noted that the connections between communicative behavior and social situations have not been adequately explored (p. 144); I theorize that focusing on L2 social situations familiar to typical EFL Japanese learners, such as communicating with peers within the EFL classroom, might lead to a clearer interpretation of foreign language speaking confidence.

Gaps in the Existing Literature

There are several gaps in the existing literature concerning both personality traits and foreign language speaking confidence. The first gap concerns the lack of

available literature about both personality and foreign language speaking confidence among a Japanese sample population. This lack of information is important because the theoretical Big Five personality traits and foreign language speaking confidence constructs are assumed to be universal in nature; that is, the psychological constructs should be generalizable to all human populations regardless of ethnicity or linguistic background. The overall dearth of studies regarding personality and foreign language speaking confidence in East Asian contexts has already led several researchers to claim that such psychological constructs are “Western” in nature and have no direct application to non-Western populations.

The second gap this study addresses concerns the statistical limitations of existing personality and foreign language speaking confidence studies. One crucial limitation is a lack of instrument validation in general and thus a lack of knowledge of the unidimensionality of the personality and foreign language speaking confidence constructs. This limitation has so far led to questionable Likert-scale questionnaire study results that are sample-specific and have proved difficult to reproduce across different samples. A second limitation concerns the treatment of Likert-scale ordinal data as interval data, including the use of combined raw scores for analyses. A third limitation concerning foreign language speaking confidence is a lack of studies with large sample sizes, leading to a lack of generalizability to

larger populations. A final limitation regarding previous use of structural equation modeling (SEM) in SLA personality trait and foreign language speaking confidence studies is the overuse of path respecification to let the data drive the model, and the lack of follow-up verification of the model with new samples (MacCallum, 1995; MacCallum, Roznowski, & Necowitz, 1992). This limitation is often overlooked even in general psychology studies; however, the use of data-driven models leads to confirmation bias (MacCallum & Austin, 2000). In other words, without validation of a respecified model, researchers tend to focus only on what they had hypothesized to be relationships between constructs, and therefore either ignore or reject alternative, equally plausible models.

A third gap concerns the number of personality traits. Although many mainstream psychologists claim that there are five basic traits underlying human personality, known as the Big Five, there is still a great deal of debate whether there truly are five traits. Numerous studies recently conducted in East Asian contexts have discovered a range of personality constructs between three to seven factors. The main difficulty determining the number of personality traits appears to stem from problems outlined in the previous section regarding statistical limitations; these limitations must be addressed for the gap concerning the number of personality traits to be adequately explored.

A fourth gap is that previous foreign language speaking confidence studies have treated both foreign language anxiety and foreign language confidence as unidimensionality constructs, although clearly both are multidimensional. This gap is important because studies that use questionnaires that have not been confirmed for construct unidimensionality typically produce results that are not reproducible or generalizable across different samples. A fifth, related gap is the inadequate numbers of parameters used in previous foreign language anxiety and foreign language speaking confidence studies, in other words, using only individual perceived foreign language self-competence and foreign language anxiety to define foreign language confidence while seemingly overlooking the possible importance of other factors such as social situations and the desire to speak. This gap is important because existing models of foreign language confidence fail to address theoretical issues in mainstream social anxiety research, such as the relationship among the so-called genetic, or trait-based, personality factors, and social environmental factors.

A final gap is the lack of information concerning the relationship among personality traits and foreign language confidence. This gap is important for two reasons. The first reason is pedagogical: Foreign language instructors instinctively recognize that students in the EFL classroom can often have conflicting personalities that adversely affect pair, group, and class activities. The second

reason is theoretical: Although social anxiety is now conceptualized as a combination of both personality and social factors, there are few studies in the SLA field that have shown definite links between certain personality traits, social situations, and foreign language speaking confidence. Both pedagogical and theoretical issues regarding the relationship between personality and foreign language speaking confidence need to be examined.

In the current study, these gaps in the literature were addressed in several ways. The first gap was addressed by the fact that this study was conducted entirely with Japanese participants in order assist in the verification of personality and anxiety generalization in a non-Western context. The *N*-size of 1,081 is one of the largest single Japanese sample populations to date in a personality or foreign language speaking confidence study.

The second gap concerning statistical limitations was treated in three ways. First, Rasch measurement analysis was used to verify the unidimensionality of the Big Five personality traits and foreign language speaking confidence measurement instruments. These validation results were performed prior to further factor analysis or structural equation modeling. Next, the use of standardized *z*-scores for Likert-scale data and the verification of statistical assumptions for all analytic techniques prior to analysis allowed more accurate analyses of the data by avoiding statistically dubious practices such as adding raw scores from non-normal

distributions. The problems associated with small sample sizes were avoided by collecting data from 1,081 participants spread across twelve different institutions and forty-nine different classes throughout Japan. Finally, the problems associated with path respecification in SEM models were avoided by dividing the data set into two random smaller sets of 539 participants each, using one set of data to test the models and the second set of data to verify the model after model respecification, and by examining two mathematically equivalent SEM models.

The third gap concerning the number of personality traits was addressed through the use of Rasch analysis to check the unidimensionality of the hypothesized constructs. Examination item fit estimates and residuals of the principal components analysis for each hypothesized construct can determine the existence of fewer or more personality traits than the hypothesized five-factor solution. This step was crucial, because only personality traits that can be validated for the study population can be used in subsequent data analysis models.

The fourth gap concerning the treatment of foreign language anxiety and foreign language confidence as unidimensional constructs was likewise addressed in six ways. First, the concept of L2 self-confidence was renamed *foreign language confidence* to more accurately reflect its nature (i.e., all confidence is ipso facto “self”; one may not have another’s confidence). Second, foreign language confidence was theorized as the interaction of genetic, or inherited personality trait

factors, and social factors. Third, the inherited personality trait factors were represented by the latent variables comprising the Big Five personality traits, and the social factors were represented by the latent variables comprising Foreign Language Classroom Speaking Confidence and Favorable Social Situations. Fourth, each constituent construct was considered only in terms of speaking situations; other aspects of foreign language confidence such as test anxiety were not examined. Fifth, all constructs were examined for construct validity through the use of Rasch analysis. Sixth, relationships among the latent variables were analyzed in three equivalent structural equation models.

The final gap concerning the relationships among personality traits and foreign language speaking confidence was addressed through the creation, testing, respecification, and confirmation of structural equation models using the latent variables comprising the Big Five personality traits, Foreign Language Classroom Speaking Confidence, and Favorable Social Situations. Although previous SLA researchers failed to find significant relationships among personality and foreign language confidence variables, it was expected that the models in this study will be able to demonstrate significant relationships in agreement with general psychology theories of social anxiety and confidence.

Statement of Purpose

The purpose of this research was to explore the relationship between factors of the Big Five personality traits and foreign language classroom speaking confidence among Japanese university students. Preliminary analysis involved the validation of two Likert-scale instruments: the Revised Big Five Factor Markers questionnaire and the Foreign Language Classroom Speaking Confidence questionnaire. It was theorized that, in contrary to previous SLA research findings, a combination of the Big Five personality traits and social situation variables would predict foreign language classroom speaking confidence.

Research Questions

1. To what degree are the Big Five personality traits valid for a Japanese university student sample?
2. To what degree is foreign language classroom speaking anxiety associated with perceived foreign language speaking self-competence and the desire to speak English?
3. To what degree are the Big Five personality trait factors associated with foreign language classroom speaking confidence?
4. To what degree does the classroom climate influence foreign language classroom speaking confidence?

CHAPTER 3

METHODS

In this chapter I describe the methodology for the study. First, I explain the characteristics of the twelve institutions around Japan from which study participants were drawn. Next, I describe the two instruments used in the study, the Revised Big Five Factor Marker questionnaire and the Foreign Language Classroom Speaking Confidence questionnaire. Following that I explain the research design and analytical techniques used, and I conclude the chapter with a justification for the models presented in the study.

Participants

The participants in the study were drawn from twelve four-year universities from the western, central, and eastern parts of Japan's main island of Honshu. The twelve institutions will be referred to in the study as School 1, School 2, School 3, School 4, School 5, School 6, School 7, School 8, School 9, School 10, School 11, and School 12, respectively.

School 1 Participants

One hundred seven participants were attending School 1, a private four-year liberal arts university for women in western Japan. All participants were first-year university students engaged in a two-year study abroad preparation program in the Department of International Studies. They were taking two 45-minute classes of English instruction per week, one of which was taught by a native speaker of English and one of which was taught by a native speaker of Japanese. The participants came from twelve intact classes of approximately eight to ten students each. English proficiency was variable, ranging from high-beginning to high-intermediate level. The university is considered highly competitive and is one of the oldest private women's universities in western Japan, consisting of approximately 5,000 students. There were 1,340 admittances for the 2007 academic year; nearly four times that number applied, for an acceptance rate between 25% and 30%.

School 2 Participants

One hundred forty-six participants were attending School 2, a private four-year women's liberal arts university in central Japan with a total enrollment of approximately 5,100 students. Eighteen participants were first-year students majoring in International Studies, and 128 participants were first-year students

majoring in management and policy studies. The international studies majors had intermediate level English proficiency and were highly motivated, while the management and policy majors had low-intermediate level English proficiency with lower motivation. The university is moderately competitive, with 2,336 admittances for the 2007 academic year out of 4,785 applications, an acceptance rate of 48.8%.

School 3 Participants

Forty-two participants came from School 3, a four-year private co-educational liberal arts university in western Japan. Regarded as one of the top universities in the region, School 3 consists of approximately 19,000 undergraduate students studying in a variety of liberal arts disciplines. The participants for this study came from two intact English integrated skills courses that met three times per week. Roughly equal numbers of the participants were majoring in law or sociology, and their average English proficiency was high intermediate as measured by the TOEFL examination (average 450 TOEFL paper-based).

School 4 Participants

Forty-three participants came from School 4, a four-year private women's liberal arts university in eastern Japan. Eleven students were majoring in English,

18 students were majoring in international studies, and 14 students were majoring in liberal studies. All participants had an English proficiency level ranging from low intermediate to intermediate. School 4 is well known as a teacher's training college and consists of approximately 3,000 undergraduates majoring in international peace studies, liberal arts, foreign languages, or cross-culture studies. The participants were taking two 90-minute English classes per week, one of which is taught by a native speaker of English.

School 5 Participants

One hundred twenty-nine participants came from School 5, one of the oldest and largest private universities in the country with branch campuses throughout eastern Japan. With a total undergraduate population of over 70,000 students, this university enjoys a reputation at the undergraduate level as excelling in sports, rather than in academic achievement; however, the university also boasts the highest number of doctorates conferred in Japan and has strong connections to several large Japan-based multinational companies. The participants in this study were all commerce majors and were generally considered to have low English proficiency and low motivation. The participants came from five intact English classes, all of which met once per week for 90 minutes with a native speaker of English.

School 6 Participants

Nineteen participants came from one intact English speaking class at a private four-year university in eastern Japan. Primarily known for foreign language studies, School 6 consists of over 8,000 undergraduate students majoring in foreign languages, law, or economics. The participants from this school were English language majors and were considered intermediate to high-intermediate English proficiency level. They were taking four 90-minute English classes per week, one of which was an oral communication class taught by a native speaker of English.

School 7 Participants

Thirty-six participants came from School 7, a medium-sized private liberal arts university in eastern Japan. School 7 consists of just under 8,000 undergraduates and is considered one of the most elite universities in the country. Many of its graduates continue to graduate school, and the school is well known as a favorite institution for members of Japanese high society and royal family members. The participants were majoring in philosophy and generally had high English proficiency. The participants came from one English oral communication class that met once per week for 90 minutes with a native speaker of English. The participants were taking one additional English reading and grammar class per week with a Japanese native speaker.

School 8 Participants

Forty participants came from two intact English conversation skills classes at a private four-year university in eastern Japan. Consisting of over 30,000 undergraduate students, School 8 is considered one of the oldest and most elite private institutions in Japan and is well known for law studies. The participants' English proficiency level was considered high intermediate. The participants were taking four English classes per week, two of which were taught by a native speaker of English. They were majoring in international studies.

School 9 Participants

Three hundred four participants were from School 9, one of the largest private universities in western Japan. The participants' English proficiency level ranged from high-beginner to low-intermediate, and their English speaking motivation was considered somewhat low. The participants came from 11 intact classes averaging 25 to 28 students each. Participants were majoring in economics, Japanese literature, sociology, or architecture. School 9 comprises over 60,000 undergraduate students across six separate campuses and is considered a "third tier" university. The first-year students were studying English four times per week. English speaking courses met twice per week for 90 minutes each time and were

taught by native English-speaking instructors, and TOEIC examination preparation courses met twice per week and were taught by native Japanese instructors.

School 10 Participants

Seventy participants came from School 10, a medium-sized public university in central Japan. The participants came from three intact classes and were majoring in pharmacology. They generally had low-intermediate English proficiency as measured by the TOEIC. School 10 is one of the oldest science universities in central Japan and consists of over 7,000 undergraduates. The participants in this study were taking two English classes per week: a speaking and listening course with a native speaker of English and a reading and writing course with a native speaker of Japanese.

School 11 Participants

Forty participants came from School 11, a small private liberal arts university in central Japan. School 11 consists of approximately 4,500 undergraduates and is considered a “last chance” university in the central Japan region. The participants were majoring in English, but generally had a low English proficiency. They were taking two English classes per week, one communicative English class with a native speaker of English and one reading class with a native speaker of Japanese.

School 12 Participants

One hundred five participants came from School 12, a private four-year liberal arts university that has a total population of approximately 21,000 undergraduate students. The participants were majoring in economics, commerce, engineering, English, psychology, sports therapy, media studies, life sciences, law, and philosophy. All participants were first-year university students taking two ninety-minute courses per week of required four-skills English instruction; one course was taught by a native speaker of English, and one was taught by a native speaker of Japanese.

The participants came from four intact classes averaging 25 to 30 students each. The participants' English proficiency ranged from false-beginner level to high-intermediate level. The university is considered highly competitive and is one of the oldest four-year private universities in western Japan. There were 6,295 admittances out of 26,298 applicants for the 2008 academic year, for an overall acceptance rate of 26.3%; however, individual department acceptance rates varied from 7.6% to 51.9%.

Summary of Participants in this Study

The total number of participants was 1,081 and consisted of 445 male students and 628 female students. Eight students did not list a gender. All of the participants

had completed at least six years of formal English study, including three years each at junior high school and senior high school. The average age of the participants was 18.87 years old. Informed consent and the right not to participate in the study was conveyed orally to the student participants by individual instructors of the various classes according to a typed script provided by the researcher both in English and in Japanese (Appendix A). Consent information indicating that the data collected would remain absolutely anonymous and confidential and would furthermore have no bearing on final course grades was also conveyed orally by the instructors. This information was additionally written in Japanese at the beginning of the measurement instruments that the participants completed (Appendix B).

Table 1

Percentage of Activities Experienced by Study Participants in the Foreign Language Classroom

Type of activity	Participating Schools											
	1	2	3	4	5	6	7	8	9	10	11	12
Teacher-fronted (i.e., lectures)	15	15	15	5	25	30	10	10	5	10	10	20
Individual work	30	10	25	5	30	30	10	10	5	20	15	20
Pair work	40	65	35	25	25	25	35	30	40	40	30	25
Group Work	15	10	25	60	20	15	30	30	40	30	30	20
Other (i.e., speeches or presentations)	0	0	0	5	0	0	15	20	10	0	15	15

Note. All numbers are percentages.

To confirm the usage of foreign language classroom activities appropriate to items in the two questionnaires, all instructors were asked to indicate, in

percentages, the approximate amount of class time devoted to teacher-fronted, individual student, pair, and group activities. The results indicated that all instructors contributing data to the study utilized pair and group activities extensively in their foreign language classrooms (Table 1).

Instrumentation

Revised Big Five Factor Markers

The Revised Big Five Factor Markers (R-B5FM) questionnaire (Appendices C and D; items used in final models are indicated by asterisks) is a revised version of the Big Five Factor Markers questionnaire, which was designed to measure the existence of the Big Five factors of personality traits and which consisted of two versions comprising 50 and 100 items, respectively (Goldberg, 1992, 1993; Goldberg, Johnson, Eber, Hogan, Ashton, Cloninger, & Gough, 2006). The Factor Marker 50-item version consisted of five factors of 10 items each: Extraversion-Introversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Imagination.

I adapted the 50-item version of the Factor Markers instrument for Japanese university students after conducting a Rasch measurement model analysis of data gathered in a pilot study with 283 participants (see the Pilot Study Analysis section of Chapter 4: Preliminary Analysis). The number of items was reduced to 37; this

form included items from both the 50-item and 100-item original versions by Greenberg (1993). Items in the original Factor Marker 50-item questionnaire were translated into Japanese by a native speaker of Japanese (Nakayama, Yamamoto, & Santiago, 2006). However, after the results of the pilot study were analyzed, the revised questionnaire was back-translated into English by a different native speaker of Japanese. After careful comparison of the English back-translations and the original English items, the Japanese translations of several items were revised and double-checked again by a third speaker of Japanese for translation accuracy. It should be noted that only one of the Emotional Stability items (EMO 4, Have frequent mood swings) seems to reflect the idea of stable or unchanging emotional states or moods. However, the construct is known as Emotional Stability in the Big Five model of human personality traits, and the label is thus used in this study as well for consistency with previous Big Five personality trait studies.

Finally, whereas a 5-point Likert scale including a middle, neutral value was used with the original Factor Markers instrument, a 6-point semantic Likert scale was used with the R-B5FM. In response to the statement, “This personality trait describes me,” the participants answered on a semantic scale from 1 to 6. The poles of the scale were labeled (1 = *Strongly disagree*, 6 = *Strongly agree*); points 2 through 5 were not labeled.

Foreign Language Classroom Speaking Confidence

The Foreign Language Classroom Speaking Confidence (FLCSC) questionnaire (Appendices E and F; items used in final models are indicated by asterisks) consisted of six separate constructs, (a) Foreign Language Classroom Speaking Anxiety, (b) Perceived Foreign Language Speaking Self-Competence, (c) Desire to Speak English, (d) Past English Classroom Experiences, (e) Current English Classroom Perception, and (f) Perceived Social Value of Speaking English, that are measured with a 6-point semantic Likert scale (1 = *Strongly disagree*, 6 = *Strongly agree*, with no labels for points 2 through 5). The constructs measured by this instrument were designed specifically for use with Japanese students, and three of the six constructs were validated in two pilot studies involving 152 and 98 university students, respectively (see the Pilot Study Analysis section of Chapter 4: Preliminary Analysis).

Foreign Language Classroom Speaking Anxiety. Items on the Foreign Language Classroom Speaking Anxiety (FLCSA) questionnaire were based on speaking anxiety items from the Foreign Language Classroom Anxiety Scale (FLCAS) (Horwitz, Horwitz, & Cope, 1986) and communication apprehension items from the Personal Report of Communication Apprehension (PRCA) (McCroskey, 1978). The FLCSA construct consisted of eleven items measuring the

participants' anxiety toward using English with classmates in a foreign language classroom setting. This construct was validated in a pilot study and had a Rasch item reliability estimate of .93.

Perceived Foreign Language Speaking Self-Competence. The Perceived Foreign Language Speaking Self-Competence (PFLSS) questionnaire was based on the speaking component of the Common European Framework of Reference (CEFR), which was devised by the Council of Europe in 2001 as part of the European Language Portfolio Project (Council of Europe, 2004), and by items concerning perceived self-competence developed by McCroskey and McCroskey (1988). The PFLSS construct was a series of ten "I can" statements designed to measure participants' perceived self-competence for speaking English. This construct was validated in a pilot study and had a Rasch item reliability estimate of .96.

Desire to Speak English. The Desire to Speak English (DSE) construct was made up of nine items designed to measure participant intentions to speak English and to improve their English speaking ability. This construct was created based on the hypothesis that, even if participants felt no anxiety while speaking a foreign language, this lack of anxiety would not lead to language learning if the

participants had no interest in speaking English or in improving their speaking ability. Items in this construct are conceptually similar to the idea of International Posture (Yashima, 2002; Yashima et al., 2004). This construct was validated in a pilot study and had a Rasch item reliability estimate of .96.

Past English Classroom Experiences. As it has been hypothesized that language learning experiences contribute to the formation of language confidence (Clément, Dörnyei, & Noels, 1994; Young, 1991), I gathered data to test this hypothesis with the Past English Classroom Experiences (PECE) construct, which was designed to measure participants' past experiences in the English classroom. Because all of the participants in the study were first-year university students, all eight items measuring PECE concerned experiences in high school English classrooms. A previous version of this instrument was piloted with 98 Japanese university students (Rasch item reliability = .89); however, three items from the instrument were revised prior to the main study and the new version of this instrument was not validated prior to the main study (see the Pilot Study Analysis section of Chapter 4: Preliminary Analysis).

Current English Classroom Perception. SLA anxiety researchers have noted that one of the primary sources of language anxiety development is the fear

of speaking in front of one's peers (Cohen & Norst, 1989; MacIntyre, 1999; Price, 1991). Thus, in the EFL situation, I theorized that the individual learner's perception of classmates' English abilities, desires, and attitudes toward speaking English would play an important role in shaping reactions to the social situations in which foreign language speaking anxiety and foreign language speaking confidence develop.

The Current English Classroom Perception (CECP) construct consisted of eight items concerning various activities in the English classroom and participant attitudes toward their current English classroom atmosphere. Because the participants were attending university courses whose instructors used a combination of pair, group, and whole class oral activities, I expected that the participants' perceptions of the current English classroom atmosphere would have a negative correlation with Foreign Language Classroom Speaking Anxiety and a positive correlation with Perceived Foreign Language Speaking Self-Competence. In other words, if participants perceived that the atmosphere provided positive support for English learning and speaking, their speaking anxiety levels would be low and their perceived speaking competence levels would be high. This instrument was piloted with 98 Japanese university students (Rasch item reliability = .83); however, all items were slightly revised, and the current version of this

instrument was not validated prior to the main study (see the Pilot Study Analysis section of Chapter 4: Preliminary Analysis).

Perceived Social Value of Speaking English. Clement et al. (2003), sought to demonstrate the relationship between frequency of L2 contact, L2 confidence, and WTC, in an ESL context. They theorized that Francophone minority language group members at the University of Ottawa, Canada, would feel more social pressure to communicate in English, while the use of French by majority language group English-speakers would be “self-attributed to personal disposition and not external normative pressure” (Clément, Baker, & MacIntyre, 2003, p. 195). The researchers examined language group identity with a twelve-item sub-scale called Subjective L2 Norms and found that the factor influenced L2 Confidence among English-speakers, but not among French-speakers.

However, unlike students in the Canadian ESL context, students in the Japanese EFL situation encounter very few opportunities for L2 contact. According to the most recent census conducted by the Japanese government, there was a record high of 1.3 million foreigners living in Japan in 2005; however, those foreigners comprised only 1.2% of the total population (Population of Japan 2005, n.d., p. 71). Moreover, the native-English speaking population of Japan was extremely small compared to the Chinese and Korean-speaking populations: There

were 466,637 Korean nationals and 346, 877 Chinese nationals, but U. S. citizens only totaled 37,417 (2.4% of the foreign population) and the number consistently dropped each year from 1995 to 2005 (Population of Japan 2005, n.d., p. 72).

These numbers are unlikely to be perfectly accurate, and other native English-speakers (from the United Kingdom, Canada, Australia, Ireland, New Zealand, and other countries) were too few to be listed in the Japan 2005 census report; however, given the numbers available, a reasonable guess is that less than 50,000 of all foreigners living in Japan are native English speakers. Even if there were an influx of foreign tourists to Japan, the vast majority of tourists to date seem to come from neighboring South Korea and China, and, indeed, the Japanese government actively promotes tourism from these countries by offering free travel visas to people from mainland China, Taiwan, and South Korea. Thus, it stands to reason that the average Japanese university student is unlikely to have much contact with speakers of English outside of the foreign language classroom.

If Clement et al.'s results indicating that social norms are not as influential as individual differences for the majority language users are accurate, then Japanese speakers' foreign language speaking confidence should be similarly influenced by their internal dispositions rather than by the normative pressure of their social context, i.e., the foreign language classroom. However, personal experience as well as anecdotal evidence has suggested that Japanese learners are strongly influenced

by social situations, and particularly by the attitudes of their classmates. Therefore, I have hypothesized that perceptions of the value placed upon speaking English by members of Japanese society strongly affects the individual Japanese EFL university learner's speaking confidence level.

To test this hypothesis, I created an eight-item instrument called Perceived Social Value of Speaking English (PSV) that was designed to measure the participants' sense of social pressure to speak English, including pressure from family, friends, classmates, teachers, companies, and the society at large. I hypothesized that PSV is positively correlated to PECE and CECP, as all three constructs are measures of societal influences on foreign language speaking confidence. The PSV instrument was not piloted or validated prior to the main study.

Research Design

Before data collection commenced, the questionnaire items were randomized into two questionnaires. The R-B5FM was on one side of an A3 piece of paper, and the FLCSC was on the other side. The participants completed the questionnaire during regular class time in June at approximately the midpoint of the spring academic semester. It took the participants approximately 20 to 25 minutes to complete the questionnaire.

Analytical Procedures

In this section, I first provide an overview of the analytical techniques used in this study. Next, I explain the procedures and models used to examine each research question. Finally, I provide a justification for why the analytical techniques used in this study represent an improvement over the techniques used in previous studies.

Overview of Analysis Techniques

The research questions were investigated using (a) Rasch analysis using WINSTEPS 3.63 software (Linacre, 2006b), (b) exploratory factor analysis (EFA) using SPSS 15 software (SPSS, 2006), (c) confirmatory factor analysis (CFA) using EQS 6.1 software (Bentler, 1995), and (d) structural equation modeling (SEM) structural regression analysis using EQS software.

Rasch analysis. The Rasch model (Rasch, 1960), also known as the one-parameter logistic model (1PL), measures the relationship between item difficulty and person ability as the ratio of success/failure for passing an item and expresses the difference in logarithm form (Embretson & Reise, 2000). The Rasch model states that the probability a person will get an item correct is logistically related to the difference between the person's ability level and the item difficulty.

The logit form of the dichotomous Rasch model is

$$\ln [P_{ni} / 1 - P_{ni}] = \theta_n - b_i$$

where

P_n = person n

i = the item being answered

θ_n = the ability level of person n

b_i = the ability level of the item (Wright, 1999).

Data is fit to the Rasch model by mathematically transforming raw scores on items into logarithms and then by placing both item responses and persons on the same log-odds scale. In other words, the unrelated percentages of correctly and incorrectly guessed items become transformed to a linear scale, which generally extends from -5 to +5 logits, with the mean difficulty level of items set to zero.

Rasch analysis of Likert-scale data is important for three reasons. First, most researchers in the social sciences treat data obtained through Likert-scale questionnaires as interval data, while in fact such data are ordinal. Researchers using the Rasch model consider Likert-scale data as ordinal; however, these ordinal data can be transformed into interval data with the Rasch Rating Scale or Partial Credit models (Andrich, 1978; Masters, 1982; Wolfe & Smith, 2007a). Second, although researchers in the social sciences often assume that each item contributes equally to the measurement of the construct, some items are much easier to endorse

than others. Rasch analysis can demonstrate the relative endorsability of items through the use of item-person maps, which display both items and persons on the same logit scale (Wilson, 2005, p. 96) and item difficulty estimates. In addition, the quality of each item can be assessed through Rasch item fit analysis such as item-measure correlations and mean-squared fit indices. Item and person responses that misfit the model can be the result of carelessness, response set answering, or item bias (Wolfe & Smith, 2007b, p. 211).

Third, researchers frequently assume that the latent constructs being measured are unidimensional, when in fact no checks of this assumption are performed. Embretson and Reise noted, “[I]t is known that the size of an internal consistency index (e.g., coefficient alpha) is irrelevant to dimensionality” (Embretson & Reise, 2000, p. 231), and Waugh and Chapman have further commented that, “...just because scores on items are correlated doesn’t mean that one has a conceptual scale of items even if there is a strong loading on a single factor” (Waugh & Chapman, 2005, p. 81). Further, Wright pointed out that classical test or “true-score” theory factor analytic techniques do not account for standardized errors (residuals) left behind after extracting the first factor. The residuals then become “noise” that reduce the validity of further factors (Wright, 1996, p. 10).

After subjecting questionnaire items designed to measure a single construct to a Rasch principal components analysis (PCA) of the residuals, resulting contrasts

of eigenvalues over a certain level indicate the possible existence of another factor, and thus would demonstrate lack of construct validity (Linacre, 2007; Smith, 2000). Thus, Rasch analysis for item fit, person fit, and construct validity is a necessary step before further analysis. In addition to checking construct validity, Rasch analysis of the data in this study answers primary research question one concerning the validity of the Big Five personality traits for a Japanese sample. The following sub-sections describe the Rasch analyses used to examine the data in this study.

Rasch Principal Components Analysis (PCA). Following data screening, a Rasch Factor Analysis (Wright, 1996) was conducted using the technique of Rasch Principal Components Analysis (PCA) of item residuals to determine which items cohered to form unidimensional constructs. In Rasch Factor Analysis, data obtained from a multidimensional instrument are examined by conducting an unrotated PCA on the item residuals that remain after extracting a linear measure. Item residuals that correlate highly (e.g., above a factor loading of .40) have variance unexplained by the Rasch model and may need further examination to determine unidimensionality of construct (Linacre, 1998; Bond & Fox, 2007; Smith, 2002).

A unidimensional construct should normally account for at least 50% of the total variance; the residual errors left over from the factor extraction should ideally be uncorrelated error with eigenvalues less than 3.0 and variance accounted for

under 10% (Linacre, 2007). Principal residual contrasts with greater eigenvalues and accounted variance indicate the presence of correlated errors and thus additional constructs in the data. All Rasch PCA of item residual analyses were conducted using Winsteps 3.63 (Linacre, 2006b).

Rasch Likert scale category functioning analysis. The same 6-point Likert scale, with “1” representing *Strongly disagree* and “6” representing *Strongly agree*, were used with the two questionnaires in this study (R-B5FM and FLCSC). In order to determine whether the 6-point Likert employed in this study was effective, the following criteria (Linacre, 2002) were considered:

1. There are at least 10 observations for each step of the scale.
2. The average person measure for each step should be higher than the average person measure of the previous step.
3. Outfit mean squares of each step should be less than 2.0.
4. There should be gaps in step difficulties of no fewer than .59 logits or 2.68 CHIPs.

In the event the above criteria were not met, Likert scale categories were combined until they met the criteria proposed by Linacre. Because logits are often difficult for the average reader to understand, the logit scale was transformed to CHIPs for the item and person measures. Also known as response probability units,

CHIPs have an arbitrary mean score of 50 for the item difficulty estimates (Smith, 2000; Wright & Stone, 1979).

Rasch item fit analysis. To determine the fit of items to the hypothesized construct, Rasch analyses use two kinds of chi-squared fit statistics (infit and outfit) that come in both unstandardized and standardized forms. Because persons closer to the level of item difficulty can give a more accurate measurement of the item performance, infit statistics are weighted to give more information about the variance from expected model by persons who are at or near the endorsability difficulty level of the item (Bond & Fox, 2007, p. 238). On the other hand, outfit statistics are not weighted, making them more easily affected by the responses of persons who find the item too easy or too difficult to endorse (Bond & Fox, 2007, p. 57). Typically, more attention is paid to infit statistics than outfit statistics.

Unstandardized *mean squared* fit statistics (MNSQ) are measured on a scale of zero to positive infinity, with an item mean-square fit statistic of 1.0 indicating perfect fit of the data to the expectations of the Rasch model (de Ayala, 2009, p. 53). Standardized *t* or *z* fit statistics (ZSTD) are similar to the normal distribution curve and have a mean of zero with a scale theoretically extending to positive and negative infinity. Item response patterns that are too predictable (overfit; < 1.0

mean squared or 2.0 z) or too erratic and unpredictable (underfit; > 1.0 mean squared or $- 2.0 z$) indicate item misfit to the model's expectations.

Guidelines for interpreting acceptable fit varies; however, the sample size can often determine which guidelines make more sense than others. The three pilot studies to the main study had relatively small sample sizes (283, 152, and 98, respectively), making ZSTD a more reasonable estimate than MNSQ for determining item fit. For the first pilot study, Smith's 2002 recommended criteria of plus or minus 2.0 ZSTD for small sample sizes was adopted, and for the second and third pilot studies, Linacre's 2007 recommended criteria of plus or minus 3.0 ZSTD was adopted. For the main study, Linacre's 2007 recommended criteria of Infit and Outfit MNSQ statistics between .5 and 1.5 logits as an indication of good item fit for Likert-scale-based models was adopted.

Confirmatory factor analysis and structural equation modeling.

Confirmatory factor analysis (CFA) is frequently used in the field of psychology as a means of examining construct validity for Likert-scale questionnaires. CFA can also be used to examine the relationships between latent variables representing psychological constructs. Generally speaking, each latent variable in a CFA model is expected to consist of at least three measurements, or parameters, with a ratio of at least 10 cases to each parameter (Kline, 2005). Structural equation modeling

(SEM) differs from CFA in that variables are predicted to influence each other in causal relationships. SEM can also use either latent variables such as Desire to Speak English, which consist of at least three observed variables, or individual observed variables such as individual questionnaire items. Observed variables that are theorized to predict other variables in a directional path are used in path analysis, while latent variables that are theorized to predict other variables are used in structural regression analysis, the SEM technique used in this study.

Modeling specification takes three forms: (a) confirmatory modeling of a single model, (b) the generation of a preliminary model that is then modified until adequate fit is achieved, and (c) the specification and evaluation of multiple a priori models. The approach taken in this study follows the advice of MacCallum and Austin (2000), who recommended that any model generation or model modification be evaluated by fitting the model to a separate sample for confirmation.

Model fit indices used in this study. Fit indices remain a controversial topic among users of SEM. The traditional fit index of Chi-squared (χ^2) has been criticized for being overly influenced by sample size, although Kline (2005) mentioned the use of “normed chi-squared” (NC), which is computed by dividing chi-squared by the degrees of freedom to adjust for sample-size (pp. 136-137). Many goodness-of-fit indices have been criticized for being too lenient as well as

sample-dependent. Hu and Bentler (1998, 1999) recommended against the use of commonly used indices such as χ^2 , goodness of fit (GFI), and adjusted goodness of fit (AGFI) due to their inconsistent sensitivity to model misspecification and sample size. Instead, they recommended reporting standardized root mean square residual (SRMR), non-normed fit index (NNFI), and root mean square error of approximation (RMSEA). MacCallum and Austin (2000) particularly encouraged the use of RMSEA, as it provides a confidence interval.

Although Bentler (2007) recommended that SEM articles should report “at most two other indices of fit, such as CFI (comparative fit index) and RMSEA...” (Bentler, 2007, p. 826), because there is no way of knowing which of these indices will be used in the future, I reported the following: χ^2 , NC, CFI, GFI, AGFI, SRMR, NNFI, and RMSEA with confidence intervals. Other statistics reported from EQS output are in line with current recommendations (MacCallum & Austin, 2000; Raykov, Tomer, & Nesselroade, 1991). All confirmatory factor analyses and structural equation models were conducted using EQS 6.1 software (Bentler, 1995).

Proposed models. In this study, three confirmatory factor analysis (CFA) models answered research questions 1 and 2 concerning the validity of the Big Five personality traits and the composition of foreign language speaking confidence. A structural regression model (SRM) was created in order to answer research

questions 3 and 4 concerning the relationships between the Big Five personality traits, social situation, and foreign language classroom speaking confidence (Figure 5). A structural regression model consists of two components: a CFA (measurement model portion) and an SEM (structural model portion). The measurement model portion must be tested prior to the structural model portion to demonstrate validity of the latent variables in the path model (Kline, 2005, p. 209).

In the structural regression model, the Big Five personality traits were hypothesized to predict both favorable social conditions and foreign language classroom speaking confidence latent variables, and favorable social conditions variables are also hypothesized to predict foreign language classroom speaking confidence variables. There is no existing empirical research that entirely supports the hypothesized model in this study; however, Clément et al. (2003) found that there was a significant predictor path between the belief that speaking French was important to others (i.e., perceived social pressure in the classroom) and confidence in speaking French in a model of L2 use among native speakers of English in Ottawa. Thus, it was theorized in the structural regression model that favorable social conditions variables influenced the desire to speak English as well as perceptions of speaking competence and foreign language speaking anxiety in the classroom.

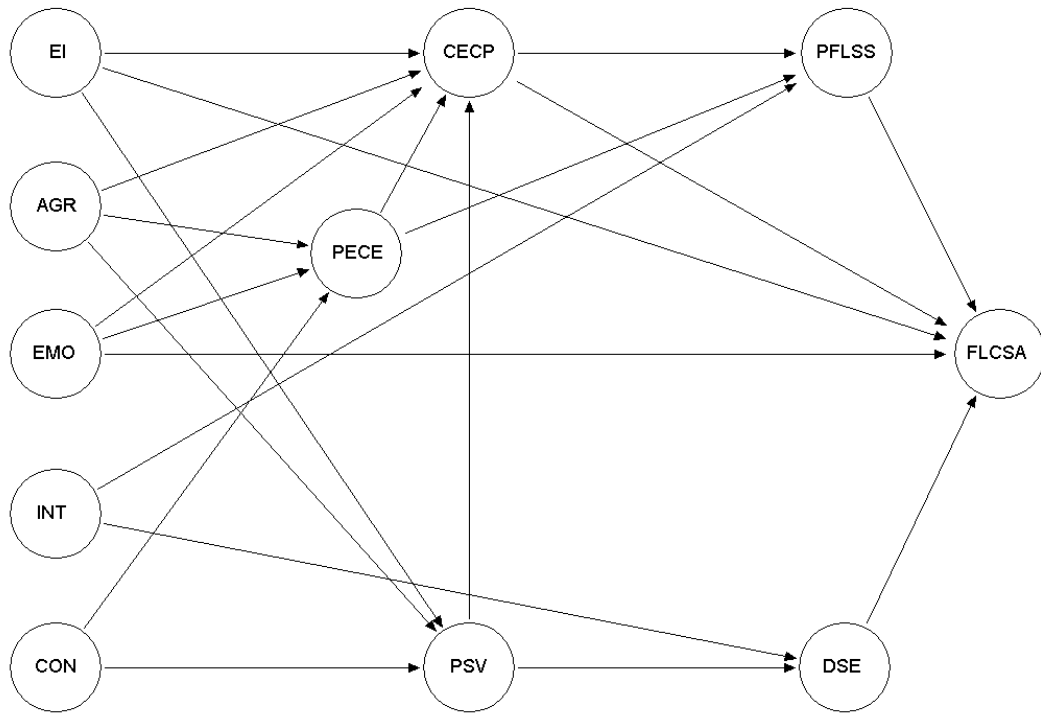


Figure 5. The structural path portion of the structural regression model. EI = Extraversion-Introversion; AGR = Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination; PECE = Positive English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

Testing the Proposed Models

In order to test the CFA and SRM, the data ($N = 1081$) were split into two random data sets of 539 participants to make it cross-validation possible (Cudeck & Browne, 1983). The first data set of 539 participants, or calibration sample, was used to verify the CFA measurement component of the structural regression model

as per the two-step rule (Bollen, 1989; Kenny, Kashy, & Bolger, 1998). All latent variables were placed into three separate CFA models examine model fit. After path respecifications, each model was tested with the second data set of 539 participants as a validation sample for cross-validation of the path invariance (Byrne, 2006, pp. 252-255). Following confirmation of the CFA measurement component of the structural regression model, the structural path component was similarly tested with the calibration sample for model fit. After path respecification, the structural path was cross-validated with the validation sample to determine path invariance.

Justification for Analysis Techniques

My use of modeling techniques in this study was an improvement over previous studies for three reasons. First, both CFA and SEM have been used by previous SLA researchers as an exploratory procedure rather than as a means of confirming a hypothesized model, leading past researchers to insert post hoc pathways into their models in order to achieve adequate model fit without using a separate validation sample for confirmation. Previous researchers have also used fewer than three observed variable predictors for latent variables, resulting in violations of multicollinearity and the empirical underidentification of the model (Kenny, 1979).

Second, previous researchers have confused CFA and SEM; for example, CFA models have been constructed using a single hypothesized construct divided into three separate measured variables to predict a latent variable, which then was used as a measured variable in an SEM path analysis. While structural regression models can be used to combine aspects of CFA and SEM models, both the measurement model and structural model components have to be separately tested for fit in order for the entire structural regression model to be evaluated (Kline, 2005, p. 233). These are steps that previous researchers did not implement.

Finally, previous researchers have not considered equivalent models. Achieving adequate model fit does not imply that the hypotheses or the causal relationships are true; however, the lack of equivalent model testing is not restricted to the field of SLA. In fact, most psychological researchers ignore the existence of alternative path models that mathematically could demonstrate fit the data (MacCallum & Austin, 2000; MacCallum, Wegener, Uchino, & Fabrigar, 1993). Considering alternative models to the hypothesized model, even if adequate fit is found for the hypothesized model, helps prevent confirmation bias (MacCallum & Austin, 2000, p. 213).

In the next chapter, I present results from the preliminary analysis. First, the results from three pilot studies conducted prior to the main study are briefly discussed. Next, the data from the two instruments in the main study are examined

using Rasch analysis and exploratory factor analysis (EFA). Finally, results from Rasch analysis and EFA are compared.

CHAPTER 4

PRELIMINARY ANALYSIS

The preliminary analysis for this study consisted of three steps. The first step involved analyzing data from the pilot studies, which were conducted prior to the main study, in order to select items for the Revised Big Five Factor Markers (R-B5FM) and Foreign Language Classroom Speaking Confidence (FLCSC) questionnaires used in the main study. The second step concerned validating both instruments by using the Rasch rating scale model to analyze the data collected from the participants in the main study. In the third step, exploratory factor analysis (EFA) was used to check normality assumptions and to provide a traditional method of examining of the hypothesized factors in the questionnaire data gathered from the participants in the main study. In the fourth step, results from Rasch analysis and EFA were compared to assist in determining which items to retain for the subsequent modeling analysis.

Pilot Study Analysis

Three pilot studies were conducted in the spring term of the 2008 academic year. In the first pilot study, the 50-item version of Greenberg's (1999) Factor Marker questionnaire, which was administered to 283 first- and second-year

students at a private Japanese university, was examined. Data were examined using EFA, Rasch item fit analysis, and a Rasch PCA of item residuals analysis. The EFA revealed overall good fit to the hypothesized five-factor model, accounting for 40.95% of the total variance. However, three Agreeableness items, one Intellect/Imagination item, and one Conscientiousness item failed to load onto their hypothesized factors at the .40 criterion, while two Agreeableness items and two Intellect/Imagination items loaded onto unintended factors at below .40.

The Rasch analysis results of the Factor Marker pilot study indicated that slightly over half of the items of each of the five hypothesized constructs fit the Rasch model, based on a fit criterion of ± 2.0 infit and outfit standardized z -scores recommended for small sample sizes (Smith, 2002). Person separation was less than 2.0 for four of the five constructs, with only Extraversion-Introversion exceeding 2.0; this finding indicated the relative homogeneity of the participants and the inability of the questionnaire items to separate them into at least two statistically distinct groups. A Rasch principal component analysis (PCA) of item residuals showed that first contrast variances for the five expected constructs ranged from 8.7% (eigenvalue = 1.9) to 15.7% (eigenvalue = 2.3), well above the ideal first contrast value of 5% suggested by Linacre (1998). None of the constructs had principal first contrasts with eigenvalues greater than 3.0, one indication of bidimensionality (Linacre, 2006a); however, the first contrast for

Intellect/Imagination approached this value, with an eigenvalue of 2.8. The Agreeableness construct showed the least defined level of unidimensionality, with only 31.4% of variance explained for the construct compared to 68.6% unexplained variance. However, the construct also had the smallest person separation (1.42) and the lowest person reliability (.67), both of which suppressed the expected variance accounted for by the Rasch model (Linacre, 2006a). Moreover, the variance accounted for by the Rasch model was less than four times that of the unexplained variance for all five constructs, showing the instrument to be an unreliable indicator of the latent constructs (Linacre, 2007). Thus, while not necessarily failing the test of unidimensionality, the constructs did not conclusively demonstrate the presence of an adequate degree of unidimensionality (Apple & Neff, in progress; Apple & Neff, 2010; Neff & Apple, 2008).

Based on the pilot study results, items that misfit their intended constructs were either deleted or revised and new items were added. The new items were based on items from similar constructs on Greenberg's (1998) 100-item version of the same five-factor personality questionnaire. For Extraversion-Introversion, five items were discarded, three of the remaining five items were reworded positively, and three new items were added. For Agreeableness, six items were discarded, two of the remaining four items were reworded positively, and four new items were added. For Conscientiousness, five items were discarded, two of the remaining five

items were reworded positively, and two new items were added. For Emotional Stability, five items were discarded and two new items were added. For Intellect/Imagination, six items were discarded, one of the remaining four items was revised, and three new items were added. The resulting R-B5FM questionnaire consisted of 37 items, which were intended to measure the five hypothesized factors of the five-factor model of human personality. Although the original items were previously translated from the original English into Japanese (Nakayama, Yamamoto, & Santiago, 2006), the item translations were revised by a native speaker of Japanese and back-translated into English by a different native speaker of Japanese to confirm the accuracy of the translation. Both speakers of Japanese were teaching English as a foreign language at the tertiary level of education in Japan and had extensive experience in SLA research writing in English.

The second and third pilot studies concerned the creation and testing of the Foreign Language Classroom Speaking Confidence (FLCSC) instrument. The second pilot study was focused on the first two constructs of Foreign Language Classroom Speaking Anxiety (FLCSA) and Perceived Foreign Language Speaking Self-Competence (PFLSS). The FLCSA originally consisted of 20 items based on the speaking anxiety items from the Foreign Language Classroom Anxiety Scale (Horwitz, Horwitz, & Cope, 1986) and communication apprehension items from the Personal Report of Communication Apprehension (McCroskey, 1978). The

PFLSS originally consisted of 16 items based on speaking competence items from the Common European Framework of Reference (Council of Europe, 2004) and the Self-Perceived Communication Competence Scale (McCroskey & McCroskey, 1988). Items from the two instruments were randomized into a single questionnaire for the pilot study.

A sample of 152 first-year students attending a private Japanese university completed the questionnaire, and the data were analyzed with Winsteps 3.63 (Linacre, 2006b). As the *N*-size was somewhat small, item fit analysis was conducted based on the slightly lenient criteria of ± 3.0 standardized *z*-scores (Linacre, 2007) in order to retain as many items and persons as possible. Seven out of 20 FLCSA items and four out of 16 PFLSS items misfit the model. However, the Rasch PCA of item residuals indicated that 75.4% of the variance for the FLCSA and 81.2% of the variance for the PFLSS was explained by the Rasch model. The first residual contrasts for Foreign Language Classroom Speaking Anxiety and Perceived Foreign Language Speaking Self-Competence explained only 3.9% (eigenvalue = 2.4) and 2.5% (eigenvalue = 2.1) of the variance, respectively, indicating that the constructs were fundamentally unidimensional. The Rasch item reliability of the items measuring Foreign Language Classroom Speaking Anxiety was .93 (item separation = 6.41), and the Rasch item reliability of the items measuring Perceived Foreign Language Speaking Self-Competence was .96 (item

separation = 8.61). The two constructs were uncorrelated, with a non-significant Pearson's correlation coefficient of $r = -.33$ (Apple, 2008).

These results seemed to agree with those reported by Matsuoka (2006), who found that Perceived Competence and Communication Apprehension were correlated at $r = -.34$ for the Japanese university students in her study (p. 105); thus, anxiety and perceived competence were not correlated strongly enough to form a higher order latent variable of linguistic confidence as suggested by previous researchers (e.g., Clément, Dörnyei, & Noels, 1994; MacIntyre, Noels, & Clément, 1997; MacIntyre, Clément, Dörnyei, & Noels, 1998a).

For the third pilot study, eight items from Foreign Language Classroom Speaking Anxiety and six items from Perceived Foreign Language Speaking Self-Competence were deleted, resulting in 12 items for Foreign Language Classroom Speaking Anxiety and 10 items for Perceived Foreign Language Speaking Self-Competence. Three additional constructs were added. Desire to Speak English (DSE) consisted of nine items and was intended to complement Foreign Language Classroom Speaking Anxiety and Perceived Foreign Language Speaking Self-Competence constructs, based on the hypothesis that students who desire to communicate in English were more likely to become anxious while speaking than those who did not want to speak. The Past English Classroom Experiences (PECE) construct (nine items) and Current English Classroom

Perception (CECP) construct (eight items) were created to examine whether classroom dynamics influenced speaking confidence.

The participants in the third pilot study were 98 first-year Japanese university students. Data were analyzed using Winsteps 3.63. The results confirmed that the Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English constructs were fundamentally unidimensional. The Rasch model explained 70.4% of the variance for Foreign Language Classroom Speaking Anxiety, with the first principal contrast explaining 5.7% (eigenvalue = 2.3); 77.5% of the variance for Perceived Foreign Language Speaking Self-Competence, with a first principal contrast of 5.5% (eigenvalue = 2.2); and 73% of the variance for Desire to Speak English, with the first principal contrast explaining 6.0% (eigenvalue = 1.8). However, the unidimensionality of Past English Classroom Experiences and Current English Classroom Perception were not as well defined. The Rasch model explained 60.8% of the variance for Past English Classroom Experiences, with the first principal contrast explaining 10.0% (eigenvalue = 2.3), and 56.9% of the variance for Current English Classroom Perception, with the first residual contrast explaining 10.0% (eigenvalue = 1.9).

Item and person reliability estimates were better overall for the three previously examined constructs of FLCSA, PFLSS, and DSE than for the new

constructs of PECE and CECP. Item reliability for the FLCSA construct was .93 (item separation = 3.65) and person reliability was .85 (person separation = 2.37). Item reliability for the PFLSS construct was .96 (item separation = 4.83) and person reliability was .82 (person separation = 2.16). Item reliability for the DSE construct was .96 (item separation = 4.96) and person reliability was .83 (person separation = 2.17). Item reliability for the PECE construct was .89 (item separation = 2.82) and person reliability was .80 (person separation = 2.02). Item reliability for the CECP construct was .83 (item separation = 2.24) and person reliability was .79 (person separation = 1.79). There were no misfitting items for the FLCSA, PFLSS, DSE, and CECP constructs. However, PECE had one misfitting item that concerned previous test scores, and removing the item from the analysis caused an additional item concerning previous course grades to misfit. An examination of the item-person maps for both Past English Classroom Experiences and Current English Classroom Perception revealed item redundancy, possibly due to similar item wording, that might have led to relatively small item separation estimates.

Following the analyses of the data gathered in the pilot study, two items about grades and test scores were eliminated from Past English Classroom Experiences and one additional item about pair work was added, resulting in eight items designed to measure Past English Classroom Experiences. All eight items designed to measure Current English Classroom Perception were retained; however, the

Japanese translations for all items were slightly revised in order to improve the wording of the items and to separate them from other items. After examining the analyses and retaining the best items for Foreign Language Classroom Speaking Anxiety and Perceived Foreign Language Speaking Self-Competence, the number of item measuring Foreign Language Classroom Speaking Anxiety was reduced to 11 and the number of items measuring Perceived Foreign Language Speaking Self-Competence was reduced to ten. Items measuring Desire to Speak English were left unchanged. A sixth construct, Perceived Social Value of Speaking English (PSV) was created (eight items); this construct was intended to further examine the participants' perceptions of the usefulness of speaking English in Japanese society in general. However, these items were not piloted prior to the main study.

Instrument Validation

In this section, I present the results from the validation of the questionnaires used in the main study. First, the background of the main study participants is described. Second, the data comprising the five hypothesized constructs from the Revised Big Five Factor Markers questionnaire are examined with the Rasch rating scale model and exploratory factor analysis. Third, the data comprising the six hypothesized constructs from the Foreign Language Classroom Speaking

Confidence questionnaire are examined using the Rasch rating scale model and exploratory factor analysis.

Background of the Main Study Participants

The Revised Big Five Factor Marker questionnaire (R-B5FM) and Foreign Language Classroom Speaking Confidence (FLCSC) questionnaire were completed by 1,081 first-year Japanese university students. The students completed the questionnaire during class time in June, 2008. The participants were studying communicative English in 49 intact classes at twelve separate universities across Japan with teachers who were native speakers of English (Table 2). Four hundred forty-five students were male (41.2%), and 628 students were female (58.1%). Eight students did not list a gender. The students were studying a wide variety of majors (Table 3); the largest numbers were majoring in Economics ($n = 225$), International Studies ($n = 183$), Commerce ($n = 164$), and Management Studies ($n = 128$), with all other majors numbering less than 100 students each. The average age of the participants was 18.87.

Prior to instrument validation analysis, data were screened for instances of excessive missing data. Cases without responses to more than 5% of the questionnaire items (i.e., more than two or three out of the total 89 questions) were eliminated as untrustworthy participants. An examination of the data revealed that

Table 2

Background of the Participants

School	Location	<i>n</i>	Male	Female
School 1	Kyoto	107	0	107
School 2	Nagoya	146	0	146
School 3	Kobe	42	13	29
School 4	Tokyo	43	0	43
School 5	Shizuoka	129	79	48
School 6	Saitama	19	5	13
School 7	Tokyo	36	10	26
School 8	Tokyo	40	14	26
School 9	Osaka	304	215	89
School 10	Gifu	70	37	30
School 11	Nagoya	40	14	26
School 12	Kyoto	105	59	45

three participants had an excessive number of missing responses; the first person failed to answer 29 items, the second person failed to answer 20 items, and the third person failed to answer six items while answering “5” or “6” to all other items.

These three participants were removed, resulting in a final *N*-size of 1,078 for both questionnaires.

The first validation step for each instrument in the main study consisted of using the Rasch rating scale model. First, because the Rasch model is unaffected by reasonable amounts of missing data, missing data points were replaced by periods (.) to inform the Winsteps software program not to process the missing information during the Rasch analyses. Second, unidimensionality was tested by a Rasch principal component analysis of item residuals using Winsteps 3.63 (Linacre, 2006b) to examine the variance accounted for by the Rasch model and the item

Table 3

Major Fields of Study of the Participants

Major	Number	Percent
International Studies	183	16.9
Management Studies	128	11.8
English	75	6.9
Law / Political Science	24	2.2
Social Welfare	22	2.0
Liberal Studies	14	1.3
Commerce	164	15.2
Philosophy	36	3.3
Economics	225	20.8
Architecture	62	5.7
Japanese Literature	28	2.6
Sociology / Cultural Studies	24	2.2
Engineering	1	.1
Pharmacology	70	6.5
Media Studies	21	1.9
Theology	1	.1
Sports Health	2	.2
Missing	1	.1
Total	1,081	100.0

residuals. Third, each construct was analyzed separately for Likert scale category functioning and Rasch item fit.

Rasch person and item separation and reliability estimates were also obtained at this stage. In a Rasch analysis, person reliability is analogous to Cronbach's alpha, because it demonstrates the consistency of individual person responses to questionnaire items (Bond & Fox, 2007, p. 284). However, because extreme scores that have no error variance are included in calculating Cronbach's alpha, the reliability estimates produced are generally higher than Rasch person reliability

estimates (Clauser & Linacre, 1999). By way of comparison, Cronbach's alpha estimates were also determined as a more traditional means of demonstrating reliability. Fourth, Rasch item-person maps were obtained for each construct to examine item difficulty levels in relation to the person ability estimates and the structure of the empirical item hierarchy.

The second validation step consisted of exploratory factor analysis (EFA) using SPSS 15 to examine the number of factors according to traditional factor analysis based on correlational matrices. The data were screened again at this stage to confirm the assumptions of normality necessary for conducting the EFA. The EFA results were compared to the Rasch analysis results to determine which items to retain for the confirmatory factor analyses (see Chapter 5: Structural Equation Model Results).

The Revised Big Five Factor Marker Questionnaire

Rasch PCA of item residuals. A Rasch Factor Analysis using Rasch principal components analysis of the item residuals was conducted to determine which items cohered to form unidimensional constructs. The R-B5FM data were hypothesized to form the five personality constructs of Extraversion-Introversion (EI), Agreeableness (AGR), Conscientiousness (CON), Emotional Stability (EMO), and Intellect/Imagination (INT).

For the first step of the Rasch Factor Analysis, all items were input at the same time to allow them to interact with each other (Table 4). The Rasch model explained 29.2% of the variance (eigenvalue = 15.2); the first contrast explained 10.0% of the variance (eigenvalue = 5.2), indicating that the data were not unidimensional. Seven items from the Extraversion-Introversion (EI) construct had high positive loadings between .37 and .62, and four items from the Agreeableness (AGR) construct had high positive loadings between .27 and .66. However, one EI item (EI 7) had a low negative loading (-.01) and three AGR items had low positive loadings (.17, .18, and .19); additionally, the EI and AGR items were not easily distinguishable from each other. Six items from Emotional Stability (EMO) had high negative loadings between -.35 and -.52; however, one EMO item (EMO 4) had a lower negative loading (-.23) and an item from the Conscientiousness construct (CON 3) was mixed with the EMO items.

In order to clarify the factor structure, only the 16 items from the hypothesized Extraversion-Introversion and Agreeableness constructs were analyzed to determine whether the Extraversion-Introversion and Agreeableness items would separate into two distinct constructs (Table 5). These items were selected because of their high positive residual loadings in the initial analysis. The Rasch model explained 51.1% (eigenvalue = 18.8), and the first contrast explained 7.0% of the variance (eigenvalue = 2.4). One Agreeableness item (AGR 5, “Am on

Table 4

*Rasch Principal Components Analysis for the Revised Big Five Factor Markers
Questionnaire (37 Items)*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
AGR 5	.66	48.83	.92	.91
EI 8	.62	50.44	.70	.71
EI 3	.59	49.96	.87	.87
EI 1	.58	53.75	.89	.88
EI 2	.55	44.19	.92	.92
AGR 8	.52	49.44	.53	.53
EI 6	.51	47.14	.98	.98
EI 5	.42	53.09	.96	.95
EI 4	.37	51.71	.85	.86
AGR 6	.35	51.63	.59	.59
AGR 7	.27	48.71	.88	.89
INT 2	.25	49.44	.87	.87
INT 6	.19	47.79	.98	.99
AGR 2	.19	46.18	.82	.82
AGR 4	.18	47.92	.75	.77
AGR 3	.17	48.64	.95	.97
INT 1	.15	47.32	.98	.98
INT 7	.05	50.64	.69	.71
INT 3	.05	53.61	1.03	1.04
INT 4	.03	49.74	.84	.85
EMO 2	-.52	49.70	1.45	1.47
EMO 3	-.51	44.51	1.22	1.21
EMO 6	-.51	51.15	1.32	1.35
EMO 1	-.48	48.28	1.37	1.37
CON 3	-.45	51.11	1.11	1.14
EMO 7	-.38	52.34	1.43	1.50
EMO 5	-.35	46.82	1.26	1.30
CON 5	-.35	52.82	1.10	1.10
CON 1	-.34	51.63	.85	.86
CON 4	-.32	52.15	.95	.96
CON 2	-.29	51.09	1.12	1.13
CON 6	-.26	51.36	1.30	1.32
AGR 1	-.23	48.10	.85	.88
EMO 4	-.23	45.06	1.26	1.29
CON 7	-.22	51.74	.87	.88
INT 5	-.15	57.27	1.61	1.68
EI 7	-.01	54.66	1.16	1.18

Note. EI = Extraversion-Introversion; AGR = Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination. Measures are Rasch CHIPS.

good terms with nearly everyone”) had a high positive loading (.66), and four Extraversion-Introversion items (EI 8, EI 3, EI 6, and EI 1) had positive loadings above .40. Item AGR 5 and the four Extraversion-Introversion items appeared to measure “interpersonal skills,” which includes initiating conversations, talking often, dealing skillfully with others, and being on good terms with people in general.

Table 5
*Rasch Principal Components Analysis for the Items Measuring
 Extraversion-Introversion and Agreeableness*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
AGR 5	.66	48.85	.87	.86
EI 8	.50	51.10	.69	.70
EI 3	.47	50.41	.89	.88
EI 6	.42	46.51	1.03	1.04
EI 1	.41	55.70	.94	.93
AGR 8	.29	49.70	.60	.61
EI 2	.18	42.48	.89	.87
AGR 6	.05	52.76	.75	.77
AGR 7	.00	48.69	1.18	1.24
AGR 1	-.54	47.83	1.36	1.49
AGR 2	-.52	45.18	.98	.98
AGR 3	-.52	48.59	1.17	1.20
EI 7	-.35	56.96	1.55	1.66
EI 4	-.23	52.87	.99	1.06
AGR 4	-.22	47.58	1.01	1.05
EI 5	-.10	54.79	1.07	1.08

Note. EI = Extraversion-Introversion; AGR = Agreeableness. Measures are Rasch CHIPs.

Three Agreeableness items (AGR 1, AGR 2, and AGR 3) had high negative loadings above .40; none of the Extraversion-Introversion items did. These three

items appeared to measure what might be termed “Good Samaritan-ship,” given that all the items were related to asking about the problems of people and becoming actively involved in their emotional well-being. Because four of the Extraversion-Introversion items cohered to form a construct, but only three Agreeableness items cohered, the Agreeableness construct could not be clearly distinguished from the Extraversion-Introversion construct. To further examine each construct, the items designed to measure Agreeableness and Extraversion-Introversion were analyzed separately to examine the construct structure, Likert scale category functioning, and item fit.

Extraversion-Introversion. The eight items hypothesized to measure Extraversion-Introversion were analyzed with the Rasch model and the Rasch PCA of item residuals was examined (Table 6). The Rasch model explained 68.4% of the variance (eigenvalue = 17.3), and the first contrast explained 8.0% of the variance (eigenvalue = 2.0). The criteria for determining unidimensionality was 50% of the variance explained by the Rasch model and a first contrast with either 10% or less variance explained or an eigenvalue of less than 3.0. Given the amount of variance explained by the Rasch model versus the variance explained by the first contrast, I concluded that the items measuring Extraversion-Introversion formed a reasonably

Table 6

*Rasch Principal Components Analysis for the Items Measuring
Extraversion-Introversion*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
EI 5	.77	53.78	.99	.99
EI 4	.74	51.64	.92	.94
EI 1	.17	54.79	.86	.84
EI 7	.07	56.20	1.70	1.84
EI 3	-.52	48.93	.90	.90
EI 2	-.48	40.30	.92	.92
EI 6	-.43	44.66	.94	.96
EI 8	-.41	49.70	.73	.75

Note. EI = Extraversion-Introversion. Measures are Rasch CHIPs.

unidimensional construct. Items EI 5 and EI 4 loaded positively above .40, while items EI 3, EI 2, EI 6, and EI 8 loaded negatively above .40. The positively loading items concerned attracting attention or being the center of attention, while the negatively loading items concerned dealing with or talking with people. Two items (EI 1, “Am the life of the party,” and EI 7, “Feel at ease in a large group of people”) failed to load above .40. Item EI 1 might have been perceived as inappropriate by the students, as the concept of “party” in the North American dormitory-based university context for which the original questionnaire was designed does not exist in Japan. Item EI 7 might have been perceived as vague, because the identity of the “group of people” was not made clear. Japanese social relations are both hierarchical as well as in-group oriented; the participants might have reacted differently depending on their interpretation of whom the group consisted (e.g., university club members or strangers in public).

Because the previous analysis of the Extraversion-Introversion and Agreeableness items indicated that item AGR 5 (“Am on good terms with nearly everyone”) had a high positive loading and cohered with Extraversion-Introversion items, a separate PCA was conducted including all eight items hypothesized to measure Extraversion-Introversion plus item AGR 5 (Table 7). The Rasch model explained 68.8% of the variance (eigenvalue = 19.8); the first contrast explained 7.6% of the variance (eigenvalue = 2.2). Items EI 5 and EI 4 had high positive loadings, and items AGR 5 and EI 3 had high negative loadings. Items EI 1 and EI 7 had slightly higher positive loadings than the first analysis, but still failed to load above .40. On the other hand, items EI 8, EI 2, and EI 6, which had loaded negatively above -.40 in the first analysis all failed to load above -.40 when item AGR 5 was included in the analysis. Thus, item AGR 5 appears to exert undue influence on the Extraversion-Introversion items that are concerned with talking to people.

As a further test of whether to include item AGR 5 as an item contributing to the Extraversion-Introversion construct, person measures were obtained for two sets of items measuring the construct and a disattenuated correlation analysis was performed. *Disattenuated* refers to removing the item residual errors on person measures from item loadings on the factor contrast, after which the person measures from positively loading items are then correlated with measures from

Table 7

*Rasch Principal Components Analysis for the Items Measuring
Extraversion-Introversion Including Item AGR 5*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
EI 5	.75	53.78	1.07	1.07
EI 4	.75	51.64	1.03	1.03
EI 1	.22	54.79	.88	.86
EI 7	.16	56.20	1.76	1.94
AGR 5	-.61	47.51	.83	.81
EI 3	-.50	48.93	.86	.86
EI 8	-.39	49.70	.70	.72
EI 2	-.34	40.30	.92	.93
EI 6	-.29	44.66	.95	.97

Note. EI = Extraversion-Introversion. Measures are Rasch CHIPs.

negatively loading items to determine whether the item responses indicate the same latent construct (Smith, E., 2002; Smith, R., 2000). A high correlation of the person measures ($> .7$) suggests construct validity (Smith, R., 2000).

First, a Rasch analysis was conducted with items AGR 5, EI 3, EI 8, EI 2 and EI 6 and the resulting person measures were exported to SPSS. Next, a separate Rasch analysis was conducted with items EI 3, EI 8, EI 2, and EI 6 (i.e., the negatively loading items without AGR 5). Person measures were exported to SPSS and a bivariate correlational analysis was conducted with both sets of person measures. The disattenuated correlation was strong ($r = .99, p < .01$), indicating that both sets of items were measuring the same construct. Thus, while strongly related to the Extraversion-Introversion items, item AGR 5 did not add anything of significance to the measurement of the Extraversion-Introversion construct and was

therefore not included with the Extraversion-Introversion items in subsequent analyses.

Agreeableness. After examining the items from the Extraversion-Introversion construct, all eight items from the hypothesized Agreeableness construct were input and the Rasch PCA of item residuals results were examined (Table 8). The Rasch model explained 48.5% variance (eigenvalue = 7.6), and the first contrast explained 12.9% (eigenvalue = 2.0) of the variance. The Rasch model explained 48.5% of the variance, the first contrast explained 12.9% of the variance, and the eigenvalue was less than 3.0. Given the amount of variance accounted for by the Rasch model versus the first contrast, and pending an examination of the item separation statistic, I concluded that the items measuring Agreeableness formed a weakly unidimensional construct. Items AGR 2, AGR 3, and AGR 1 loaded positively above .40, while items AGR 8, AGR 5, and AGR 6 loaded negatively above .40. The positively loading items were related to actively investigating the emotional states of others, while the negatively loading items were concerned with the ability to make others feel relaxed or comforted.

Two items (AGR 7, “Have a good word for everyone,” and AGR 4, “Take time out for others”) failed to load above .40. For both these items, the wording might have been too vague; students might have had difficulty interpreting whether

Table 8

Rasch Principal Components Analysis for the Items Measuring the Agreeableness Construct

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
AGR 2	.66	46.12	.97	.96
AGR 3	.58	49.94	1.19	1.22
AGR 1	.51	49.09	1.27	1.31
AGR 8	-.57	51.19	.65	.66
AGR 5	-.54	50.23	1.10	1.09
AGR 6	-.50	54.58	.70	.70
AGR 7	-.35	50.05	1.12	1.13
AGR 4	-.03	48.80	.97	.99

Note. AGR = Agreeableness. Measures are Rasch CHIPs.

“everyone” and “others” consisted of people from their recognized social groups (e.g., clubs, classes, families) or strangers outside their social groups. Item AGR 4 in particular loaded extremely weakly, possibly due to the wording of “take time,” which implies a deliberate course of action. This item bears a relationship to the North American (and Christian) concept of the good Samaritan who helps everyone regardless of social status or social relationship. Japanese are unlikely to help random strangers, because they would then feel a sense of reciprocal social obligation or debt (*on-giri*) that would create a strong personal connection or bond (*en*) that might not be warranted depending on the depth of personal familiarity with the other person (Davies & Ikeno, 2002; Hendry, 2003; Wierzbicka, 1997).

Because item AGR 5 had a high positive loading in the previous analysis of the Extraversion-Introversion items, a separate Rasch PCA of item residuals was conducted to determine whether removing the item would improve the

Table 9

Rasch Principal Components Analysis for the Items Measuring the Agreeableness Construct Excluding Item AGR 5

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
AGR 2	.65	46.00	.96	.94
AGR 3	.55	49.98	1.16	1.17
AGR 1	.44	49.09	1.23	1.25
AGR 8	-.65	51.28	.71	.72
AGR 6	-.54	54.79	.77	.78
AGR 7	-.49	50.08	1.16	1.17
AGR 4	-.16	48.78	.99	.99

Note. AGR = Agreeableness. Measures are Rasch CHIPs.

measurement of the Agreeableness construct (Table 9). In this analysis, the Rasch model explained 48.6% of the variance (eigenvalue = 6.6), and the first contrast explained 14.1% of the variance (eigenvalue = 1.9). The items with positive loadings were virtually unaffected, while the negatively loading items had slightly higher loadings, resulting in item AGR 7 reaching the .40 criterion. Rasch person measures were obtained for the construct both with and without item AGR 5. The measures had a strong disattenuated correlation ($r = .99$, $p < .01$), indicating that they measured the same construct. Thus, it was determined that removing item AGR 5 had no discernable effect on the measurement of Agreeableness.

After the Extraversion-Introversion and Agreeableness constructs were extracted, the remaining items measuring Emotional Stability, Conscientiousness, and Intellect/Imagination were analyzed again (Table 10). The Rasch model explained 29.5% of the variance (eigenvalue 8.8), and the first contrast explained

12.7% of the variance (eigenvalue = 3.8). The Emotional Stability items had high positive loadings and cohered to form one construct. A separate analysis was subsequently conducted with only the seven items measuring the hypothesized Emotional Stability construct.

Emotional Stability. All seven items from the hypothesized Emotional Stability construct were input and the Rasch PCA of item residual results were examined (Table 10). The Rasch model explained 59.1% variance (eigenvalue = 10.1), and the first residual contrast explained 10.6% of the variance (eigenvalue = 1.8), and the eigenvalue for the first residual contrast was 1.8, which was less than the 3.0 criterion. Given the amount of variance accounted for by the Rasch model and the small eigenvalue of the first contrast, I concluded that the items measuring Emotional Stability formed a fundamentally unidimensional construct.

Items EMO 5, EMO 6, and EMO 3 loaded positively above .40, while items EMO 2 and EMO 7 loaded negatively above .40. The positively loading items seemed related to anxiety or worry, while the negatively loading items seemed related to stronger, more aggressive emotions. Two items (EMO 4, “Have frequent mood swings,” and EMO 1, “Often feel blue”) failed to load above .40. Students might have interpreted these items as passived or depressed emotions unrelated to active or aggressive emotions such as worry or anger.

Table 10

Rasch Principal Components Analysis for the Items Measuring Emotional Stability, Conscientiousness, and Intellect/Imagination

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
EMO 1	.65	48.08	1.19	1.18
EMO 2	.59	49.46	1.18	1.18
EMO 5	.55	46.68	1.15	1.19
EMO 6	.55	50.85	1.14	1.16
EMO 7	.53	52.00	1.25	1.28
EMO 3	.51	44.44	1.05	1.02
EMO 4	.49	44.98	1.15	1.17
CON 5	-.47	52.46	.92	.91
CON 2	-.47	50.80	.95	.95
CON 7	-.47	51.41	.77	.77
INT 7	-.41	50.36	.69	.70
INT 4	-.41	49.50	.82	.84
CON 1	-.40	51.31	.69	.69
CON 6	-.36	51.05	1.16	1.17
INT 3	-.31	53.22	.97	.98
CON 4	-.28	51.81	.83	.84
CON 3	-.27	50.82	.90	.91
INT 6	-.20	47.62	1.03	1.04
INT 5	-.18	56.76	1.44	1.45
INT 2	-.13	49.21	.91	.91
INT 1	-.03	47.16	.98	.97

Note. EMO = Emotional Stability; CON = Conscientiousness. INT = Intellect/Imagination. Measures are Rasch CHIPs.

Table 11

Rasch Principal Components Analysis for the Items Measuring Emotional Stability

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
EMO 5	.65	48.24	1.02	1.06
EMO 6	.57	53.57	.99	1.01
EMO 3	.49	45.33	.96	.95
EMO 2	-.61	51.78	.96	1.00
EMO 7	-.59	55.06	1.08	1.10
EMO 4	-.25	46.02	1.15	1.19
EMO 1	-.20	50.00	.89	.92

Note. EMO = Emotional Stability. Measures are Rasch CHIPs.

Table 12

*Rasch Principal Components Analysis for the Items Measuring
Intellect/Imagination and Conscientiousness*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
INT 2	.70	47.88	1.02	1.04
INT 1	.64	45.43	1.16	1.16
INT 3	.43	52.72	1.03	1.02
INT 6	.38	45.97	1.15	1.16
INT 5	.32	56.93	1.60	1.62
INT 7	.28	49.29	.69	.71
INT 4	.21	48.23	.85	.87
CON 2	-.56	49.80	.97	.96
CON 5	-.51	51.81	.94	.93
CON 6	-.49	50.11	1.25	1.30
CON 3	-.47	49.83	1.03	1.10
CON 4	-.41	51.02	.95	.97
CON 1	-.36	50.42	.74	.74
CON 7	-.29	50.55	.78	.78

Note. INT = Intellect/Imagination; CON = Conscientiousness. Measures are in Rasch CHIPs.

Next, the remaining items hypothesized to measure Intellect/Imagination (INT) and Conscientiousness (CON) were analyzed (Table 12). The Rasch model explained 40.3% of the variance (eigenvalue = 9.4), and the first residual contrast explained 12.3% of the variance (eigenvalue = 2.9). All items from the Intellect/Imagination construct had positive loadings, and all items from the Conscientiousness construct had negative loadings, indicating that items from the two constructs formed two distinctive constructs. To confirm construct unidimensionality, the items hypothesized to measure the Intellect/Imagination construct and the items hypothesized to measure the Conscientiousness construct were analyzed separately using the Rasch PCA of item residuals.

Intellect/Imagination. The seven items from the Intellect/Imagination construct were input into Winsteps and the Rasch PCA of item residuals results were examined (Table 13). The Rasch model explained 53.7% of the variance (eigenvalue = 8.1), and the first residual contrast explained 13.3% of the variance (eigenvalue = 2.0). Although the first contrast variance was higher than the criterion value of 10%, given the amount of variance accounted for by the Rasch model and the small eigenvalue of the first contrast, I concluded that the items measuring the Intellect/Imagination formed a reasonably unidimensional construct. Items INT 1, and INT 2 loaded positively above .40, while items INT 5, INT 4, and INT 3 loaded negatively above .40. The positively loading items were related to abstract thinking or imagination, while the negatively loading items were related to intellectual qualities. One item (INT 7, “Can handle a lot of information”) loaded at .35,

Table 13

Rasch Principal Components Analysis for the Items Measuring Intellect/Imagination

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
INT 1	.80	45.20	.98	.96
INT 2	.78	48.11	.80	.80
INT 6	.20	45.83	1.15	1.16
INT 5	-.52	58.75	1.52	1.50
INT 4	-.42	48.52	.95	1.00
INT 3	-.40	53.82	.94	.92
INT 7	-.35	49.78	.76	.80

Note. INT = Intellect/Imagination. Measures are in Rasch CHIPs.

slightly below the .40 criterion value. The vague wording of the item might have confused the participants because it can be interpreted as referring to organization skills or to intellectual capacities. Item INT 6 (“Love to think of new ways to do things”) loaded weakly at .20.

Conscientiousness. The seven items hypothesized to measure Conscientiousness were input into Winsteps and the Rasch PCA of item residuals analysis was conducted (Table 14). The Rasch model explained 56.7% of the variance (eigenvalue = 9.2), and the first residual contrast explained 13.1% of the variance (eigenvalue = 2.1). Although the first contrast variance was higher than the criterion value of 10%, given the amount of variance accounted for by the Rasch model and the weak eigenvalue of the first contrast, I concluded that the items measuring Conscientiousness formed a fundamentally unidimensional construct. Items CON 6 and CON 2 loaded positively above .40, while items CON 4 and CON 5 loaded negatively above .40. The positively loading items were related to keeping physical objects in order, while the negatively loading items were related to planning. Two items (CON 1, “Am prepared,” and CON 7, “Continue my work until it’s perfect”) loaded at .36, slightly below the .40 criterion value. One item (CON 3, “Like rules”) loaded weakly at -.15. For each item, the wording might have been too vague. Students might have wondered “am prepared...for what?”,

“continue...which work?”, or “like rules...about what?” when responding to the items.

Table 14

Rasch Principal Components Analysis for the Items Measuring Conscientiousness

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CON 6	.85	49.47	1.25	1.26
CON 2	.83	49.09	.91	.90
CON 4	-.47	50.67	1.02	1.04
CON 5	-.46	51.72	.93	.92
CON 1	-.36	49.88	.86	.88
CON 7	-.36	50.05	.94	.99
CON 3	-.15	49.11	1.05	1.09

Note. CON = Conscientiousness. Measures are in Rasch CHIPs.

Rasch analysis of the five constructs comprising the R-B5FM. This section presents Rasch analysis results for each of the five constructs hypothesized to comprise the Big Five (Extraversion-Introversion, Agreeableness, Emotional Stability, Intellect/Imagination, and Conscientiousness). Each construct was analyzed separately for Likert-scale category functioning, and categories were combined when necessary. Items comprising each construct were also analyzed for item difficulty and fit to the construct. Item-person maps were also generated for each construct to provide a visual representation of the location of persons and items on the construct and to view the empirical item hierarchy.

Extraversion-Introversion. The Likert scale category functioning was examined for the eight items measuring Extraversion-Introversion (Table 15). The minimum of 10 observations per category was met, as the smallest number of observations was 833 (category 1). The outfit MNSQ statistic for all categories was well below the 2.0 criterion. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a 6-point scale. The smallest gap between the fourth and fifth thresholds ($\tau_4 = 10.39$, $\tau_5 = 13.37$) was 2.98 CHIPs, slightly above the criterion, and there were no disordered thresholds.

Rasch item statistics were obtained for the Extraversion-Introversion construct (Table 16). Item EI 7 had an infit MNSQ of 1.70 and outfit MNSQ of 1.84, indicating misfit to the model. The other items showed good fit and ranged in measure difficulty from 40.30 (EI 2) to 56.20 (EI 7). Point-measure correlations, which demonstrate the relative strength of the item in relation to other items measuring the construct, were high (.69-.74). Rasch person reliability (separation) was .81 (2.06). The separation figure is barely above the minimum recommended level of 2.0, indicating the similarity of person ability estimates for the participants. The Rasch item reliability (separation) was 1.00 (15.49), indicating that the item difficulty estimates were well separated in relation to their standard errors.

Because one item (EI 7) misfit the construct, a separate Rasch item analysis was conducted without this item. The results are displayed in Table 17. Infit and

Table 15

Category Structure Functioning for Extraversion-Introversion

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	833	(10)	1.15	1.12	None	(-30.36)
2 Disagree	1513	(18)	.97	.96	-16.83	-14.65
3 Somewhat disagree	2261	(26)	.87	.92	-8.88	-3.72
4 Somewhat agree	1911	(22)	.84	.85	1.95	5.24
5 Agree	1154	(13)	1.05	1.09	10.39	14.42
6 Strongly agree	902	(10)	1.05	1.16	13.37	(28.05)

Note. Structure measure and category measure scores are reported in Rasch CHIPs.

Table 16

Rasch Statistics for the Items Measuring Extraversion-Introversion

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
EI 7	56.20	.32	1.70	9.9	1.84	9.9	.44
EI 5	53.78	.31	.99	-.1	.99	-.2	.70
EI 6	44.66	.31	.94	-1.4	.96	-1.0	.72
EI 4	51.64	.31	.92	-1.8	.94	-1.5	.69
EI 2	40.30	.32	.92	-1.9	.92	-1.7	.69
EI 3	48.93	.30	.90	-2.4	.90	-2.4	.71
EI 1	54.79	.31	.86	-3.4	.84	-3.9	.74
EI 8	49.70	.30	.73	-7.1	.75	-6.4	.73

Note. EI = Extraversion-Introversion; PMC = Point-measure correlation. Measures are Rasch CHIPs.

outfit MNSQ statistics changed slightly for all items. Whereas prior to removing EI 7, infit and outfit standardized z -scores were all negative, after removing EI 7 z -scores for the items spread more evenly from positive to negative. Point-measure correlations for all items increased slightly. However, person measures from the construct with the item EI 7 were correlated strongly with person measures from the construct with the item removed ($r = .98, p < .01$), demonstrating that the items

Table 17

Rasch Statistics for the Items Measuring Extraversion-Introversion Excluding Item EI 7

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
EI 5	55.40	.33	1.12	2.8	1.11	2.4	.71
EI 4	52.93	.33	1.05	1.2	1.03	1.3	.70
EI 2	39.80	.34	1.04	.9	1.05	1.2	.74
EI 6	44.85	.33	1.02	.4	1.03	.6	.72
EI 3	49.79	.33	1.01	.2	1.01	.3	.72
EI 1	56.57	.34	.92	-1.9	.89	-2.7	.76
EI 8	50.68	.33	.84	-4.1	.87	-3.2	.74

Note. EI = Extraversion-Introversion; PMC = point-measure correlation. Measures are Rasch CHIPs.

including EI 7 measured the same construct as the items excluding EI 7. Without EI 7, the Rasch item reliability was 1.00 (separation 16.26), and the Rasch person reliability was .81 (2.20). Although item and person separation estimates were slightly higher than that of the items including EI 7, reliability estimates showed no changes. Thus, despite the slight improvement in the Rasch item statistics, there were no discernable differences due to removing EI 7 and therefore the item was retained for further analysis.

The item-person map for the Extraversion-Introversion construct (Figure 6) shows the CHIP scale for both persons and items on the left side. The scale ranges from 10 to 90, with an item mean of 50. Participants are represented on the left side of the item-person map as number signs (#) and periods (.), with each # standing for 10 persons and periods standing for one person. The items measuring

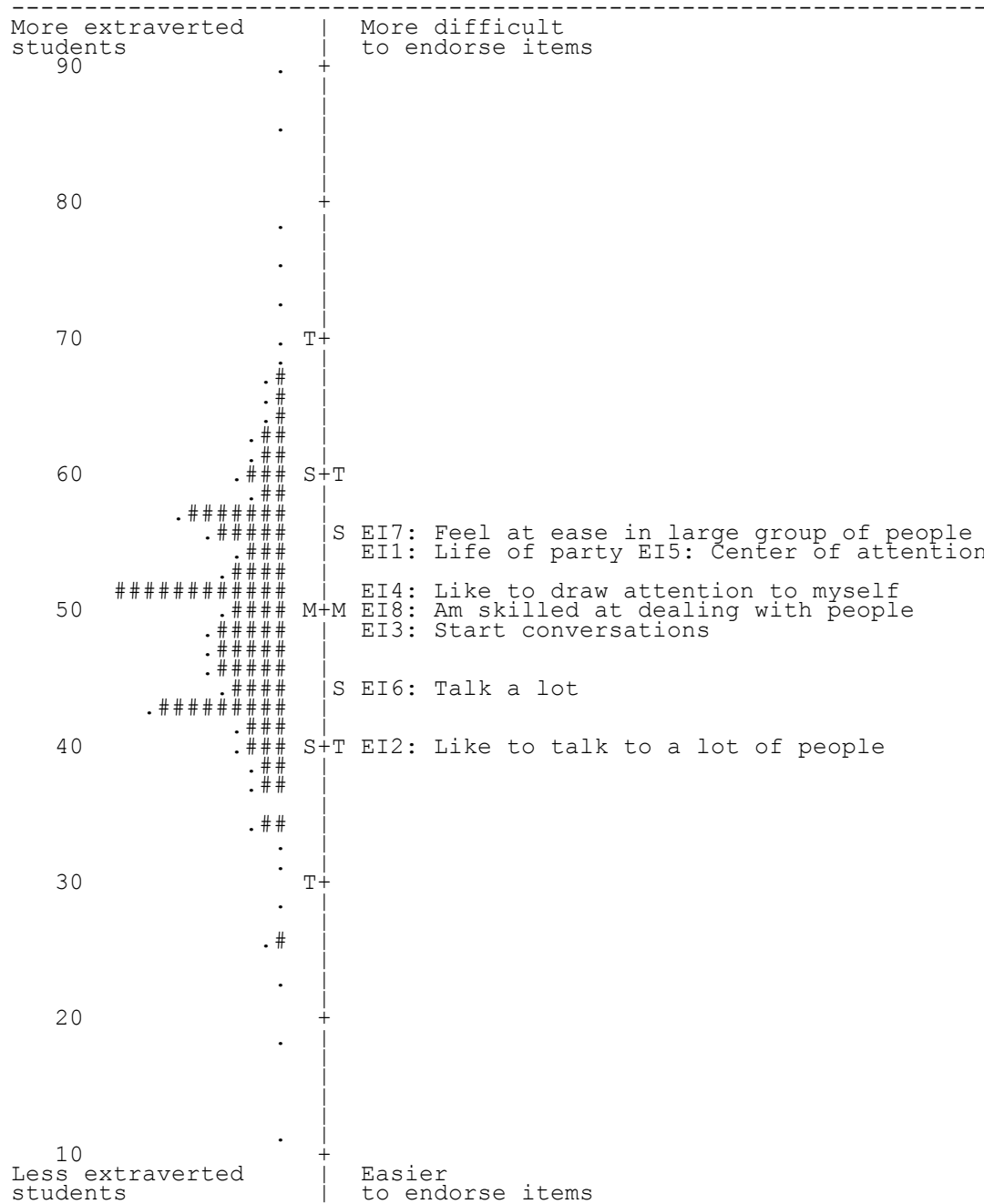


Figure 6. The item-person map of the Extraversion-Introversion construct. Each # represents 10 persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

Extraversion-Introversion are displayed on the right side of the vertical line.

Persons are placed along the line according to their ability estimates, and items are

placed according to their endorsement difficulty level. A person at the mean (M) of person ability on the left side has a 50% chance of endorsing an item of the same difficulty level (M) on the right side of the vertical line (Bond & Fox, 2007, p. 280). Items located above the item mean are more difficult to endorse, so that a person located at the person mean has less probability of endorsing the item. The person ability mean ($M = 50.01$, $SD = 10.28$) and mean item endorsability difficulty ($M = 50.00$, $SD = 5.03$) showed that the items on the whole were well targeted for the sample. The most difficult item to endorse was EI 7 (“Feel at ease in a large group of people,” difficulty measure = 56.20); however, this item also exhibited overfit to the model, an indication that the item was not able to discriminate between persons of varying levels of the construct. The easiest item to endorse was EI 2 (“Like to talk to a lot of people,” difficulty measure = 40.30).

Agreeableness. Likert scale category functioning was examined for the eight items of the Agreeableness construct (Table 18). The minimum of 10 observations per category was met, as the smallest number of observations was 419 (category 1). Outfit MNSQ for all categories was well below the required 2.0. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a 6-point

Table 18

Category Structure Functioning for Agreeableness

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	419	(5)	1.16	1.19	None	(-29.56)
2 Disagree	960	(11)	.98	1.00	-15.24	-15.12
3 Somewhat disagree	2251	(26)	.86	.86	-11.01	-4.88
4 Somewhat agree	2546	(30)	.84	.87	.15	4.65
5 Agree	1570	(18)	.92	.92	10.02	15.08
6 Strongly agree	820	(10)	1.16	1.18	16.07	(30.00)

Note. Measures are Rasch CHIPs.

scale. The smallest gap between the first and second thresholds ($\tau_1 = -15.24$, $\tau_2 = -11.01$) was 4.23 CHIPs, well above the criterion, and there were no disordered thresholds.

Rasch statistics were obtained for items measuring Agreeableness (Table 19). All items showed good fit to the model and ranged in measure difficulty from 46.12 (AGR 2) to 54.58 (AGR 6). Point-measure correlations were moderately high for seven of the eight items (.59-.70); however, the point-measure correlation for item AGR 1 was relatively low (.46), a possible indication that the item was not contributing as much as other items to the construct. The Rasch item reliability estimate (separation) was .98 (6.87), indicating that the item difficulty estimates were well separated in relation to their standard errors. The Rasch person reliability estimate (separation) was .74 (1.67). The reliability figure is not high, but still adequate; however, the separation figure is below the minimum recommended level of 2.0, indicating the similarity of the person ability estimates for this construct.

Table 19

Rasch Statistics for the Items Measuring Agreeableness

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
AGR 1	49.09	.31	1.27	6.1	1.31	6.8	.46
AGR 3	49.94	.31	1.19	4.4	1.22	4.9	.60
AGR 7	50.05	.31	1.12	2.8	1.13	3.0	.60
AGR 5	50.23	.31	1.10	2.4	1.09	2.2	.66
AGR 4	48.80	.31	.97	-.8	.99	-.2	.59
AGR 2	46.12	.31	.97	-.6	.96	-1.0	.65
AGR 6	54.58	.31	.70	-8.3	.70	-8.0	.68
AGR 8	51.19	.31	.65	-9.7	.66	-9.3	.70

Note. AGR = Agreeableness; PMC = point-measure correlation. Measures are Rasch CHIPs.

A comparison of the person ability mean ($M = 52.43$, $SD = 8.64$) and item endorsability difficulty ($M = 50.00$, $SD = 2.23$) showed that the items on the whole were slightly below the average ability of the sample, with all items tightly grouped around the mean (Figure 7). The most difficult item to endorse was AGR 6 (“Make people feel at ease,” difficulty measure = 54.58). The easiest item to endorse was AGR 2 (“Feel concern for others,” difficulty measure = 46.12).

Emotional Stability. Likert scale category functioning was examined for the seven items of the Emotional Stability construct (Table 20). The minimum of 10 observations per category was met, as the smallest number of observations was 582 (category 1). Outfit for all categories was below the required 2.0. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a

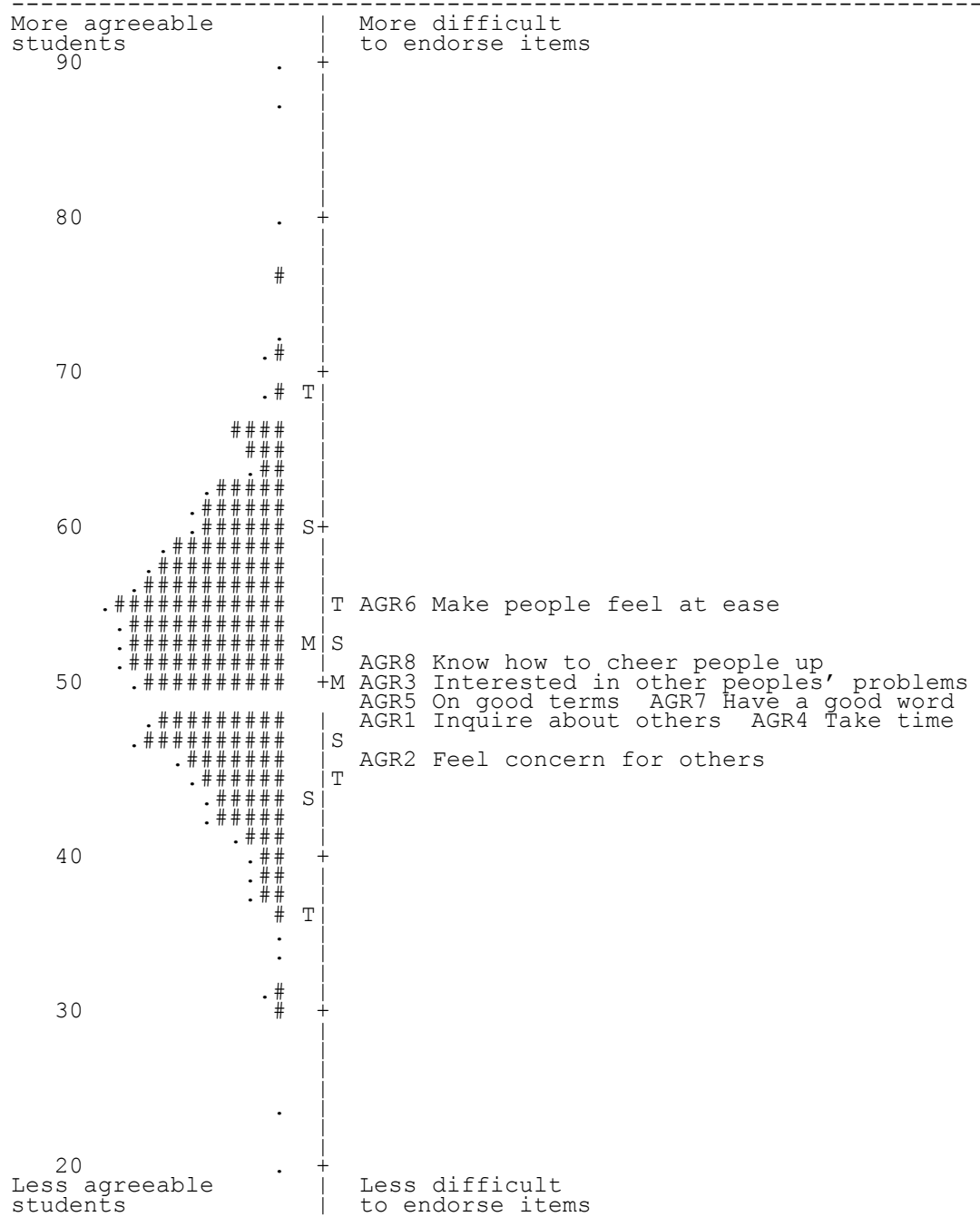


Figure 7. The item-person map of the Agreeableness construct. Each # represents six persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

6-point scale. The smallest gap between the second and third thresholds ($\tau_2 = -5.77$, $\tau_3 = -.14$) was 5.63 CHIPS, above the criterion, and there were no disordered thresholds.

Rasch item fit statistics for the Emotional Stability construct showed good fit for all items (Table 21). Item difficulty ranged from 45.33 (EMO 3) to 55.06 (EMO 7). Point-measure correlations were moderately high for all items (.52-.67). The Rasch item reliability estimate (separation) was .99 (12.03), indicating that the item difficulty estimates were well separated in relation to their standard errors. The Rasch person reliability estimate (separation) was .71 (1.56). The reliability figure is adequate; however, the separation figure is below the minimum recommended level of 2.0, indicating the similarity of person ability estimates for this construct.

A comparison of the person ability mean ($M = 52.69$, $SD = 9.09$) and item endorsability difficulty ($M = 50.00$, $SD = 3.43$) showed that the items on the whole were somewhat easily endorsable by many of the participants (Figure 8). The persons formed a double-peaked distribution curve, with each peak at the mid points between the mean and first standard deviation. One peak is at the same difficulty level as item EMO 5 (“Panic easily”), while the second peak is at the same difficulty level as EMO 7 (“Get angry easily”). Thus, despite the low person separation (1.56), based on the item-person map, the participants are divisible into two groups: those who are susceptible to vulnerable emotions (i.e., panic, worry)

Table 20

Category Structure Functioning for Emotional Stability

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	582	(8)	1.10	1.15	None	(-25.87)
2 Disagree	1050	(14)	.96	.98	-11.87	-11.80
3 Somewhat disagree	1483	(20)	.90	.90	-5.77	-3.54
4 Somewhat agree	1662	(22)	.90	.97	-.14	3.24
5 Agree	1582	(21)	.97	.98	5.02	11.78
6 Strongly agree	1075	(14)	1.10	1.13	12.76	(26.42)

Note. Measures are Rasch CHIPs.

Table 21

Rasch Statistics for the Items Measuring Emotional Stability

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
EMO 4	46.02	.29	1.15	3.4	1.19	4.2	.52
EMO 7	55.06	.28	1.08	1.9	1.10	2.2	.62
EMO 5	48.24	.28	1.02	.5	1.06	1.4	.58
EMO 6	53.57	.27	.99	-.3	1.01	.3	.61
EMO 2	51.78	.27	.96	-1.0	1.00	.0	.65
EMO 3	45.33	.29	.96	-1.0	.95	-1.1	.62
EMO 1	50.00	.27	.89	-2.7	.92	-2.0	.67

Notes. EMO = Emotional Stability; PMC = point-measure correlation. Measures are Rasch CHIPs.

(Ashton et al., 2006) versus those who are susceptible to strong aggressive emotions (i.e., anger). The most difficult item to endorse was EMO 7 (“Get angry easily,” difficulty measure = 55.06). The easiest item to endorse was EMO 3 (“Worry about things,” difficulty measure = 45.33).

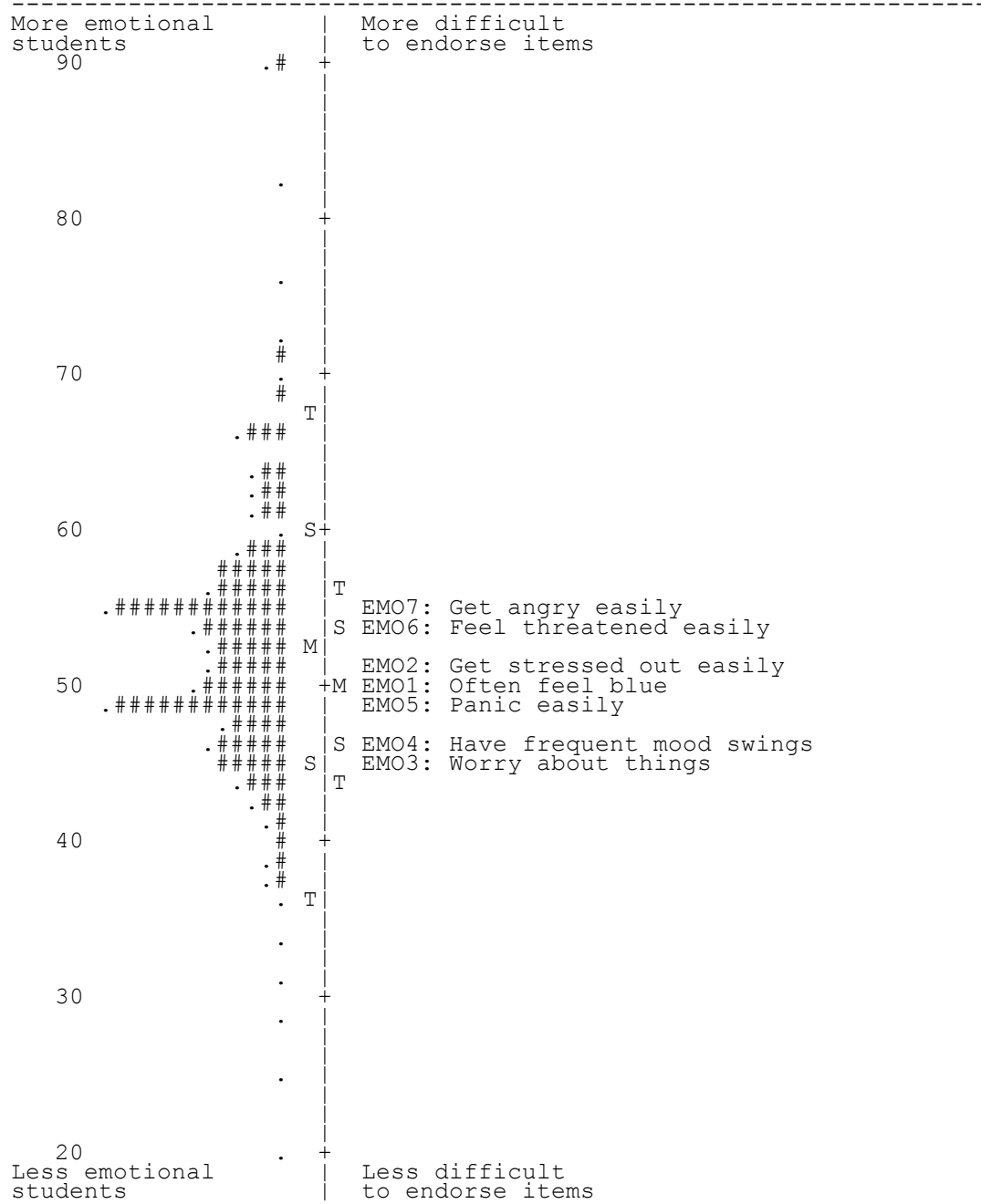


Figure 8. The item-person map of the Emotional Stability construct. Each # represents ten persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

Intellect/Imagination. Likert scale category functioning was examined for the seven items of the Intellect/Imagination construct (Table 22). The minimum of 10 observations per category was met, as the smallest number of observations was 631 (category 6). Outfit for all categories was well below the required 2.0. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a 6-point scale. The smallest gap between the fourth and fifth thresholds ($\tau_4 = 8.74$, $\tau_5 = 13.10$) was 4.36 CHIPs, well above the criterion, and there were no disordered thresholds.

Rasch item fit statistics for the Intellect/Imagination construct showed good fit for six of the seven items (Table 23). Item measure difficulty ranged from 45.20 (INT 1) to 58.75 (INT 5). Point-measure correlations were moderately high for all items (.54-.68). The infit mean squared of one item (INT 5, “Love to read challenging materials”) barely misfit the model (Infit MNSQ = 1.52; Outfit MNSQ = 1.50). The Rasch item reliability (separation) was 1.00 (14.36), indicating that the item difficulty estimates were well separated in relation to their error estimates. The Rasch person reliability estimate (separation) was .66 (1.41). The reliability figure is somewhat low, and the separation figure is well below the minimum recommended level of 2.0, indicating the similarity of person ability estimates for the construct.

Table 22

Category Structure Functioning for Intellect/Imagination

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	713	(9)	1.09	1.10	None	(-28.20)
2 Disagree	1277	(17)	.96	.98	-14.21	-13.54
3 Somewhat disagree	1979	(26)	.86	.87	-8.40	-3.85
4 Somewhat agree	1809	(24)	.92	.97	.77	4.35
5 Agree	1085	(14)	.88	.89	8.74	13.47
6 Strongly agree	631	(8)	1.21	1.22	13.10	(27.45)

Note. Measures are in Rasch CHIPs.

A separate analysis with item INT 5 temporarily removed revealed no significant differences in item fit (Table 24). When person measures from the construct measured by all seven items were correlated with person measures from the construct measured with the item INT 5 removed, the disattenuated correlation was very high ($r = .97, p < .01$), indicating that the items were measuring the same construct. Without item INT 5, the Rasch person reliability estimate (separation) was .66 (1.38) and the Rasch item reliability estimate (separation) was .99 (9.99). These statistics were virtually unchanged, indicating that removing INT 5 did not significantly affect the measurement of the construct.

A comparison of the person ability mean ($M = 49.51, SD = 7.98$) and item endorsability difficulty ($M = 50.00, SD = 4.43$) showed that the items on the whole were targeted well for the sample; however, four items (INT 4, INT 2, INT 6, and INT 1) fell within the first standard deviation below the mean, indicating that the items were slightly easy for many of the participants to endorse (Figure 9).

Table 23

Rasch Statistics for the Items Measuring Intellect/Imagination

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
INT 5	58.75	.31	1.52	9.9	1.50	9.9	.54
INT 6	45.83	.29	1.15	3.6	1.16	3.8	.54
INT 4	48.52	.29	.95	-1.2	1.00	-.1	.54
INT 1	45.20	.29	.98	-.5	.96	-1.0	.65
INT 3	53.82	.29	.94	-1.5	.92	-2.1	.62
INT 7	49.77	.29	.76	-6.4	.80	-5.2	.60
INT 2	48.11	.29	.80	-5.4	.80	-5.3	.68

Note. INT = Intellect/Imagination; PMC = point-measure correlation. Measures are Rasch CHIPs.

Table 24

Rasch Statistics for the Items Measuring Intellect/Imagination Excluding Item INT 5

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
INT 6	46.98	.30	1.19	4.5	1.19	4.5	.57
INT 3	55.91	.31	1.14	3.2	1.11	2.6	.59
INT 4	49.96	.30	1.04	.9	1.06	1.5	.56
INT 1	46.29	.30	.97	-.8	.95	-1.2	.69
INT 7	51.36	.31	.85	-3.9	.87	-3.2	.60
INT 2	49.50	.30	.79	-5.5	.80	-5.3	.71

Note. INT = Intellect/Imagination; PMC = point-measure correlation. Measures are Rasch CHIPs.

However, there are two substantial gaps in item difficulty level between the mean and the second standard deviation, indicating that participants with more ability on the construct might not be precisely measured by the existing item hierarchy. An examination of the person measures revealed that three persons at the top of the person ability hierarchy and seven persons at the bottom of the person ability hierarchy had more than double the standard error mean of 3.73, for a total of 10

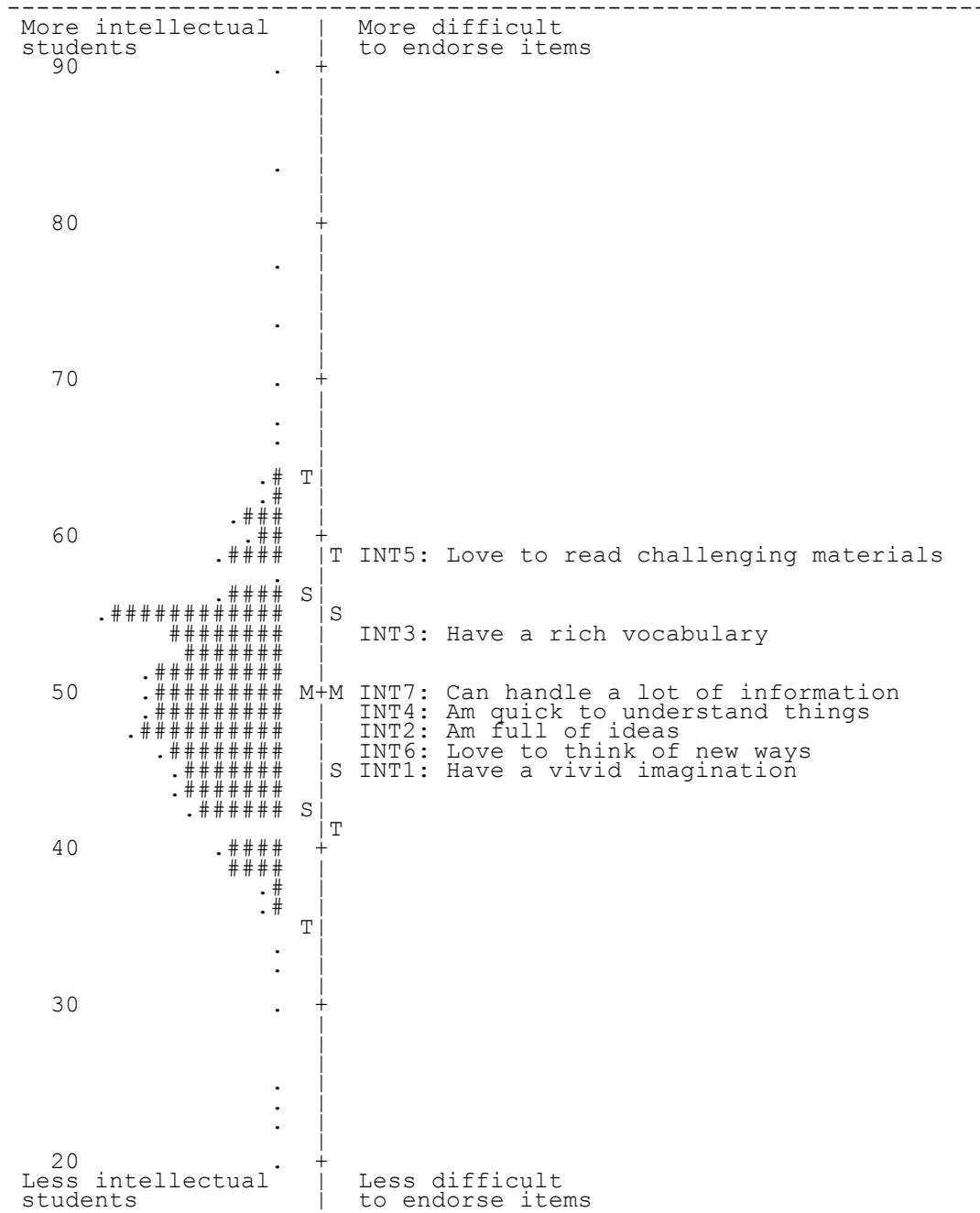


Figure 9. The item-person map of the Intellect/Imagination construct. Each # represents eight persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

persons or .01% of the total *N*-size. Thus, despite the graphic depiction in Figure 9, the construct was well targeted for the sample. The most difficult item to endorse was INT 5 (“Love to read challenging materials,” difficulty measure = 58.75). The easiest item to endorse was INT 1 (“Have a vivid imagination,” difficulty measure = 45.20).

Conscientiousness. Likert scale category functioning was examined for the Conscientiousness construct (Table 25). The minimum of 10 observations per category was met, as the smallest number of observations was 468 (category 6). Outfit for all categories was well below the required 2.0. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a 6-point scale. The smallest gap between the third and fourth thresholds ($\tau_3 = 1.51$, $\tau_4 = 8.32$) was 6.81 CHIPs, well above the criterion, and there were no disordered thresholds.

Table 25

Category Structure Functioning for Conscientiousness

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	722	(10)	1.07	1.06	None	(-30.40)
2 Disagree	1430	(19)	.88	.89	-16.87	-14.76
3 Somewhat disagree	2127	(28)	.88	.89	-8.90	-4.11
4 Somewhat agree	1688	(23)	.89	.90	1.51	4.75
5 Agree	1022	(14)	.95	.97	8.32	14.56
6 Strongly agree	468	(6)	1.28	1.36	15.94	(29.59)

Note. Measures are in Rasch CHIPs.

Rasch item fit statistics for the Conscientiousness construct showed good fit for all items (Table 26). Item measure difficulty ranged from 49.09 (CON 2) to 51.72 (CON 5). Point-measure correlations were high for all items (.63-.71). The Rasch item reliability (separation) was good at .87 (2.60), but some item difficulty estimates were similar, and this slightly suppressed the reliability estimate. The Rasch person reliability estimate (separation) was .76 (1.79). The reliability estimate is adequate, but the separation figure is slightly below the minimum recommended level of 2.0, indicating the similarity of person ability estimates for the construct.

Table 26

Rasch Statistics for the Items Measuring Conscientiousness

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
CON 6	49.47	.31	1.25	5.6	1.26	5.9	.65
CON 3	49.11	.30	1.05	1.2	1.09	2.1	.64
CON 4	50.67	.31	1.02	.5	1.04	.9	.63
CON 7	50.05	.31	.94	-1.5	.99	-.3	.65
CON 5	51.72	.31	.93	-1.7	.92	-2.0	.69
CON 2	49.09	.30	.91	-2.1	.90	-2.4	.71
CON 1	49.88	.31	.86	-3.5	.88	-2.9	.67

Note. CON = Conscientiousness; PMC = point-measure correlation. Measures are Rasch CHIPs.

A comparison of the person ability mean ($M = 47.82$, $SD = 10.27$) and item endorsability difficulty ($M = 50.00$, $SD = .87$) showed that the item difficulty estimates were slightly higher than the person ability estimates (Figure 10). The

items were tightly grouped around the item difficulty mean. The most difficult item to endorse was CON 5 (“Do things according to a plan,” difficulty measure = 51.72). The easiest item to endorse was CON 2 (“Put things back in their proper places,” difficulty measure = 49.09).

Summary of Rasch results for the R-B5FM constructs. According to the Rasch PCA results, the Rasch model explained more than 50% of the variance for four of the five constructs comprising the Revised Big Five Factor Markers questionnaire (Table 27). This figure is greater than Linacre’s 2007 criterion for unidimensionality. The Rasch model explained less than 50% of the variance for the Agreeableness construct (variance explained = 48.8%). However, Agreeableness had a person separation index of 1.67, which might have suppressed the variance due to a lack of clear separation of the participant responses into two distinct groups. One item from the Agreeableness construct (AGR 5, “Am on good terms with nearly everyone”) had a high positive loading with the Extraversion-Introversion items; however, subsequent analyses failed to demonstrate any significant contribution from the AGR 5 item with items measuring the Extraversion-Introversion construct. There was also no significant improvement to the items measuring the Agreeableness construct after excluding item AGR 5.

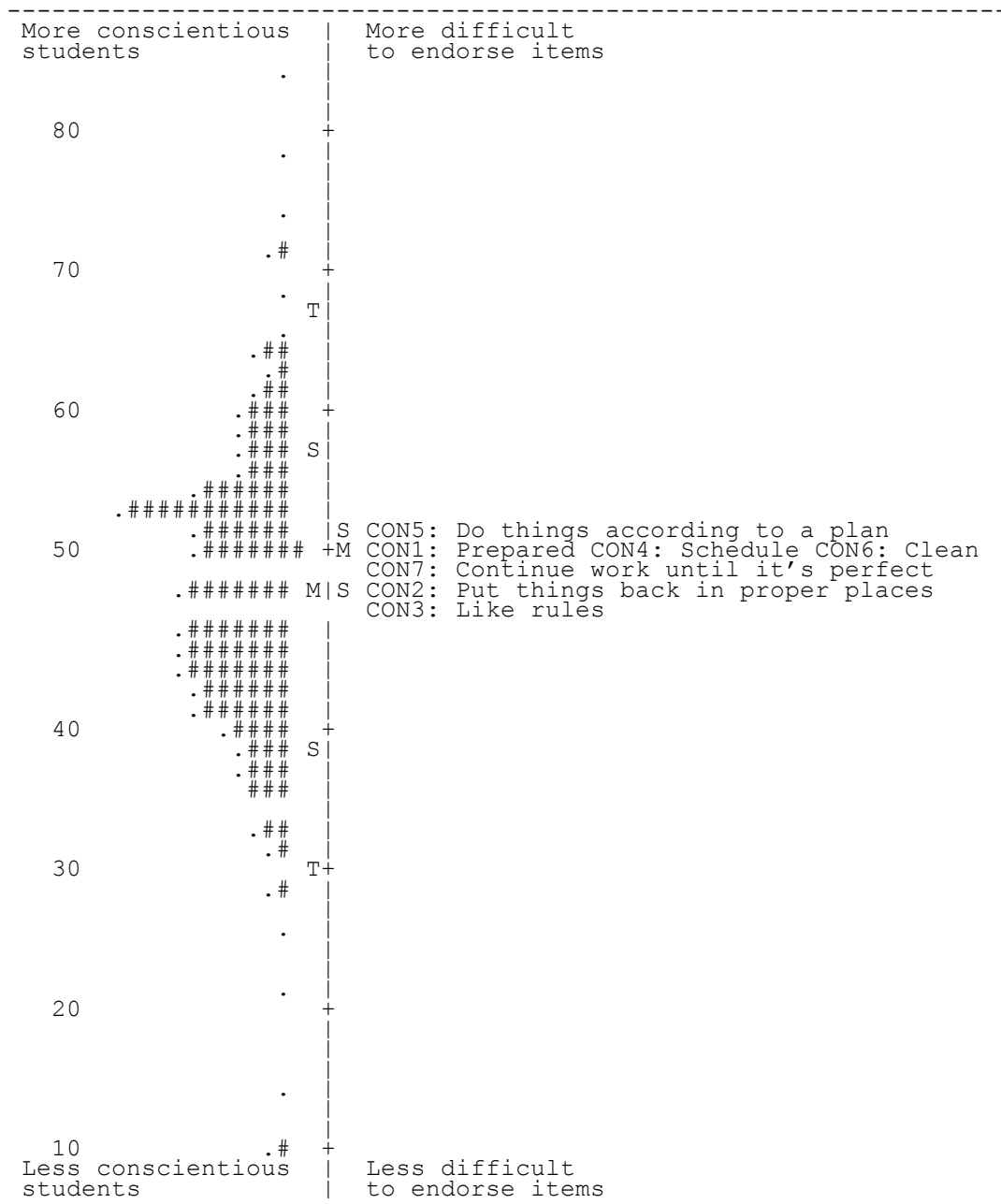


Figure 10. The item-person map of the Intellect/Imagination construct. Each # represents nine persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

The first residual contrasts for the five constructs explained variances ranging from 8.0% to 13.3%. Although only the first contrast for Extraversion-Introversion was under the 10% criterion set as an indication for the potential existence of a secondary dimension, the eigenvalues for all the first contrasts were under the criteria of 3.0. Thus, based on Linacre’s 2007 criteria of 50% or more variance

Table 27

Explained and Unexplained Variance for the Constructs Including Variance Explained by the First Contrast

Construct	Rasch Eigenvalue	Variance Explained	Unexplained Variance	First Contrast Eigenvalue	First Contrast Explained Variance
EI	17.3	68.4%	31.6%	2.0	8.0%
AGR	7.6	48.8%	51.2%	2.0	12.9%
EMO	10.1	59.1%	40.9%	1.8	10.6%
INT	8.1	53.7%	46.3%	2.0	13.3%
CON	9.2	56.7%	43.3%	2.1	13.1%

Note. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; INT = Intellect/Imagination; CON = Conscientiousness.

explained by the Rasch model, less than 10% explained by the first contrast, or a first contrast eigenvalue of less than 3.0, all the constructs can be considered fundamentally unidimensional.

Rasch item analysis for the five hypothesized constructs revealed that two items misfit their intended constructs: EI 7 (“Feel at ease in a large group of people”) had an infit/outfit index of 1.70/1.84, and INT 5 (“Love to read challenging materials”) had an infit/outfit index of 1.52/1.50. However, the removal

of the misfitting items did not substantially affect the Rasch person and item reliability estimates or separation statistics of the constructs. None of the fit statistics of the remaining items in each construct was substantially affected, and the Rasch PCA results indicated that there was no substantial improvement in the unidimensionality of either construct. Thus, the removal of items EI 7 and INT 5 was not warranted at this stage.

Table 28

Item and Person Separations and Reliabilities for the Constructs Comprising the Revised Big Five Factor Marker Questionnaire

Construct	Item separation	Item reliability	Person separation	Person reliability	Cronbach's Alpha
EI	15.49	1.00	2.06	.81	.86
AGR	6.87	.98	1.67	.74	.78
EMO	12.03	.99	1.56	.71	.77
INT	14.36	1.00	1.41	.66	.71
CON	2.60	.87	1.79	.76	.81

Note. EI = Extraversion-Introversion; AGR = Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination.

The Rasch person reliability estimates ranged from .66 to .83, and Cronbach's alpha estimates ranged from .71 to .86 (Table 28). Extraversion-Introversion had the highest reliability estimates (person reliability = .81, α = .86).

Intellect/Imagination had the lowest reliability estimates (person reliability = .66, α = .71). The Rasch item reliability estimates ranged from .87 to 1.00.

Conscientiousness had the lowest the lowest item reliability estimate, .87.

After concluding the Rasch PCA of item residuals and Rasch item analyses for the constructs measured by the Revised Big Five Factor Marker questionnaire, Rasch person measures were obtained for each construct. The measures were correlated in SPSS 15.0 (Table 29). Several constructs were correlated significantly ($p < .01$). Items measuring Extraversion-Introversion correlated with Agreeableness at $r = .72$, correlated with Intellect/Imagination at $r = .41$, and correlated with Conscientiousness at $r = .13$. Items measuring Agreeableness had weak correlations with Emotional Stability ($r = .13$) and Conscientiousness ($r = .13$) and a moderate correlation with Intellect/Imagination ($r = .39$). Intellect/Imagination had a weak correlation with Conscientiousness ($r = .27$).

Table 29

Correlational Matrix of Rasch Person Measures for Five Constructs Comprising the Revised Big Five Factor Markers Questionnaire

Construct	EI	AGR	EMO	INT	CON
Extraversion-Introversion	—				
Agreeableness	.72*	—			
Emotional Stability	.05	.13*	—		
Intellect/Imagination	.41*	.39*	-.01	—	
Conscientiousness	.13*	.24*	.04	.27*	—

Note. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; INT = Intellect/Imagination; CON = Conscientiousness. $N = 1078$.

* = $p < .01$

Exploratory Factor Analysis. The second step in the preliminary analysis of the data involved conducting an exploratory factor analysis of the data to confirm the presence of the five hypothesized factors for the sample population. All items

were included in the analysis, after which the results were compared to the Rasch analysis to assist in determining which items to retain for further confirmatory factor analysis. As recommended by Kaiser (1974), a minimum eigenvalue of 1.0 was used as the criterion for identifying factors. A minimum loading of .40 was used as a criterion for identifying good performing items (Stevens, 2009).

Data screening. Descriptive statistics were obtained for the Revised Big Five Factor Marker questionnaire items (Table 30). Several items, most notably those measuring Emotional Stability, approached 1.0 kurtosis; however, Tabachnick and Fidell (2007) noted that problems in variance analysis due to positive kurtosis or negative kurtosis diminish with sample sizes of over 100 and 200, respectively (p. 80). The *N*-size was found to vary between 1,070 and 1,078, depending on the item; such large *N*-sizes made problems associated with kurtosis highly unlikely for the sample population. There were 80 missing data points that accounted for 0.2% of the total 37,730 data points in the data set (1078 cases times 35 items). Because this percentage was well below the maximum 5% of missing data points recommended for maintaining the structure of correlational matrices in factor analysis procedures (Tabachnick & Fidell, 2007, p. 63), the mean raw scores for each item were substituted in the place of the missing data points prior to running the EFA and

Table 30

Descriptive Statistics for the Revised Big Five Factor Marker Questionnaire Items

Item	Description	<i>N</i>	<i>M</i>	<i>SD</i>	Skew	Kurt.
AGR 1	Inquire about others' well-being	1,070	3.84	1.23	-.26	-.30
INT 2	Am full of ideas	1,078	3.63	1.32	.05	-.67
EMO 2	Get stressed out easily	1,074	3.59	1.52	-.08	-1.00
EI 4	Like to draw attention to myself	1,077	3.28	1.34	.21	-.52
CON 6	Like to clean up things	1,077	3.33	1.50	.12	-.96
EI 7	Feel at ease in a large group of people	1,077	2.84	1.35	.46	-.39
CON 5	Do things according to a plan	1,075	3.11	1.37	.22	-.72
INT 5	Love to read challenging materials	1,075	2.49	1.45	.76	-.38
AGR 2	Feel concern for others	1,077	4.12	1.30	-.45	-.37
EMO 4	Have frequent mood swings	1,078	4.29	1.35	-.49	-.49
EI 2	Like to talk to a lot of people	1,075	4.41	1.33	-.53	-.45
AGR 3	Am interested in other people's problems	1,076	3.75	1.38	-.18	-.59
EMO 1	Often feel blue	1,074	3.81	1.51	-.20	-.97
INT 1	Have a vivid imagination	1,074	3.95	1.40	-.19	-.84
CON 7	Continue my work until it's perfect	1,076	3.28	1.30	.07	-.68
INT 6	Love to think of new ways to do things	1,077	3.88	1.36	-.11	-.79
EMO 5	Panic easily	1,074	4.03	1.40	-.32	-.72
EI 8	Am skilled at dealing with people	1,075	3.47	1.28	.06	-.47
AGR 4	Take time out for others	1,074	3.86	1.21	-.10	-.34
CON 4	Follow a daily schedule	1,074	3.21	1.31	.21	-.54
EMO 3	Worry about things	1,074	4.36	1.40	-.65	-.39
INT 7	Can handle a lot of information	1,073	3.44	1.16	.20	-.12
CON 3	Like rules	1,076	3.37	1.36	.14	-.65
EI 6	Talk a lot	1,076	3.98	1.44	-.12	-.93
AGR 5	Am on good terms with nearly everyone	1,078	3.72	1.41	-.09	-.78
AGR 6	Make people feel at ease	1,077	3.29	1.15	.18	-.05
EI 3	Start conversations	1,076	3.55	1.38	.07	-.76
EMO 6	Feel threatened easily	1,076	3.37	1.42	.16	-.83
INT 4	Am quick to understand things	1,077	3.58	1.22	-.06	-.36
CON 2	Put things back in their proper places	1,076	3.37	1.41	.12	-.70
EI 5	Like to be the center of attention	1,076	3.07	1.40	.33	-.63
AGR 7	Have a good word for everyone	1,075	3.74	1.33	-.11	-.61
EMO 7	Get angry easily	1,077	3.18	1.50	.20	-.93
CON 1	Am prepared	1,076	3.29	1.27	.08	-.49
AGR 8	Know how to cheer people up	1,076	3.63	1.14	-.02	-.14
EI 1	Am the life of the party	1,078	2.97	1.37	.40	-.51
INT 3	Have a rich vocabulary in my native language	1,078	2.99	1.30	.36	-.43

Note. EI = Extraversion-Introversion; AGR = Agreeableness; CON =

Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination.

CFA. The data were also examined for statistical outliers (z -scores ≥ 3.29) using factor scores; however, none was identified in the data sample, demonstrating normality and linearity of the variables. The Kaiser-Meyer-Olkin measure of sampling adequacy was .877, supporting the factorability of the R -matrix. Bartlett's Test of Sphericity, which examines the existence of multicollinearity and singularity of the correlational matrix, was significant at .000. Thus, normality assumptions of the variables were checked and confirmed for the data prior to factor analysis.

Factor analysis results. Because prior Big Five personality research assumed the existence of five orthogonal factors, factor extraction was conducted using principal axis factoring with varimax orthogonal rotation with Kaiser normalization to aid in the interpretation of the factor loadings (Table 31). A total variance of 49.16% was accounted for by the five factors. However, several items did not load onto their intended factors. All items measuring Extraversion-Introversion and Agreeableness loaded onto the same factor. One item intended to load onto the Extraversion-Introversion factor (EI 7) failed to load above the .40 criterion. One item intended to load onto the Agreeableness construct (AGR 1) loaded into the

Table 31

Exploratory Factor Analysis Results of the Revised Big Five Factor Marker Questionnaire (Five-Factor Solution)

Item	Item description	Factor					h^2
		1	2	3	4	5	
AGR 5	Am on good terms with nearly everyone	.77	.13	-.18	.00	.04	.64
EI 8	Am skilled at dealing with people	.74	.11	-.12	.11	.06	.59
EI 3	Start conversations	.71	.07	-.08	.07	.01	.52
EI 6	Talk a lot	.70	-.01	.07	.06	.16	.52
EI 2	Like to talk to a lot of people	.69	-.03	.08	.06	-.14	.51
EI 1	Am the life of the party	.68	.01	-.04	.20	.20	.31
AGR 8	Know how to cheer people up	.67	.15	-.02	.22	.05	.52
EI 5	Like to be the center of attention	.58	.01	.05	.18	.01	.37
EI 4	Like to draw attention to myself	.56	-.07	.16	.19	.02	.38
AGR 6	Make people feel at ease	.56	.24	-.03	.17	-.05	.40
AGR 2	Feel concern for others	.53	.08	.31	.02	-.30	.47
AGR 3	Am interested in other people's problems	.49	.00	.30	.00	-.21	.37
AGR 4	Take time out for others	.44	.12	.10	.08	-.14	.25
AGR 7	Have a good word for everyone	.43	.09	.04	.27	-.09	.27
INT 6	Love to think of new ways to do things	.33	.10	.01	.32	-.13	.24
EI 7	Feel at ease in a large group of people	.31	.13	.11	-.17	-.06	.16
CON 5	Do things according to a plan	.03	.67	-.02	.06	.08	.46
CON 2	Put things back in their proper places	.05	.63	-.02	.06	-.18	.44
CON 1	Am prepared	.07	.63	.07	.20	.16	.48
CON 7	Continue my work until it's perfect	.15	.59	.02	.21	.05	.41
CON 3	Like rules	-.03	.58	.10	-.01	-.01	.35
CON 5	Follow a daily schedule	.13	.57	.11	.01	.03	.35
CON 6	Like to clean up things	.07	.54	.01	.02	-.17	.33
EMO 3	Worry about things	.04	.21	.65	-.03	-.04	.47
EMO 1	Often feel blue	.00	.02	.64	.01	.14	.43
EMO 6	Feel threatened easily	-.06	.04	.58	-.09	-.05	.35
EMO 2	Get stressed out easily	-.10	.03	.56	.11	.37	.47
EMO 5	Panic easily	.08	-.03	.54	-.09	-.07	.31
EMO 4	Have frequent mood swings	.07	-.16	.41	.17	.15	.25
AGR 1	Inquire about others' well-being	.22	.15	.39	-.06	-.17	.25
INT 2	Am full of ideas	.33	-.03	.06	.67	-.05	.55
INT 1	Have a vivid imagination	.29	.00	.17	.63	-.07	.51
INT 3	Have a rich vocabulary in my native language	.15	.17	-.12	.48	.10	.31

(Table 31 continues)

(Table 31 continued)

Item	Item description	Factor					h^2
		1	2	3	4	5	
INT 5	Love to read challenging materials	-.04	.12	-.04	.40	.04	.18
INT 7	Can handle a lot of information	.26	.32	-.12	.36	.08	.32
INT 4	Am quick to understand things	.21	.29	-.14	.33	.00	.25
EMO 7	Get angry easily	.03	.00	.46	.01	.53	.50
Eigenvalues		6.02	3.05	2.72	2.04	.89	
Percentage of variance		16.26	8.23	7.36	5.51	2.42	

Note. EI = Extraversion-Introversion; AGR = Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination; h^2 = item commonalities. The mean of raw scores was substituted for 80 missing data points. Loadings above the .40 criterion are in bold face. $N = 1078$.

Emotional Stability factor. Item EMO 7, which was intended to load onto the Emotional Stability factor, loaded onto a separate, and likely spurious, factor, while items EMO 2 and EMO 4 failed to load above .40 onto any factor. Moreover, only two items intended to load onto the Intellect/Imagination construct loaded onto their intended factor, while the remaining five Intellect/Imagination items failed to load above .40. Most important, the fifth factor failed the Kaiser criterion of having an eigenvalue over 1.0 and had only one item above the .40 criterion. Thus, according to the EFA results, the Big Five did not consist of five factors for this sample.

Because the fifth factor from the five-factor solution failed the Kaiser value criterion, a four-factor solution was produced including items from all five

Table 32

*Exploratory Factor Analysis Results of the Revised Big Five Factor Marker
Questionnaire (Four-Factor Solution)*

Item	Item description	Factor				h^2
		1	2	3	4	
AGR 5	Am on good terms with nearly everyone	.74	.11	-.19	.04	.61
EI 8	Am skilled at dealing with people	.72	.10	-.13	.14	.56
EI 2	Like to talk to a lot of people	.71	-.02	.04	.02	.51
EI 3	Start conversations	.70	.07	-.09	.08	.51
EI 6	Talk a lot	.67	-.02	.08	.10	.47
AGR 8	Know how to cheer people up	.66	.14	-.03	.24	.51
EI 1	Am the life of the party	.65	-.01	-.02	.26	.50
EI 5	Like to be the center of attention	.58	.01	.05	.17	.37
EI 4	Like to draw attention to myself	.57	-.07	.16	.17	.38
AGR 2	Feel concern for others	.57	.11	.24	-.08	.40
AGR 6	Make people feel at ease	.56	.24	-.05	.16	.40
AGR 3	Am interested in other people's problems	.52	.02	.25	-.08	.34
AGR 4	Take time out for others	.46	.13	.07	.04	.23
AGR 7	Have a good word for everyone	.45	.09	.03	.23	.26
INT 6	Love to think of new ways to do things	.36	.11	-.01	.27	.21
EI 7	Feel at ease in a large group of people	.31	.14	.08	-.18	.16
CON 5	Do things according to a plan	.02	.66	-.02	.11	.44
CON 2	Put things back in their proper places	.07	.64	-.05	.02	.42
CON 1	Am prepared	.06	.61	.09	.26	.45
CON 3	Like rules	-.03	.59	.09	.00	.35
CON 7	Continue my work until it's perfect	.15	.58	.03	.24	.41
CON 5	Follow a daily schedule	.12	.56	.10	.03	.34
CON 6	Like to clean up things	.09	.55	-.02	-.02	.31
EMO 1	Often feel blue	.01	.03	.66	-.01	.43
EMO 3	Worry about things	.07	.22	.64	-.09	.47
EMO 2	Get stressed out easily	-.11	.02	.61	.15	.04
EMO 6	Feel threatened easily	-.03	.06	.56	-.15	.34
EMO 5	Panic easily	.11	-.01	.52	-.16	.31
EMO 7	Get angry easily	-.01	-.03	.50	.11	.26

(Table 32 continues)

(Table 32 continued)

Item	Item description	Factor				h^2
		1	2	3	4	
EMO 4	Have frequent mood swings	.07	-.17	.44	.16	.25
AGR 1	Inquire about others' well-being	.25	.17	.35	-.13	.23
INT 2	Am full of ideas	.36	-.03	.07	.59	.48
INT 1	Have a vivid imagination	.33	.01	.18	.54	.43
INT 3	Have a rich vocabulary in my native language	.14	.15	-.08	.52	.32
INT 5	Love to read challenging materials	-.04	.11	-.01	.41	.19
INT 7	Can handle a lot of information	.26	.31	-.10	.40	.33
INT 4	Am quick to understand things	.21	.28	-.13	.34	.25
Eigenvalues		6.09	3.01	2.68	2.04	
Percentage of Variance		16.47	8.14	7.24	5.51	

Note. EI = Extraversion-Introversion; AGR = Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination; h^2 = item commonalities. The mean of raw scores was substituted for 80 missing data points. Loadings above the .40 criterion are in bold face. $N = 1,078$.

hypothesized constructs (Table 32). The resulting solution accounted for 45.45% of the variance; however, once again all but one of the Agreeableness items loaded onto the same factor as the Extraversion-Introversion items, with AGR 1 failing to load above .40 on any factor. One of the Extraversion-Introversion items (EI 7) and two of the six Intellect/Imagination items (INT 6 and INT 4) failed to load onto any factor above the .40 criterion.

Given the weak eigenvalue of the fifth factor in the five-factor solution, I determined that the four-factor solution was the most appropriate for this sample. An oblique rotation using direct oblimin was requested for the four factors and the

Table 33

Factor Correlation Matrix for the Four Factors Comprising the Revised Big Five Factor Marker Questionnaire

Factor	EI/AGR	CON	EMO	INT
Extraversion-Introversion/Agreeableness	—			
Conscientiousness	.25*	—		
Emotional Stability	.11*	.04	—	
Intellect/Imagination	.26*	.17*	-.08*	—

Note. EI/AGR = Extraversion-Introversion/Agreeableness; CON = Conscientiousness; EMO = Emotional Stability; INT = Intellect/Imagination. $N = 1,078$.

* $p < .05$.

resulting factor correlation matrix was examined (Table 33). Five correlations were significant. Extraversion-Introversion/Agreeableness had weak correlations with Conscientiousness, $r = .25$, Emotional Stability, $r = .11$, and Intellect/Imagination, $r = .26$. Conscientiousness had a weak correlation with Intellect/Imagination, $r = .17$. Emotional Stability had a weak negative correlation with Intellect/Imagination, $r = -.08$. All correlations were under the value of $r = .32$ recommended as the criterion for using oblique rotation by Tabachnick & Fidell (2007, p. 646).

Comparison of the Rasch analysis and EFA results. The Rasch PCA results gave a somewhat different picture of the hypothesized Big Five constructs compared to the EFA results. According to the Rasch PCA of item residuals results, the Extraversion-Introversion and Agreeableness constructs were essentially

independent despite being strongly correlated at $r = .72$. According to the EFA, the items measuring Extraversion-Introversion and Agreeableness measured the same factor. Both techniques identified several problem items (Table 34). Some items were identified by both Rasch PCA and EFA as misfitting their intended constructs (Table 35). Based on Rasch PCA and EFA loadings, only two items were agreed upon by both techniques: EI 7 (“Feel at ease in a large group of people”) and INT 6 (“Love to think of new ways to do things”). However, the Rasch item analysis additionally showed that item AGR 1 (“Inquire about others’ well-being”) had a point-measure correlation of .49, a possible indication that the item was not contributing as much to the Agreeableness construct as the other items.

Both the Rasch PCA of item residuals and the EFA identified EI 7, AGR 1, and INT 6 as problematic items. The first item, EI 7 (“Feel at ease in a large group of people”), was meant to measure Extraversion-Introversion; however, it is not clear from the wording what sort of people the participants would feel at ease with. The second, item AGR 1 (“Inquire about others’ well-being”), is similarly vague about what “other” means in relation to the participants. The nature of personal relationships in Japan is highly dependent upon two important social factors: the *kohai* (junior) and *senpai* (senior) hierarchical relationship between younger/older or novice/expert people, and the *uchi/soto* (“inside/outside”) group relationship between those who share common personal history and those who do not.

Table 34

A Comparison of the Big Five Constructs Identified by Rasch PCA of Item Residuals and EFA

Construct/Factor	Rasch PCA	EFA
	Items loading onto factor	Items loading onto factor
Extraversion-Introversion	EI items 2, 3, 4, 5, 6, 8	EI items 1, 2, 3, 4, 5, 6
		AGR items 2, 3, 4, 5, 6, 7, 8
Agreeableness	AGR items 1, 2, 3, 5, 6, 8	(All items loaded onto same factor as EI items)
Emotional Stability	EMO items 2, 3, 5, 6, 7	EMO items 1, 2, 3, 4, 5, 6, 7
Intellect/Imagination	INT items 1, 2, 3, 4, 5	INT items 1, 2, 3, 5, 7
Conscientiousness	CON items 2, 4, 5, 6	CON items 1, 2, 3, 4, 5, 6, 7

Notes. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; INT = Intellect/Imagination; CON = Conscientiousness. Loadings were based on a minimum loading criterion of .40. $N = 1,078$.

The third item, INT 6 (“Love to think of new ways to do things”), reads literally in Japanese “I like to try to do new things” (*atarashii yarikata wo tamesu no ha suki da*). This item failed to load above .40 on the Intellect/Imagination construct during Rasch PCA of item residuals analysis and it loaded below the .40 criterion onto the Extraversion-Introversion factor for both the five-factor and four-factor solutions

Table 35

Problematic Items Identified by Rasch PCA of Item Residuals and EFA

Construct/Factor	Rasch PCA	EFA
	Items failing to load	Items failing to load
Extraversion-Introversion	EI 1, EI 7	EI 7
Agreeableness	AGR 4, AGR 7 ^a	AGR 1
Emotional Stability	EMO 1, EMO 4	(All items loaded)
Intellect/Imagination	INT 6, INT 7	INT 4, INT 6
Conscientiousness	CON 1, CON 3, CON 7	(All items loaded)

Notes. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; INT = Intellect/Imagination; CON = Conscientiousness. Loadings were based on a minimum loading criterion of .40. $N = 1,078$.

^a = Rasch item analysis also identified item AGR 1 as possibly problematic.

in the EFA. It is possible that the participants interpreted this item as indicative of an outgoing personality who is not afraid to try new things and thus not afraid to make mistakes. Based on the results of the Rasch PCA of item residuals and the EFA, I decided that the three items EI 7, AGR 1, and INT 6 were candidates for removal during the confirmatory factor analysis phase.

Summary of instrument validation for the R-B5FM questionnaire.

Although the Big Five personality traits are hypothesized to exist across cultures, problems emerged during examination of items measuring the Extraversion-Introversion, Agreeableness, and Intellect/Imagination constructs in both of the Rasch PCA of item residuals and the EFA. Extraversion-Introversion and Agreeableness correlated strongly in the Rasch PCA of item residuals analysis,

and the items measuring Extraversion-Introversion and Agreeableness loaded onto the same factor in the EFA. Thus, research question one regarding the validity of the Big Five for a Japanese population has been answered: The Big Five do not appear to exist for the sample Japanese population in this study. The preliminary data analysis results indicate that of the five hypothesized human personality traits comprising the Big Five, only four—Extraversion-Introversion, Emotional Stability, Conscientiousness, and Intellect/Imagination—are valid for the sample.

In the following section, I examine the six hypothesized constructs of the Foreign Language Classroom Speaking Confidence questionnaire. Data analysis procedures follow the same instrument validation steps as previously outlined in the Background of the Main Study Participants section.

The Foreign Language Classroom Speaking Confidence Questionnaire

Rasch PCA of item residuals. In order to test the dimensionality of the items in the FLCSC, all items were analyzed using a Rasch PCA of item residuals (Table 36). The Rasch model explained a total variance of 54.4% (eigenvalue = 62.1), and the first residual contrast explained 7.7% of the variance (eigenvalue = 8.8). All eleven items from the Foreign Language Classroom Speaking Anxiety (FLCSA) construct had high positive loadings above .40, indicating that they cohered to form a construct. Items from other hypothesized constructs were mixed; items from the

FLCSA were input again separately for a further PCA of item residuals analysis to confirm their dimensionality.

Foreign Language Classroom Speaking Anxiety. The eleven items hypothesized to measure the Foreign Language Classroom Speaking Anxiety construct were input into the Rasch model and the Rasch PCA of item residuals results were examined (Table 37). The Rasch model explained 78.2% of the variance (eigenvalue = 39.4); the first residual contrast explained 4.0% of the variance (eigenvalue = 2.0). These results indicated that the 11 FLCSA items formed a fundamentally unidimensional construct. Items FLCSA 7 and FLCSA 10 had positive residual loadings above .40, while items FLCSA 2, FLCSA 11, and FLCSA 6 had negative residual loadings above .40. The positively loading items concerned social anxiety about being laughed at, while the negatively loading items concerned physical nervousness about speaking.

Six items (FLCSA 1, "I'm worried that other students in class speak English better than I do," FLCSA 3, "I'm worried that my partner speaks better English than I do," FLCSA 8, "I'm worried about making mistakes when I speak English with a partner," FLCSA 9, "I feel tense when I have to speak English with a classmate in a pair," FLCSA 5, "I'm worried about making mistakes while

Table 36

Rasch Principal Components Analysis for the Items Measuring the Six Foreign Language Classroom Speaking Confidence Constructs (52 Items)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
FLCSA 8	.75	55.14	1.24	1.24
FLCSA 9	.75	54.03	1.38	1.45
FLCSA 5	.74	52.81	1.32	1.33
FLCSA 10	.74	57.14	1.44	1.49
FLCSA 11	.69	50.28	1.41	1.50
FLCSA 7	.67	60.29	1.39	1.45
FLCSA 2	.67	49.25	1.46	1.49
FLCSA 4	.64	53.57	1.21	1.31
FLCSA 1	.64	54.26	1.27	1.26
FLCSA 6	.62	49.86	1.09	1.10
FLCSA 3	.49	53.90	1.27	1.43
PSV 4	.11	43.94	1.10	1.16
PSV 6	.11	44.48	.97	1.07
PECE 6	.07	52.67	1.31	1.31
PECE 7	.06	52.76	1.18	1.20
PSV 2	.01	46.37	.83	.85
CECP 1	.00	49.45	1.02	1.08
PFLSS 2	-.52	48.60	.75	.78
PFLSS 5	-.51	56.69	.88	.88
PFLSS 4	-.50	54.15	.73	.73
PFLSS 8	-.49	56.48	1.12	1.11
DSE 7	-.45	46.63	.88	.87
CECP 7	-.45	48.79	.64	.66
PFLSS 3	-.43	52.65	.74	.75
DSE 6	-.42	46.75	.61	.62
DSE 3	-.42	44.34	.95	.91
DSE 8	-.40	46.80	.86	.84
CECP 6	-.40	46.78	.90	.96
DSE 9	-.39	46.72	.81	.80
CECP 8	-.38	47.76	.67	.68
PFLSS 6	-.38	56.49	.97	.95
PFLSS 1	-.37	56.05	.94	.95
DSE 5	-.37	43.56	1.21	1.12
PFLSS 7	-.36	55.30	.81	.84
DSE 1	-.36	44.86	.90	.86
CECP 2	-.31	46.66	.80	.84
DSE 4	-.30	38.65	.85	.72
DSE 2	-.27	36.94	1.07	.86
CECP 3	-.22	45.63	.82	.86

(Table 36 continues)

(Table 36 continued)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CECP 5	-.21	45.07	.83	.85
PSV 3	-.21	48.56	.67	.68
CECP 4	-.14	47.07	.90	.94
PSV 8	-.13	53.58	.69	.68
PECE 3	-.10	53.96	1.00	1.01
PECE 1	-.09	54.59	1.12	1.09
PECE 8	-.09	44.37	1.12	1.15
PECE 2	-.08	54.28	1.05	1.04
PECE 4	-.07	61.28	1.31	1.26
PECE 5	-.06	55.23	1.14	1.14
PSV 5	-.05	45.36	.90	.97
PSV 7	-.04	42.11	.92	.95
PSV 1	-.03	49.08	1.13	1.16

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. Measures are in Rasch CHIPs

Table 37

Rasch Principal Components Analysis for the Items Measuring Foreign Language Classroom Speaking Anxiety

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
FLCSA 7	.64	60.07	.98	.96
FLCSA 10	.59	55.36	.91	.87
FLCSA 1	.39	50.94	1.05	1.04
FLCSA 3	.32	50.41	1.42	1.66
FLCSA 8	.19	52.33	.73	.71
FLCSA 2	-.61	43.05	1.22	1.25
FLCSA 11	-.54	44.69	1.03	1.10
FLCSA 6	-.51	44.03	.97	.97
FLCSA 9	-.13	50.58	.85	.84
FLCSA 5	-.07	48.68	.83	.88
FLCSA 4	-.05	49.86	1.10	1.45

Note. FLCSA = Foreign Language Classroom Speaking Anxiety. Measures are in Rasch CHIPs.

speaking English,” and FLCSA 4, “I start to panic when I speak in English with a classmate in a pair”) failed to load above .40.

After extracting the items measuring Foreign Language Classroom Speaking Anxiety, the remaining 41 items were analyzed (Table 38). The Rasch model explained 62.6% of the variance (eigenvalue = 68.6), and the first residual contrast explained 5.0% of the variance (eigenvalue = 5.5). All the items with high positive loadings above .40 were designed to measure Past English Classroom Experiences. To clarify the structure of Past English Classroom Experiences, a separate Rasch PCA of item residuals was conducted with the items hypothesized to measure Past English Classroom Experiences.

Past English Classroom Experiences. The eight items hypothesized to measure Past English Classroom Experiences were input into the Rasch model and the Rasch PCA of item residuals results were examined (Table 39). The Rasch model explained 80.8% of the variance (eigenvalue = 33.6); the first contrast explained 4.3% of the variance (eigenvalue = 1.8). These results indicated that the eight items measuring Past English Classroom Experiences formed a fundamentally unidimensional construct. Items PECE 6 and PECE 7 had high positive loadings above .40, and items PECE 3, PECE 4, and PECE 1 had high negative loadings above -.40. The items with positive loadings were related to the overall classroom

Table 38

Rasch Principal Components Analysis of Item Residuals for Items Measuring Desire to Speak English, Past English Classroom Experiences, Perceived Foreign Language Speaking Self-Competence, Current English Classroom Perception, and Perceived Social Value of Speaking English (41 Items)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
PECE 7	.62	54.36	1.39	1.43
PECE 4	.61	64.36	1.42	1.41
PECE 5	.61	57.27	1.29	1.29
PECE 3	.58	55.77	1.11	1.11
PECE 2	.54	56.15	1.18	1.17
PECE 1	.54	56.52	1.27	1.22
PECE 6	.45	54.25	1.58	1.61
PFLSS 8	.34	58.76	1.09	1.12
PFLSS 1	.33	58.26	.99	1.01
PFLSS 5	.32	58.99	.83	.84
PFLSS 7	.25	55.00	.85	.90
PFLSS 4	.24	56.00	.71	.71
PFLSS 6	.24	58.77	1.00	.99
PFLSS 3	.19	54.23	.75	.76
PECE 8	.16	44.57	1.24	1.26
PFLSS 2	.04	49.46	.72	.75
PSV 1	.02	50.01	1.35	1.40
PSV 6	.01	44.69	1.17	1.39
DSE 1	-.48	45.13	.94	.90
DSE 9	-.46	47.27	.86	.85
DSE 8	-.46	47.37	.90	.89
DSE 3	-.44	44.53	.96	.90
DSE 2	-.43	36.35	1.07	.87
DSE 5	-.42	43.65	1.25	1.19
DSE 4	-.41	38.19	.85	.77
CECP 2	-.41	47.20	.84	.95
DSE 7	-.39	47.17	.90	.89
CECP 8	-.39	48.47	.70	.72
CECP 3	-.39	46.01	.90	1.02
DSE 6	-.35	47.30	.63	.63
PSV 3	-.32	49.41	.77	.79
CECP 7	-.32	49.68	.64	.67
PSV 7	-.29	42.02	1.05	1.17
CECP 5	-.27	45.37	.91	.95
CECP 6	-.24	47.34	.90	.96
PSV 5	-.24	45.70	1.05	1.23

(Table 38 continues)

(Table 38 continued)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CECP 4	-.20	47.67	1.02	1.23
PSV 2	-.19	46.87	1.00	1.11
PSV 8	-.16	55.34	.84	.84
PSV 4	-.12	44.08	1.32	1.52
CECP 1	-.01	50.46	1.24	1.45

Note. PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. Measures are in Rasch CHIPs.

atmosphere, while the items with negative loading were concerned with speaking in English in class. Three items (PECE 8, “I got along well with classmates in my high school English classes,” PECE 5, “I did enjoyable and interesting activities in my high school English classes,” and PECE 2, “I enjoyed doing pair work in my high school English classes”) failed to load above the .40 criterion. Because Japanese high schools divide English language instruction into separate grammar,

Table 39

Rasch Principal Components Analysis for the Items Measuring Past English Classroom Experiences

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
PECE 6	.71	48.46	1.29	1.37
PECE 7	.60	48.61	.82	.84
PECE 8	.19	35.60	1.53	1.46
PECE 5	.05	52.46	.93	.93
PECE 3	-.60	50.53	.71	.70
PECE 4	-.53	61.84	1.18	1.16
PECE 1	-.40	51.46	1.01	1.06
PECE 2	-.33	51.05	.81	.82

Note. PECE = Past English Classroom Experiences. Measures are in Rasch CHIPs.

reading, and speaking classes, the respondents might have been confused as to which class PECE 8 referred. Likewise, they might have wondered to which “interesting activities” and in which English class the item PECE 2 referred. Answers to PECE 2 might have depended too much on the existence of or the frequency of pair work in students’ high school classes. Three students handwrote the words *nakatta* (“We didn’t have any”) and two students handwrote the words *shinakatta* (“We didn’t do any”) in the margin next to the item PECE 2 without circling an answer.

After extracting the items measuring PECE, the remaining 33 items were analyzed (Table 40). The Rasch model explained 64.7% of the variance (eigenvalue = 60.4); the first contrast explained 4.7% of the variance (eigenvalue = 4.4). All the items with high positive loadings above .40 were DSE items, while all the items with high negative loadings above -.40 were PFLSS items. To confirm construct unidimensionality, the items hypothesized to measure the Desire to Speak English and the items hypothesized to measure the Perceived Foreign Language Speaking Self-Competence were analyzed separately using Rasch PCA of item residuals.

Table 40

Rasch Principal Components Analysis for the Items Measuring Perceived Foreign Language Speaking Self-Competence, Desire to Speak English, Current English Classroom Perception, and Perceived Social Value of Speaking English (33 Items)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
DSE 1	.64	45.97	.96	.92
DSE 3	.58	45.30	.97	.90
DSE 8	.58	48.48	.93	.92
DSE 5	.54	44.32	1.27	1.18
DSE 2	.51	36.37	1.07	.83
DSE 4	.49	38.35	.84	.83
DSE 7	.48	48.25	.92	.91
DSE 9	.42	48.36	.88	.86
PSV 7	.29	42.52	1.12	1.24
DSE 6	.27	48.40	.65	.65
PSV 4	.20	44.80	1.46	1.84
PSV 2	.16	47.91	1.10	1.26
PSV 5	.13	46.60	1.14	1.37
PSV 1	.08	51.46	1.54	1.66
PSV 3	.05	50.78	.80	.81
PSV 6	.03	45.48	1.35	1.68
PFLSS 4	-.55	58.25	.80	.80
PFLSS 8	-.50	61.38	1.25	1.32
PFLSS 5	-.50	61.65	.96	.97
PFLSS 3	-.45	56.24	.85	.86
PFLSS 2	-.44	50.83	.78	.80
PFLSS 7	-.42	57.13	.98	1.05
PFLSS 1	-.42	60.81	1.13	1.19
PFLSS 6	-.32	61.39	1.12	1.12
CECP 4	-.25	48.81	1.11	1.24
CECP 1	-.20	51.96	1.40	1.59
CECP 7	-.17	51.08	.68	.68
CECP 3	-.16	46.95	.94	1.01
CECP 5	-.15	46.23	.98	1.00
CECP 6	-.14	48.43	.96	.99
CECP 2	-.13	48.29	.86	.92
CECP 8	-.09	49.72	.73	.73
PSV 8	-.07	57.50	.91	.91

Note. DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. Measures are in Rasch CHIPs.

Desire to Speak English. The nine items hypothesized to measure Desire to Speak English were input into the Rasch model and the Rasch PCA of item residuals results were examined (Table 41). The Rasch model explained a total variance of 84.2% (eigenvalue = 47.9); the first principal contrast explained 3.0% of the variance (eigenvalue = 1.7). These results indicated that the nine DSE items formed a fundamentally unidimensional construct. Items DSE 6 and DSE 9 had high positive loadings above .40. Three items, items DSE 1, DSE 5, and DSE 8, had high negative loadings above -.40. The items with positive loadings were related to the the desire to speak English in the classroom, while the items with negative loading were concerned with speaking English to non-Japanese people either at their university or in another country.

Four items (DSE 4, “I want to be able to express my opinions and ideas in English,” DSE 2, “I want to become a good speaker of English,” DSE 7, “I want to find opportunities to speak English outside class,” and DSE 3, “I want to speak English with people from different countries”) failed to load above the .40 criterion. These items might have contained vague wording compared to other items, which specifically focused on making friends or visiting English-speaking countries. It is possible that the participants did not have a strong perception of the need for self-expression or for speaking English with non-native speakers of English.

Table 41

Rasch Principal Components Analysis for the Items Measuring Desire to Speak English

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
DSE 6	.65	55.97	1.10	1.24
DSE 9	.43	55.88	1.08	1.10
DSE 4	.30	39.12	.93	1.05
DSE 2	.21	35.95	1.15	1.01
DSE 7	.19	55.71	.93	.97
DSE 1	-.63	51.74	.89	.84
DSE 5	-.49	48.94	1.33	1.18
DSE 8	-.41	56.09	.82	.83
DSE 3	-.38	50.60	.88	.85

Note. DSE = Desire to Speak English. Measures are in Rasch CHIPs.

Perceived Foreign Language Speaking Self-Competence. The eight items hypothesized to measure Perceived Foreign Language Speaking Self-Competence were input into Winsteps and the Rasch PCA of item residuals results were examined (Table 42). The Rasch model explained a total variance of 80.6% (eigenvalue = 33.3), and the first contrast explained 4.1% of the variance (eigenvalue = 1.7). These results indicated that the eight items measuring Perceived Foreign Language Speaking Self-Competence formed a fundamentally unidimensional construct. Items PFLSS 8, PFLSS 5, and PFLSS 4 had high positive loadings above .40, and items PFLSS 7 and PFLSS 1 had high negative loadings above .40. The items with positive loadings concerned presenting in front of a group of classmates in class, while the items with negative loading concerned

Table 42

Rasch Principal Components Analysis for the Items Measuring Perceived Foreign Language Speaking Self-Competence

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
PFLSS 8	.70	54.86	1.08	1.01
PFLSS 5	.62	55.34	.82	.80
PFLSS 4	.47	49.68	.71	.70
PFLSS 7	-.56	47.79	1.05	1.10
PFLSS 1	-.40	53.94	1.03	.98
PFLSS 6	-.26	54.90	1.32	1.35
PFLSS 3	-.20	46.33	1.00	1.06
PFLSS 2	-.10	37.15	1.00	.97

Note. PFLSS = Perceived Foreign Language Speaking Self-Competence. Measures are in Rasch CHIPs.

speaking English outside the classroom. Three items (PFLSS 6, “I can easily join in a conversation in a group of native English speakers,” PFLSS 3, “I can talk about my hobbies in English during pair work with a classmate,” and PFLSS 2, “I can introduce myself in English during pair work to a classmate”) failed to load above the .40 criterion. Item PFLSS 6 was the most difficult item for students to endorse (difficulty measure = 54.90) and might have been outside the realm of their experience. Items PFLSS 3 (difficulty measure = 46.33) and PFLSS 2 (difficulty measure = 37.15) were the two easiest items to endorse. The items referred to activities that Japanese students normally encounter during their junior high school English classes; participants might have found the items inappropriate for their educational level and therefore inadequate indicators of their ability to speak English.

After extracting the items measuring DSE and PFLSS, the remaining 16 items were analyzed (Table 43). The Rasch model explained 52.5% of the variance (eigenvalue = 17.7); the first contrast explained 9.0% of the variance (eigenvalue = 3.0). All the eight items with high positive loadings above .40 were CECP items, and all items with high negative loadings above .40 were PSV items. To confirm construct unidimensionality, items hypothesized to measure Current English Classroom Perception and items hypothesized to measure Perceived Social Value of Speaking English were analyzed separately with a Rasch PCA of item residuals.

Table 43

Rasch Principal Components Analysis of Item Residuals for the Items Measuring Current English Classroom Perception and Perceived Social Value of Speaking English (16 Items)

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CECP 2	.65	49.61	.71	.71
CECP 3	.63	48.19	.76	.75
CECP 7	.59	52.59	.64	.63
CECP 8	.56	51.17	.64	.64
CECP 5	.37	47.41	.85	.87
CECP 4	.37	50.17	.95	.97
CECP 6	.29	49.77	.94	.96
PSV 3	.03	52.26	.78	.78
PSV 8	.02	59.36	.92	.92
PSV 4	-.56	45.88	1.40	1.55
PSV 1	-.50	53.00	1.64	1.69
PSV 6	-.49	46.60	1.25	1.41
PSV 7	-.43	43.45	1.10	1.11
PSV 2	-.36	49.21	1.08	1.15
PSV 5	-.25	47.81	1.09	1.15
CECP 1	-.12	52.53	1.34	1.47

Note. CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. Measures are in Rasch CHIPs.

Current English Classroom Perception. The eight items hypothesized to measure the Current English Classroom Perception construct were input into the Rasch model and the Rasch PCA of item residuals results were examined (Table 44). The Rasch model explained 64.8% of the variance (eigenvalue = 14.7); the first contrast explained 7.6% of the variance (eigenvalue = 1.7). These results indicated that the eight CECP items formed a fundamentally unidimensional construct. Item CECP 1 had a high positive loading above .40, and items CECP 2, CECP 3, and CECP 7 had high negative loadings. The sole positive loading item concerned a comfort level associated with speaking Japanese in the classroom; this item was the most difficult to endorse (difficulty measure = 54.61) and also misfit the model (Infit mean square = 1.89, outfit mean squared = 1.97). The items with negative

Table 44

Rasch Principal Components Analysis for the Items Measuring Current English Classroom Perception

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CECP 1	.76	54.61	1.89	1.97
CECP 6	.24	49.23	1.10	1.11
CECP 4	.17	49.80	1.11	1.12
CECP 5	.11	45.89	.97	1.01
CECP 2	-.63	49.00	.71	.70
CECP 3	-.57	47.00	.75	.72
CECP 7	-.41	53.26	.71	.72
CECP 8	-.38	51.21	.71	.69

Note. CECP = Current English Classroom Perception. Measures are in Rasch CHIPS.

loadings were related to the comfort level in the classroom. Four items (CECP 6, “My current English class makes me feel relaxed,” CECP 4, “There are many opportunities to speak with classmates in my current English class,” CECP 5, “I get along with my current English classmates,” and CECP 8, “I enjoy doing pair work in my current English class”) did not load above the .40 criterion.

Perceived Social Value of Speaking English. The eight items hypothesized to measure Perceived Social Value of Speaking English were input into Winsteps and the Rasch PCA of item residuals results were examined (Table 45). The Rasch model explained a total variance of 61.2% (eigenvalue = 12.6.), and the first contrast explained 10.2% of the variance (eigenvalue = 2.1). Although the first contrast explained slightly more than the criterion of 10% variance explained, the first contrast eigenvalue was well below the criterion of 3.0. Therefore, I concluded that the eight items measuring Perceived Social Value of Speaking English formed a fundamentally unidimensional construct. Two items (PSV 3 and PSV 8) had high positive loadings above .40, and three items (PSV 4, PSV 7, and PSV 6) had high negative loadings above -.40. The items with positive loadings were concerned with classmates speaking English, and the items with negative loadings were related to the value attached to English speaking skills by Japanese society. Three items (PSV 2, “My friends think that it’s cool to speak English,” PSV 5, “My English teachers

hope that I improve my English speaking ability,” and PSV 1, “My parents hope I become a good speaker of English”) failed to load above the .40 criterion. Based on the failure of the two items concerned with parents and teachers to load strongly on the construct, it could be inferred that the participants paid little attention to the possible wants and hopes of those older than their classroom peers.

Table 45

Rasch Principal Components Analysis for the Items Measuring Perceived Social Value of Speaking English

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
PSV 3	.76	52.61	.79	.78
PSV 8	.73	59.79	.94	.94
PSV 2	.09	49.51	.91	.92
PSV 5	.08	48.10	1.00	1.01
PSV 4	-.69	46.13	1.13	1.14
PSV 7	-.50	43.65	.96	.92
PSV 6	-.49	46.87	1.07	1.17
PSV 1	-.03	53.35	1.27	1.25

Note. PSV = Perceived Social Value of Speaking English. Measures are in Rasch CHIPs.

However, the failure of PSV 2, concerning the image attached to speaking English by their peers, is somewhat confusing, given that similar items about classmates (PSV 3 and PSV 8) had high loadings. Because the items related to a greater society outside the classroom all loaded strongly, it is possible that the participants perceived English speaking from an instrumental orientation. That is, they perceived the need for speaking English in terms of a future job, or as a means

of benefiting society once they graduate from university. Likewise, they might also have perceived fellow classmates' desire to speak English from a similar perspective, meaning that the question of whether classmates find speaking English "cool" was simply not an issue of concern for most of the participants.

Rasch analysis of the six constructs comprising the FLCSC. In this section I present Rasch analysis results for each of the six constructs hypothesized to comprise the higher order latent variables of Foreign Language Classroom Speaking Confidence (Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English) and Favorable Social Conditions (Past English Classroom Experiences, Current English Classroom Environment, and Perceived Social Value of Speaking English). Each construct was analyzed separately for Likert-scale category functioning, and categories were combined when necessary. Items comprising each construct were also analyzed for item difficulty and fit to the construct. Item-person maps were also generated for each construct for a visual representation of the location of persons and items on the construct.

Foreign Language Classroom Speaking Anxiety. Likert scale category functioning was examined for Foreign Language Classroom Speaking Anxiety

(Table 46). The minimum of 10 observations per category was met, as the smallest number of observations was 1128 (category 6). Outfit for all categories was well below the required 2.0. Separation between adjacent thresholds was greater than the required 2.68 CHIPs for a 6-point scale. The smallest gap between the first and second thresholds ($\tau_1 = -12.51$, $\tau_2 = -6.81$) was 5.70 CHIPs, well above the criterion, and there were no disordered thresholds.

Next, Rasch item statistics were obtained for the Foreign Language Classroom Speaking Anxiety construct (Table 47). One item (FLCSA 3, “I’m worried that my partner speaks better English than I do”) had an infit mean square statistic of 1.42 and an outfit mean square statistic of 1.66, indicating possible misfit to the model. Other items showed good fit and ranged in measure difficulty from 43.05 (FLCSA 2) to 60.07 (FLCSA 7). Point-measure correlations, which demonstrate the relative strength of the item in relation to other items in the construct, were high (.60-.76).

Table 46

Category Structure Functioning for Foreign Language Classroom Speaking Anxiety

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	2091	(18)	1.07	1.19	None	(-26.58)
2 Disagree	2257	(19)	.91	.90	-12.51	-12.45
3 Somewhat disagree	2436	(21)	.86	.84	-6.81	-3.78
4 Somewhat agree	2216	(19)	.92	.98	-.41	3.55
5 Agree	1580	(13)	1.04	1.16	6.69	12.47
6 Strongly agree	1128	(10)	1.19	1.26	13.04	(26.95)

Note. Measures are in Rasch CHIPs.

Table 47

Rasch Analysis for the Items Measuring Foreign Language Classroom Speaking Anxiety

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
FLCSA 3	50.41	.29	1.42	9.0	1.66	9.9	.60
FLCSA 4	49.86	.29	1.10	2.4	1.45	8.7	.61
FLCSA 2	43.05	.29	1.22	4.9	1.25	5.0	.66
FLCSA 11	44.69	.29	1.03	.8	1.10	2.2	.69
FLCSA 1	50.94	.29	1.05	1.2	1.04	.9	.70
FLCSA 7	60.07	.34	.98	-.3	.96	-.6	.67
FLCSA 6	44.03	.29	.97	-.8	.97	-.7	.70
FLCSA 10	55.36	.31	.91	-2.0	.87	-2.7	.72
FLCSA 5	48.68	.29	.83	-4.3	.88	-2.8	.75
FLCSA 9	50.58	.29	.85	-3.7	.84	-3.7	.74
FLCSA 8	52.33	.30	.73	-7.0	.71	-6.9	.76

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PMC = point-measure correlation. Measures are in Rasch CHIPS.

The Rasch person reliability estimate (separation) was .87 (2.58). The separation figure was above the minimum recommended level of 2.0, indicating the existence of two distinct groups of persons as measured by the construct. The Rasch item reliability estimate (separation) was 1.00 (15.39), indicating that the item difficulty estimates were well separated in relation to their error estimates.

To examine whether removing the misfitting item FLCSA 3 would improve overall item fit, FLCSA 3 was temporarily removed and the remaining 10 items were analyzed again (Table 48). Item FLCSA 4 slightly misfit the model (Infit mean squared = 1.14, outfit mean squared = 1.54). All other items showed good fit to the model. The Rasch person reliability estimate (separation) was .86 (2.52), and

the item reliability estimate (separation) was 1.00 (16.66). These statistics were not significantly changed from prior to removing item FLCSA 3. Disattenuated person measures for the construct both with and without the item FLCSA 3 were strongly correlated ($r = .99, p < .01$), indicating that deleting item FLCSA 3 resulted in no significant change to the person ability estimates. The Rasch PCA of item residuals analysis was conducted again without item FLCSA 3, and the results were examined (Table 49). The Rasch model explained 76.3% of the variance (eigenvalue = 32.3), compared to 78.2% of the variance with item FLCSA 3 included. The first contrast explained 4.2% of the variance (eigenvalue = 1.8), compared to 4.0% of the variance with the item FLCSA 3 included. The positive loading items had slightly higher loadings, and negative loading items had slightly

Table 48

Rasch Statistics for the Items Measuring Foreign Language Classroom Speaking Anxiety Excluding Item FLCSA 3

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
FLCSA 4	49.90	.30	1.14	3.3	1.54	9.9	.62
FLCSA 1	51.04	.30	1.22	5.0	1.29	5.8	.68
FLCSA 2	42.64	.30	1.24	5.3	1.27	5.4	.67
FLCSA 11	44.40	.30	1.05	1.1	1.15	3.3	.70
FLCSA 7	60.76	.35	1.03	.6	1.01	.3	.68
FLCSA 6	43.69	.30	1.01	.3	1.01	.3	.70
FLCSA 10	55.75	.32	.95	-1.1	.92	-1.8	.72
FLCSA 5	48.64	.30	.85	-3.6	.89	-2.4	.76
FLCSA 9	50.66	.30	.87	-3.3	.85	-3.5	.75
FLCSA 8	52.52	.31	.77	-6.0	.75	-5.8	.76

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PMC = point-measure correlation. Measures are in Rasch CHIPS.

Table 49

*Rasch Principal Components Analysis for the Foreign Language Classroom
Speaking Anxiety Excluding Item FLCSA 3*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
FLCSA 7	.75	63.52	.95	1.02
FLCSA 10	.72	57.44	.90	.88
FLCSA 4	.08	48.67	1.20	1.45
FLCSA 1	.07	52.02	1.17	1.25
FLCSA 2	-.49	41.14	1.27	1.29
FLCSA 11	-.46	43.11	1.06	1.19
FLCSA 6	-.44	41.48	1.09	1.17
FLCSA 5	-.19	48.78	.85	.83
FLCSA 9	-.11	50.75	.94	.90
FLCSA 8	.00	53.10	.75	.76

Note. FLCSA = Foreign Language Classroom Speaking Anxiety. Measures are in Rasch CHIPs.

lower loadings, but all of the items that loaded above the .40 criterion. Because there was no improvement to the measurement of the construct or other items due to removing the item, FLCSA 3 was retained for future analyses.

A comparison of the mean of person ability ($M = 46.84$, $SD = 11.77$) and item endorsability difficulty ($M = 50.00$, $SD = 4.78$) showed that the items were slightly difficult to endorse for some of the participants (Figure 11). The most difficult item to endorse was FLCSA 7 (“I’m afraid my partner will laugh when I speak English with a classmate in a pair,” difficulty measure = 60.07). The easiest item to endorse was FLCSA 2 (“I feel nervous speaking English in front of the entire class,” difficulty measure = 42.64).

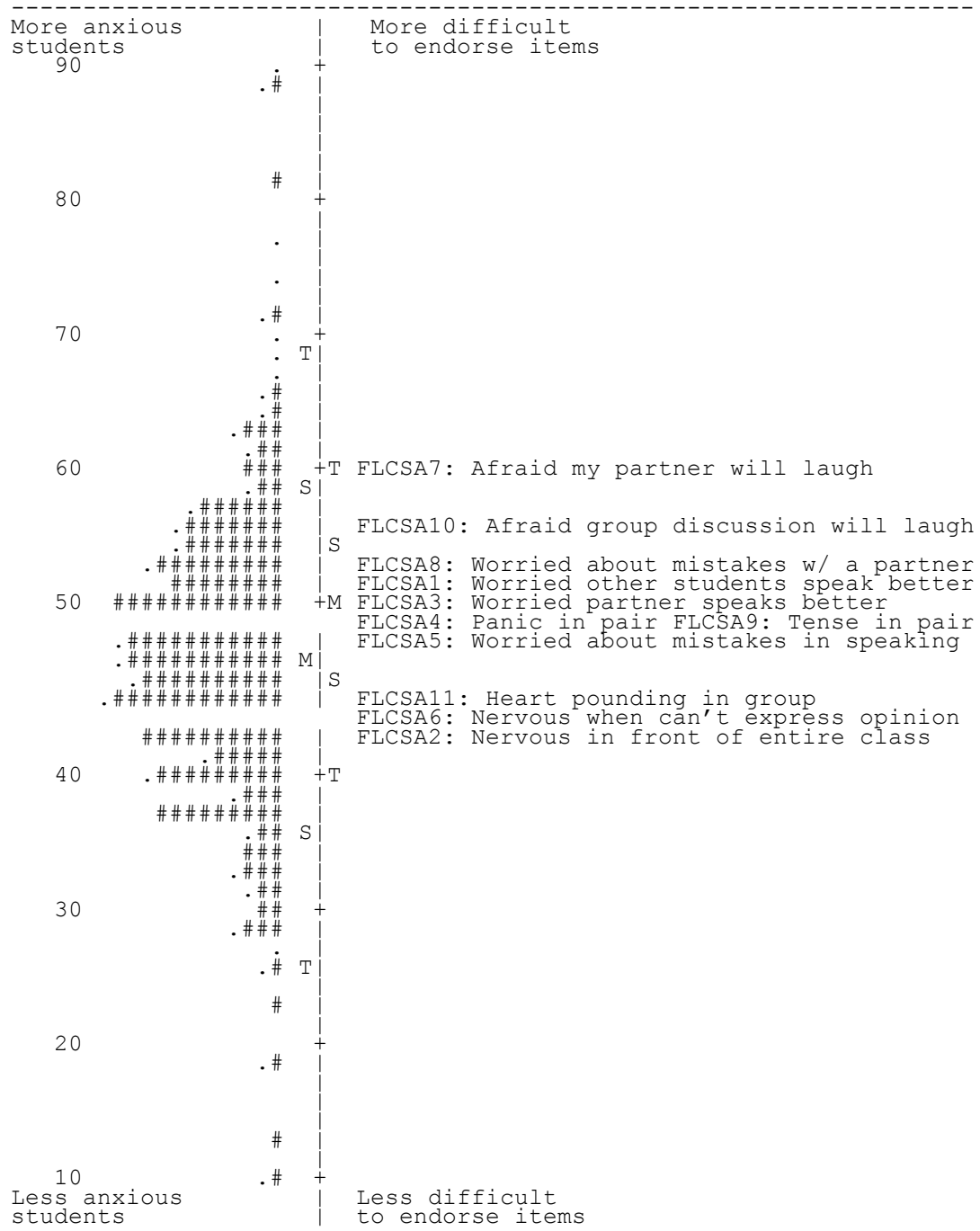


Figure 11. The item-person map of the Foreign Language Classroom Speaking Anxiety construct. Each # represents six persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

Past English Classroom Experiences. Likert scale category functioning was examined for items measuring Past English Classroom Experiences (Table 50). The minimum of 10 observations per category was met, as the smallest number of observations was 942 (category 6). The smallest gap between the fourth and fifth thresholds ($\tau_4 = 7.99$, $\tau_5 = 9.77$) was 1.78 CHIPs, below the criterion of 2.68 CHIPs for a 6-point scale. The categories *Agree* and *Strongly Agree* were combined into one category and the Likert scale category functioning was examined again (Table 51). The results indicated that the separation between the third and fourth thresholds ($\tau_3 = 6.34$, $\tau_4 = 8.20$) was 1.86 CHIPs, which was below the criterion. Categories 4 and 5 were again combined, and the analysis was conducted again (Table 52). The results showed clear separation between all categories, with the smallest gap between the second and third thresholds ($\tau_2 = .11$, $\tau_3 = 7.39$) being 7.28 CHIPs. Therefore, the original 6-point scale was reduced to a 4-point scale. The Outfit MNSQ statistic for all categories was well below the required 2.0, and there were no disordered thresholds.

Next, Rasch item statistics were obtained for the items measuring Past English Classroom Experiences (Table 53). One item (PECE 8, “I got along well with classmates in my high school English classes”) had a standard error of .66 and an infit MNSQ statistic of 1.62, indicating possible misfit to the model. Item PECE 8 was additionally much easier to endorse (difficulty measure = 28.14) than the

Table 50

Category Structure Functioning for Past English Classroom Experiences (6-point Scale)

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	1635	(19)	.99	1.03	None	(-26.22)
2 Disagree	1578	(19)	.83	.82	-11.81	-12.18
3 Somewhat disagree	1862	(22)	.88	.83	-7.84	-3.18
4 Somewhat agree	1384	(16)	.91	1.02	1.89	4.08
5 Agree	942	(11)	.93	.94	7.99	12.06
6 Strongly agree	1043	(12)	1.45	1.58	9.77	(24.84)

Note. Measures are in Rasch CHIPs

Table 51

Category Structure Functioning for Past English Classroom Experiences (5-point Scale)

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	1635	(20)	.97	1.05	None	(-23.96)
2 Disagree	1578	(19)	.87	.87	-9.85	-9.37
3 Somewhat disagree	1862	(22)	.91	.88	-4.69	.32
4 Somewhat agree	1384	(17)	.94	1.09	6.34	9.57
5 Agree	1850	(22)	1.23	1.31	8.20	(23.11)

Note. Measures are in Rasch CHIPs. Categories *Agree* and *Strongly disagree* were combined into Category 5.

Table 52

Category Structure Functioning for Past English Classroom Experiences (4-point Scale)

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	1635	(20)	.97	1.08	None	(-21.14)
2 Disagree	1578	(20)	.91	.91	-7.50	-5.83
3 Somewhat disagree	1862	(23)	.92	.90	.11	5.86
4 Agree	2922	(36)	1.15	1.30	7.39	(21.09)

Note. Measures are in Rasch CHIPs. Categories *Somewhat agree*, *Agree*, and *Strongly agree* were combined into Category 4.

other items and had a relatively low point-measure correlation of .45, indicating that it was contributing less to the measurement of the construct than the other items. The other items showed good fit to the model and ranged in measure difficulty from 48.97 (PECE 6) to 66.35 (PECE 4). Point-measure correlations were high (.64-.76). The Rasch person reliability estimate (separation) was .79 (1.93). The reliability estimate was adequate, and the separation figure was slightly below the minimum recommended level of 2.0, indicating the similarity of many of the person ability estimates as measured by the construct. The Rasch item reliability estimate (separation) was 1.00 (19.64), indicating that the item difficulty estimates were well separated in relation to their error estimates.

Table 53

Rasch Statistics for the Items Measuring the Past English Classroom Experiences Construct

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
PECE 8	28.14	.66	1.62	7.3	1.39	2.8	.45
PECE 6	48.97	.42	1.36	7.3	1.45	7.0	.64
PECE 4	66.35	.44	1.17	3.4	1.26	3.5	.70
PECE 1	52.73	.41	.94	-1.4	.98	-.3	.73
PECE 5	53.92	.40	.93	-1.8	.96	-.8	.73
PECE 7	48.04	.42	.85	-3.5	.83	-3.1	.73
PECE 2	51.41	.41	.84	-3.8	.81	-3.7	.75
PECE 3	50.45	.41	.77	-5.7	.72	-5.7	.76

Note. PECE = Past English Classroom Experiences; PMC = point-measure correlation. Measures are in Rasch CHIPs.

To examine whether removing the misfitting item PECE 8 would improve the overall item fit, item PECE 8 was temporarily removed and the remaining 10 items were analyzed again (Table 54). The remaining items showed good fit to the Rasch model. The Rasch person reliability estimate (separation) was .76 (1.80), and the item reliability estimate (separation) was .99 (13.39). Compared to the items prior to removing item PECE 8, the person statistics were not significantly changed. However, because item PECE 8 had been the easiest item to endorse, removing the item significantly decreased the item separation statistic. Disattenuated person measures for the construct both with and without item PECE 8 were strongly correlated ($r = .98, p < .01$), indicating that deleting the item had virtually no effect on the person ability estimates. A Rasch PCA of item residuals analysis was conducted again without item PECE 8, and the results were examined (Table 55). The Rasch model explained 75.6% of the variance (eigenvalue = 21.6), compared to 80.8% of the variance with item PECE 8 included. The first contrast explained 6.2% of the variance (eigenvalue = 1.8), compared to 4.3% of the variance with item PECE 8 included. Item PECE 1 (“I felt encouraged to speak English in my high school English classes”), which had formerly loaded negatively above .40, failed to load above the .40 criterion. Because there was no improvement to the construct or other items due to removing the item, PECE 8 was retained for future analyses.

Table 54

Rasch Statistics for the Items Measuring the Past English Classroom Experiences Construct Excluding Item PECE 8

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
PECE 6	45.64	.43	1.42	8.3	1.46	7.2	.66
PECE 4	63.83	.45	1.19	3.8	1.24	3.5	.69
PECE 5	50.86	.41	.95	-1.1	.96	-.8	.74
PECE 1	49.61	.42	.96	-1.0	.95	-.9	.74
PECE 7	44.64	.43	.90	-2.3	.89	-2.0	.74
PECE 2	48.22	.42	.89	-2.5	.87	-2.5	.75
PECE 3	47.20	.43	.82	-4.3	.80	-4.0	.76

Note. PECE = Past English Classroom Experiences; PMC = point-measure correlation. Measures are in Rasch CHIPS.

Table 55

Rasch Principal Components Analysis for the Items Measuring the Past English Classroom Experiences Excluding Item PECE 8

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
PECE 6	.74	45.64	1.42	1.46
PECE 7	.62	44.64	.90	.89
PECE 5	.21	50.86	.95	.96
PECE 3	-.58	47.20	.82	.80
PECE 4	-.47	63.83	1.19	1.24
PECE 2	-.36	48.22	.89	.87
PECE 1	-.32	49.61	.96	.95

Note. PECE = Past English Classroom Experiences. Measures are in Rasch CHIPS.

A comparison of the mean of person ability ($M = 56.81$, $SD = 16.47$) and item endorsability difficulty ($M = 50.00$, $SD = 9.84$) indicated that the items were less difficult to endorse than the ability level of the participants (Figure 12). The most difficult item to endorse was PECE 4 (“I spoke English frequently with classmates

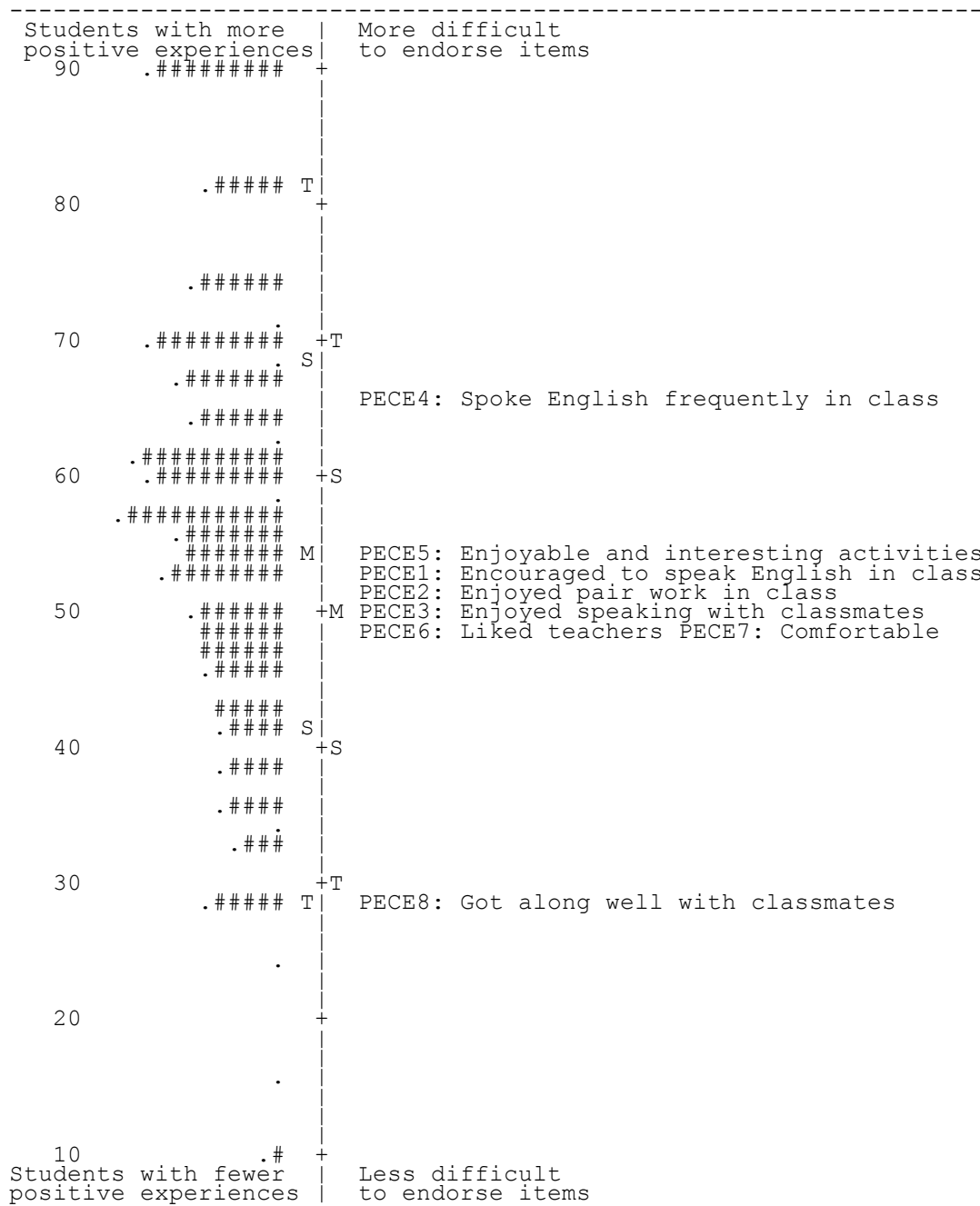


Figure 12. The item-person map of the Past English Classroom Experiences construct. Each # represents eight persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

during my high school English classes,” difficulty measure = 66.35). The easiest item to endorse was PECE 8 (“I got along well with classmates in my high school English classes,” difficulty measure = 28.14).

Desire to Speak English. Likert scale category functioning was examined for Desire to Speak English (Table 56). The minimum of 10 observations per category was met, as the smallest number of observations was 275 (category 1). The smallest gap between the first and second thresholds ($\tau_1 = -18.04$, $\tau_2 = -11.06$) was 7.01 CHIPs, well above the criterion of 2.68 CHIPs. The Outfit MNSQ statistic for all categories was below the required 2.0; however, category 1 (*Strongly disagree*) approached this value (Outfit MNSQ = 1.95). There were no disordered thresholds.

Table 56

Category Structure Functioning for Desire to Speak English

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	275	(3)	1.63	1.95	None	(-31.78)
2 Disagree	569	(7)	1.14	1.29	-18.07	-16.24
3 Somewhat disagree	1234	(15)	.94	.98	-11.06	-4.88
4 Somewhat agree	1657	(20)	.88	.84	.91	5.40
5 Agree	1802	(21)	.89	.84	11.05	16.09
6 Strongly agree	2935	(35)	.94	.95	17.18	(31.08)

Note. Measures are in Rasch CHIPs.

Next, Rasch item statistics were obtained for the items measuring Desire to Speak English (Table 57). All items showed good fit and ranged in measure

difficulty from 35.95 (DSE 2) to 56.09 (DSE 8). Point-measure correlations were high (.63-.83). The Rasch person reliability estimate (separation) was .86 (2.51). The separation figure was well above the minimum recommended level of 2.0, indicating the existence of two distinct groups of persons as measured by the construct. The Rasch item reliability estimate (separation) was 1.00 (17.12), indicating that the item difficulty estimates were well separated in relation to their error estimates.

Table 57

Rasch Statistics for the Items Measuring the Desire to Speak English Construct

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
DSE 5	48.94	.39	1.33	6.2	1.18	2.9	.73
DSE 6	55.97	.37	1.10	2.1	1.24	4.5	.78
DSE 2	35.95	.50	1.15	2.5	1.01	.1	.63
DSE 9	55.88	.37	1.08	1.7	1.10	1.9	.79
DSE 4	39.12	.46	.93	-1.3	1.05	.7	.68
DSE 7	55.71	.37	.93	-1.4	.97	-.7	.81
DSE 1	51.74	.38	.89	-2.4	.84	-3.0	.80
DSE 3	50.60	.38	.88	-2.7	.85	-2.8	.79
DSE 8	56.09	.37	.82	-4.0	.83	-3.6	.83

Note. DSE = Desire to Speak English; PMC = point-measure correlation. Measures are in Rasch CHIPS.

A comparison of the mean of person ability ($M = 68.19$, $SD = 19.35$) and item endorsability difficulty ($M = 50.00$, $SD = 7.15$) indicated that the items were less difficult to endorse than the ability level of the participants (Figure 13). The most difficult item to endorse was DSE 6 (“I want to speak English as much as possible

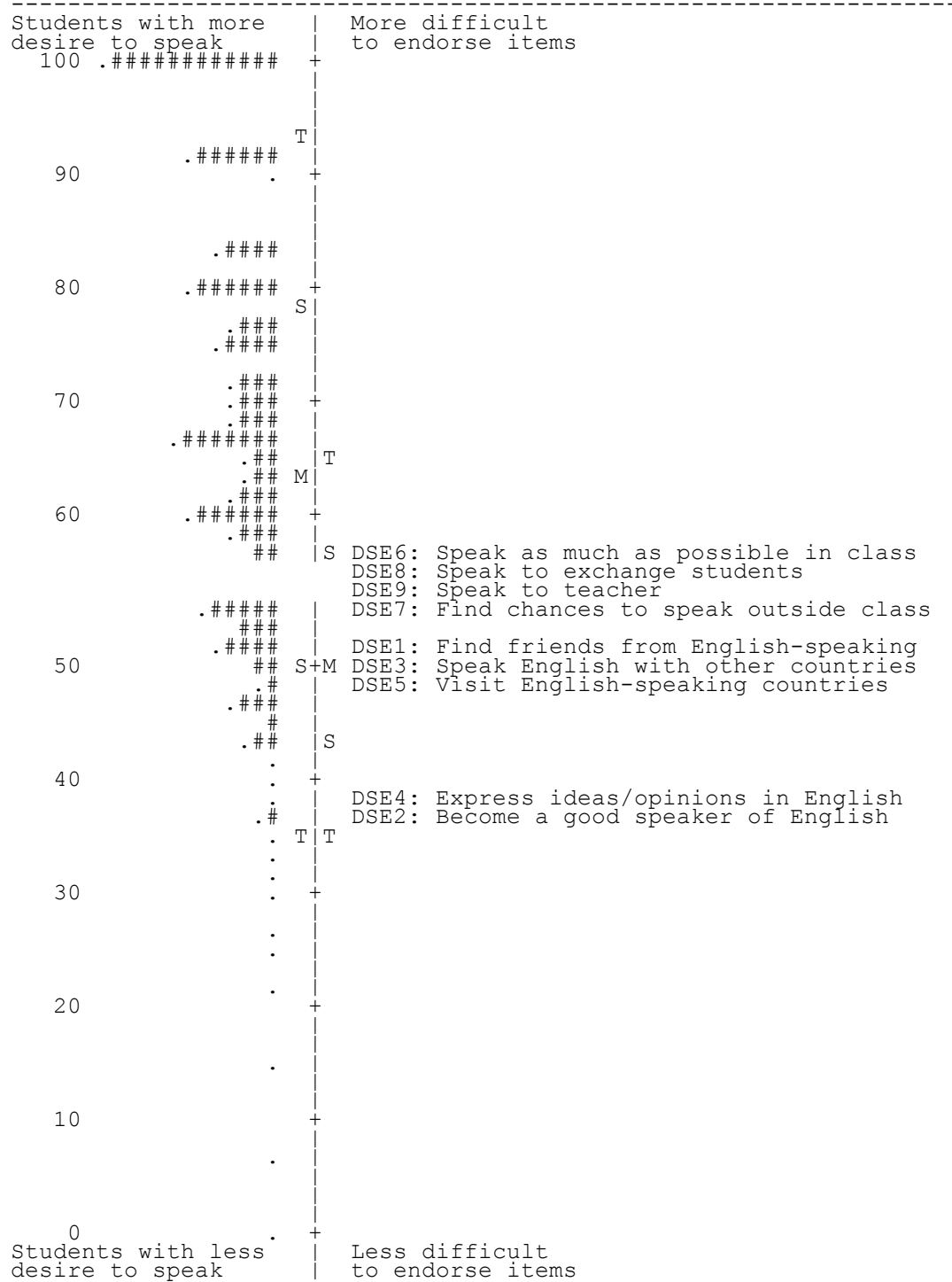


Figure 13. The item-person map of the Desire to Speak English construct. Each # represents ten persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

in my English class,” difficulty measure = 55.97). The easiest item to endorse was DSE 2 (“I want to become a good speaker of English,” difficulty measure = 35.95).

Perceived Foreign Language Speaking Self-Competence. Likert scale category functioning was examined for Perceived Foreign Language Speaking Self-Competence (Table 58). The minimum of 10 observations per category was met, as the smallest number of observations was 623 (category 6). The smallest gap between the fourth and fifth thresholds ($\tau_4 = 12.99$, $\tau_5 = 19.41$) was 6.42 CHIPs, well above the criterion of 2.68 CHIPs. The Outfit MNSQ statistic for all categories was below the required 2.0, and there were no disordered thresholds.

Next, Rasch item statistics were obtained for the items measuring Perceived Foreign Language Speaking Self-competence (Table 59). All items showed good fit and ranged in measure difficulty from 37.13 (PFLSS 2) to 55.34 (PFLSS 5). Point-measure correlations were high (.71-.83). The Rasch person reliability

Table 58
Category Structure Functioning for Perceived Foreign Language Speaking Self-Competence

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	1276	(15)	1.03	1.02	None	(-35.17)
2 Disagree	1973	(23)	.88	.85	-22.37	-17.77
3 Somewhat disagree	2068	(24)	.85	.83	-10.09	-5.08
4 Somewhat agree	1673	(20)	.95	.98	.06	6.17
5 Agree	839	(10)	.96	.97	12.99	17.78
6 Strongly agree	623	(7)	1.48	1.52	19.41	(33.22)

Note. Measures are in Rasch CHIPs.

Table 59

Rasch Statistics for the Items Measuring Perceived Foreign Language Speaking Self-Competence

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
PFLSS 6	54.91	.37	1.32	6.7	1.35	6.8	.71
PFLSS 7	47.80	.35	1.05	1.2	1.10	2.2	.75
PFLSS 8	54.84	.37	1.08	1.8	1.01	.3	.79
PFLSS 3	46.33	.35	1.00	-.1	1.06	1.4	.75
PFLSS 1	53.97	.36	1.03	.7	.98	-.4	.78
PFLSS 2	37.13	.35	1.00	.1	.97	-.6	.76
PFLSS 5	55.34	.37	.82	-4.5	.80	-4.7	.81
PFLSS 4	49.68	.35	.71	-7.5	.70	-7.4	.83

Note. PFLSS = Perceived Foreign Language Speaking Self-Competence; PMC = point-measure correlation. Measures are in Rasch CHIPS.

estimate (separation) was .88 (2.73). The separation figure was well above the minimum recommended level of 2.0, indicating the existence of two distinct groups of persons as measured by the construct. The Rasch item reliability estimate (separation) was 1.00 (15.82), indicating that the item difficulty estimates were well separated in relation to their error estimates.

A comparison of the mean of person ability ($M = 44.65$, $SD = 16.16$) and item endorsability difficulty ($M = 50.00$, $SD = 5.87$) indicated that the items were more difficult to endorse than the ability level of the participants (Figure 14). The most difficult item to endorse was PFLSS 5 (“I can give an English presentation as part of a group in front of my class,” difficulty measure = 55.34). The easiest item to

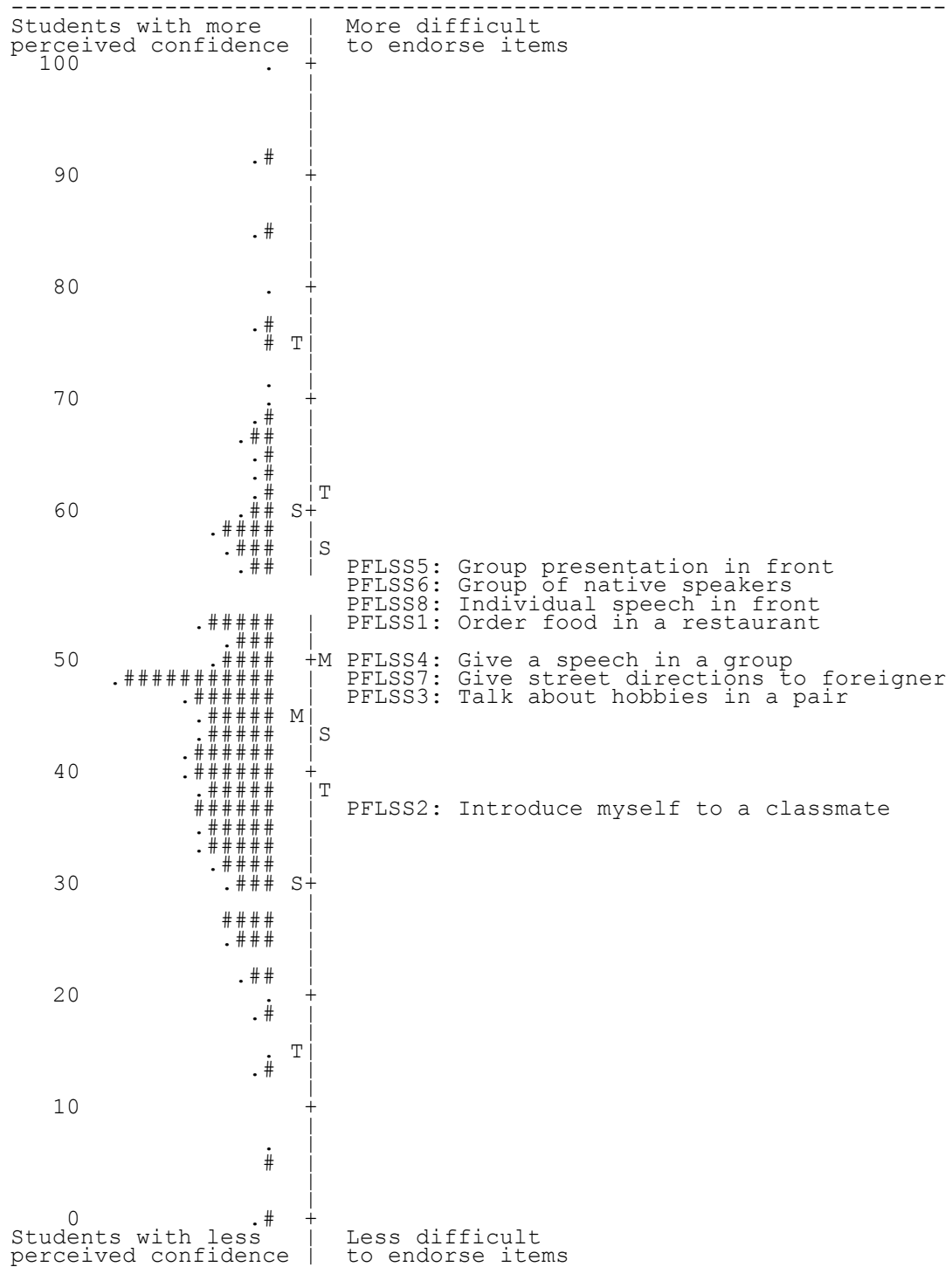


Figure 14. The item-person map of Perceived Foreign Language Speaking Self-Competence. Each # represents eight persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

endorse was PFLSS 2 (“I can introduce myself in English during pair work to a classmate,” difficulty measure = 37.13).

Current English Classroom Perception. Likert scale category functioning was examined for Current English Classroom Perception (Table 60). The minimum of 10 observations per category was met, as the smallest number of observations was 197 (category 1). The smallest gap between the fourth and fifth thresholds ($\tau_4 = 12.74$, $\tau_5 = 17.41$) was 4.67 CHIPs, well above the criterion of 2.68 CHIPs. The Outfit MNSQ statistic for all categories was below the required 2.0 criterion, and there were no disordered thresholds.

Next, Rasch item statistics were obtained for the items measuring Current English Classroom Perception (Table 61). Item CECP 1 (“I feel comfortable expressing my opinions in Japanese in my current English class”) misfit the model (Infit MNSQ = 1.89, Outfit MNSQ = 1.97) and had a relatively low point-measure correlation of .48. Other items showed good fit and ranged in measure difficulty from 45.89 (CECP 5) to 53.26 (CECP 7). Point-measure correlations were high (.66-.79). The Rasch person reliability estimate (separation) was .81 (2.08). The separation figure was well above the minimum recommended level of 2.0, indicating the existence of two distinct groups of persons as measured by the construct. The Rasch item reliability estimate (separation) was .98 (7.51), indicating

Table 60

Category Structure Functioning for Current English Classroom Perception

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	197	(2)	1.53	1.60	None	(-32.09)
2 Disagree	586	(7)	1.12	1.18	-18.36	-16.66
3 Somewhat disagree	1590	(19)	.90	.92	-11.42	-5.33
4 Somewhat agree	2471	(29)	.81	.78	-.37	5.54
5 Agree	1837	(22)	.92	.92	12.74	16.74
6 Strongly agree	1717	(20)	1.01	1.02	17.41	(31.60)

Note. Measures are in Rasch CHIPs.

Table 61

Rasch Statistics for the Items Measuring Current English Classroom Perception

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
CECP 1	54.61	.33	1.89	9.9	1.97	9.9	.48
CECP 4	49.80	.34	1.11	2.5	1.12	2.5	.66
CECP 6	49.23	.34	1.11	2.3	1.11	2.3	.66
CECP 5	45.89	.35	.97	-.6	1.01	.2	.68
CECP 3	47.00	.35	.75	-6.3	.72	-6.7	.75
CECP 7	53.26	.34	.71	-7.4	.72	-7.1	.78
CECP 2	49.00	.34	.71	-7.6	.70	-7.5	.78
CECP 8	51.21	.34	.71	-7.6	.69	-7.7	.79

Note. CECP = Current English Classroom Perception; PMC = point-measure correlation. Measures are in Rasch CHIPs.

that the item difficulty estimates were fairly well separated in relation to their error estimates.

However, item CECP 1, which was the most difficult item to endorse (difficulty measure = 54.61), severely misfit the model and might have exerted a disproportional pull on the other positively loading items. Therefore, CECP 1 was temporarily removed and the Rasch statistics for the remaining items were

examined (Table 62). The item statistics changed slightly, resulting in better item fit for items CECP 3, CECP 8, CECP 7, and CECP 2, and slightly worse fit for items CECP 6, CECP 4, and CECP 5. Item difficulty measures also changed minimally, but the hierarchy of item difficulty remained unchanged. Point-measure correlations also improved slightly. The Rasch person reliability estimate (separation) was .83 (2.20), compared to .81 (2.08) prior to removing item CECP 1. The Rasch item reliability estimate (separation) was .98 (6.86), compared to .98 (7.51) prior to removing item CECP 1.

Table 62

*Rasch Statistics for the Items Measuring Current English Classroom Perception
Excluding Item CECP 1*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
CECP 6	49.85	.38	1.34	7.0	1.38	7.4	.67
CECP 4	50.55	.38	1.30	6.3	1.35	6.9	.69
CECP 5	45.84	.39	1.13	2.9	1.18	3.5	.70
CECP 3	47.18	.38	.84	-3.9	.81	-4.3	.77
CECP 8	52.26	.37	.82	-4.3	.82	-4.3	.80
CECP 7	54.73	.37	.78	-5.6	.78	-5.6	.81
CECP 2	49.59	.38	.75	-6.2	.70	-6.2	.80

Note. CECP = Current English Classroom Perception; PMC = point-measure correlation. Measures are in Rasch CHIPS.

An additional PCA of item residuals analysis was conducted on the remaining seven CECP items to determine whether there would be any changes in construct dimensionality (Table 63). The Rasch model explained 68.2% of the variance

(eigenvalue = 15.0), compared to 64.8% prior to removing item CECP 1. The first contrast explained 7.0% of the variance (eigenvalue = 1.5), compared to 7.6% prior to removing item CECP 1. Two items (CECP 2 and CECP 3) had high positive loadings above .40, and two items (CECP 4 and CECP 5) had high negative loadings above .40. The items with positive loadings were related to comfort level in the classroom, and the items with negative loadings were related to interacting directly with classmates. Three items (CECP 8, “I enjoy doing pair work in my current English class,” CECP 7, “I enjoy speaking in English with classmates in my current English class,” and CECP 6, “My current English class makes me feel relaxed”) failed to load above the .40 criterion.

Table 63

Rasch Principal Components Analysis for the Items Measuring Current English Classroom Perception Excluding Item CECP 1

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
CECP 2	.70	49.59	.75	.74
CECP 3	.69	47.18	.84	.81
CECP 8	.13	52.26	.82	.82
CECP 4	-.52	50.55	1.30	1.35
CECP 5	-.51	45.84	1.13	1.18
CECP 7	-.13	54.73	.78	.78
CECP 6	-.06	49.85	1.34	1.38

Note. CECP = Current English Classroom Perception. Measures are in Rasch CHIPs.

According to the PCA of item residuals results, the removal of item CECP 1 did not affect the loadings of items CECP 2 and CECP 3. However, items CECP 4 and CECP 5, which did not load above .40 when including item CECP 1 in the analysis, loaded above .40 after removing CECP 1. Likewise, items CECP 7 and CECP 8, which had formerly loaded above .40, failed to load above .40 after removing item CECP 1. A correlation of disattenuated person measures for the construct both with and without the item CECP 1 showed a strong correlation ($r = .97, p < .01$), indicating that deleting item CECP 1 made virtually no difference to the person ability estimates.

A comparison of the mean of person ability ($M = 59.67, SD = 12.82$) and item endorsability difficulty ($M = 50.00, SD = 2.76$) indicated that the items were less difficult to endorse than the ability level of the participants (Figure 15). All the items fell beneath the mean ability of the participants. The most difficult item to endorse was item CECP 1 (“I feel comfortable expressing my opinions in Japanese in my current English class,” difficulty measure = 54.61). The easiest item to endorse was item CECP 5 (“I get along with my current English classmates,” difficulty measure = 45.84).

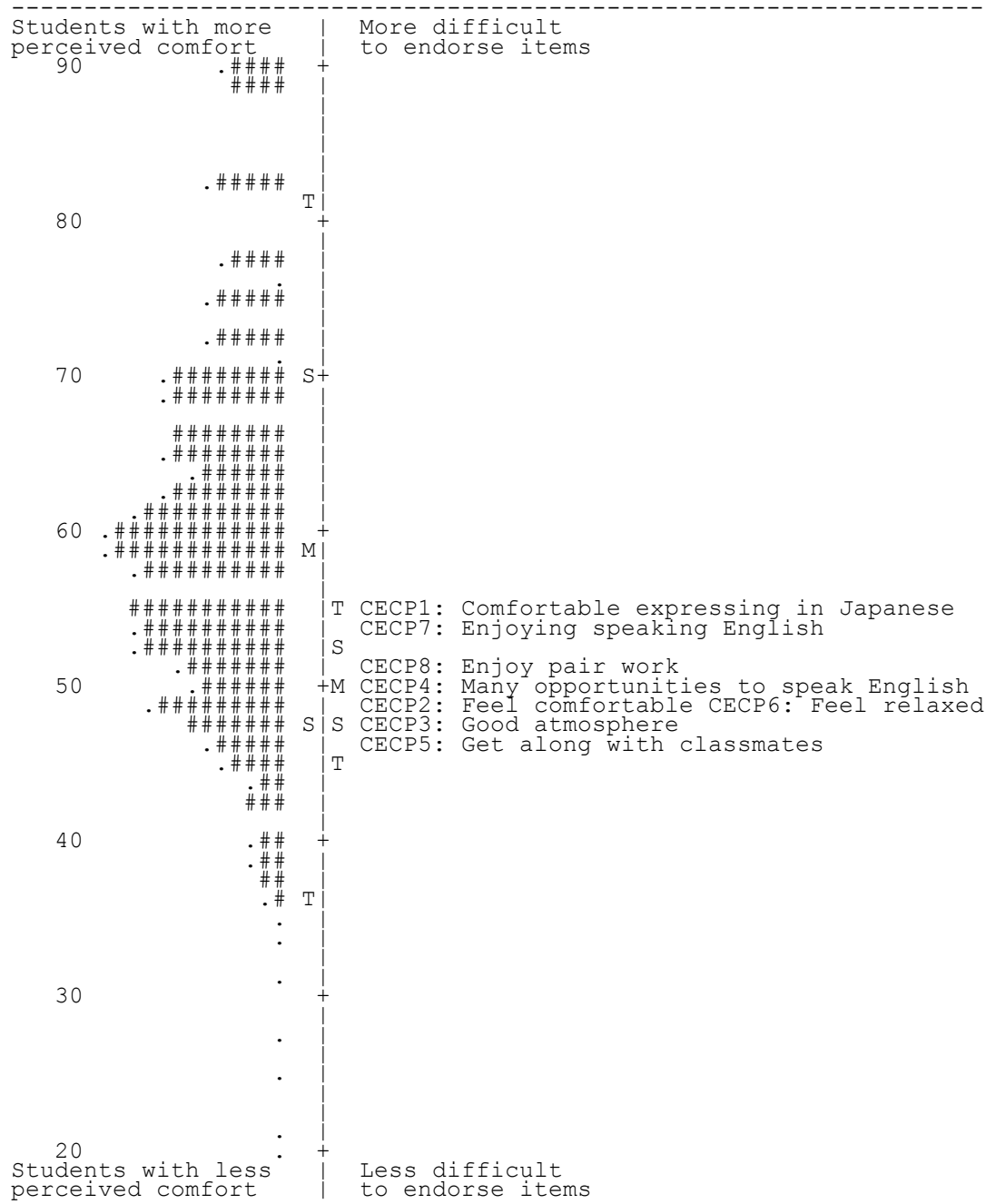


Figure 15. The item-person map of the Current English Classroom Perception construct. Each # represents five persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

Perceived Social Value of Speaking English. Likert scale category

functioning was examined for Perceived Social Value of Speaking English (Table 64). The minimum of 10 observations per category was met, as the smallest number of observations was 359 (category 1). The smallest gap between the first and second thresholds ($\tau_1 = -10.45$, $\tau_2 = -8.54$) was 1.91 CHIPS, which was under the criterion of 2.68 CHIPS. Therefore, Category two (*Disagree*) and Category three (*Somewhat disagree*) were combined and the analysis was conducted again (Table 65). The results indicated that the gaps between the second, third, and fourth thresholds were less than 2.68 CHIPS ($\tau_2 = 4.77$, $\tau_3 = 6.88$, $\tau_4 = 9.08$). Category four (*Somewhat agree*) and Category five (*Agree*) were combined and the Likert scale category functioning analysis was conducted again (Table 66). The resulting 4-point scale had thresholds that surpassed the minimum 2.68 CHIPS criterion ($\tau_1 = -21.06$, $\tau_2 = -1.79$, $\tau_3 = 22.85$). The Outfit MNSQ statistic for all categories was below the 2.0 criterion, and there were no disordered thresholds. Therefore, in the place of the original 6-point Likert scale, a 4-point Likert scale was used for this construct.

Next, Rasch item statistics were obtained for the items measuring Perceived Social Value of Speaking English (Table 67). All items showed good fit and ranged in measure difficulty from 44.74 (PSV 7, “Speaking English will help me become an internationally-minded person”) to 58.64 (PSV 8, “My classmates want to speak

Table 64

*Category Structure Functioning for Perceived Social Value of Speaking English
(6-point Scale)*

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	359	(4)	1.45	1.53	None	(-25.39)
2 Disagree	636	(7)	.93	1.00	-10.45	-12.13
3 Somewhat disagree	1443	(17)	.84	.84	-8.54	-3.64
4 Somewhat agree	1960	(23)	.89	.90	.53	3.65
5 Agree	1927	(23)	.83	.86	7.72	12.08
6 Strongly agree	2187	(26)	1.07	1.07	10.74	(25.44)

Note. Measures are in Rasch CHIPs.

Table 65

*Category Structure Functioning for Perceived Social Value of Speaking English
(5-point Scale)*

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	359	(4)	1.39	1.46	None	(-31.76)
2 Disagree	2079	(24)	.96	1.02	-20.73	-8.72
3 Somewhat agree	1960	(23)	.92	.88	4.77	3.97
4 Agree	1927	(23)	.85	.87	6.88	10.44
5 Strongly agree	2187	(26)	1.04	1.09	9.08	(23.75)

Note. Measures are in Rasch CHIPs. Categories *Disagree* and *Somewhat disagree* were combined into Category 2.

in English during class”). Item PSV 6 (“Japanese companies think highly of workers who have English speaking ability”) had a relative low point-measure correlation (.49); the other items had moderate point-measure correlations (.56-.66). The Rasch person reliability estimate (separation) was .71 (1.55). The separation figure was below the minimum recommended level of 2.0, indicating the similarity of many of the person ability estimates as measured by the construct. The Rasch

Table 66

*Category Structure Functioning for Perceived Social Value of Speaking English
(4-point Scale)*

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure Measure	Category Measure
1 Strongly disagree	359	(4)	1.40	1.35	None	(-32.11)
2 Disagree	2079	(24)	.92	.94	-21.06	-10.02
3 Agree	3887	(46)	.90	.88	-1.79	5.51
4 Strongly agree	2187	(26)	1.04	1.08	22.85	(21.28)

Note. Measures are in Rasch CHIPs. Categories *Disagree* and *Somewhat disagree* were combined into Category 2. Categories *Somewhat agree* and *Agree* were combined into Category 3.

Table 67

*Rasch Statistics for the Items Measuring the Perceived Social Value of Speaking
English Construct*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
PSV 1	52.45	.26	1.21	5.1	1.19	4.4	.61
PSV 6	47.62	.26	1.03	.9	1.15	3.4	.49
PSV 4	46.45	.27	1.06	1.5	1.08	1.9	.56
PSV 8	58.64	.28	1.05	1.3	.98	-.3	.57
PSV 5	48.44	.26	.97	-.7	.98	-.5	.60
PSV 2	49.47	.26	.94	-1.6	.97	-.8	.63
PSV 7	44.74	.27	.89	-2.7	.90	-2.2	.59
PSV 3	52.19	.26	.89	-3.0	.86	-3.6	.66

Note. PSV = Perceived Social Value of Speaking English; PMC = point-measure correlation. Measures are in Rasch CHIPs.

item reliability estimate (separation) was 1.00 (14.98), indicating that the item difficulty estimates were well separated in relation to their error estimates. A comparison of the mean of person ability ($M = 59.67$, $SD = 12.82$) and item

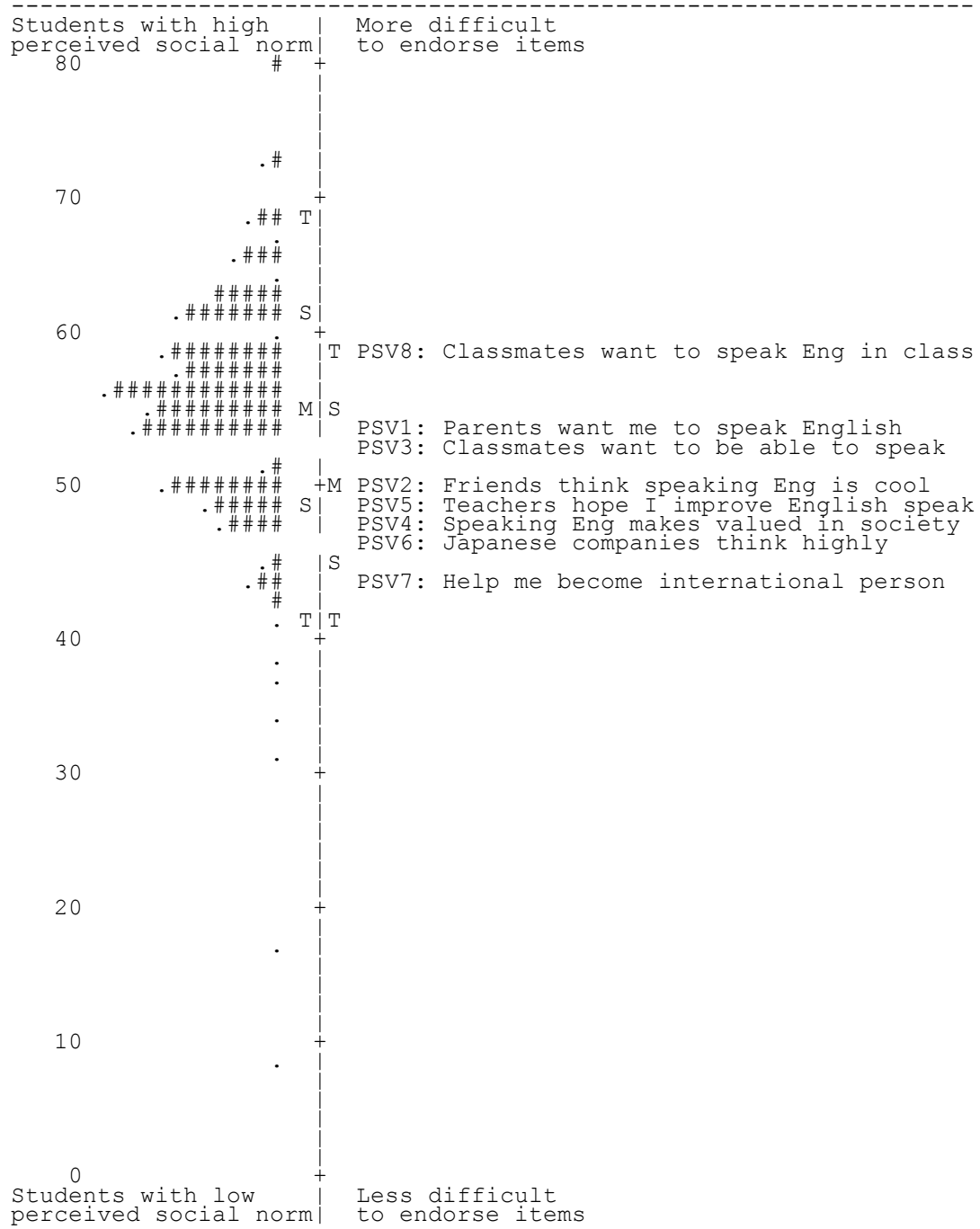


Figure 16. The item-person map of the Perceived Social Value of Speaking English construct. Each # represents ten persons. Each . represents one person. M stands for mean. S stands for one standard deviation from the mean. T stands for two standard deviations from the mean.

endorsability difficulty ($M = 50.00$, $SD = 4.09$) indicated that the items were relatively easy to endorse for many of the participants (Figure 16). All the items except for one (item PSV 8, “My classmates want to speak English during class”) fell beneath the mean ability of the participants. The most difficult item to endorse was item PSV 8 (“My classmates want to speak English during class,” difficulty measure = 58.64). The easiest item to endorse was item PSV 7 (“Speaking English will help me become an internationally-minded person,” difficulty measure = 44.74).

Summary of Rasch results for the FLCSC constructs. According to the Rasch PCA of item residual results, the Rasch model explained variance greater than Linacre’s 2007 unidimensional criterion of 50% for all six constructs comprising the Foreign Language Classroom Speaking Confidence questionnaire (Table 68). The first contrast of the Perceived Social Value of Speaking English construct explained 10.2% of the variance, slightly above the 10% criterion set as an indication of the existence of a secondary dimension within the item residuals. However, the Rasch eigenvalue was 2.1, which was well below the 3.0 criterion. Moreover, the Rasch model explained 61.2% of the variance for the construct, indicating that the items measuring Perceived Social Value of Speaking English formed a fundamentally unidimensional construct.

First residual contrasts for the other five constructs had explained variances ranging from 3.0% to 7.6%, and the eigenvalues were all under the 3.0 criterion. Thus, based on Linacre's 2007 criteria of 50% or more variance explained by the Rasch model, less than 10% variance explained by the first contrast, or a first contrast eigenvalue of less than 3.0, all the constructs should be considered fundamentally unidimensional.

Table 68

Explained and Unexplained Variance for the Constructs Including Variance Explained by the First Contrast

Construct	Rasch Eigenvalue	Variance Explained	Unexplained Variance	First Contrast Eigenvalue	First Contrast Explained Variance
FLCSA	39.4	78.2%	21.8%	2.0	4.0%
PECE	33.6	80.8%	19.2%	1.8	4.3%
DSE	47.9	84.2%	15.8%	1.7	3.0%
PFLSS	33.3	80.6%	19.4%	1.7	4.1%
CECP	14.7	64.8%	35.2%	1.7	7.6%
PSV	12.6	61.2%	38.8%	2.1	10.2%

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PECE = Past English Classroom Experiences; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English.

After the Rasch PCA of item residuals analysis, a rating scale analysis was conducted on the six constructs. Likert scale category functioning analysis revealed that the six-point scale was not functioning adequately for the Past English Classroom Experiences and Perceived Social Value of Speaking English constructs.

For Past English Classroom Experiences, Categories 4 (*Somewhat agree*), 5 (*Agree*), and 6 (*Strongly agree*) were combined into one category to create a 4-point scale.

For Perceived Social Value of Speaking English, Categories 2 (*Disagree*) and 3 (*Somewhat disagree*) were combined into one category, and Categories 4 (*Somewhat agree*) and 5 (*Agree*) were combined into one category, resulting in a 4-point scale.

The Rasch item analysis for the six hypothesized constructs revealed that three items misfit their intended constructs: item FLCSA 3 (“I’m worried that my partner speaks better English than I do”), item PECE 8 (“I got along well with classmates in my high school English classes”), and item CECP 1 (“I feel comfortable expressing my opinions in Japanese in my current English class”). After temporarily removing the items from their respective constructs, Rasch item analysis and PCA were conducted again and the results were examined to determine whether item removal improved the constructs. Removing item FLCSA 3 did not improve the measurement of Foreign Language Classroom Speaking Anxiety, and removing item PECE 8 did not improve the measurement of Past English Classroom Experiences. However, removing item CECP 1 resulted in greater variance accounted for by the Rasch model, improved item fit for other items measuring the Current English Classroom Perception, and improved Rasch

person reliability and separation. The item was noted as a possible problem item and was retained for further analysis.

All six constructs had acceptable Rasch person and item reliability estimates (Table 69). Rasch person reliability estimates ranged from .71 to .88, with Cronbach's alpha estimates ranging from .76 to .93. Perceived Foreign Language Speaking Self-Competence had the highest reliability estimate (person reliability = .88, $\alpha = .93$). Perceived Social Value of Speaking English had the lowest reliability estimate (person reliability = .71, $\alpha = .76$). Item separation was greatest for Past English Classroom Experiences (19.64). Current English Classroom Perception had the smallest spread of items, with an item separation of 7.51.

Table 69

Item and Person Separations and Reliabilities for the Constructs Comprising the Foreign Language Classroom Speaking Confidence Questionnaire

Construct	Item separation	Item reliability	Person separation	Person reliability	Cronbach's Alpha
FLCSA	15.39	1.00	2.58	.87	.91
PECE	19.64	1.00	1.93	.79	.86
DSE	17.12	1.00	2.51	.86	.93
PFLSS	15.82	1.00	2.73	.88	.93
CECP	7.51	.98	2.08	.81	.87
PSV	14.98	1.00	1.55	.71	.76

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PECE = Past English Classroom Experiences; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English.

After concluding the Rasch PCA of item residuals, rating scale, and Rasch item analyses for the constructs measured by the Foreign Language Classroom Speaking Confidence questionnaire, Rasch person measures were obtained for each construct. The measures were correlated in SPSS 15.0 (Table 70). Several constructs were correlated significantly ($p < .01$). Items measuring Foreign Language Classroom Speaking Anxiety had weak correlations with Desire to Speak English, $r = .10$, and

Table 70

Correlational Matrix of Rasch Person Measures for the Six Constructs Comprising the Foreign Language Classroom Speaking Confidence Questionnaire

Constructs	FLCSA	PECE	DSE	PFLSS	CECP	PSV
Foreign Language Classroom Speaking Anxiety	—					
Past English Classroom Experiences	.09*	—				
Desire to Speak English	.10*	.27*	—			
Perceived Foreign Language Speaking Self-Competence	-.11*	.43*	.52*	—		
Current English Classroom Perception	-.03	.20*	.55*	.46*	—	
Perceived Social Value of Speaking English	.22*	.29*	.62*	.38*	.55*	—

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PECE = Past English Classroom Experiences; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. $N = 1,078$.

* = $p < .01$

Perceived Foreign Language Speaking Self-Competence, $r = -.11$. Items measuring Past English Classroom Experiences had a weak correlation with Current English Classroom Perception, $r = .20$, and a moderate correlation with Perceived Foreign Language Speaking Self-Competence, $r = .43$. Desire to Speak English had a strong correlation with Perceived Social Value of Speaking English, $r = .62$. Current English Classroom Perception had moderate correlations with Perceived Foreign Language Speaking Self-Competence, $r = .46$, and Perceived Social Value of Speaking English, $r = .55$.

Exploratory Factor Analysis. The second step in the preliminary analysis of the data involved conducting an exploratory factor analysis to confirm the presence of the five hypothesized factors for the sample population. All items were included in the analysis, after which the results were compared to the Rasch analysis to assist in determining which items to retain for further confirmatory factor analysis. As recommended by Kaiser (1974), a minimum eigenvalue of 1.0 was used as the criterion for identifying factors. A minimum loading of .40 was used a criterion for identifying good performing items (Stevens, 2009).

Data Screening. Descriptive statistics were obtained (Table 71). Several items, most notably those measuring Desire to Speak English (DSE), exceeded 1.0

Table 71

Descriptive Statistics for the Foreign Language Classroom Speaking Confidence Questionnaire

Item	Description	N-size	M	SD	Skew	Kurt.
FLCSA 3	I'm worried that my partner speaks better English than I do.	1,076	3.13	1.58	.22	-1.04
PFLSS 3	I can talk about my hobbies during pair work with a classmate	1,075	3.33	1.34	.15	-.67
DSE 4	I want to be able to express my opinions and ideas in English	1,076	5.31	1.03	-1.74	3.10
PECE 5	I did enjoyable and interesting activities in my high school English classes	1,077	2.91	1.54	.49	-.77
PSV 6	Japanese companies think highly of workers who have English speaking ability	1,072	4.67	1.18	-.80	.32
CECP 4	There are many opportunities to speak with classmates in my current English class	1,070	4.29	1.29	-.39	-.53
FLCSA 7	I'm afraid my partner will laugh when I speak English with a classmate in a pair	1,077	2.19	1.31	1.13	.66
PECE 6	I liked my high school English teachers	1,078	3.33	1.67	.16	-1.13
PFLSS 5	I can give an English presentation as part of a group in front of my class	1,077	2.68	1.39	.65	-.26
DSE 7	I want to find opportunities to speak English outside class	1,076	4.35	1.52	-.56	-.74
CECP 2	I feel comfortable in my current English class	1,077	4.35	1.28	-.35	-.56
PSV 3	My classmates want to be able to speak in English	1,074	4.04	1.36	-.10	-.81
PFLSS 7	I can give street directions in English to a foreigner	1,073	3.23	1.37	.24	-.66
DSE 2	I want to become a good speaker of English	1,075	5.43	1.01	-2.10	4.42
CECP 5	I get along with my current English classmates	1,077	4.59	1.24	-.59	-.32
FLCSA 4	I start to panic when I speak in English with a classmate in a pair	1,077	3.18	1.37	.18	-.68
PECE 2	I enjoyed doing pair work in my high school English classes	1,068	3.06	1.51	.34	-.77
PSV 1	My parents hope I become a good speaker of English	1,076	3.95	1.59	-.29	-.99

(Table 71 continues)

(Table 71 continued)

Item	Description	N-size	M	SD	Skew	Kurt.
DSE 3	I want to speak English with people from different countries	1,077	4.69	1.46	-.85	-.37
FLCSA 10	I'm afraid that others in a group discussion will laugh if I speak English	1,077	2.61	1.47	.68	-.44
PSV 4	Speaking English will help me become a valued member of society	1,078	4.74	1.30	-.97	.42
PECE 7	High school English classes had a comfortable atmosphere	1,078	3.32	1.53	.18	-.90
PFLSS 1	I can order food in English in a restaurant	1,078	2.77	1.43	.64	-.38
CECP 7	I enjoy speaking in English with classmates in my current English class	1,078	4.00	1.31	-.04	-.69
PSV 7	Speaking English will help me become an internationally-minded person	1,077	4.97	1.14	-1.14	1.10
FLCSA 9	I feel tense when I have to speak English with a classmate in a pair	1,077	3.10	1.52	.21	-.96
DSE 6	I want to speak English as much as possible in my English class	1,077	4.34	1.32	-.36	-.63
PECE 8	I got along well with classmates in my high school English classes	1,077	4.68	1.35	-.89	.09
PFLSS 8	I can give an individual speech in English in front of my class	1,076	2.71	1.53	.68	-.52
CECP 3	The current English class atmosphere is good	1,074	4.50	1.23	-.46	-.43
PSV 2	My friends think that it's cool to speak English	1,072	4.39	1.32	-.47	-.42
FLCSA 6	I feel nervous when I can't express my opinion in English	1,072	3.82	1.48	-.24	-.84
CECP 1	I feel comfortable expressing my opinions in Japanese in my current class	1,078	3.89	1.36	-.20	-.61
DSE 1	I want to find friends from English-speaking countries	1,078	4.62	1.47	-.74	-.54
CECP 8	I enjoy doing pair work in my current English class	1,072	4.17	1.32	-.30	-.52
PFLSS 2	I can introduce myself in English during pairwork with a classmate	1,078	4.04	1.37	-.28	-.65
PECE 1	I felt encouraged to speak English in my high school English classes	1,076	3.01	1.59	.44	-.82
PSV 5	My English teachers hope that I improve my English speaking ability	1,077	4.55	1.29	-.72	-.03

(Table 71 continues)

(Table 71 continued)

Item	Description	N-size	M	SD	Skew	Kurt.
FLCSA 8	I'm worried about making mistakes when I speak English with a partner	1,073	2.92	1.48	.40	-.77
PECE 3	I enjoyed speaking with classmates in my high school English classes	1,067	3.11	1.48	.36	-.67
DSE 5	I want to visit English-speaking countries	1,077	4.79	1.52	-1.08	.02
CECP 6	My current English class makes me feel relaxed	1,074	4.33	1.26	-.37	-.54
FLCSA 2	I feel nervous speaking English in front of the entire class	1,076	3.93	1.56	-.32	-.91
DSE 8	I want to speak English with foreign exchange students at my university	1,074	4.32	1.51	-.48	-.83
FLCSA 5	I'm worried about making mistakes while speaking English	1,074	3.31	1.57	.13	-1.05
PFLSS 4	I can give an English speech in a group of my classmates	1,076	3.08	1.36	.36	-.43
PECE 4	I spoke English frequently with classmates during my high school English classes	1,077	2.07	1.31	1.33	1.26
FLCSA 11	I can feel my heart pounding when it's my turn to speak English in a group	1,077	3.75	1.53	-.16	-.99
DSE 9	I want to speak in English to my English teacher	1,071	4.34	1.46	-.47	-.77
PFLSS 6	I can easily join in a conversation in a group of native English speakers	1,077	2.70	1.39	.61	-.29
PSV 8	My classmates want to speak in English during class	1,076	3.18	1.34	.39	-.39
FLCSA 1	I'm worried that other students in class speak English better than I do	1,078	3.07	1.59	.30	-1.00

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English.

kurtosis; however, Tabachnick and Fidell (2007) noted that problems in variance analysis due to positive kurtosis or negative kurtosis diminish with sample sizes of over 100 and 200, respectively (p. 80). The *N*-size was found to vary between 1,067

and 1,078, depending on the item; such large *N*-sizes made problems associated with kurtosis highly unlikely for the sample population. There were 131 missing data points, or 0.2% of the total 56,056 data points (1,078 cases times 52 items). Because this percentage was well below the maximum 5% of missing data points recommended for maintaining the structure of correlational matrices in factor analysis procedures (Tabachnick & Fidell, 2007, p. 63), the mean raw scores for each item were substituted in the place of missing data points prior to EFA and CFA. The data were also examined for statistical outliers (z -scores ≥ 3.29). Three persons were discovered to be above this number; two persons were statistical outliers on all Desire to Speak English items, while one person was a statistical outlier on all Desire to Speak English items and Current English Classroom Perception items. However, given the sample size of 1,078, the number of outliers was smaller than could be accounted for by random chance (Stevens, 2009, p. 14). Therefore, I decided to retain the three persons for further analyses. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .942 and Bartlett's Test of Sphericity was significant at .000, supporting factorability of the matrix. Thus, normality assumptions of the variables were checked and confirmed for the data prior to factor analysis.

Factor analysis. After confirming the normality assumptions, a factor analysis was conducted using principal axis factoring. Because the factors were hypothesized to have large correlations, oblique rotation using direct oblimin was conducted. A six-factor solution was produced (Table 72). One PSV item failed to load onto the intended factor and four PSV items failed to load above the .40 criterion onto any factor. Item PSV 3 (“My classmates want to be able to speak in English”) loaded above the .40 criterion onto the Current English Classroom Perception factor. Item PSV 1 (“My parents want me to be a good speaker of English”), item PSV 2 (“My friends think that it’s cool to speak English”), item PSV 5 (“My English teachers hope that I improve my English speaking ability”), and item PSV 8 (“My classmates want to speak in English during class”) failed to load onto the intended factor above the .40 criterion. Item CECP 1 (“I feel comfortable expressing my opinions in Japanese in my current English class”) failed to load onto any factor above the .40 criterion. All other items loaded onto their intended factors at above the .40 criterion.

After the EFA, factor scores were obtained from the six factors and the scores were correlated (Table 73). Several correlations were statistically significant. Desire to Speak English correlated strongly with Perceived Social Value of Speaking English, $r = .60$, and correlated weakly with Foreign Language Classroom Speaking Anxiety, $r = .11$. Foreign Language Classroom Speaking

Table 72

Six-Factor Solution for Exploratory Factor Analysis of the Foreign Language Classroom Speaking Confidence Questionnaire

Item	Item Description	Factor						h^2
		1	2	3	4	5	6	
DSE 8	I want to speak English with foreign exchange students at my university	.89	.01	.03	-.03	-.03	-.09	.73
DSE 3	I want to speak English with people from different countries	.85	.00	-.01	-.09	-.08	.02	.73
DSE 1	I want to find friends from English-speaking countries	.84	.03	.03	-.02	.01	.03	.72
DSE 7	I want to find opportunities to speak English outside class	.78	-.02	.01	-.02	-.12	-.01	.68
DSE 5	I want to visit English-speaking countries	.78	-.04	-.03	-.07	-.02	.07	.61
DSE 9	I want to speak in English to my English teacher	.69	.06	.01	.15	-.08	-.08	.61
DSE 4	I want to be able to express my opinions and ideas in English	.66	.00	.03	.05	.07	.13	.55
DSE 2	I want to become a good speaker of English	.63	.00	.02	.06	.11	.13	.49
DSE 6	I want to speak English as much as possible in my English class	.53	.04	.03	.17	-.18	.05	.58
FLCSA 8	I'm worried about making mistakes when I speak English with a partner	-.02	.80	.01	-.02	-.03	.01	.63
FLCSA 5	I'm worried about making mistakes while speaking English	-.01	.77	-.01	.02	.02	.05	.61
FLCSA 9	I feel tense when I have to speak English with a classmate in a pair	.00	.76	-.02	-.05	.02	-.05	.58
FLCSA 10	I'm afraid that others in a group discussion will laugh if I speak English	-.05	.75	-.02	-.10	-.04	.01	.57
FLCSA 1	I'm worried that other students in class speak English better than I do	-.03	.71	.02	.05	-.09	.10	.53
FLCSA 7	I'm afraid my partner will laugh when I speak English with a classmate in a pair	-.03	.69	.02	-.14	-.08	-.03	.48
FLCSA 6	I feel nervous when I can't express my opinion in English	.11	.66	-.05	.12	.02	.00	.49

(Table 72 continues)

(Table 72 continued)

Item	Item Description	Factor						h^2
		1	2	3	4	5	6	
FLCSA 11	I can feel my heart pounding when it's my turn to speak English in a group	.01	.66	.00	.06	.15	-.05	.46
FLCSA 2	I feel nervous speaking English in front of the entire class	-.01	.60	.00	.11	.17	.00	.41
FLCSA 4	I start to panic when I speak in English with a classmate in a pair	.03	.57	.06	-.08	.09	-.05	.36
FLCSA 3	I'm worried that my partner speaks better English than I do	.04	.56	.03	.05	-.05	.08	.35
PECE 7	High school English classes had a comfortable atmosphere	-.09	.00	.79	-.01	.04	.01	.58
PECE 3	I enjoyed speaking with classmates in my high school English classes	.06	-.05	.77	-.02	-.02	.02	.63
PECE 2	I enjoyed doing pair work in my high school English classes	.02	-.03	.75	.03	.01	.05	.59
PECE 5	I did enjoyable and interesting activities in my high school English classes	-.01	.03	.69	-.04	-.14	-.01	.66
PECE 6	I liked my high school English teachers	.03	.10	.64	.09	.04	-.10	.42
PECE 1	I felt encouraged to speak English in my high school English classes	.07	.04	.63	-.10	-.16	.06	.57
PECE 4	I spoke English frequently with classmates during my high school English classes	-.04	.00	.54	-.19	-.30	.01	.46
PECE 8	I got along well with classmates in my high school English classes	.08	-.09	.44	.24	.19	.12	.32
CECP 3	The current English class atmosphere is good	.03	.01	-.02	.83	.05	-.09	.65
CECP 2	I feel comfortable in my current English class	.07	-.06	-.02	.78	.02	.00	.66
CECP 8	I enjoy doing pair work in my current English class	.13	.00	.07	.71	-.04	.02	.66
CECP 7	I enjoy speaking in English with classmates in my current English class	.12	-.05	.10	.69	-.11	-.03	.69
CECP 4	There are many opportunities to speak with classmates in my current English class	-.10	.02	-.01	.63	-.09	.04	.40

(Table 72 continues)

(Table 72 continued)

Item	Item Description	Factor						h^2
		1	2	3	4	5	6	
CECP 5	I get along with my current English classmates	.00	.01	.04	.62	-.03	.08	.44
CECP 6	My current English class makes me feel relaxed	.18	-.28	.00	.45	-.05	.03	.42
PSV 3	My classmates want to be able to speak in English	.19	.15	-.01	.41	-.19	.09	.47
PSV 8	My classmates want to speak in English during class	.14	.19	.04	.36	-.25	-.01	.39
CECP 1	I feel comfortable expressing my opinions in Japanese in my current English class	-.02	.03	.00	.26	-.12	.09	.12
PFLSS 4	I can give an English speech in a group of my classmates	.01	-.01	.01	.17	-.77	.00	.70
PFLSS 5	I can give an English presentation as part of a group in front of my class	.02	-.07	.10	.08	-.74	-.01	.69
PFLSS 8	I can give an individual speech in English in front of my class	.01	-.08	.10	.06	-.74	-.05	.66
PFLSS 1	I can order food in English in a restaurant	.09	.02	.05	-.06	-.74	-.03	.61
PFLSS 7	I can give street directions in English to a foreigner	.09	-.03	.04	.04	-.63	-.01	.50
PFLSS 6	I can easily join in a conversation in a group of native English speakers	.14	-.06	.06	-.04	-.61	.00	.48
PFLSS 3	I can talk about my hobbies during pair work with a classmate	.05	-.07	.03	.14	-.60	.06	.51
PFLSS 2	I can introduce myself in English during pair work with a classmate	.04	-.08	.02	.32	-.54	.06	.57
PSV 4	Speaking English will help me become a valued member of society	.00	.04	-.03	-.06	.02	.71	.48
PSV 7	Speaking English will help me become an internationally-minded person	.17	.01	-.08	.03	.01	.57	.44
PSV 6	Japanese companies think highly of workers who have English speaking ability	-.04	-.04	.07	.00	.06	.51	.25
PSV 1	My parents hope I become a good speaker of English	.13	.09	.07	-.05	-.21	.34	.29
PSV 2	My friends think that it's cool to speak English	.19	.14	.01	.16	-.06	.28	.31

(Table 72 continues)

(Table 72 continued)

Item	Item Description	Factor						h^2
		1	2	3	4	5	6	
PSV 5	My English teachers hope that I improve my English speaking ability	.19	.05	.00	.22	.00	.26	.28

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. The mean of the raw score was substituted for 131 missing data points. Loadings above the .40 level are in bold face. $N = 1,078$.

Anxiety correlated weakly with Perceived Foreign Language Speaking Self-Competence, $r = .11$. Past English Classroom Experiences had a moderate negative correlation with Perceived Foreign Language Speaking Self-competence, $r = -.43$, and a weak correlation with Current English Classroom Perception, $r = .18$. Several correlations were above the $r = .32$ value recommended as the criterion for using oblique rotation by Tabachnick and Fidell (2007, p. 646), justifying the use of direct oblimin to interpret the factor loadings.

Table 73

Factor Correlation Matrix of the Six Factors Comprising the Foreign Language Classroom Speaking Confidence Questionnaire

Factor name	Factor					
	DSE	FLCSA	PECE	CECP	PFLSS	PSV
Desire to Speak English	—					
Foreign Language Classroom Speaking Anxiety	.11*	—				
Positive English Classroom Experiences	.28*	.05	—			
Current English Classroom Perceptions	.52*	-.03	.18*	—		
Perceived Foreign Language Speaking Self-competence	-.47*	.11*	-.43*	-.33*	—	
Perceived English Speaking Social Norms	.60*	.16*	.28*	.39*	-.14*	—

Note. DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PFLSS = Perceived Foreign Language Speaking Self-Competence; PSV = Perceived Social Value of Speaking English. $N = 1078$.

* $p < .01$

Comparison of the Rasch analysis and EFA results. The Rasch PCA results were compared to the EFA results for the six hypothesized constructs comprising the Foreign Language Classroom Speaking Confidence questionnaire. All six constructs were identified by both the Rasch PCA item residuals analysis and the EFA. Some items were identified by both Rasch analysis and the EFA as misfitting their intended constructs (Table 74). Based on Rasch PCA of item

residuals and EFA loadings, four items were identified as poorly-performing items by both techniques.

Both Rasch PCA and EFA identified items PSV 1, PSV 2, PSV 5, and item CECP 1 as problematic items. Three of the items—PSV 1 (“My parents hope I become a good speaker of English”), PSV 2 (“My friends think that it’s cool to speak English”), and PSV 5 (“My English teachers hope that I improve my English speaking ability”)—were meant to measure the Perceived Social Value of Speaking English construct. However, the remaining items were focused either on the use of English outside the immediate classroom context or the desire of current classmates to use English. In other words, students perceived the items measuring the construct as indicating the instrumental value of learning English in Japanese society once they graduate from university. The opinions of parents and teachers were obviously not salient for students who desired to improve their English for the purpose of future employment opportunities (items PSV 1 and PSV 5). Even the possible image attached to speaking English by peers (item PSV 2) was not as important as the possible career benefits from speaking English.

The final item, CECP 1 (“I feel comfortable expressing my opinions in Japanese in my current English class”), was originally meant to examine whether students felt comfortable in class even if they used Japanese (their native language) with classmates. However, the word “Japanese” seemed to attract undue attention,

because the item had too much influence on the loading of other items in the Rasch PCA of item residuals and failed to load in the EFA. Removing the item

Table 74

A Comparison of the Foreign Language Classroom Speaking Confidence Constructs Identified by Rasch PCA of Item Residuals and EFA

Construct/Factor	Rasch PCA	EFA
	Items loading onto factor	Items loading onto factor
Foreign Language Classroom Speaking Anxiety	FLCSA items 2, 6, 7, 10, 11	FLCSA items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11
Past English Classroom Experiences	PECE items 1, 3, 4, 6, 7	PECE items 1, 2, 3, 4, 5, 6, 7, 8
Desire to Speak English	DSE items 1, 5, 6, 8, 9	DSE items 1, 2, 3, 4, 5, 6, 7, 8, 9
Perceived Foreign Language Speaking Self-Competence	PFLSS items 1, 2, 4, 5, 7, 8	PFLSS items 1, 2, 3, 4, 5, 6, 7, 8
Current English Classroom Perceptions	CECP items 1, 2, 3, 7, 8	CECP items 2, 3, 4, 5, 6, 7
		PSV items 3
Perceived Social Value of Speaking English	PSV items 3, 6, 7, 8	PSV items 4, 6, 7

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. *N* = 1,078. Loadings were based on a minimum loading criterion of .40.

Table 75

Problematic Items Identified by Rasch PCA of Item Residuals and EFA

Construct/Factor	Rasch PCA	EFA
	Items failing to load	Items failing to load
Foreign Language Classroom Speaking Anxiety	FLCSA 1, FLCSA 3, FLCSA 4, FLCSA 5, FLCSA 8, FLCSA 9	(All items loaded)
Past English Classroom Experiences	PECE 2, PECE 5, PECE 8	(All items loaded)
Desire to Speak English	DSE 2, DSE 3, DSE 4, DSE 7	(All items loaded)
Perceived Foreign Language Speaking Self-Competence	PFLSS 3, PFLSS 6	(All items loaded)
Current English Classroom Perception	CECP 4, 5, 6 ^a	CECP 1
Perceived Social Value of Speaking English	PSV 1, PSV 2, PSV 4, PSV 5	PSV 1, PSV 2, PSV 5, PSV 8

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English. $N = 1,078$. Loadings were based on a minimum loading criterion of .40.

^a = Rasch item analysis also identified item CECP 1 as possibly problematic.

improved the variance explained by the Rasch model and the item statistics for remaining items in the Rasch item analysis. Based on the results of the Rasch PCA of item residuals and EFA, I decided that the four items, PSV 1, PSV 2, PSV 5, and CECP 1, were candidates for removal during the confirmatory factor analysis phase (see Chapter 5: Structural Equation Results).

Summary of Preliminary Analysis Findings

The Rasch item analysis of the R-B5FM and FLCSC instruments indicated that five out of 89 items misfit their intended constructs. The Rasch PCA of item residuals analysis indicated questionable areas of unidimensionality among the eleven constructs, and the EFA confirmed that one of the five Big Five personality traits, Agreeableness, was indistinguishable from Extraversion-Introversion by the participants in this sample. The personality trait Intellect/Imagination was measured by only three items, making it questionable though marginally acceptable for further CFA and SEM analysis. On the other hand, five of the six Foreign Language Classroom Speaking Confidence constructs were robust, with each construct consisting of at least five to seven items with factor loadings above .60. The sixth construct, Perceived Social Value of Speaking English, consisted of only three items, one of which had a loading of only .47, making it a weak, although marginally acceptable factor.

In the next chapter, I answer the four research questions posed at the end of Chapter 2 by examining the latent variables from the Revised Big Five Factor Marker questionnaire and the Foreign Language Classroom Speaking Confidence questionnaire using confirmatory factor analysis (CFA) and structural equation modeling (SEM). First, three separate CFA were conducted to confirm the measurement portion of the SEM model by testing the ten latent variables

comprising the SEM components of the Big Five personality traits, Favorable Social Conditions, and Foreign Language Classroom Speaking Confidence. Second, all ten latent variables were placed into a structural regression model and the structural portion was tested (see Chapter 3: Methods). Third, two alternate structural regression models were tested to demonstrate why the initial model was preferred over mathematically equivalent models. The CFA models answered the first and second research questions, while the structural regression model answered the third and fourth research questions.

CHAPTER 5

STRUCTURAL EQUATION MODEL RESULTS

This chapter presents the results from testing the structural regression models, which answer the four research questions posed at the end of Chapter 2. First, a confirmatory factor analysis (CFA) was conducted to validate the measurement model portion of the structural regression models. The measurement model portion consisted of the ten hypothesized latent variables that comprised the second-order variables Big Five Personality Traits, Favorable Social Conditions, and Foreign Language Speaking Confidence. The CFA of the Big Five Personality Traits confirmed the preliminary results (see Chapter 4) and answered the first research question regarding the validity of the Big Five for a Japanese sample. The CFA of Foreign Language Speaking Confidence answered the second research question regarding the relationship between its three constituent constructs.

Second, after confirming the measurement model portion, the structural portion of the structural regression model was tested. All ten hypothesized latent variables were placed into a causal path with the Big Five Personality Traits as independent indicator variables and the three Favorable Social Conditions variables as both dependent variables and indicator variables of the three Foreign Language Classroom Speaking Confidence variables. The structural regression model

answered the third and fourth research questions regarding the relationship between the Big Five, Favorable Social Conditions, and Foreign Language Classroom Speaking Confidence. Finally, two alternate structural regression models were examined and compared to the original hypothesized model to demonstrate why the initial structural regression model should be preferred despite the possibility of other mathematically equivalent models (Kline, 2005, p. 153).

Confirmatory Factor Analysis Procedures

All confirmatory factor analyses were conducted using EQS 6.1 software (Bentler, 1995). All data were previously screened and normality assumptions were confirmed during the EFA phase of preliminary instrument validation analysis (see Chapter 4: Preliminary Analyses for details). The CFA testing procedure consisted of six steps. First, the data set was randomized into two data sets of 539 cases each. Second, all questionnaire items were used to test the original hypothesized models using the first half of the data set as a calibration sample. Third, items were removed based on the results of the Rasch analysis and EFA results, and the models were then tested again using the first half of the data set. Fourth, items were deleted and item paths were respecified based on recommendations of the LaGrange Multiplier and Wald tests, in conjunction with a theoretical rationale for the changes (Boomsa, 2000), and the models were tested again. Fifth, the new model

was tested using the second half of the data set as a validation sample and the results were tabled and compared to the results from the first half. Finally, cross-validation was conducted by constraining covariance paths in the CFA and the SEM to determine invariance across independent samples from the same population (Byrne, 2006, p. 252).

Model fit was assessed through the use of several commonly reported fit indices: chi-squared (χ^2), comparative fit index (CFI), goodness of fit index (GFI), adjusted goodness of fit index (AGFI), standardized root mean residuals (SRMR), non-normed fit index (NNFI), and root mean square error of approximation (RMSEA) with 90% confidence intervals. Ideally, the χ^2 value should be as low as possible; although the large sample size in the main study likely precludes the chance of the chi-squared index having much meaning as an absolute fit value, it is presented here for the sake of completeness. Normed-chi-squared (NC) was also computed, with values under 5 considered as acceptable (Pedhazur & Pedhazur-Schmelkin, 1991; Schumacker & Lomax, 2004). CFI, GFI, AGFI, and NNFI values above .95 indicated good fit for the models in this study, and values between .90 and .95 indicated adequate fit. Kenny (2010) stated that .90 was a lower bound of good fit for CFI and GFI but also noted that the “current consensus” was not to use GFI and AGFI due to their sample size dependency; however, because GFI and AGFI continue to be reported commonly in journal articles, they

were reported here for the sake of completeness (Byrne, 2006). SRMR and RMSEA, fit indices that are not affected by sample size, should ideally be .06 or below (Hu & Bentler, 1999; Schreiber, Stage, King, Nora, et al., 2006). However, Kenny (2010) indicated an upper bound of .08 for SRMR, while both Kenny and Byrne (2001) indicated an upper bound of .08 for RMSEA. For this study, SRMR and RMSEA indices below .08 were considered to show good fit.

The first CFA model was designed to examine the indicator variables of the five hypothesized constructs of the Revised Big Five Factor Marker questionnaire (Extraversion-Introversion, Agreeableness, Emotional Stability, Conscientiousness, and Intellect/Imagination) that comprised the Big Five Personality Traits. The second and third CFA models examined the six hypothesized constructs of the Foreign Language Classroom Speaking Confidence questionnaire. The second CFA model examined the indicator variables of the three hypothesized constructs of Foreign Language Speaking Confidence (Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English). The third CFA model examined the indicator variables of the three hypothesized constructs of Favorable Social Conditions (Past English Classroom Experiences, Current English Classroom Perceptions, and Perceived Social Value of Speaking English). The three CFA models corresponded to the heuristic

representation of the arrangement of latent variables in the hypothesized causal structure (Figure 17).

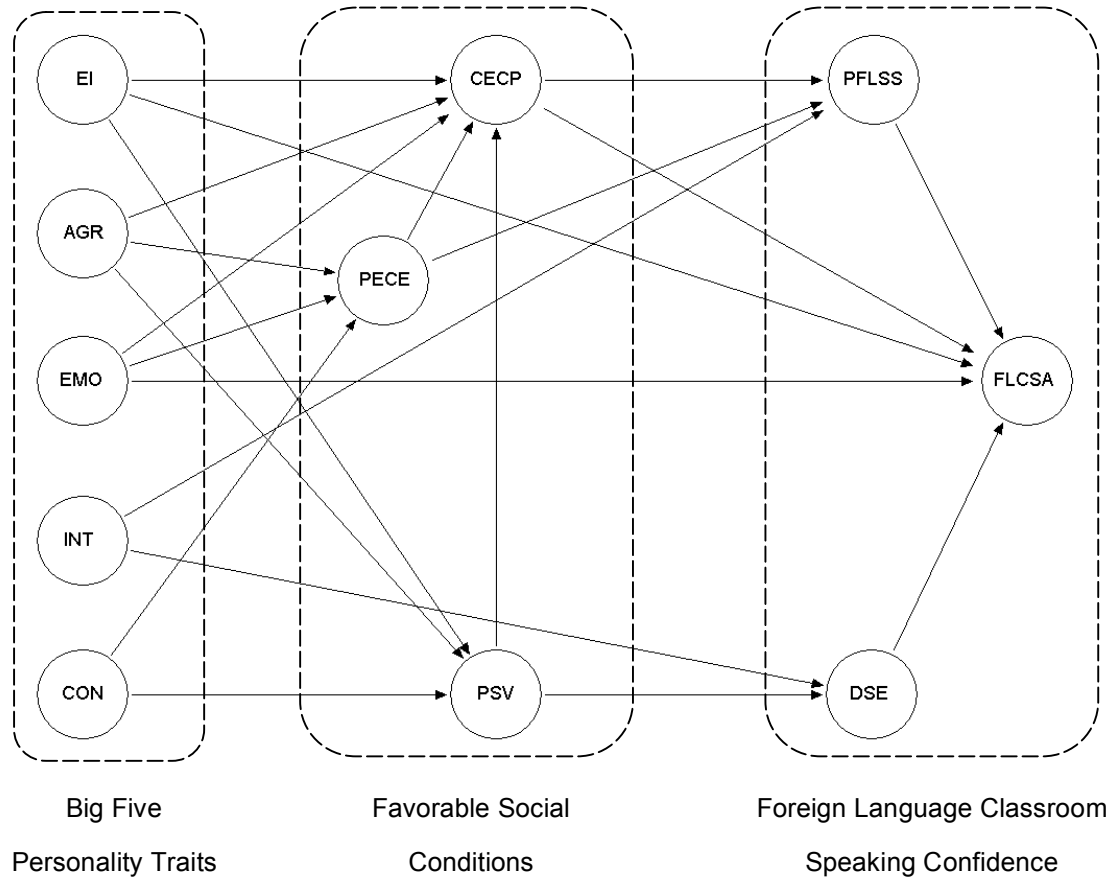


Figure 17. Hypothesized model of a causal structure for foreign language classroom speaking anxiety. The CFA measurement model portion is represented by the dotted-line ovals. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; INT = Intellect/Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

Research Question One: Validity of the Big Five Personality Traits

The first research question asked, “To what degree are the Big Five personality traits valid for a Japanese university student sample?” Although this research question was answered by the Rasch and EFA results (see Chapter 4: Preliminary Results), the first CFA model provided additional confirmation of those results.

For the CFA model designed to examine constructs from the Revised Big Five Factor Marker (R5BFM) questionnaire, data from 34 items (excluding the three items EI 7, AGR 1, and INT 6 that were identified by the Rasch analysis and EFA as misfitting items) were analyzed using the covariance matrix **S**. Fit was not good, with only the standardized root mean residuals (SRMR) and root mean square error of approximation (RMSEA) approaching adequate fit, $\chi^2(517) = 2382.872$ ($p = .000$), CFI = .73, GFI = .78, AGFI = .75, SRMR = .08, NNFI = .70, RMSEA = .08 (.079-.085) (Figure 18). The free parameter to case ratio was 78 free parameters to 539 cases, or 6.91, well below the recommended ratio of at least 10 to 1 (Kline, 2005). Finally, the assumption of lack of multicollinearity was violated by the extremely large correlation of Extraversion-Introversion and Agreeableness ($r = .89$). This violation of multicollinearity confirmed that five constructs could not be identified for the Big Five in this sample and thus confirmed the preliminary results and the answer to research question one: The Big Five personality traits are not valid for the Japanese university student sample in this study.

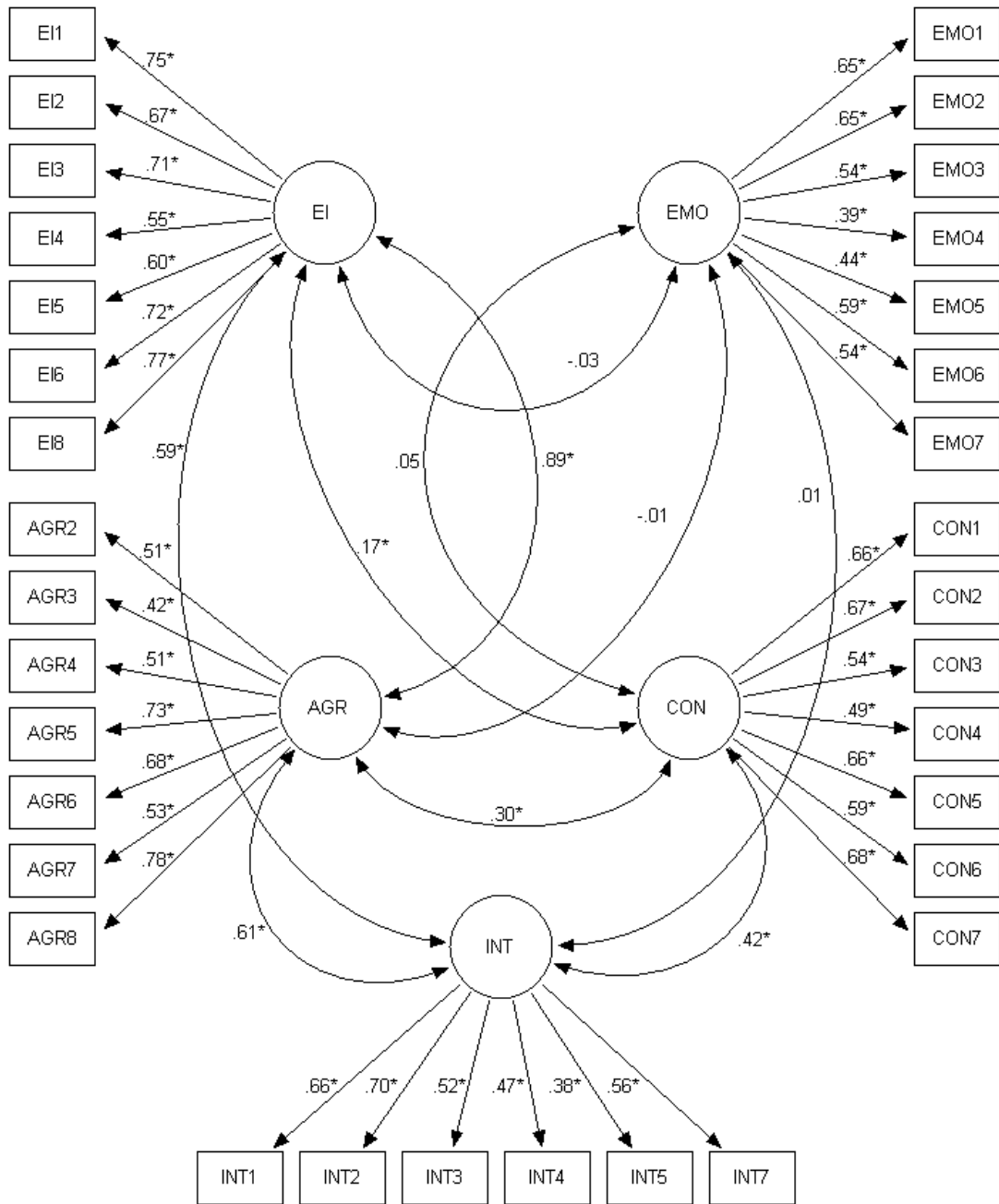


Figure 18. Initial confirmatory factor analysis model of the Big Five Personality Traits, consisting of 34 items. EI = Extraversion-Introversion; AGR = Agreeableness; EMO = Emotional Stability; CON = Conscientiousness; INT = Intellect/Imagination; CFI = .73, GFI = .78, AGFI = .75, SRMR = .08, NNFI = .70, RMSEA = .08 (.079-.085).

* = $p < .05$.

Following the failure of the initial model, the model was respecified according to the “single-step modification approach,” in which individual measurement variables were removed from the model one at a time to clarify the structure of the latent variables (Boomsa, 2000; Kaplan & Wenger, 1993). As recommended by McDonald and Ho (2002), model modifications were made only when removing variables or modifying paths could be justified on a theoretical basis. For example, the Extraversion-Introversion and Agreeableness constructs could not be clearly distinguished; however, not all the items from both constructs were logically related to the same concept. Some items were clearly related to the concept of “being talkative,” while other items were related to the concepts of “needing attention” or “feeling.”

After removing items one at a time and retrying the model, the items for the EI construct were related to speaking to people (EI 3, EI 6, EI 8, and AGR 5). Not coincidentally, all four items were identified by the Rasch PCA analysis as cohering together (Table 6). For the participants in this study, talking to other people was clearly indicative of an outgoing, extraverted personality type. Several different versions of a model that included the Agreeableness construct were attempted; however, no matter which combination of Agreeableness-related measured variables were included, the Agreeableness construct always had an extremely large correlation with Extraversion-Introversion. This was in agreement

with the preliminary analyses, which indicated the difficulty of separating the items measuring Extraversion-Introversion and Agreeableness. Thus, the final version of the CFA model did not include a separate latent variable for Agreeableness. In accordance with the nature of the four items retained, the construct was renamed Extraversion.

Other constructs were likewise examined and tested after removing one item at a time. Emotional Stability was eventually measured by four items (EMO 1, EMO 2, EMO 3, and EMO 6), which were related to vulnerable, passive emotions rather than angry, aggressive emotions. Intellect/Imagination was the most difficult construct to reconcile, as only the items concerned with imagination (INT 1, INT 2, and INT 6) cohered strongly enough to form a distinct construct in the model. Surprisingly, despite the fact that INT 6 (Love to think of new ways to do things) failed to load onto the Intellect/Imagination construct in both the Rasch analysis and the EFA, the CFA model had better fit when INT 6 was included. Although INT 6 had a relatively low loading (.41) in the CFA model, the item was clearly indicative of an imaginative personality and was therefore retained. Because all three items were concerned with imagination rather than intellect, the construct was renamed Imagination (IMA).

The Conscientiousness construct eventually consisted of three items (CON 1, CON 5, and CON 7), all of which were related to a planning, perfectionist

personality. Other items were more related to physical cleaning and organizing or following rules and schedules. Although it might seem that the items concerning rules (CON 3) and schedules (CON 4) should be included in the construct, one should keep in mind that all the participants were university students who essentially had no choice but to follow rules and schedules in their daily academic lives. Thus, for the study participants, following rules and keeping a daily schedule might have been less an indication of conscientiousness and more an indication of necessity.

The final CFA model for the Big Five thus consisted of four latent variables (Extraversion, Emotional Stability, Imagination, and Conscientiousness) and 14 measured variables (Figure 18). There were 34 free parameters, resulting in a parameter to case ratio of 15.85, well above the recommended level of 10 to 1 (Kline, 2005). This model displayed good fit with the calibration sample; further path respecifications based on recommendations from the LaGrange Multiplier test and the Wald test did not result in an improvement in fit indices. The second randomized half of the data was used as a validation sample to cross-validate the model after path respecification ($n = 539$). Both calibration and validation samples showed good fit for SRMR and RMSEA, and good or adequate fit for CFI and GFI (Table 76).

Cross-validation was conducted to test invariance across samples, with equality constraints placed upon the covariance paths. Good or adequate fit was

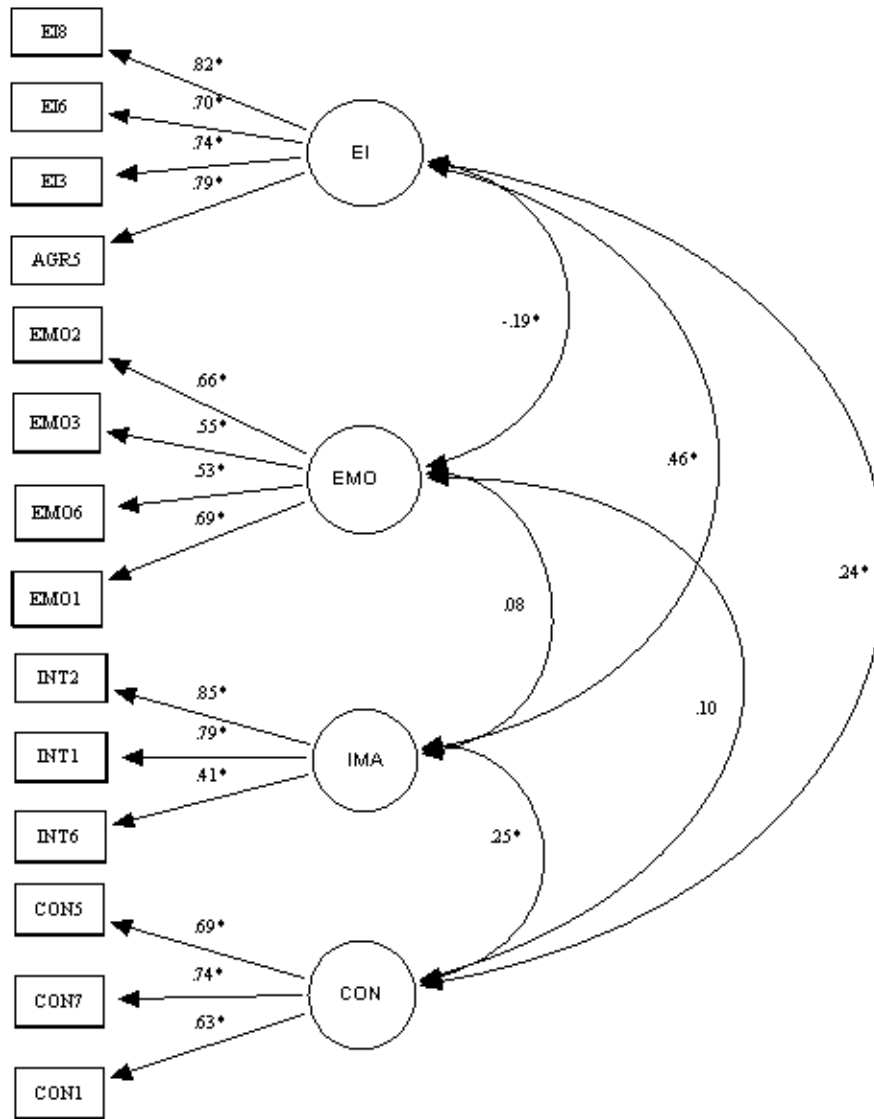


Figure 19. Final confirmatory factor analysis model of the Big Five Personality Traits, consisting of 14 items. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness.

* = $p < .05$.

Table 76

Comparison of Fit Indices for the Calibration Sample and Validation Sample for the Big Five Personality Traits

Sample	χ^2 (71)	NC	CFI	GFI	AGFI	SRMR	NNFI	RMSEA [90% CI]
Calibration sample	178.37	2.51	.95	.96	.94	.05	.94	.05 [.043-.063]
Validation sample	292.03	4.11	.91	.93	.89	.07	.88	.08 [.067-.085]

Note. The χ^2 was significant, $p = .000$. Each sample consisted of $n = 539$. NC = normed chi-squared; CFI = comparative fit index; GFI = goodness of fit index; AGFI = adjusted goodness of fit index; SRMR = standardized root mean residual; NNFI = non-normed fit index; RMSEA = root mean square error of approximation.

Table 77

Cross-Validation Statistics for Latent Variable Paths in the Confirmatory Factor Analysis of the Big Five Personality Traits

Path	Calibration Sample beta weight	Validation sample beta weight	Chi-squared	Probability
EI \leftrightarrow EMO	-.19	-.12	2.435	.119
EI \leftrightarrow IMA	.46	.42	.029	.865
EI \leftrightarrow CON	.24	.20	.007	.935
EMO \leftrightarrow IMA	.08	.09	.273	.601
EMO \leftrightarrow CON	.10	.13	.742	.389
IMA \leftrightarrow CON	.25	.23	.024	.877

Note. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness.

demonstrated on all indices except for chi-squared, $\chi^2(148) = 474.029$, NC = 3.20, CFI = .93, GFI = .94, AGFI = .92, SRMR = .06, NNFI = .91, RMSEA = .06

(.058-.070). All covariance paths were invariant between the two samples (Table 77).

Research Question Two: Foreign Language Classroom Speaking Confidence

The second research question asked, “To what degree is foreign language classroom speaking anxiety associated with perceived foreign language speaking self-competence and the desire to speak English?” This question was answered by examining the relationship between the three Foreign Language Classroom Speaking Confidence latent variables, Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English, in the second CFA model.

For the initial CFA model, data from 28 items were analyzed using the covariance matrix **S**. Fit was not good, with only SRMR and RMSEA showing adequate fit, $\chi^2 (347) = 1394.22 (p = .000)$, CFI = .89, GFI = .83, AGFI = .81, SRMR = .07, NNFI = .88, RMSEA = .076 (.071-.079) (Figure 20). There were 60 free parameters to 539 cases for a ratio of 8.98, which was above the minimum ratio of at least 5 to 1 but below the recommended 10 to 1 ratio (Kline, 2005). Although the model had adequate fit according to the SRMR and RMSEA results, the lack of fit on other indices and the small ratio of parameters to cases suggested that the initial model lacked generalizability between samples.

In order to improve the model, the single-step modification approach of removing one parameter at a time and re-examining model fit was used. The final model consisted of 18 items (Figure 21). The FLCSA latent variable was measured with seven items (FLCSA 2, FLCSA 5, FLCSA 6, FLCSA 8, FLCSA 9, FLCSA 10, and FLCSA 11). Three items (FLCSA 3, FLCSA 4, and FLCSA 7) not included in the measurement of FLCSA asked the participants to compare themselves to classmates during classroom pair work; because the teachers in the study allowed students to choose their speaking partners, it is possible that the participants felt far less anxiety about speaking in a pair. The final item not included (FLCSA 1) asked the participants to compare their speaking ability with other classmates; thus, the item was more related to perceived speaking ability than to speaking anxiety.

The PFLSS latent variable consisted of six items (PFLSS 2, PFLSS 3, PFLSS 4, PFLSS 5, PFLSS 7, and PFLSS 8). The two items not included asked the participants whether they could order food in a restaurant (PFLSS 1) or speak with a group of native English speakers (PFLSS 6). In addition to occurring outside the language classroom, the activities were outside the realm of most participants' English-speaking experiences. The DSE latent variable consisted of five items (DSE 1, DSE 3, DSE 5, DSE 7, and DSE 8), all of which concerned the desire to use English outside the classroom, either in Japan or in foreign countries, with non-Japanese speakers of English. Of the four items not included, two did not

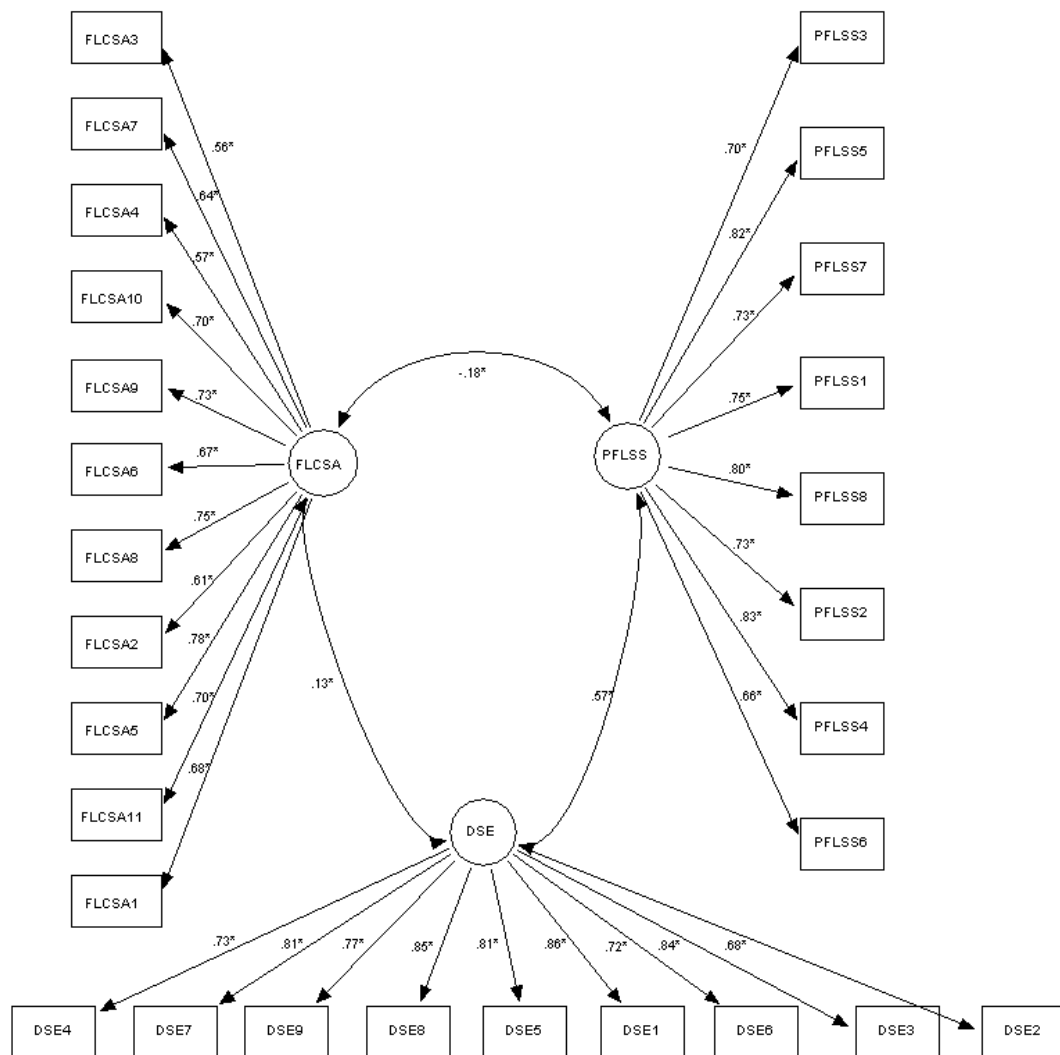


Figure 20. Initial confirmatory factor analysis model of Foreign Language Classroom Speaking Confidence latent variables, consisting of 28 items. FLCSA = Foreign Language Classroom Speaking Anxiety; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence; CFI = .89, GFI = .83, AGFI = .81, SRMR = .07, NNFI = .88, RMSEA = .076 (.071-.079).

* = $p < .05$.

mention specific goals for speaking (DSE 2, “I want to be a good speaker of English,” and DSE 4, “I want to be able to express my opinions and ideas in English”). One item (DSE 6) asked the participants whether they wanted to speak English in the classroom, and one item (DSE 9) asked whether they wanted to speak to their teacher.

There were 39 free parameters in the final model, for a parameter to case ratio of 13.82, which was above the recommended 10 to 1 ratio. Overall fit was good or adequate. The validation sample was used to test the model, and the results confirmed the model fit. Both calibration and validation samples showed good fit for SRMR and RMSEA, and good or adequate fit for CFI and NNFI (Table 78).

Cross-validation was conducted to test invariance across samples, with equality constraints placed upon the covariance paths. Good or adequate fit was demonstrated on four indices, although the confidence interval around RMSEA indicated the possibility of misfit, $\chi^2(301) = 1359.25$, $NC = 4.52$, $CFI = .92$, $GFI = .87$, $AGFI = .83$, $SRMR = .06$, $NNFI = .91$, $RMSEA = .08 (.076-.085)$. All covariance paths were invariant between the two samples (Table 79).

An examination of the correlations of the three latent variables answered the second research question. Foreign Language Classroom Speaking Anxiety had a weak positive correlation with Desire to Speak English ($r = .11$) and a weak ($r = -.19$).

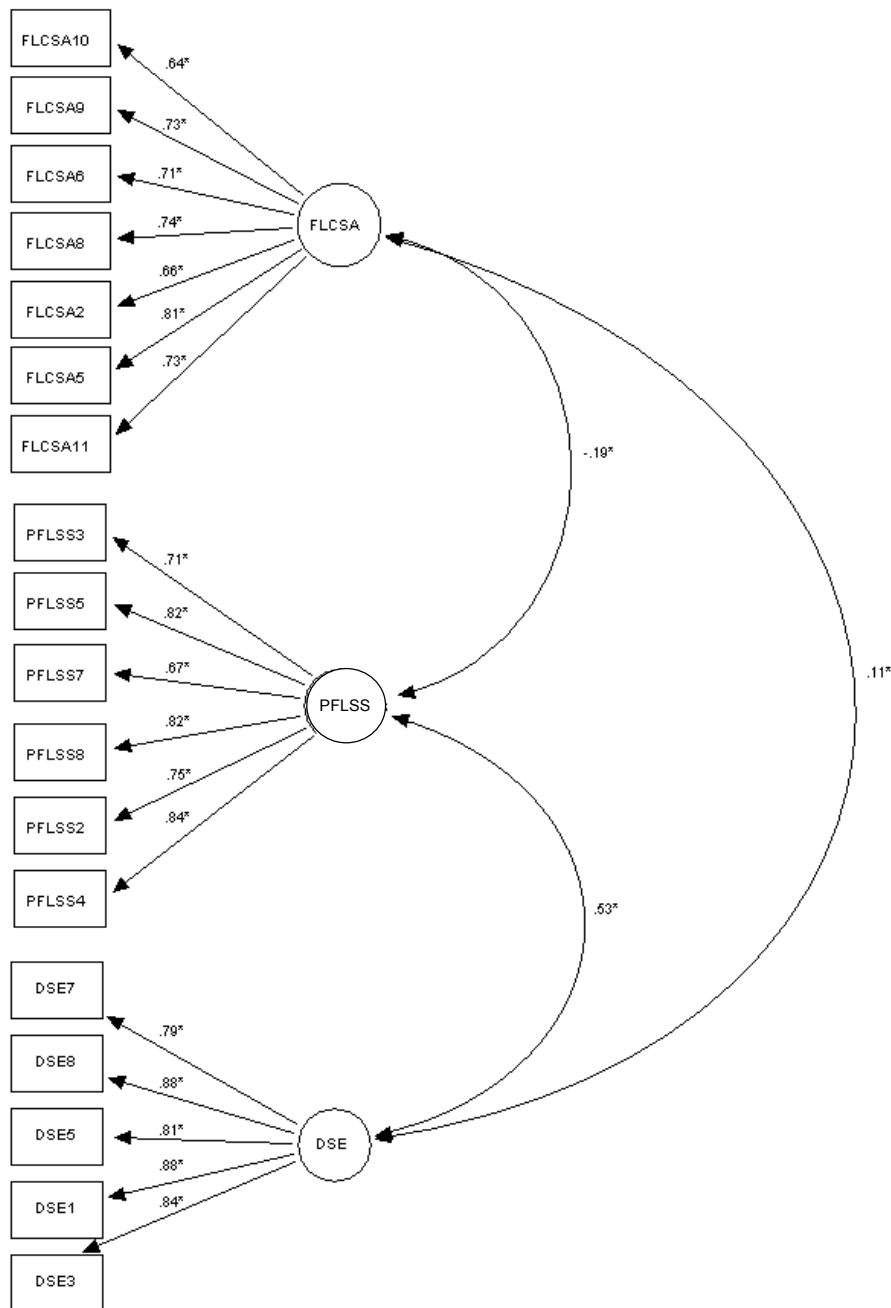


Figure 21. Final confirmatory factor analysis model of Foreign Language Classroom Speaking Confidence latent variables, consisting of 16 items. Beta weights are from the calibration sample. FLCSA = Foreign Language Classroom Speaking Anxiety; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence.

* = $p < .05$.

Table 78

Comparison of Fit Indices for the Calibration Sample and Validation Sample for Foreign Language Classroom Speaking Confidence

Sample	χ^2 (132)	NC	CFI	GFI	AGFI	SRMR	NNFI	RMSEA [90% CI]
Calibration Sample	408.84	3.10	.95	.92	.89	.06	.94	.06 [.056-.069]
Validation Sample	509.74	3.86	.94	.89	.86	.06	.93	.07 [.066-.080]

Note. The χ^2 was significant, $p = .000$. Each sample consisted of $n = 539$. NC = normed chi-squared; CFI = comparative fit index; GFI = goodness of fit index; AGFI = adjusted goodness of fit index; SRMR = standardized root mean residual; NNFI = non-normed fit index; RMSEA = root mean square error of approximation.

Table 79

Cross-Validation Statistics for Latent Variable Paths in the Confirmatory Factor Analysis of Foreign Language Classroom Speaking Confidence

Path	Calibration Sample beta weight	Validation sample beta weight	Chi-squared	Probability
FLCSA \leftrightarrow PFLSS	-.19	-.18	.695	.405
FLCSA \leftrightarrow DSE	.11	.08	1.319	.251
PFLSS \leftrightarrow DSE	.53	.53	.003	.957

Note. FLCSA = Foreign Language Classroom Speaking Anxiety; DSE = Desire to Speak English; PFLSS = Perceived Foreign Language Speaking Self-Competence.

negative correlation with Perceived Foreign Language Speaking Self-Competence

Perceived Foreign Language Speaking Self-competence and Desire to Speak

English were strongly correlated ($r = .53$). The results confirm the hypothesis

presented in Chapter 2 that foreign language speaking confidence includes Desire

to Speak English, rather than being made up of only Foreign Language Anxiety and Perceived Competence.

Confirmatory Factor Analysis: Favorable Social Conditions

Although there was no specific research question regarding the relationship of the three latent Favorable Social Conditions variables (Past English Classroom Experiences, Current English Classroom Perception, and Perceived Social Value of Speaking English), it was necessary to confirm the measurement model prior to testing the final structural regression model for all ten latent variables in the study.

For the initial CFA model, data from 20 items were analyzed using the covariance matrix **S**. As suggested by the Rasch analysis and EFA results, four items (CECP 1, PSV 1, PSV 2, and PSV 5) were excluded from the initial model. Fit was not good according to most of the fit indices, with only SRMR showing adequate fit, $\chi^2 (167) = 867.97 (p = .000)$, CFI = .85, GFI = .85, AGFI = .81, SRMR = .08, NNFI = .83, RMSEA = .09 (.082-.094) (Figure 22). There were 43 free parameters to 539 cases for a ratio of 12.53, which was above the recommended 10 to 1 ratio. CECP and PSV were strongly correlated ($r = .74$) and posed a risk of violating the assumption of multicollinearity. Two items (PSV 4, $\beta = .28$; PSV 6, $\beta = .22$) also failed to load above .40 on the PSV variable.

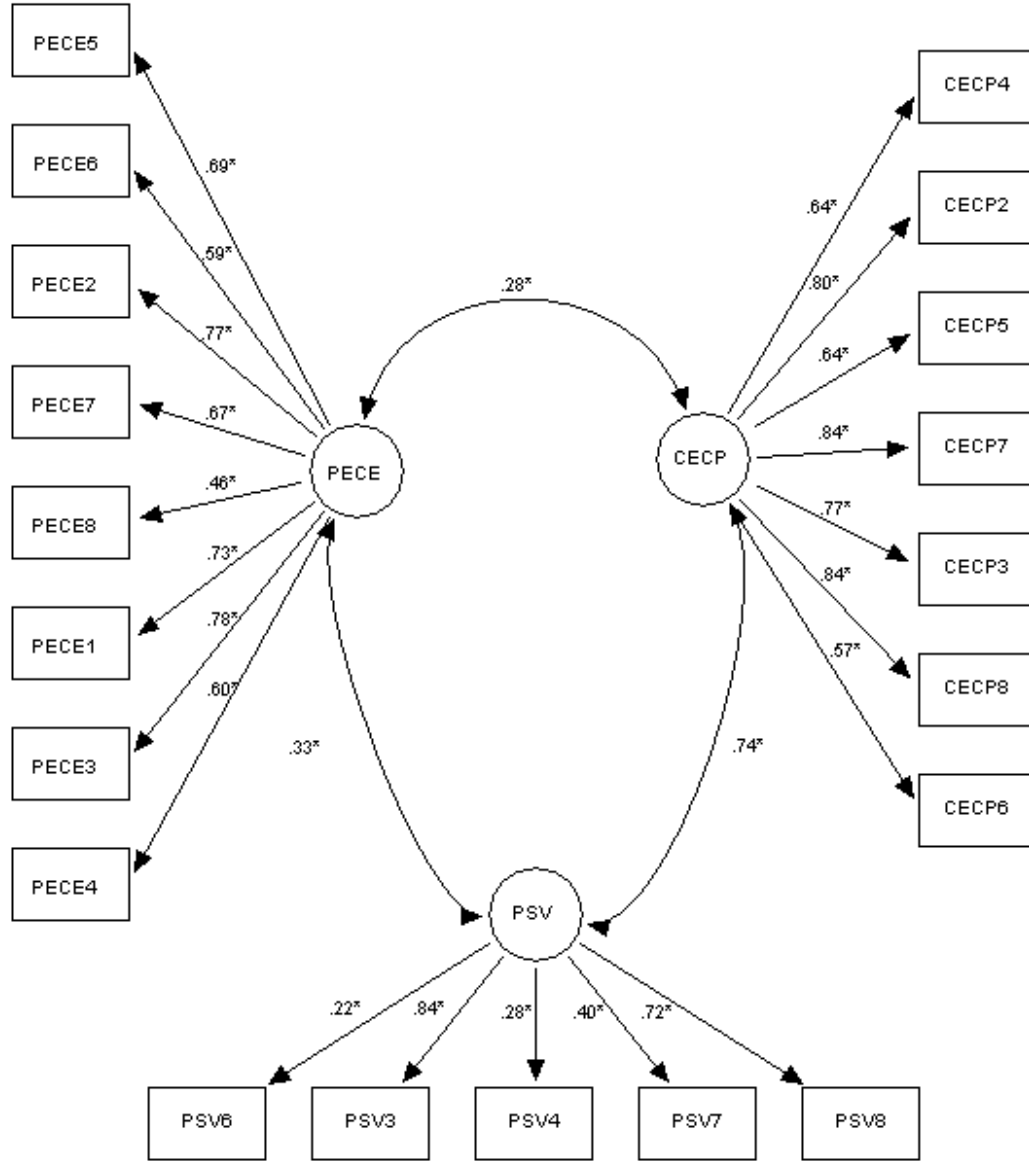


Figure 22. Initial confirmatory factor analysis model of Favorable Social Conditions latent variables, consisting of 20 items. PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; CFI = .85, GFI = .85, AGFI = .81, SRMR = .08, NNFI = .83, RMSEA = .09 (.082-.094).

* = $p < .05$

In order to improve model fit, the single-step modification approach of removing one parameter at a time and re-examining model fit was used. The final model consisted of 12 items (Figure 23). The PECE latent variable was measured by four items (PECE 1, PECE 3, PECE 4, and PECE 7) and concerned the speaking atmosphere of the participants' high school English classrooms. Items not included concerned classroom activities (PECE 2, PECE 5), teachers (PECE 6), and relationships with classmates (PECE 8). The CECP latent variable was measured by five items (CECP 2, CECP 3, CECP 5, CECP 7, and CECP 8) and concerned the general speaking atmosphere in the current English classroom. Three items were not included. Item CECP 1 concerned comfort level when speaking in Japanese, Item CECP 4 ("There are many opportunities to speak with classmates...") was worded vaguely, and CECP 6 ("My current English class makes me feel relaxed") was redundant, as it was essentially the same as item CECP 2 ("I feel comfortable in my current English class").

Finally, the PSV latent variable consisted of three items (PSV 4, PSV 6, and PSV 7), which were concerned with the perception of the value of speaking English by Japanese society. Three items not included (PSV 2, PSV 3, and PSV 8) concerned the desire of friends and classmates to speak English. These three items could not form a variable on their own: one item (PSV 2) asked whether the participants thought their friends found it "cool" to speak English, one item (PSV

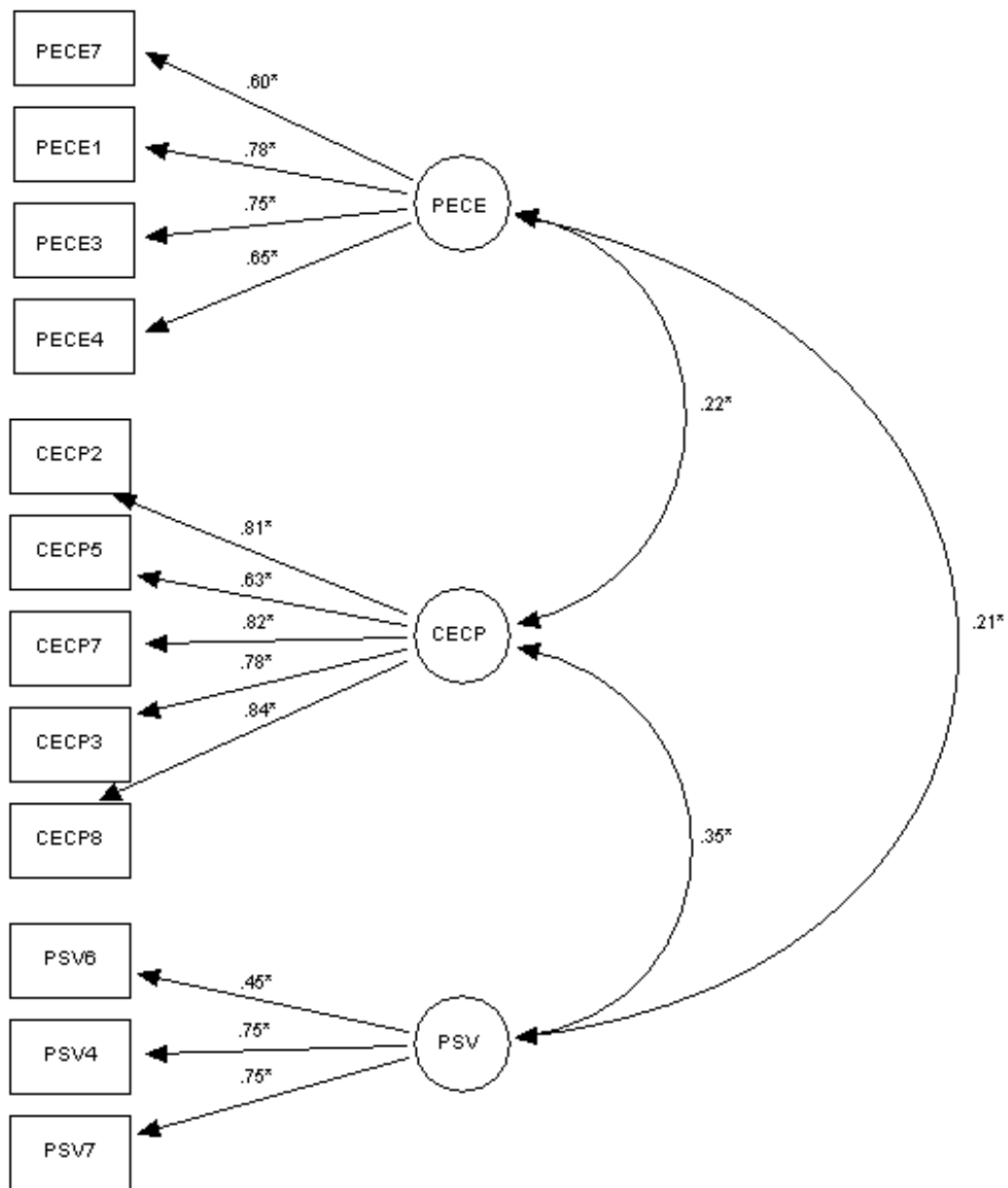


Figure 23. Final confirmatory factor analysis model of Favorable Social Conditions latent variables, consisting of 12 items. Beta weights are from the calibration sample. PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English.

* = $p < .05$

3) concerned classmates' desire to speak English, and one item (PSV 8) was about classmates' desire to become able to speak English. Both of the two remaining items (PSV 1, My parents hope I become a good speaker of English, and PSV 5, My English teachers hope that I improve my English speaking ability) concerned the perceptions of older authority figures.

There were 27 free parameters in the final model, for a parameter to case ratio of 19.96, well above the recommended 10 to 1 ratio. Overall fit to the model was good or adequate for the calibration sample. The validation sample was used to test the model, and the results confirmed the model fit. Both calibration and validation samples showed good fit for SRMR and RMSEA, and good or adequate fit for CFI, GFI, AGFI, and NNFI (Table 80).

Cross-validation was conducted to test invariance across samples, with equality constraints placed upon the covariance paths. Good or adequate fit was demonstrated on all indices except chi-squared, although the confidence interval around RMSEA indicated the possibility of misfit, $\chi^2(105) = 435.199$, $NC = 4.14$, $CFI = .94$, $GFI = .94$, $AGFI = .90$, $SRMR = .06$, $NNFI = .94$, $RMSEA = .08$ (.069-.084). All covariance paths were invariant between the two samples (Table 81).

Table 80

Comparison of Fit Indices for the Calibration Sample and Validation Sample for Favorable Social Conditions

Sample	χ^2 (51)	NC	CFI	GFI	AGFI	SRMR	NNFI	RMSEA [90% CI]
Calibration Sample	215.52	4.23	.94	.94	.91	.06	.92	.08 [.067-.088]
Validation Sample	219.08	4.30	.94	.93	.90	.06	.92	.08 [.068-.089]

Note. The χ^2 was significant, $p = .000$. Each sample consisted of $n = 539$. NC = normed chi-squared; CFI = comparative fit index; GFI = goodness of fit index; AGFI = adjusted goodness of fit index; SRMR = standardized root mean residual; NNFI = non-normed fit index; RMSEA = root mean square error of approximation.

Table 81

Cross-Validation Statistics for Latent Variable Paths in the Confirmatory Factor Analysis of Favorable Social Conditions

Path	Calibration Sample beta weight	Validation sample beta weight	Chi-squared	Probability
CECP \leftrightarrow PECE	.22	.19	.212	.645
CECP \leftrightarrow PSV	.35	.32	.121	.728
PECE \leftrightarrow PSV	.21	.18	.063	.802

Note. PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English.

Summary of the Confirmatory Factor Analysis Findings

The three CFA models confirmed the measurement portion of the structural regression model concerning the relationship between the Big Five personality traits, Favorable Social Conditions variables, and foreign language classroom

speaking confidence variables. The first research question regarding the validity of the Big Five personality traits for a Japanese university student sample was answered in the negative by the preliminary Rasch and EFA results, and these results were confirmed by the first CFA model. The second research question regarding the relationship among the variables comprising Foreign Language Classroom Speaking Confidence was answered through the correlation of the three latent variables, Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English, in the second CFA model. The remaining component of the structural regression measurement portion was confirmed for the three latent variables comprising Favorable Social Conditions in the third CFA model.

In the next section, I present the results of the structural regression models, which answer the third and fourth research questions posed at the end of Chapter 2.

Structural Equation Modeling

This section presents findings from the structural portion of the hypothesized structural regression model for the Big Five Personality Traits, Favorable Social Conditions, and Foreign Language Classroom Speaking Confidence variables. Based on the confirmatory factor analyses of the measurement model, the structural regression model was revised to include only four of the Big Five personality traits.

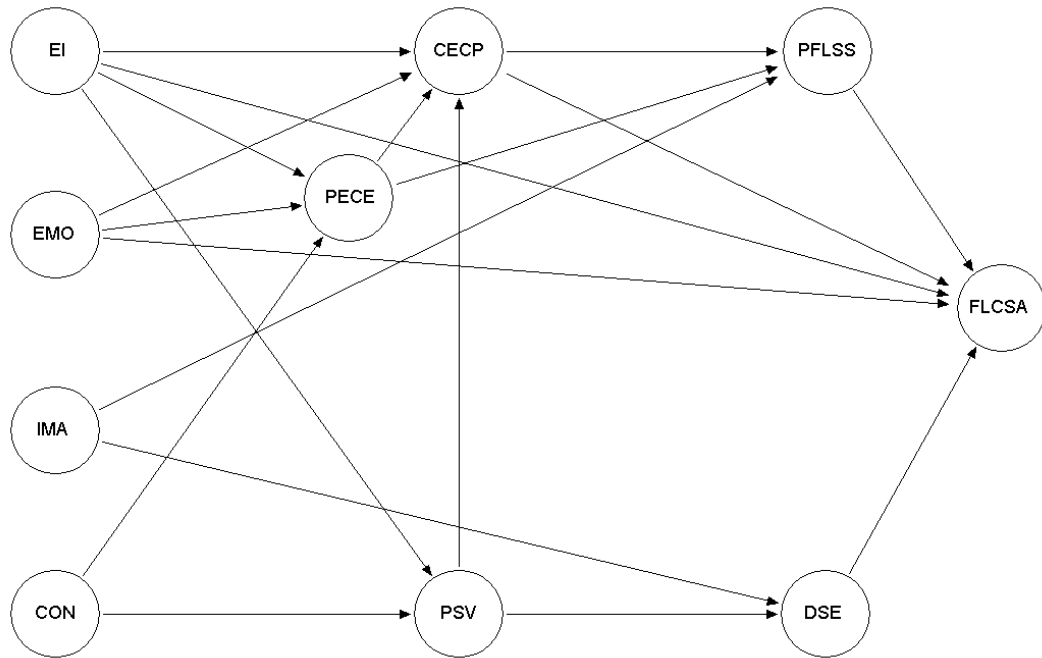


Figure 24. Revised structural regression model of a causal structure for foreign language classroom speaking anxiety. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

The latent variable structural path to be tested is presented in Figure 24. As with the CFA models, the initial path model was tested using the calibration sample and item paths were respecified using the single-step modification approach based on recommendations of the LaGrange Multiplier and Wald tests if theoretically

justifiable. Following path respecification, the revised structural path model was cross-validated using the validation sample. After fit statistics were obtained, two alternate path models were also analyzed to avoid confirmation bias (MacCallum & Austin, 2000, p. 213). The justifications for the paths in the structural regression model are specified in the following sections.

Justification for the Big Five Personality Trait Variable Paths

The paths from the four personality trait variables (Extraversion, Emotional Stability, Imagination, and Conscientiousness) were justified as follows:

1. The path from Extraversion to Foreign Language Classroom Speaking Anxiety was based on previous research that indicated a strong negative influence between extraversion and anxiety (MacIntyre & Charos, 1996).
2. The paths from Extraversion to both Past English Classroom Experiences and Current English Classroom Perception were based on the hypothesis that extraverts excel at social interactions in classroom settings (MacIntyre & Noels, 1994).
3. The path between Emotional Stability and Foreign Language Classroom Speaking Anxiety was supported by mainstream social anxiety research

(e.g., Kagen et al., 1991), in which anxiety consisted of inherited personal traits and social components.

4. The path from Imagination to Perceived Foreign Language Speaking Self-Competence was based on previous research that indicated a moderate positive influence between intellect and perceptions of competence (MacIntyre & Charos, 1996).
5. The path from Conscientiousness to Past English Classroom Experiences was based on the hypothesis that participants with good organization skills and study habits have obtained better results in English classes in high school due to better preparedness.
6. The path from Conscientiousness to Perceived Social Value of Speaking English was based on the hypothesis that participants who learn in a methodical, planned fashion are more aware of social situations both inside and outside the classroom and pay more attention to future goals after their education ends.

MacIntyre and Charos's 1996 model included a path between emotional stability and integrativeness, based on the hypothesis that people who do not have the desire to integrate into an L2 community would become uncomfortable talking with L2 native speakers. However, based on a lack of evidence for Japanese students to desire integration into an L2 community, as well as the lack of evidence

that desire influences personality traits, I reasoned that there should logically be no path between Emotional Stability and Desire to Speak English. Paths between the Big Five personality traits and other latent variables were not hypothesized due to a lack of support in the literature supporting such influences.

Justification for the Favorable Social Conditions Variable Paths

The paths among the three Favorable Social Conditions variables (Past English Classroom Experiences, Current English Classroom Perception, and Perceived Social Value of Speaking English) were justified as follows:

1. The path from Past English Classroom Experiences to Current English Classroom Perception was based on the hypothesis that participants who had previously enjoyed their high school English classes would also enjoy their current university classes, leading to a more positive classroom perception for classmates.
2. The path from Past English Classroom Experiences to Perceived Foreign Language Speaking Self-Competence was based on previous literature hypothesizing that previous foreign language learning experiences influence one's perception of the degree of competence in that foreign language (MacIntyre et al., 1998).

3. The path from Current English Classroom Perception to Perceived Foreign Language Speaking Self-Competence was based on previous literature that suggested that positive peer support in the foreign language classroom might encourage a sense of competence (Clement et al., 1994; MacIntyre et al., 1998).
4. The path from Current English Classroom Perception to Foreign Language Classroom Speaking Anxiety was based on previous research indicating that poor relationships or negative experiences within the foreign language classroom lead to increased anxiety (Clement et al., 1994; MacIntyre et al., 1998).
5. The path from Perceived Social Value of Speaking English to Current English Classroom Perception was based on a suggestion in previous research that participants who perceive value for English outside the immediate classroom learning situation participate actively in classroom activities, creating a classroom atmosphere supportive of English speaking (Clément et al., 1994).
6. The path from Perceived Social Value of Speaking English to Desire to Speak English was based on the hypothesis that participants who perceive value for English outside the classroom have a greater desire to

find opportunities to speak English outside the foreign language classroom.

Paths from Past English Classroom Experiences and Current English Classroom Perception to Desire to Speak English, Past English Classroom Experiences and Perceived Social Value of Speaking English to Foreign Language Classroom Speaking Anxiety, and Perceived Social Value of Speaking English to Perceived Foreign Language Speaking Self-Competence were not hypothesized, due to a lack of support in the existing literature.

Justification for the Foreign Language Classroom Speaking Confidence

Variable Paths

The paths among the three foreign language speaking confidence variables (Desire to Speak English, Perceived Foreign Language Speaking Self-Competence, and Foreign Language Classroom Speaking Anxiety) were justified as follows:

1. The path from Perceived Foreign Language Speaking Self-Competence to Foreign Language Classroom Speaking Anxiety was based on the hypothesis that, because previous literature posited that communicative competence leads to L2 confidence (MacIntyre et al., 1998), perceptions of lack of speaking competence should logically lead to anxiety.

2. The path from Desire to Speaking English to Foreign Language Classroom Speaking Anxiety was based on the hypothesis that participants who do not care about speaking English feel no anxiety over making mistakes, whereas participants who want to speak English might feel anxiety.

A path between Desire to Speak English and Perceived Foreign Language Speaking Self-Competence was not hypothesized due to a lack of support in the existing literature.

Research Question Three: Personality and Speaking Confidence

Research question three asked, “To what degree are the Big Five personality trait factors associated with foreign language speaking confidence?” To answer this question, the structural model shown in Figure 24 was tested using the calibration sample. Initial fit was not good for most fit indices, with only RMSEA showing adequate fit, $\chi^2 (885) = 2179.77 (p = .000)$, CFI = .89, GFI = .84, AGFI = .82, SRMR = .10, NNFI = .88, RMSEA = .05 (.049-.055).

The Wald test suggested the removal of six non-significant directional paths:

1. from Extraversion (EI) to Foreign Language Classroom Speaking Anxiety (FLCSA),

2. from Emotional Stability (EMO) to Current English Classroom Perception (CECP),
3. from EMO to Past English Classroom Experiences (PECE),
4. from Imagination (IMA) to Desire to Speak English (DSE),
5. from PECE to CECP, and
6. from CECP to FLCSA.

The LaGrange Multiplier test suggested the addition of two directional paths from Perceived Foreign Language Speaking Self-Competence (PFLSS) to DSE and from CECP to DSE, and one covariance path between EI and IMA. The paths recommended for removal were removed one at a time as per the single-step modification approach of removing one parameter at a time and re-examining model fit. The final model consisted of 104 free parameters for a parameter to case ratio of 5.18, barely above the minimum recommended ratio of 5 to 1 (Figure 25). Adequate fit was obtained for CFI and NNFI, and good fit was obtained for SRMR and RMSEA. The respecified path model was analyzed again using the validation sample, and the results were compared. Good fit was found for SRMR and RMSEA (Table 82).

Cross-validation was conducted to test invariance across samples, with equality constraints placed upon the variable paths. Good or adequate fit was

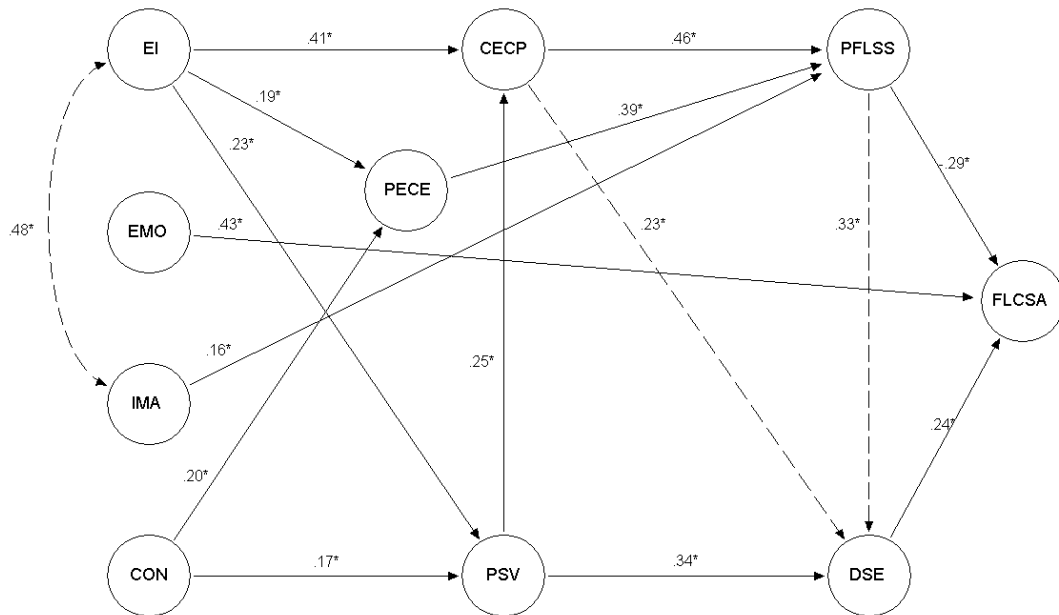


Figure 25. Structural regression model 1 of a causal structure for foreign language classroom speaking anxiety, after model path respecification. Dashed lines indicate paths added to the model. Beta weights are from the calibration sample. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

demonstrated on four fit indices, $\chi^2(1788) = 4118.828$, $NC = 2.30$, $CFI = .90$, $GFI = .85$, $AGFI = .83$, $SRMR = .07$, $NNFI = .90$, $RMSEA = .05 (.047-.051)$. One path (PSV → DSE) was noninvariant, with a beta weight of .34 on the calibration sample and .47 on the validation sample (Table 83).

Research question three was answered by examining the paths between the latent variables comprising the Big Five personality traits and the latent variables

Table 82

Comparison of Fit Indices for the Calibration Sample and Validation Sample for Structural Regression Model 1

Sample	χ^2 (884)	NC	CFI	GFI	AGFI	SRMR	NNFI	RMSEA [90% CI]
Calibration Sample	1915.63	2.17	.91	.86	.84	.07	.90	.046 [.044-.049]
Validation Sample	2184.92	2.47	.89	.84	.82	.07	.88	.052 [.049-.055]

Note. The χ^2 was significant, $p = .000$. Each population sample consisted of $n = 539$. NC = normed chi-squared; CFI = comparative fit index; GFI = goodness of fit index; AGFI = adjusted goodness of fit index; SRMR = standardized root mean residual; NNFI = non-normed fit index; RMSEA = root mean square error of approximation.

comprising Foreign Language Classroom Speaking Confidence. Emotional Stability (EMO) had a medium effect on Foreign Language Classroom Speaking Anxiety (FLCSA), $\beta = .43$. Imagination (IMA) had a small effect on Perceived Foreign Language Speaking Self-Competence (PFLSS), $\beta = .16$.

On the other hand, the hypothesized relationships between Extraversion and Foreign Language Classroom Speaking Anxiety and between Imagination and Desire to Speak English were not significant. Thus, two of the four Big Five personality traits (Emotional Stability and Imagination) had significant effects on the Foreign Language Classroom Speaking Confidence variables, while the remaining two Big Five personality traits (Extraversion and Conscientiousness) had indirect influences, as they were filtered through the Favorable Social Conditions variables.

Table 83

Cross-Validation Statistics for Latent Variable Paths in the Structural Regression Model 1

Path	Calibration Sample beta weight	Validation sample beta weight	Chi-squared	Probability
EI → CECP	.41	.38	.006	.939
PSV → CECP	.25	.23	.674	.412
EI → PECE	.19	.19	2.282	.131
CON → PECE	.20	.15	.094	.759
EI → PSV	.23	.28	2.852	.091
CON → PSV	.17	.10	1.324	.250
IMA → PFLSS	.16	.15	.468	.494
CECP → PFLSS	.46	.39	2.988	.084
PECE → PFLSS	.39	.43	.176	.675
CECP → DSE	.23	.22	.000	.985
PSV → DSE	.34	.41	4.423	.035
PFLSS → DSE	.33	.38	.032	.857
EMO → FLCSA	.43	.44	.116	.734
PFLSS → FLCSA	-.29	-.29	.311	.577
DSE → FLCSA	.24	.21	.025	.874
EI ↔ IMA	.48	.46	1.490	.222

Note. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

Research Question Four: Classroom Climate and Speaking Confidence

Research question four asked, “To what degree does the classroom climate influence foreign language classroom speaking confidence?” To answer this question, I examined structural path model 1 for the correlations among the among the Favorable Social Conditions variables, Past English Classroom Experience, Current English Classroom Perception, and Perceived Social Value of Speaking

English, and the Foreign Language Classroom Speaking Confidence variables, Perceived Foreign Language Speaking Self-Competence, Desire to Speak English, and Foreign Language Classroom Speaking Anxiety.

Although not a significant predictor of speaking anxiety, Current English Classroom Perception (CECP) was found to have a medium to large effect on Perceived Foreign Language Speaking Self-Competence (PFLSS), $\beta = .46$, and a medium influence on Desire to Speak English (DSE), $\beta = .23$. Positive English Classroom Experience (PECE) likewise had a medium effect on PFLSS, $\beta = .39$, and Perceived Social Value of Speaking English (PSV) had a medium effect on Desire to Speak English (DSE), $\beta = .34$. Although Past English Classroom Experiences did not influence Current English Classroom Perception, Perceived Social Value of Speaking English had a small to medium effect on Current English Classroom Perception, $\beta = .25$; thus, Perceived Social Value of Speaking English indirectly influenced Perceived Foreign Language Speaking Self-Competence through Current English Classroom Perception.

Equivalent Models

It has been argued that social science researchers need to evaluate mathematically equivalent models in order to strengthen support for their original models and to avoid confirmation bias (Hoyle & Panter, 1995; MacCallum &

Austin, 2000; McDonald & Ho, 2002). Although there can be dozens of equivalent models in a simple path model, for a larger model, such as the one in this study, the path directions are not arbitrary. However, MacCallum stated unequivocally, “[I]n SEM, one is not free to ignore the presence of equivalent models and to assume that one’s specified model provides the valid explanation of the data” (MacCallum, 1995, p. 31).

Therefore, I tested the two most likely equivalent models in which the paths among the Foreign Language Classroom Speaking Confidence variables (Desire to Speak English, Perceived Foreign Language Speaking Self-Competence, and Foreign Language Classroom Speaking Anxiety) differed from the original, preferred model. The beta weights of paths among the Favorable Social Conditions variables, PECE, CECP, and PSV, and the foreign language classroom speaking confidence variables, PFLSS, DSE, and FLCSA, were examined and compared among the three models.

Equivalent model 2. In structural regression equivalent model 2, Desire to Speak English (DSE) predicts Perceived Foreign Language Speaking Self-Competence (PFLSS) (Figure 26). This is the opposite direction of the path between the two variables in the structural regression model 1 (Figure 25), in which Perceived Foreign Language Speaking Self-Competence predicted Desire to Speak

English in a post-hoc, data-driven path. The fit indices for the equivalent model 2 in Figure 26 are identical to structural regression model 1 in Figure 25. Because the original hypothesized model did not include the post hoc path between PFLSS and DSE, it is worth examining the differences that occur, depending on the direction of the influence.

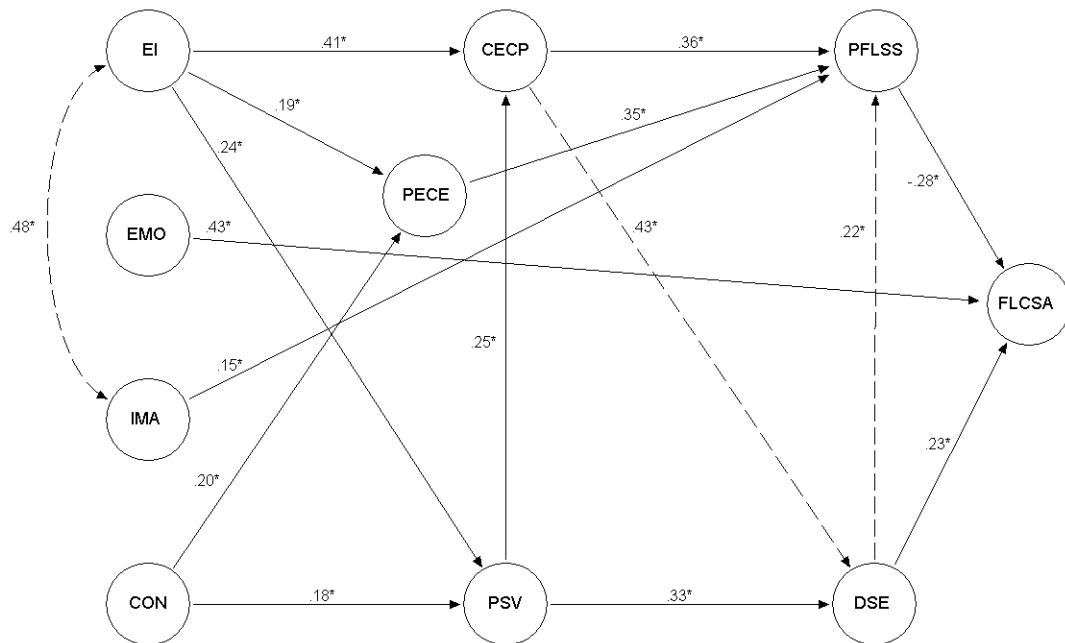


Figure 26. Structural regression equivalent model 2 of a causal structure for foreign language classroom speaking anxiety, in which DSE influences PFLSS. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

In structural regression model 1, Perceived Foreign Language Speaking Self-Competence influenced Desire to Speak English, $\beta = .33$; when the path direction was reversed in the equivalent model 2, Desire to Speak English influences Perceived Foreign Language Speaking Self-Competence, $\beta = .22$. This change additionally alters the strength of the influence of the Favorable Social Conditions variables on the Foreign Language Classroom Speaking Confidence variables. The influence of Current English Classroom Perception on Desire to Speak English was .23 in Model 1, but increased to .43 in Model 2. The influence of Current English Classroom Perception on Perceived Foreign Language Speaking Self-Competence was .46 in Model 1, but decreased to .36 in Model 2. The strengths of the paths from Perceived Social Value of Speaking English to Desire to Speak English (.35 in Model 1, .33 in Model 2), from Past English Classroom Experiences to Perceived Foreign Language Speaking Self-Competence (.39 in Model 1, .35 in Model 2), and from Desire to Speak English to Foreign Language Classroom Speaking Anxiety (.24 in Model 1, .23 in Model 2) changed only slightly.

Equivalent model 3. In structural regression equivalent model 3, both Perceived Foreign Language Speaking Self-competence (PFLSS) and Foreign Language Classroom Speaking Anxiety predict Desire to Speak English (DSE)

(Figure 27). Similar to the differences between Model 1 and Model 2, the path direction change in Model 3 also altered the strength of other paths in the model. Although the influence of Desire to Speak English on Foreign Language Classroom Speaking Anxiety was .24 in Model 1, when the path direction was reversed in Model 3 the influence of Foreign Language Classroom Speaking Anxiety on Desire to Speak English decreased to .17. The influence of Perceived Foreign Language

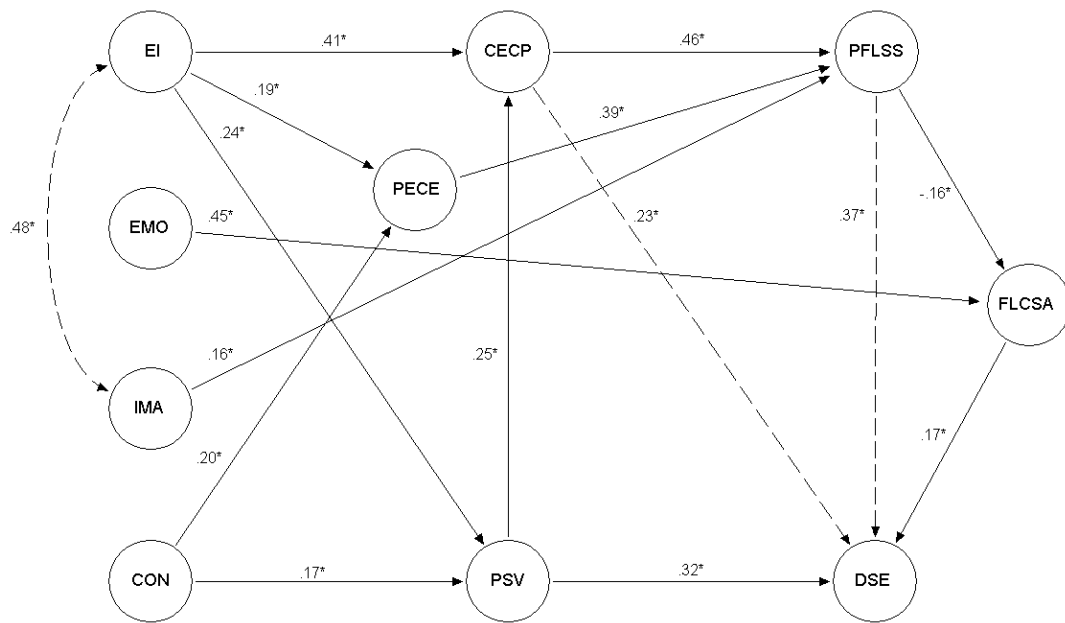


Figure 27. Structural regression equivalent model 3 of a causal structure for foreign language classroom speaking anxiety, in which FLCSA influences DSE. EI = Extraversion; EMO = Emotional Stability; IMA = Imagination; CON = Conscientiousness; PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

Speaking Self-competence on Foreign Language Classroom Speaking Anxiety was -.29 in Model 1, but decreased to -.16 in Model 3. The strength of Perceived Foreign Language Speaking Self-Competence on Desire to Speak English became slightly larger (.33 in Model 1, .37 in Model 3); however, the strength of the path from Perceived Social Value of Speaking English to Desire to Speak English (.35 in Model 1, .32 in Model 3) became slightly smaller. The strengths of the paths from Perceived Social Value of Speaking English to Current English Classroom Perception, from Current English Classroom Perception to Desire to Speak English, from Current English Classroom Perception to Perceived Foreign Language Speaking Self-Competence, and from Past English Classroom Experiences to Perceived Foreign Language Speaking Self-Competence did not change.

Choosing a model. Because all three structural regression models are mathematically equivalent, choosing which model best represents the relationships among the variables is a matter of deciding which variables are logically more likely to influence others. A comparison of the path strengths among the three equivalent models is displayed in Table 84.

In the original hypothesized model, a path was not proposed between PFLSS and DSE; however, it is plausible that students who feel more competent speaking English will want to speak more in the future should the opportunity arise. The

Table 84

Path Strengths among the Favorable Social Conditions and Foreign Language Classroom Speaking Confidence Variables of the Three Equivalent Regression Models

Path	Model 1	Model 2	Model 3
PSV → CECP	.25	.25	.25
PSV → DSE	.35	.33	.32
PECE → PFLSS	.39	.35	.39
CECP → PFLSS	.46	.36	.46
CECP → DSE	.23	.43	.23
PFLSS → FLCSA	-.29	-.26	-.16
PFLSS → DSE	.33	—	.37
DSE → PFLSS	—	.22	—
DSE → FLCSA	.24	.23	—
FLCSA → DSE	—	—	.17

Note. All path strengths are given in beta (β) weights. Beta weights are based on analysis using the calibration sample. PECE = Past English Classroom Experiences; CECP = Current English Classroom Perception; PSV = Perceived Social Value of Speaking English; PFLSS = Perceived Foreign Language Speaking Self-Competence; DSE = Desire to Speak English; FLCSA = Foreign Language Classroom Speaking Anxiety.

opposite is also true; students who feel less competent speaking English are likely to feel less desire to speak in the future. This is the situation in Model 1, in which PFLSS predicts DSE.

On the other hand, it is less plausible that the desire to speak English leads students to feel more competent (Model 2). It is possible that, given the opportunity

to speak English, wanting to speak English leads to perceived competence, assuming the students experienced positive results. However, this outcome is less logical than perceived competence leading to a desire to speak. Thus, the relationship of the variables in Model 1 is logically preferable to those of Model 2.

In the original hypothesized model, it was also theorized that DSE influenced FLCSA, stemming from the hypothesis that students who want to speak English feel more anxious about performing well or having their intended communications understood. In Model 3, the opposite possibility is that speaking anxiety leads to a desire to speak. Because the model shows the path as a positive influence, it can be taken as an indication of positive, facilitative speaking anxiety, that is, anxiety about speaking increases the desire to speak English. However, because the negative influence of PFLSS on FLCSA remains negative, albeit somewhat smaller, regardless of the path direction between DSE and FLCSA, the speaking anxiety represented in the model is clearly not facilitative. Thus, it seems unlikely that experiencing speaking anxiety in the English classroom leads students to want to speak more English. To sum up, the original modified structural regression model 1 shown in Figure 25 is preferable to the mathematically equivalent models shown in Figure 26 and Figure 27.

Summary of the Structural Regression Models

In this section, research questions three and four are answered by examining the structural regression model results. The Big Five personality traits were found both to directly and indirectly influence the variables comprising foreign language classroom speaking confidence. Emotional Stability had a moderate effect on Foreign Language Classroom Speaking Anxiety, and Imagination affected Perceived Foreign Language Speaking Self-Competence to a lesser extent. The variables comprising Favorable Social Conditions influenced foreign language classroom speaking confidence variables to varying degrees. The strongest influences were from Current English Classroom Perception and Past English Classroom Experiences on Perceived Foreign Language Speaking Self-Competence. Contrary to the original hypothesized model, Past English Classroom Experiences did not have a significant influence on Current English Classroom Perception. Other additions to the hypothesized model were a weak influence of Current English Classroom Perception on Desire to Speak English, and a moderate influence of Perceived Foreign Language Speaking Self-Competence on Desire to Speak English.

In the next chapter, I provide a more detailed interpretation of the findings. I also offer theoretical implications of the results and pedagogical implications that English classroom teachers in Japan might find useful for their teaching practices.

CHAPTER 6

DISCUSSION

In this chapter, I restate the research questions and summarize the results of the study. Following the summary, I interpret the results of the study, particularly in comparison to previous research into the connection between personality traits and foreign language anxiety. Finally, I speculate on the theoretical implications for second language acquisition concerning foreign language classroom speaking confidence and present the pedagogical implications of the study.

Research Question One: Validity of the Big Five

The first research question concerned the validity of the Big Five personality traits for a Japanese university student population. The Big Five personality traits are theorized to consist of the five constructs of Extraversion-Introversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Imagination. The first research question was answered by triangulation three analytical techniques (Rasch analysis, exploratory factor analysis, and confirmatory factor analysis) conducted on data obtained from 1,081 Japanese university students in response to a 37-item questionnaire designed to measure the Big Five personality traits.

Summary of the Results for Research Question One

All three analyses indicated that it was difficult to separate the Extraversion-introversion and Agreeableness constructs. The Rasch analysis indicated that one Agreeableness item (AGR 5) cohered with the Extraversion-Introversion items that were concerned with talkativeness. The Rasch model accounted for a relatively low percentage of variance in the Agreeableness construct, while the first residual contrast explained a high percentage of unexplained variance, a possible indication that the construct was not unidimensional as hypothesized. In the EFA, all the EI and AGR items loaded onto the same factor; thus, it was not possible to distinguish between the Extraversion-Introversion and Agreeableness constructs. The CFA confirmed this finding, as the Extraversion-Introversion and Agreeableness constructs correlated at .89, which was a violation of the assumption of multicollinearity. The Rasch analysis and CFA additionally agreed that only the Intellect/Imagination items concerned with imagination cohered strongly enough to form a distinct construct.

Although the remaining three personality trait constructs of the Big Five, Emotional Stability, Conscientiousness, and Intellect/Imagination, were confirmed to be relatively unidimensional by the Rasch analysis, a comparison of the Rasch analysis and EFA results indicated a few problematic items that might not have measured their intended constructs. An initial CFA demonstrated poor model fit,

which required path respecification involving a careful examination of item loadings onto their intended latent variables. After considering the results of the Wald test and LaGrange multiplier test, the three or four items with the highest factor loadings that had a logical relationship to the hypothesized construct were retained. After path respecification, the CFA model was tested again with the validation sample. Good fit was found on two fit indices (SRMR and RMSEA) and adequate fit was found on two additional fit indices (CFI and GFI).

In sum, the results did not support the existence of five factors comprising the Big Five for the participants in this study. In the place of the five hypothesized personality traits, the results indicated the existence of the following four personality traits: Extraversion, Emotional Stability, Conscientiousness, and Imagination.

Interpretation of the Results for the Validity of the Big Five

The finding that the Extraversion and Agreeableness items formed a single construct agrees with the results obtained from Chinese samples. For example, Cheung et al. (2001) found support for a six-factor model using the Chinese Personality Assessment Instrument (CPAI) in three separate samples, which included a factor they termed Interpersonal Relatedness. The researchers claimed that not only did items from the Extraversion and Agreeableness factors from the

NEO-PI-R and NEO-FFI overlap among Chinese participants, but also the Big Five factors did not correlate with Interpersonal Relatedness. Therefore, the Big Five was not adequate for participants from a “collectivist cultural context” (Cheung et al., 2001, p. 426).

The finding also highlights the inherent difference in the place of an “altruistic” or “agreeableness” factor in the Big Five and the Big Three models of personality traits. Proponents of the PEN model have often claimed that the Agreeableness factor was simply a facet of the Psychotic factor from the Big Three (Eysenck, 1991). It should be pointed out that temperament-based models (Buss & Plomin, 1975; Eysenck & Eysenck, 1985) generally ignore Agreeableness, in addition to “Openness to New Experiences,” on the basis that the factors are mere descriptions of behavior and not innate psychological traits. It is telling that, in a Big Five study that used CFA to compare three Big Five measurement instruments (John & Srivastava, 1999), Agreeableness and Neuroticism had a moderate negative correlation, $r = -.37$. In the same study, one entire facet from Agreeableness cross-loaded on Extraversion as well, demonstrating a lack of unidimensionality in the constructs comprising the hypothetical Big Five.

Mainstream psychologists have suggested that the links between Extraversion and Agreeableness are the product of “evolution,” that is, that the altruistic nature of human beings has evolved through the necessity of having to live in groups for

survival (Hogan, 1983) or to optimize reciprocal social exchanges for improved reproductive competition (Cosmides, 1989). The assumption is that the underlying constructs are valid, and that although Extraversion and Agreeableness are “obliquely related,” they can be refined through a process of factor analysis rotation. However, this notion does not explain why the two constructs should be perceived as the same construct by the Japanese university participants in this study.

One possible reason that the participants were unable to distinguish extraversion and agreeableness might be the nature of the participants’ understanding of appropriate personal relationships. In a Western context, people who take an interest in the welfare of others (“The Good Samaritan”) are evaluated more positively than those who do not. However, in Japan, the hierarchy of personal relationships is stressed, starting with the language one uses with one’s own family members. As Japanese children mature, they are introduced to the “web of human relations” (*ningen-kankei*), which includes the difference between the concepts of *uchi* (“in-group”) and *soto* (“out-group”). People who are related in some way to oneself—by membership in a group such as a family, school, work, or social club—are part of one’s “in-group.” People who do not participate in joint activities are considered “out-group” people to whom one has no relation at all (Hendry, 2003).

In fact, one reason often given for the difficulty Japanese encounter when dealing with “out-group” people is that there is no explicit “senior-junior” (*senpai-kohai*) relationship evident when meeting people with whom there are no prior shared experiences. The sense of knowing one’s social “senior” (i.e., someone who is more experienced than oneself) is derived through “shared activities in the same institution over a period of time,” particularly during the teens and early twenties (Nakane, 1970, p. 134). High school students in Japan are particularly subject to the *senpai-kohai* relationship; *kohai* are expected to use polite language (*keigo*) and to bow to *senpai* even when meeting them outside the school grounds, and *kohai* on sports teams in high school must perform menial tasks such as preparing equipment or even washing sports uniforms for *senpai*, without being allowed to participate in games themselves. Sugimoto (2010) has gone so far as to call this social aspect of Japanese high school education as a “quasi-military age-based hierarchy” (p. 142).

The participants in the current study were all first-year university students; thus, at the time of questionnaire implementation halfway through their first semester in a new social environment, they were still in the process of adjusting to the transition from a role as high school *senpai* to university *kohai*. Thus, when study participants read questionnaire items such as “Feel at ease in a large group of people” (EI 7) and “Make people feel at ease” (AGR 6), they might first have

wondered, “Who are these people? How am I related to them? Am I older or younger than them? Are they family relatives, soccer club members, fellow classmates, teachers, or total strangers? Who is the “senior”? Who is the “junior”? How am I expected to act in this group?” Without knowing the relative ages of the “people” referred to in the items, the study participants might have found themselves at a loss for an answer.

Similarly, other items that asked participants whether they “know how to cheer people up” (AGR 8), are “interested in other peoples’ problems,” and “inquire about others” (AGR 1) might seem simply like “busy-body” behavior to the participants. “Why should I get involved in the affairs of other people whom I may barely know, if at all?” is a thought that the participants might have had when reading such items. Showing untoward interest in strangers, or the opposite, showing the desire for interest from others, could be perceived as uncomfortable, or even unhealthy, social behavior. Helping total strangers or demanding attention from strangers at a party could potentially be seen as demonstrating a lack of awareness of the social *senpai-kohai* hierarchy, the insistence that the individual is more important than anyone else, or simply as an uncomfortable action if the party included people from an outside group. For example, for Japanese university participants in the study, membership in a university social club, or “circle,” has traditionally been an important stepping-stone for developing relationships among

senpai who may be able to offer them guidance about university life as well as employment advice upon graduation. Thus, the study participants may feel hesitant or reluctant to show too much interest in an “outside person” who has no obvious social relation to them through membership in their university groups.

At this juncture, I should point out that it is speculation on my part whether the nature of participant sample (i.e., Japanese university students) is related to the model results, because I have no comparison sample from a non-Japanese university student population. It would also be remiss of me not to acknowledge that the nature of the Japanese social hierarchy and school system has been hotly debated since the 1990s, particularly regarding the existence of Hofstede’s (1980) concept of individualism/collectivism. For example, Kubota (1999) has argued that studies concerning the teaching of English writing in Japan have focused too much on Japan as “Other.” Additionally, Takano and Osaka (1999) found in a review of fourteen cross-cultural studies that only one explicitly supported the view that Japanese were more “collective” than Americans, and that one study was by Hofstede (1980). Focusing on the existence of “subcultural diversity” in Japanese society, Sugimoto (2010) criticized the “group model of Japanese society” (p. xi, 4) as an empirically unsupported *Nihonjin-ron* (theory about Japanese-ness) based on the assumption of cultural homogeneity.

On the other hand, a number of personality studies in China (e.g., Cheung & Leung, 1998; Zhang & Bond, 1998) have demonstrated the inadequacy of Big Five-derived personality trait inventories to measure what they term “Chinese tradition.” Other psychologists (e.g., Church & Lonner, 1998; Markus & Kitayama, 1998) have speculated that “Western” society-created personality instruments view personality as stable and unchanging, while “Asian” society-created instruments follow the “Confucian” view of a “dynamic personality,” in other words, that personality continuously changes to meet the social need of the individual (Cross & Markus, 1994, pp. 391-392).

Without delving too deeply into the issue (which would necessitate a separate study), I will simply note that, based on the results from the Big Five instrument in this study, it is difficult to determine which of two possibilities are the cause of the missing Agreeableness factor: the differences between the Japanese sample and American samples for whom the instrument was originally designed, or the fact that the personality trait instrument was poorly constructed to begin with, due to a lack of an in-depth examination of construct unidimensionality. However, given that the exploratory factor analysis structure results of the pilot study essentially agreed with the main study that Extraversion and Agreeableness could not be separated, despite a substantial revision of questionnaire items, I am inclined to believe that the results indicate the existence of characteristics of the Japanese

sample in the study that explain the failure of the Big Five to emerge and that bear further examination (see Chapter 7: Suggested Future Research Directions).

Based on the results, it seems clear that someone who talks a lot within the appropriate social group was seen as an extroverted personality by the study participants. While someone who wants to be the “center of attention” (EI 5) or “likes to draw attention” (EI 4) is an unwelcome personality in the Japanese context, talking a lot is apparently acceptable and even desired behavior in the language classroom. Thus, the results agree with the conceptualization of “ego-seeking” behavior not as Extraversion (as in the Big Five Factor Model) but as Psychoticism (as in the Big Three model). In the case of the Japanese university participants in this study, seeking attention likely differed from being talkative, perhaps because of the perception that someone who wants to be the center of attention shows a certain sense of impulsivity, lack of restraint, and resistance to the social norm of avoiding standing out (Weaver, 1998; Zuckerman, Kuhlman, & Camac, 1988).

In contrast to the Extraversion and Agreeableness constructs, the Emotional Stability construct did not correlate strongly with any of the other personality trait constructs. This result is in agreement with nearly all previous personality trait research, in which Emotional Stability or Neuroticism is one of the most robust personality dimensions regardless of which personality trait model is being used (Angleitner & Ostendorf, 1994). Despite the robustness of the construct shown by

the Rasch analysis and EFA, Emotional Stability, as it was operationalized in the CFA, ultimately consisted of only four items that referred to vulnerable or passive emotional states (i.e., worry, stress, and melancholy) rather than an aggressive emotional state (i.e., anger). Thus, the results also agree with previous research that separated Emotional Stability into two related forms of “positive emotionality” and “negative emotionality” (Ashton & Lee, 2001). In addition, the CFA items were not so much an indication of “stability,” as suggested by the traditional construct label of Emotional Stability, as a tendency towards fearful or negative emotions. Therefore, I suggest that the label *Negative Emotionality* or *Neuroticism* is more appropriate for this construct.

Conscientiousness, as it emerged in this study, also agreed with previous research. The Rasch analysis indicated that Conscientiousness correlated weakly with Agreeableness ($r = .24$) and Intellect/Imagination ($r = .27$), and in the final CFA, Conscientiousness had similar correlations to Extraversion ($r = .24$) and Imagination ($r = .25$). These results are in agreement with previous studies in both Korea and Japan (Hahn et al., 1999; Kashiwagi, 2002). While proponents of the Big Five have argued that Agreeableness and Conscientiousness are distinct dimensions of personality, proponents of the Big Three system of psychoticism, extraversion, and neuroticism (PEN) have often pointed to correlations between Conscientiousness and Agreeableness as indications that the two constructs are

lower levels of the higher-order factor term Psychoticism (Eysenck, 1998). However, the results of this study show an overlap of slightly less than 6% between Conscientiousness and Extraversion and slightly more than 6% between Conscientiousness and Imagination. With such low overlap percentages, it is difficult to see how the constructs could be conceived of as indicating the same personality trait. Thus, the results agree with the Big Five proponents who have maintained that Conscientiousness, Extraversion, and Imagination are substantially different in nature.

However, the CFA results disagree as to the specific nature of the construct items. The 100 “synonym cluster” (Goldberg, 1990) and 100 “unipolar adjective” (Goldberg, 1992) Big Five questionnaires included the following descriptions of Conscientiousness: dignity, caution, punctuality, logic, conventionality, forgetfulness, sloth, aimlessness, disorganized. Rather than descriptions of personality traits, these words are more like descriptions of behavior. Having a formal manner (dignity), using critical thinking skills (logic), and not remembering where you left your car keys (forgetfulness) seem more like learned habits rather than unchanging, innate personality traits.

In fact, the final CFA for the Big Five personality traits included only three Conscientiousness items, all of which were related to perfectionism and planning; items concerned with organization skills (CON 2) and keeping things neat and tidy

(CON 6) were correlated to other items in the EFA, but their residuals correlated only with each other in the Rasch analysis. They also failed to cohere strongly enough to other items to form a construct in the CFA. It is worth reiterating that the participants were all first-year university students who by virtue of being students had to follow preplanned daily course schedules and they were socially obligated to follow rules of conduct. Moreover, because the majority of the participants had recently graduated from high school (average age = 18.87) and had completed the questionnaire in the middle of their first term at university, it is likely that they were still in the process of encountering new people (students, teachers, neighbors), learning the differences between high school and university life (fewer restrictions, more personal responsibility) and negotiating their social identities as first-year university students.

However, because a cross-sectional design was used, it is not certain whether the participants would have answered differently had they taken the questionnaire just a few months prior to entering university (see Chapter 7: Future Research Directions). Because most personality trait studies involve large samples of university students, it would not be surprising if similarly worded items in other Big Five personality trait questionnaires likewise correlated with Conscientiousness items, yet failed to form a coherent construct.

The first indication that the final personality trait, Intellect/Imagination, might be multidimensional was the very name of the construct. The Rasch analysis revealed the construct to have the largest first contrast variance explained (13.3%), and the Rasch PCA of item residuals showed that the items with positive and negative residual loadings were logically distinct; all imagination-related items loaded positively and all the intellect-related items loaded negatively.

Although the Openness or Intellect/Imagination construct did not appear in a few previous studies (e.g., Cheung et al., 2001; Ortiz et al., 2007; Wang et al., 2005), the Imagination factor did appear in this study. The Intellect/Imagination construct had medium correlations with both Extraversion ($r = .41$) and Agreeableness ($r = .39$) during the preliminary Rasch analysis, and the Imagination factor correlated strongly with Extraversion in the CFA ($r = .46$) and in the structural regression model ($r = .48$). This result is similar to that found between Intellect/Imagination and Extraversion ($r = .46$) in a previous Japan-based study (Kashiwagi, 2002) and between Openness to New Experiences and Extraversion ($r = .25$) in a Multitrait-Multimethod (MTMM) study conducted in a North-American context comparing three different instruments measuring the Big Five (John & Srivastava, 1999, p. 119).

However, only three of the original Intellect/Imagination items (INT 1, INT 2, and INT 6) were viable in the final CFA and structural regression model. All three

items were identified in the Rasch PCA of item residuals as positive loading items related to abstract or imaginative thinking. The remaining four items were concerned with intellectual skills: “Have a rich vocabulary” (INT 3), “Am quick to understand things” (INT 4), “Love to read challenging materials” (INT 5), and “Can handle a lot of information” (INT 7) seem less like descriptions of personality traits and more like descriptions of intellectual capacity. In other words, the “Intellect” half of “Intellect/Imagination” is a measure of perceived ability or aptitude, not personality.

The results of this study clearly demonstrate that the so-called “Factor V” of the Big Five, whether termed “Intellect/Imagination” or “Openness to New Experiences,” would also become valid for the Japanese university population if the items were clearly delineated into a single, unidimensional construct of Imagination. I strongly suspect that the main reason why Factor V has proven difficult to replicate across cultures is precisely because Factor V is not a unidimensional construct as assumed by the Big Five factor model.

Personality psychologists have overlooked or even ignored the problem of unidimensionality in the factors comprising their personality trait models. Graziano (1994), while theorizing that the Agreeableness trait might be “even more salient as a personality dimension to persons reared in collectivist societies” (p. 340), noted that the concept of Agreeableness “may not be unidimensional” (p. 341). However,

he also dismissed the lack of dimensionality as a problem, because there was a “major link in adults between Agreeableness...and Extraversion” due to the two factors “overlapping characteristics and facets” (p. 342). This “overlap” is no doubt due to the fact that the Big Five was a creation of exploratory factor analysis, and not a product of careful experimentation with a theoretical underpinning (Block, 1995; Digman, 1990, 1996; Eysenck, 1991; Goldberg, 1992). Proponents of the Big Five and the PEN model based their arguments almost entirely on factor analysis using various rotational analyses to clarify the factor loadings, and when examining samples in different parts of the world, Big Five personality trait researchers assume the existence *a priori* of a set number of factors (Eysenck, 1994; Goldberg, 1994; Zuckerman, 1994; Zuckerman, Kuhlman, & Camac, 1988).

Overall, the results for the Big Five personality traits in the present study clearly demonstrate that relying solely on EFA and Cronbach’s alpha to determine the nature of personality trait constructs can lead researchers down the garden path to thinking that they have demonstrated the existence of five personality trait constructs by looking for what they already assumed to be present in the data. As Eysenck has bluntly put it, “Factor analysis is a necessary but not a sufficient method for arriving at the basic dimensions of humanity” (Eysenck, 1994, p. 38).

Research Question Two: Foreign Language Classroom Speaking Confidence

The second research question concerned the degree to which Foreign Language Classroom Speaking Anxiety was related to Perceived Foreign Language Speaking Self-Competence and Desire to Speak English. All three constructs were hypothesized to comprise foreign language classroom speaking confidence.

Although previous research considered foreign language confidence a combination of the two concepts of anxiety and perceived competence (e.g., Clément et al, 1980; Clément et al, 1994; MacIntyre et al., 1998a), the Desire to Speak English construct was added in this study based on the hypothesis that participants who felt no desire to speak in English would logically not experience anxiety in speaking English due to a lack of interest in successful language speaking outcomes. The inclusion of three variables, rather than only two, was further justified on the grounds that any higher order latent variable such as foreign language confidence ought to consist of at least three related constructs in order to avoid empirical model underidentification (Chen et al., 2001; Kenny et al., 1998). The second research question was answered by conducting a CFA with Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English.

Summary of the Results for Research Question Two

In the CFA model, Foreign Language Classroom Speaking Anxiety was found to have a small negative correlation to Perceived Foreign Language Speaking Self-Competence, $r = -.19$. The Foreign Language Classroom Speaking Anxiety construct also correlated weakly with Desire to Speak English, $r = .11$. These correlations were similar to those found in the preliminary Rasch analysis, in which Foreign Language Classroom Speaking Anxiety displayed a weak negative correlation with Perceived Foreign Language Speaking Self-Competence ($r = -.11$) and a weak positive correlation with Desire to Speak English ($r = .10$). After path respecification of the model, good fit was found on two fit indices (SRMR and RMSEA) and adequate fit was found on two additional fit indices (CFI and NNFI) using the validation sample. Thus, the results indicated that Foreign Language Classroom Speaking Anxiety for the participants in this study was only weakly related to the participants' Perceived Foreign Language Speaking Self-Competence and Desire to Speak English.

Interpretation of the Results for Foreign Language Classroom Speaking

Confidence

The results of the CFA contradict previous research into foreign language confidence insofar as foreign language anxiety was not related strongly enough to

perceived self-competence for the two factors to form a higher order latent variable of foreign language confidence. The strength of the correlations between Desire to Speak English and Perceived Foreign Language Speaking Self-Competence ($r = .53$) demonstrates support for the theory that the concept of foreign language classroom speaking confidence must include a measure of the desire of participants to use the foreign language. However, because the strength of the relationships among Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English is different between the CFA model and the final SRM, it is also possible that foreign language classroom speaking confidence includes more than the three hypothesized constructs.

It should be pointed out that the Desire to Speak English items in the final CFA model (DSE 1, DSE 3, DSE 5, DSE 7, DSE 8) are conceptually similar to the idea of international posture (Yashima, 2002). This suggests that a missing component of speaking confidence until now could be an attitudinal or motivational construct such as integrative orientation (Gardner, 1988; Gardner & MacIntyre, 1993) or friendship orientation (Clément & Kruidenier, 1983, 1985). Another missing component could be learning beliefs, especially efficacy (Bandura, 1982), concerning learning how to speak English, which might change as the participants experience positive or negative results from speaking opportunities in the foreign

language classroom. Still another possibility is the existence of social factors specific to the foreign language classroom, such as fear of negative social evaluation by peers or factors related to overall classroom dynamics that might affect how participants view their English speaking competence.

The results of the structural regression model indicated that Foreign Language Classroom Speaking Anxiety had a moderate effect on Perceived Foreign Language Speaking Self-Competence, with a negative beta weight of $\beta = -.29$. This finding is similar to that of Matsuoka (2006), who found that perceived self-competence and communication apprehension (i.e., foreign language anxiety) for Japanese university students were moderately correlated at $r = -.34$. The relationship between Desire to Speak English and Foreign Language Classroom Speaking Anxiety in the SRM was positive, but the strength of the effect was almost the same ($\beta = .24$) as that between Perceived Foreign Language Speaking Self-Competence and Foreign Language Classroom Speaking Anxiety. The path between Perceived Foreign Language Speaking Self-Competence and Desire to Speak English ($\beta = .33$) additionally supported the CFA model results by showing the connections between Perceived Foreign Language Speaking Self-Competence and Desire to Speak English.

Previous researchers have explored the connection between the perception of competence and anxiety (e.g., MacIntyre et al., 1997; MacIntyre et al., 1998a);

however, the fact that the correlation between the two variables in the current study is relatively weak indicates that the nature of foreign language speaking confidence is more complicated than previously assumed. The strong correlation in the CFA between Desire to Speak English and Perceived Foreign Language Speaking Self-Competence ($r = .53$) suggests that there is a strong element of motivational desire necessary for study participants to feel confident in speaking English. Additionally, the structural regression model indicates that not only does the desire to speak lead directly to anxiety about speaking, but that perceptions of competence in speaking English leads both to anxiety and the desire to speak. Because the social situation variables influence Perceived Foreign Language Speaking Self-Competence and Desire to Speak English, the results can be construed as evidence that Foreign Language Classroom Speaking Confidence is not a psychological variable that occurs in isolation within individuals; instead, it is directly related to a stable personality trait that can be heightened or subdued in accordance with the support or lack of support from peers in the foreign language classroom.

The fact that the relationships among the three components of foreign language classroom speaking confidence become stronger in the structural regression model compared to the correlations in the CFA indicates that the personality and social situational variables in the structural regression model play

an important role in determining the English speaking confidence of participants in the study. This finding is in agreement with suggestions made in previous studies (e.g., MacIntyre & Gardner, 1994; Price, 1991) that “language anxiety arises when speaking in a social context” (MacIntyre, 1995, p. 246). Thus, it is difficult to examine the concept of foreign language classroom speaking confidence solely based on three latent constructs that are supposed to exist within an individual. This concept is elaborated on further during the discussion of the relationships among the social situation and foreign language classroom speaking confidence variables (see the Research Question Four: Classroom Climate and Speaking Confidence section).

Research Question Three: Personality and Speaking Confidence

The third research question concerned the degree to which the Big Five personality trait factors impact upon foreign language speaking confidence in the classroom. Extraversion and Emotional Stability were hypothesized to influence Foreign Language Classroom Speaking Anxiety, and Imagination was hypothesized to influence Perceived Foreign Language Speaking Self-Competence and Desire to Speak English. Conscientiousness was not hypothesized to directly influence any of the three foreign language classroom speaking confidence variables. The third research question was answered by conducting a structural

regression model with the Big Five personality traits as exogenous variables, the social situation latent variables as interceding endogenous variables, and the foreign language classroom speaking confidence latent variables as outcome variables.

Summary of the Results for Research Question Three

The results of the structural regression model revealed that two of the four hypothesized paths between the Big Five personality trait variables and the foreign language classroom speaking confidence variables were significant. Emotional Instability had a moderate influence on Foreign Language Classroom Speaking Anxiety, $\beta = .43$, and Imagination had a weak influence on Perceived Foreign Language Speaking Self-competence, $\beta = .16$. Extraversion and Conscientiousness did not directly influence the foreign language classroom speaking anxiety variables. Thus, the results show that of the Big Five personality traits, Emotional Stability has the strongest influence on speaking anxiety in the foreign language classroom, while the remaining three traits have little to no direct influence.

Interpretation of the Results for the Relation between Personality and Speaking Confidence

Contrary to the expectation that Extraversion would have a significant direct influence on Foreign Language Classroom Speaking Anxiety, the results indicate that Extraversion has an indirect influence as mediated by social situation variables (see the following section). This also contradicts previous findings in which extraversion had a direct negative influence on anxiety (MacIntyre & Charos, 1996). It was previously theorized that extraverts' preference for social activity would lead to lower anxiety (MacIntyre & Noels, 1994); the lack of a significant path between Extraversion and Foreign Language Classroom Speaking Anxiety in the current results points to the importance of the social situation in determining whether extraverted or introverted individuals experience more or less anxiety in the foreign language classroom.

For example, if an introverted student wants to speak English in the foreign language classroom, but is surrounded by extraverted classmates who have little desire to speak English or who dominate pair and group conversations so much so that the introverted student has little opportunity to speak English, the introverted student might experience a decreased sense of competence in speaking English, which in turn leads to a lessened desire to speak English and an increase in speaking anxiety in the classroom. Simply being introverted is not enough to cause

anxiety in the classroom: The perception of speaking competence and desire to speak are contingent upon the perceptions and desires of other classmates, who can either support the individual's speaking experience or detract from it.

Likewise, the finding that Emotional Stability had a moderate, direct influence on Foreign Language Classroom Speaking Anxiety, and no significant influence on perceptions of the social situations in the classroom, contradicts previous SLA research in which personality traits were not theorized to have a direct influence on foreign language anxiety (MacIntyre, 1999; MacIntyre & Charos, 1996; MacIntyre & Gardner, 1989). On the other hand, the results strongly support the mainstream view in the fields of psychology and communication studies that social anxiety is composed of a personality-based component that can be acted upon positively or negatively by environmental influences leading to an alleviation of or an increase in anxious behaviors (Leary & Kowalski, 1995; McCroskey & Beatty, 1997). Thus, Japanese university students who are by nature more prone to emotional instability may behave in a more anxious manner in the foreign language classroom if they have previous experiences of failure in speaking English in high school. However, while not able to completely erase all traces of anxiety, the students might be in a position to control their anxious behavior through positive peer support in the current English classroom from classmates who offer encouragement and help with English language speaking tasks.

The finding that Imagination directly influences Perceived Foreign Language Speaking Self-Competence is in agreement with previous research in which the Big Five personality trait of Intellect was a cause of Perceived Competence (MacIntyre & Charos, 1996). However, the direct influence of Imagination was fairly weak ($\beta = .16$), whereas the influence of social situation variables was much stronger (see the following section). Imagination also had a sizeable correlation ($r = .48$) with Extraversion, which had an indirect influence on Perceived Foreign Language Speaking Self-Competence. The personality trait of Conscientiousness was not theorized to have any direct influences on foreign language classroom speaking confidence variables.

Thus, the results indicate partial agreement with existing models that most personality trait variables have an indirect, rather than a direct, influence on foreign language confidence variables (MacIntyre & Charos, 1996; MacIntyre et al., 1998). The main point of disagreement with previous models—the link between Emotional Stability and Foreign Language Classroom Speaking Anxiety—has not been attested in previous research. This link should therefore be regarded as a major finding of the current study with implications for both researchers and practitioners (see the Theoretical Implications and Pedagogical Implications sections).

Research Question Four: Classroom Climate and Speaking Confidence

The fourth research question concerned the degree to which the classroom climate influences foreign language speaking confidence in the classroom. In the structural regression path model, the favorable social conditions variables of Past English Classroom Experiences, Current English Classroom Perception, and Perceived Social Value of Speaking English were hypothesized as interceding variables leading to the foreign language classroom speaking confidence variables of Desire to Speak English, Perceived Foreign Language Speaking Self-Competence, and Foreign Language Classroom Speaking Anxiety. To answer the research question, the paths between the favorable social conditions variables and foreign language classroom speaking confidence variables were examined.

Summary of the Results for Research Question Four

An examination of the structural regression path model results indicated several significant paths between favorable social conditions variables and foreign language classroom speaking confidence variables. The strongest relationship was the influence of Current English Classroom Perception on Perceived Foreign Language Speaking Self-Competence. In a data-driven path, Current English Classroom Perception additionally had a small influence on Desire to Speak English. Past English Classroom Experiences had a moderate influence on

Perceived Foreign Language Speaking Self-Competence; however, the hypothesized connection between the past and current English classroom situations was not significant. Perceived Social Value of Speaking English had a small to moderate influence on Current English Classroom Perception and a moderate influence on Desire to Speak English.

Interpretation of the Results for the Influence of Classroom Climate on Foreign Language Classroom Speaking Confidence

The results highlight the importance of the perception of social conditions in which communication arises for the Japanese participants in this study: lacking members of an identifiable English-speaking community with which to interact outside the foreign language classroom, the most salient social condition for the participants is the immediate English as a Foreign Language classroom. Although none of the favorable social conditions variables directly influence Foreign Language Classroom Speaking Anxiety, they have significant impact on the participants' Desire to Speak English and Perceived Foreign Language Speaking Self-Competence. For example, Perceived Social Value of Speaking English leads directly to Desire to Speak English and also indirectly to Desire to Speak English through Current English Classroom Perception. Thus, even if a particular study participant were to see no value for speaking English outside the university English

classroom, he or she is influenced by perceptions of the value that English classroom classmates place on speaking English. Students who otherwise might not feel that speaking English is valuable or useful might find themselves desiring to speak English due to the positive attitudes of classmates; in other words, the results of the study indicate the presence of positive peer pressure in the foreign language classroom.

In a similar vein, the personality trait Imagination has a direct, but fairly weak, influence on Perceived Foreign Language Speaking Self-Competence, while the social situation variables of Positive English Classroom Experiences and Current English Classroom Perception have moderate direct influences on Perceived Foreign Language Speaking Self-Competence. Thus, even if an individual student perceives himself or herself as not having a strong Imagination, positive experiences in the high school English classroom can lead to a sense of competence at speaking English. Similarly, because the influence of Current English Classroom Perception is stronger than that of Positive English Classroom Experiences, a student with negative experiences in high school may discover that his or her negative experiences of speaking English in previous classes become counterbalanced by the perception of a positive, supportive classroom atmosphere in the current university English classroom. This finding makes the role of the foreign language teacher crucial, as the teacher is instrumental in organizing the

language learning activities and forming the overall classroom atmosphere (see the Pedagogical Implications section).

The results of the present study provide causal evidence of the connection between group dynamics in the foreign language classroom and foreign language classroom speaking confidence. The paths between Past English Classroom Experiences and Perceived Foreign Language Speaking Self-Competence support the theory that self-confidence can be affected both positively and adversely by previous experiences using the foreign language (MacIntyre et al., 1998). Likewise, the path between Current English Classroom Perception and Perceived Foreign Language Speaking Self-Competence provides evidence of the suggestion by Clément et al. (1994), that the presence of a “good classroom atmosphere” (p. 442) provides increased support for confidence due to the encouragement by fellow students to actively use the foreign language.

On the other hand, while Clément et al. theorized that “bad” foreign language experiences directly lead to an increase in anxiety, the results of the present study indicate a more complicated picture. In the model, the favorable social conditions variables indirectly influence Foreign Language Classroom Speaking Anxiety through a combination of Perceived Foreign Language Speaking Self-Competence and Desire to Speak English. Thus, the results suggest a tripartite origin of foreign language classroom speaking confidence. One component is the existence of

emotional stability or neuroticism in the individual student's personality. The second component is the existence or lack of existence of the desire to speak, which is dependent partly on perceptions of the attitudes of current classmates and partly on individual perceptions of the value of English outside the classroom. The third component is the sense of increased or decreased perception of the ability to communicate in the foreign language, which in turn is informed by a combination of past experiences in the language classroom and the support or lack thereof from current classmates. These three components of foreign language classroom speaking confidence stand in stark contrast to existing SLA theories regarding foreign language anxiety (see the Theoretical Implications section).

The concept of Perceived Social Value of Speaking English in this study should be commented on, as it differs from that used in previous research. The concepts of "Context" (MacIntyre & Charos, 1996) and "L2 Norms" (Clément et al., 2003) were defined in French as a Second Language contexts as the degree to which participants used their L1 and L2 at home and at work. In both studies, the social conditions variable led to a confidence variable (Perceived Competence in MacIntyre and Charos, and L2 Confidence for Anglophones in Clément et al.). In the present study, the items measuring Perceived Social Value of Speaking English concern the study participants' *perceptions* of the value placed on English speaking ability by members of the greater social community outside the immediate English

classroom, rather than the *frequency* of their English speaking outside the classroom. It should be noted, however, that the study participants were not asked whether they had study abroad experience or had lived abroad prior to entering the university (see the Limitations of the Study section); however, given the small percentage of the population of Japan that speak English, it is highly unlikely that the participants have had many opportunities to speak English outside their classrooms, other than the chance encounter with an English-speaking tourist. Thus, this difference from the “Context” variables in ESL studies was necessitated by the participants’ overall lack of foreign language use opportunities outside the classroom.

This difference in “context” may help explain the lack of the nonvariant path between the Perceived Social Value of Speaking English and Desire to Speak English in the final structural regression model. Although there was a moderate correlation for both samples, the beta weight of the path changed from $\beta = .34$ for the calibration sample to $\beta = .41$ for the validation sample. This slight, but statistically significant, change may be due to individual student differences in instrumental or integrative motivation for speaking English outside the foreign language classroom. The fact that the participants indicated a strong desire to communicate in English outside the classroom, as evidenced by the items measuring Desire to Speak English, additionally supports the hypothesis that a

missing element of the model may be a motivational construct (see the discussion regarding the Desire to Speak English construct in the Interpretation of the Results for Foreign Language Classroom Speaking Confidence section).

Whether the motivation to speak English is primarily instrumental (e.g., to get a job or to earn a qualification) or integrative (e.g., interest in foreign cultures or interest in speaking to foreigners) cannot be determined from the existing construct items and model. One possibility for the lack of nonvariance in the path between Perceived Social Value of Speaking English and the Desire to Speak English is that individual participants' motivations to speak English may have changed upon entering university. As high school students, the participants may have had primarily instrumental motivation for English learning (i.e., getting a high score on the English section of a certain university's entrance exam). However, upon entering university, some of the participants may have turned their thoughts toward life after formal education, which may consist of both instrumental and integrative motivation to speak English.

For example, of the top two easy to endorse items in the PSV construct, one was instrumental (PSV 6, "Japanese companies think highly of workers who have English speaking ability," Rasch item difficulty measure = 47.62; $M = 4.67$; $SD = 1.18$) and one was integrative (PSV 7, "Speaking English will help me become an internationally-minded person," Rasch item difficulty measure = 44.74; $M = 4.97$;

$SD = 1.14$). The items measuring Perceived Social Value of Speaking English may thus indicate the beginning of motivational change for the participants' views toward speaking a foreign language. It seems plausible that, as the participants progress through their university studies, their individual desires and motivations to speak English may change as they gain a clearer understanding of their careers after university education. Because of the cross-sectional nature of the study, I can only speculate on this point concerning motivation and can simply note it as a potential future direction for further study regarding the relationship between classroom climate and foreign language speaking confidence (see Chapter 7: Suggestions for Future Research).

One final comment concerns two items intended to measure Positive English Classroom Experiences. Item PECE 4 (“I spoke English frequently with classmates during my high school English classes”) was the most difficult item to endorse (Rasch item difficulty measure = 66.35; $M = 2.07$; $SD = 1.31$), and was answered by 1,077 participants. On the other hand, PECE 3 (“I enjoyed speaking with classmates in my high school English classes”) was easier to endorse (Rasch item difficulty measure = 50.45; $M = 3.11$; $SD = 1.48$), but it was answered by 1,067 participants. The difficulty in endorsing PECE 4, as well as the low mean, indicate the relative lack of experience that Japanese university students have speaking English in high school classes prior to entering university. The high standard

deviation of PECE 3, in addition to the fact that eleven participants failed to answer the item, is also indicative of a lack of communicative teaching in Japanese high schools. This finding agrees with previous research findings that high school teachers in Japan were not using communicative teaching approaches in high school English classes, even in classes designed ostensibly for improving oral communicative ability, despite nearly two decades of educational policy changes and pressure from MEXT (MEXT, 2007; Nishino, 2009; see also Chapter 2: The Japanese Context, Educational influences).

For individuals who have little to no opportunities for communicative interaction with English-speaking people outside the normal classroom, interactions inside the EFL classroom are crucial to the development of confidence in speaking English. Because Positive English Classroom Experiences has a significant influence on Perceived Foreign Language Speaking Self-Competence, the participants who did not experience pair work or speaking with classmates in English during high school are clearly at a disadvantage in developing their speaking confidence. Fortunately, the direct influence of Current English Classroom Perception on Perceived Foreign Language Speaking Self-Competence and Desire to Speak English demonstrates that negative experiences in the high school English classroom can be mitigated by a supportive, positive classroom atmosphere in the current university English classroom.

The model indicates that, as predicted by theories of social anxiety, foreign language classroom speaking anxiety for the Japanese university participants is at least partly attributable to personality traits. However, the other foreign language classroom speaking confidence variables of Perceived Foreign Language Speaking Self-Competence and Desire to Speak English are influenced by events and experiences with classmates in the participants' current foreign language classrooms. In the next section I discuss theoretical implications of the results, especially how they can be applied toward a reconceptualization of foreign language anxiety and foreign language speaking confidence.

Theoretical Implications

According to the structural regression model in this study, the primary influence of personality traits on foreign language classroom speaking confidence is indirect and is channeled through social situations in which foreign language users find themselves in the foreign language classroom. This finding thus agrees with the theory proposed in the heuristic pyramid model of MacIntyre et al. (1998a), in which personality traits comprise the bottom, most enduring layer of the pyramid and have an indirect influence on other non-linguistic elements of L2 communication. However, there is a significant difference between the theory

proposed in the heuristic pyramid and the structural regression model results in the current study.

On one hand, two of the four Big Five personality traits appear to have only indirect influences on foreign language speaking confidence. Given the strength of the effect of Extraversion on the Current English Classroom Perception variable ($\beta = .41$), it is plausible that, although extroverted students do not necessarily feel more or less anxiety in the foreign language classroom, their presence and outgoing nature can help fellow classmates feel more comfortable and relaxed. Extroverts also seem to have slightly more positive experiences in their high school English classes, and seem slightly more aware of the value of speaking English outside the classroom. In their current English classrooms at university, perhaps extroverts perform a role as “mood-makers,” who can raise classmates’ spirits or lessen tensions that can arise in the foreign language classroom.

Participants with higher levels of Conscientiousness are more likely than their peers to have more positive past experiences in their high school English classes, perhaps due to having studied more diligently or being more effectively or consistently organized in their studies and are also more likely to be aware of the value of speaking English outside the immediate educational context of the foreign language classroom. Thus, although not directly influencing foreign language classroom speaking confidence, extroverted, conscientious students have a social

advantage in the classroom that can lead them to have more confidence and to help their classmates have more confidence through improved class relations.

On the other hand, Imagination has a small, but significant direct effect on the participants' perceptions of their own English speaking ability. Although a significant influence, the effect on Perceived Foreign Language Speaking Self-Competence ($\beta = .16$) is far smaller than that of Positive English Classroom Experiences ($\beta = .39$). Given the nature of the items measuring the Imagination construct, it is possible that participants with higher levels of Imagination are not actually more competent at speaking English but simply perceive themselves as being more competent. By far the strongest direct influence between personality traits and Foreign Language Classroom Speaking Confidence, however, is the relationship between Emotional Stability and Foreign Language Classroom Speaking Anxiety, which clearly indicates that more emotional unstable participants tend to experience more anxiety speaking English in the foreign language classroom. The results thus suggest a complicated construction of Foreign Language Classroom Speaking Confidence, one that is influenced both by theoretically unchangeable individual personality traits and by malleable social classroom-related variables.

The finding that foreign language classroom speaking confidence consists of both trait and state components agrees with the opinions of mainstream

psychologists that social anxiety (which includes foreign language anxiety and thus confidence) includes both genetic (i.e., personality and/or neurological) as well as social (i.e., state or situation-specific) components. This again corroborates the heuristic pyramid model of MacIntyre et al. (1998a), which also theorized trait and state components of self-confidence. However, whereas the pyramid model uses the term L2 Self-Confidence to describe the trait component and State Communicative Self-Confidence for the effect of previous language usage, the structural regression model indicates that Foreign Language Classroom Speaking Anxiety is a unitary concept that is partly trait-informed and partly situationally informed. This finding contradicts previous research in which researchers speculated that foreign language anxiety was a kind of “situation-specific anxiety,” which was characterized as a “trait-like” tendency to be anxious in certain situations but not in others (MacIntyre & Gardner, 1989, 1991).

Moreover, the CFA model of foreign language classroom speaking confidence demonstrates that the higher order construct theorized to consist of two variables should consist of three variables—Perceived Foreign Language Speaking Self-Competence, Foreign Language Classroom Speaking Anxiety, and Desire to Speak English. Finally, the structural regression model demonstrates that individual foreign language learners can experience lower or higher foreign language classroom speaking confidence contingent upon the individual’s relationships and

interactions within the foreign language classroom and in direct relation to their individual level of Emotional Stability.

One major difference between the results of the current study and the results from previous studies about foreign language confidence (e.g., Clément, Dörnyei, & Noels, 1994; MacIntyre, Noels, & Clément, 1997) is that the constructs measured in the current study consisted of items concerning foreign language speaking within the classroom, while the constructs in previous studies consisted of items about foreign language tests, responding to direct questions from the teacher, and using the foreign language in situations outside the classroom. The differences between the results from the CFA and structural regression model of Foreign Language Classroom Speaking Confidence and the path model study of MacIntyre and Charos (1996) might additionally stem from three sources.

The first source is the difference in the ratio of *N*-size and number of items, which was less than 1:1 in MacIntyre and Charos and more than 33:1 for the current study. Stevens (2009) recommended a case to item ratio of at least 10:1 in order for the results to be generalizable across samples. Moreover, statistical modeling techniques such as path modeling and structural equation modeling require large participant numbers in relation to the number of items and parameters in order to reduce the measurement error of the intended constructs. By reducing measurement errors, the models can provide more accurate results that lead to a

greater precision of predicted outcomes during future iterations with different samples. Thus, a small case to item ratio leads to questionable study results that lack external validity because they are likely not to be reproducible.

The second source of differences between the results in MacIntyre and Charos and the present study is the composition of the anxiety construct. L2 Anxiety was measured with two items in MacIntyre and Charos, while the Foreign Language Classroom Speaking Anxiety construct in the current study consisted of seven items in the final CFA and structural regression model. When two items are used to define a construct, a high correlation between the two items simply indicates that the items are essentially measuring the same concept (and therefore are virtually identical), resulting in a conceptual, if not a statistical, violation of multicollinearity (Kenny, 1979). Therefore, at least three items should be used to define a construct, and four or more items are preferable in order to measure the construct with a greater degree of precision along the endorsability scale.

The third source of differences is the directional influence of the variable components of L2 confidence. Anxiety was previously defined as influencing perceived competence (MacIntyre, 1994; MacIntyre & Charos, 1996), whereas in the current study Foreign Language Classroom Speaking Anxiety is an outcome of both Perceived Foreign Language Speaking Self-Competence and Desire to Speak English. Although an equivalent model indicated that the data could support the

idea of Foreign Language Classroom Speaking Anxiety influencing Desire to Speak English, the equivalent model was discounted due to the negative nature of the anxiety construct. Similarly, it is unlikely that the presence of Foreign Language Classroom Speaking Anxiety would lead to greater perceived competence at speaking English. Thus, anxiety, and not perceived competence, is logically the final outcome of the model in this study.

Due to the strength of the influence of Emotional Stability on Foreign Language Classroom Speaking Anxiety, the model in this study suggests that Foreign Language Classroom Speaking Anxiety might not be not a situation-specific anxiety that only occurs in a foreign language learning context. Because Foreign Language Classroom Speaking Anxiety, and by extension Foreign Language Classroom Speaking Confidence, is at least partly influenced by personality traits, it is likely that students who demonstrate anxiety when speaking English in their classrooms are also anxious outside the classroom depending on their interactions with those around them.

A final theoretical implication for SLA researchers concerns the practice of adding raw scores obtained from Likert-scale questionnaires. The Rasch item analyses for Past English Classroom Experiences and Perceived Social Value of Speaking English indicated that the six-point Likert scale was not used by enough participants; the category responses needed to be reduced to four categories for

those two constructs, indicating that a participant answering “6” on a PECE or PSV item did not necessarily mean to endorse an item six times the value of a “1.” Moreover, the Rasch item-person maps produced for each of the original eleven constructs in the study showed that the items did not contribute equally to their intended constructs. Therefore, the results clearly demonstrate that Likert-scale based data are rank ordinal numbers, and not true interval data. Simply adding together the raw scores from items that are indicated as loading onto the same factor in exploratory factor analysis—or, even worse, are simply assumed to measure the same hypothetical latent construct without even conducting a factor analysis of some sort—produces inaccurate estimates of the constructs within the study sample.

Pedagogical Implications

The influence on foreign language classroom speaking confidence of experiences speaking in previous English classrooms, perceptions of the value of speaking English, and perceptions of classmates’ attitudes toward speaking English in the current study implies that foreign language classroom speaking confidence is not something that occurs solely within the mind of the solitary foreign language learner. On the contrary, the results indicate that although negative experiences can lead to a lowered sense of speaking self-competence and lack of awareness of the

value of English in the “real” world outside the classroom can lead to a lowered desire to speak English, a supportive classroom atmosphere can lead to a greater sense of competence in speaking English, as well as an increased desire to communicate in English. The results thus indicate that, although there is a strong individual personality trait element of foreign language classroom speaking confidence that cannot be altered, there are other classroom-based, social elements within the locus of control that can offset an individual student’s internal tendency toward anxiety and lack of confidence.

Based on the evidence provided by the model, I can speculate as to a likely chain of events for language learners at Japanese universities who develop anxiety towards speaking in English. Imagine that a student with little or no experience speaking English in her high school English classes in Japan suddenly finds herself in a class of between 25 to 30 unknown classmates in a Japanese university classroom that requires her to speak conversational English. The student quickly realizes that most of her classmates do not particularly value speaking English, are withdrawn and untalkative during class, and generally do not enjoy each other’s company. Lacking social support from classmates, the student suffers from a decreased sense of self-competence speaking English, loses the desire to speak English in the classroom, and loses speaking confidence.

On the other hand, a student with little to no experience speaking English in her high school classroom who finds herself surrounded by classmates who enjoy talking with each other and generally get along well will gain an increased sense of speaking self-competence due to positive peer support and increased opportunities within the classroom to practice speaking English. Keying off her classmates' camaraderie, she finds herself having an increased desire to speak, which, combined with her newfound perception of competence at speaking, leads to fewer feelings of anxiety when speaking English and a greater sense of foreign language classroom speaking confidence.

In my view, the foreign language classroom teacher should play a crucial role in creating a supportive, communicative atmosphere within the foreign language classroom that lowers tensions among classmates, encourages communication, and allows for an increase in the perception of speaking self-competence. Although in this study the items directly concerned with the teacher-student relationship were not included in the final model, items concerned with the relationship between pairs of students and among students in groups were indicative of a positive, supportive classroom environment. Foreign language teachers have enormous power over the style of classroom activities; therefore, teachers who want to improve the classroom group dynamics and encourage a communicative atmosphere can use a

wide variety of pair and group work activities that allow classmate to interact with each other for an optimal amount of English speaking opportunities.

The opposite, of course, is also true; teachers that choose not to employ pair and group activities and instead rely on whole class activities or teacher-fronted activities can find themselves facing classrooms in which there is little cooperation or communication among classmates and thus no peer support for individual students who have a tendency toward anxious behavior. Such a classroom environment can create an oppressive or competitive atmosphere that raises anxiety, lowers self-competence and the desire to speak English, and damages any sense of confidence in speaking English, even among students who might be otherwise emotionally stable.

Based on the study results, it seems that despite the Japanese Ministry of Education's best intentions, a majority of university students in the study experienced little to no communication in English during their high school English classes. Because the structural regression model shows a clear connection between experiences speaking English in high school English classes and perceptions of present speaking competence, students with a lack of speaking experience in high school English classes are at a distinct disadvantage compared to their university peers who do have such experience. More importantly from a speaking confidence perspective, considering the desire shown by the study participants to speak in

English, to visit foreign countries, and to interact with English-speaking foreigners at their university and abroad, the apparent fact that secondary school students in are not provided with the opportunity within the foreign language classroom to practice speaking in English, creates a classroom scenario in which a positive classroom atmosphere cannot arise and students feel less confident and more anxious about speaking in English.

The apparent lack of opportunities to practice speaking English with classmates in a supportive high school English classroom, as reported by the study participants, decreases the chances of Japanese university students to improve their English-speaking confidence before they have a chance to interact with members of an English-speaking community outside the language classroom. The continued focus in Japanese high school English classes on non-communicative English pedagogy encourages a sense of incompetence and lack of confidence in speaking English.

In the next chapter, I provide a brief overall summary of the study, followed by a short discussion of the limitations of the study, suggestions for future research, and final conclusions.

CHAPTER 7

CONCLUSION

In this chapter, I summarize the findings of the study. Following the summary, I briefly describe the limitations of the study, before proceeding to offer suggestions for possible future research. Finally, I conclude the study with some thoughts on the issue of foreign language speaking anxiety and confidence.

Summary of the Findings

In this study, I posed four research questions concerning the relationship of personality traits to foreign language classroom speaking confidence for Japanese university students: (a) the validity of the so-called Big Five personality traits for a Japanese university sample, (b) the relationships among foreign language classroom speaking anxiety, perceived foreign language speaking self-competence and the desire to speak English, (c) the influence of the Big Five personality traits on foreign language classroom speaking confidence variables, and (d) the influence of social situation variables on foreign language classroom speaking confidence.

Prior to the main study, I conducted three pilot studies to create the two measurement instruments used in the main study. The first pilot study examined the Factor Markers questionnaire (Greenberg, 1993); based on the analysis results,

several items for each of the five hypothesized constructs were revised, the number of items were reduced from 50 to 37, and the questionnaire was renamed the Revised Big Five Factor Markers questionnaire. The second and third pilot studies were conducted in order to create and test the Foreign Language Classroom Speaking Confidence questionnaire used in the main study. The second pilot study tested two constructs, and the third study tested three additional constructs. In total, ten constructs of the eleven constructs in the main study were created and tested in the three pilot studies.

First, a Rasch analysis was used to examine the construct dimensionality and item fit for the hypothesized Big Five personality traits. The results demonstrated that four of the five personality trait constructs were valid for the sample in the study. Confirmatory factor analysis of the Big Five personality traits further verified the validity of the Extraversion, Emotional Stability, Conscientiousness, and Imagination constructs.

Second, the relationship among the three hypothesized variables composing foreign language classroom speaking confidence were examined using confirmatory factor analysis. The resulting model fit indices supported the hypothesis that three related constructs, Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to

Speak English, comprised speaking confidence in the foreign language classroom for the sample in the study.

Third, after the measurement model portion for all ten constructs was validated, the structural regression model results indicated that two of the four Big Five personality traits directly influenced Foreign Language. Emotional Stability moderately influenced Foreign Language Classroom Speaking Anxiety ($\beta = .43$), and Imagination weakly influenced Perceived Foreign Language Speaking Self-Competence ($\beta = .16$). Although previous studies demonstrated the influence of Imagination on speaking competence, the direct influence of Emotional Instability on speaking anxiety was not attested prior to the study. The direct role that personality plays in determining levels of speaking anxiety among EFL students can thus be regarded as a major finding of this study.

Fourth, the structural regression model results indicated that social situation variables directly influence foreign language classroom speaking confidence for the sample. The strongest influences were that of Past English Classroom Experiences ($\beta = .39$) and Current English Classroom Perception ($\beta = .46$) on Perceived Foreign Language Speaking Self-Competence. The results thus demonstrated the importance of a supportive classroom atmosphere to encourage students' sense of speaking competence, thereby increasing their desire to communicate and reducing anxiety in the foreign language classroom.

Overall results of the study pointed not only to the links between personality traits and speaking confidence, but also the degree to which classmates can positively affect speaking confidence in the foreign language classroom. Despite these findings, it was difficult for the study participants to endorse questionnaire items regarding the frequency of pair work and communication with classmates in their high school English classes. Although the study participants indicated that they got along well with classmates, had a basically supportive classroom atmosphere, and showed an intense desire to speak in English, they also indicated an overall lack of communicative activities and the absence of encouragement to speak in their high school English classes. For students of a foreign language who have virtually no contact with native speakers of the foreign language outside the classroom, it is extremely difficult, if not impossible, to gain a sense of speaking competence without the opportunity to practice English communication skills inside the classroom by speaking to classmates.

Based on the study findings, I can therefore suggest that the dearth of speaking opportunities in high school classes has prevented many of the study participants from perceiving themselves as competent speakers of English. This sense of not being capable at speaking due to a lack of practice inside the classroom has led many of the emotionally unstable study participants, who may have already experienced anxiety in non-foreign language speaking situations, to experience

anxiety speaking English. Thus, although the study results indicate that personality is an important factor, the lack of speaking practice in high school English classes is an equally important factor in determining the levels of English speaking confidence among Japanese EFL students.

If the Ministry of Education, Culture, Sports, Science, and Technology is sincere in its goal of fostering Japanese students who can actively use English for communication, the following two changes must take place in high school English instruction in Japan:

1. The frequency of pair work and group activities must be increased in order to give Japanese learners of English more opportunities to practice speaking English.
2. Teachers must encourage their students to speak English, without fear of mistakes, in order to decrease the chances for speaking anxiety and to increase a sense of speaking competence, and thereby to increase the desire and the confidence to speak English.

Limitations

There are seven limitations to this study. First, all the participants in the study were first-year students at four-year Japanese private universities. The results might not be generalizable to non-Japanese samples in other language learning contexts.

The questionnaires in the study should be implemented in English as a Second Language contexts as well as other English as a Foreign Language contexts for cross-sample verification of the constructs and structural regression model.

Second, even with the large sample size ($N = 1,081$), the case to parameter ratio barely met the recommended ratio of 10 to 1. To meet the ideal ratio of 20 to 1, and to use the two-step rule of dividing the sample into calibration and validation sample, a sample size of approximately 3,500 would be necessary.

Third, study participants were not examined for English proficiency levels prior to implementation of the study. Although perceptions of proficiency are more important than actual English proficiency for determining levels of speaking confidence, the lack of English proficiency measures made it impossible to examine any possible connections between English proficiency and speaking anxiety.

Fourth, many of the scales were considerably reduced in terms of number of items measuring each construct. For example, Conscientiousness, Imagination, and Perceived Social Value of Speaking English ultimately consisted of three items each, less than half the original number of items. Although the items that eventually measured the construct during the CFA and SEM stages of analysis may have proved the strongest and most salient items for measuring their respective

constructs, fewer items per construct may have resulted in decreased measurement precision due to increased standard errors for person ability measures.

Fifth, the dependent variables and independent variables in the CFA and SEM are based on data from a single method of measurement. Thus, there is a possibility that some of the intercorrelation among the variables results from a method effect.

Sixth, the data in the study come from self-reported questionnaire items. As such, there is a possibility that the data were affected by social desirability (Paulhus, 1991); in other words, the participants may have responded as they thought they should respond or were expected to respond.

Finally, the study participants were not asked whether they had study abroad experience or had lived abroad prior to entering university. On the other hand, there has been a steady decrease in the past ten years in the numbers of Japanese students studying in English-speaking countries (Ford, May 19, 2010; “Fewer Japanese students studying abroad,” 2010). In fact, according to some sources the number of Japanese high school students studying in the United States was even less than the number of students from Costa Rica (Vistawide, 2009). Thus, despite the limitation and given the sample size ($N = 1081$), it is highly unlikely that the number of students with study abroad experiences in junior or high school would have been large enough to lead to statistically significant differences in the study results.

Suggestions for Future Research

Although the study answered four research questions regarding the connections among personality traits social situation, and foreign language classroom speaking confidence, there are still many possible avenues for future exploration concerning related issues.

First, as noted previously (Chapter 6), because the study participants were all Japanese university students, there is no way of knowing whether the results concerning personality traits will hold true for other Asian populations or for “Western” populations. Therefore, a logical future research topic is to implement the Revised Big Five Factor Marker questionnaire in a different “Confucian-influenced” culture (such as China or Korea) and compare the model results to those in this study.

A second possible future research topic is whether the latent traits in this study remain the same over time. The study participants, all first year university students, were entering a defining moment of their social lives. To test whether personality traits and speaking confidence remain stable, a possible future study would examine the same sample longitudinally across two or three years of university life.

Third, although the study revealed the relationship among favorable social conditions variables, the role that the classroom plays is not well understood. Because individual classrooms tend toward the mean of student attitudes and

behaviors, future research could examine differences in foreign language speaking confidence among several classrooms through the use of hierarchical linear modeling. Similarly, examining whether classrooms differ according to institutional preference could also be examined. For example, do students at Japanese “top-tier” universities experience more or less foreign language speaking anxiety than students at “lower-tier” universities? Follow-up interviews or in-depth ethnographic studies could also shed light on the reasons for differences among foreign language classrooms and educational institutions.

Fourth, the data in this study came entirely from a single method, i.e., Likert-scale based, self-report questionnaires. Future research into this area should consider adding at least one additional type of data collection method to avoid method effect. For example, participants could be asked to assess their classmates, or teachers could be asked to assess their students. The addition of multiple methods of data collection could enable a multitrait-multimethod (MTMM) analysis in order to disentangle method from trait variance (Kenny & Kashy, 1992; Ozer, 1999). Also, an observed measure, such as how many times participants actually speak in the foreign language classroom, could be added for a data subset, i.e., one or two individual classes, in order to increase the validity of study findings.

Finally, further study into the variables comprising Foreign Language Classroom Speaking Confidence is warranted. Future studies should add to the

model by investigating the interaction of motivation and learner beliefs with Foreign Language Classroom Speaking Anxiety, Perceived Foreign Language Speaking Self-Competence, and Desire to Speak English.

Final Conclusions

When I was in Vietnam for a teaching seminar, a Vietnamese teacher asked me why Japanese visitors to her country didn't speak English to her. An assistant language teacher at a high school in southern Nara Prefecture in Japan, I was often told by my colleagues that Japanese students were too shy to speak English in the classroom. Japanese friends from all walks of life have apologized to me for their lack of English speaking ability, claiming that they are not good precisely because they are Japanese.

Based on personal experiences from living in largely in rural or semi-rural areas of Japan for over a decade, I know these over-generalizations about the shyness and poor English speaking ability to be true to a certain extent, but also false and not true to a large degree, depending on the person and the situation. As someone who has experienced a great deal of speaking anxiety in both his native language and in his second and third languages, I can say with certainty that many Japanese students of English are not shy, are good at speaking English, and hope to become even better speakers in the future.

I strongly feel that EFL teachers in Japan are in a very good position to increase the English speaking confidence of their students by setting the rhythm and pace of classroom activities, through the use of pair work and group discussion in which the importance of communicating one's ideas is given clear priority over correctness of speech, and through the creation of supportive, encouraging classroom atmospheres. Speaking confidence is partly due to (theoretically) unchangeable personality, but it is also partly due to (very malleable) perceptions of the people around you. As a teacher of future speakers of English as a foreign language, I find it very encouraging that speaking confidence can be increased through cooperation and support from fellow classmates, and thus ultimately from language teachers who can provide speaking opportunities and a supportive classroom atmosphere for language learners.

REFERENCES

- Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal*, 78(2), 155-168.
- Allaman, J. D., Joyce, C. S., & Crandall, V. C. (1972). The antecedents of social desirability response tendencies of children and young adults. *Child Development*, 43, 1135-1160.
- Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38, 665-680.
- Angleitner, A., & Ostendorf, F. (1994). Temperament and the Big Five factors of personality. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 69-90). Hillsdale, NJ: Erlbaum.
- Apple, M. (2008, May). Developing an anxiety Rasch. Paper presented at the PanSIG 2008 Conference, Doshisha University, Kyoto, Japan.
- Apple, M., & Neff, P. (In review). Using the Rasch Rating Scale model to examine the Big Five Factor Marker questionnaire for a Japanese university population. Unpublished manuscript.
- Apple, M., & Neff, P. (2010, June). Validating a Big Five questionnaire for a Japanese university population. Paper presented at the Rasch 2010 Conference. Copenhagen Business School, Copenhagen, Denmark.

- Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality, 15*(5), 327-353.
- Ashton, M. C., Lee, K., de Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality, 40*(6), 851-875.
- Ashton, M. C., Perugini, M., de Vries, R. E., Boies, K., Lee, K., Szarota, P., Blas, L. D., & Raad, B. D. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality & Social Psychology, 86*(2), 353-366.
- Atkins, P. W. B., & Baddeley, A. D. (1998). Working memory and distributed vocabulary learning. *Applied Psycholinguistics, 19*, 537-552.
- Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning, 50*(2), 311-341.
- Baumeister, R. F., & Tice, D. M. (1990). Anxiety and social exclusion. *Journal of Social and Clinical Psychology, 9*, 165-195.
- Bell, I. R., Jasnoski, M. L., Kagan, J., & King, D. S. (1990). Is allergic rhinitis more frequent in young adults with extreme shyness? A preliminary survey. *Psychosomatic Medicine, 52*, 517-525.
- Bentler, P. M. (1995). EQS structural equations program [Computer software]. Encino, CA: Multivariate Software, Inc.

- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*(5), 825-829.
- Berberian, K. E., & Snyder, S. S. (1982). The relationship of temperament and stranger reaction for younger and older infants. *Merrill-Palmer Quarterly, 28*, 79-94.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin, 117*, 185-215.
- Bollen, K. A. (1989). *Structural equation with latent variables*. New York: Wiley.
- Bond, M. H., Nakazato, H., & Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese person perception. *Journal of Cross-Cultural Psychology, 6*(3), 346-357.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.
- Boomsa, A. (2000). Reporting analyses of covariance structures. *Structural Equation Modeling, 7*(3), 461-483.
- Brown, J. D., Robson, G., & Rosenkjar, P. R. (2001). Personality, motivation, anxiety, strategies, and language proficiency of Japanese students. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition* (pp. 361-391). Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i.
- Busch, D. (1982). Introversiion-extraversiion and the EFL proficiency of Japanese students. *Language Learning, 32*, 109-132.

- Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York: Wiley.
- Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Carrell, P. L., Prince, M. S., & Astika, G. G. (1996). Personality types and language learning in an EFL context. *Language Learning*, 46(1), 75-99.
- Cattell, R. B. (1943). The description of personality: Basic traits resolved into clusters. *Journal of Abnormal and Social Psychology*, 38, 476-506.
- Cattell, R. B. (1973). *Personality and mood by questioning*. San Francisco: Jossey-Bass.
- Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning*, 25, 153-161.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods and Research*, 29, 468-508.
- Cheung, F. M., Cheung, S. F., Wada, S., & Zhang, J.-X. (2003). Indigenous measures of personality assessment in Asian countries: A review. *Psychological Assessment*, 15(3), 280-289.
- Cheung, F. M., & Leung, K. (1998). Indigenous personality measures—Chinese examples. *Journal of Cross-Cultural Psychology*, 29(1), 233-248.
- Cheung, F. M., Leung, K., Zhang, J.-X., Sun, H.-F., Gan, Y.-Q., Song, W.-Z., & Xie, D. (2001). Indigenous Chinese personality constructs: Is the

Five-Factor Model complete? *Journal of Cross-Cultural Psychology*, 32(4), 407-433.

Church, S. E., & Lonner, W. J. (1998). The cross-cultural perspective in the study of personality: Rationale and current research. *Journal of Cross-Cultural Psychology*, 29(1), 32-62.

Clark, A., & Trafford, J. (1995). Boys into modern languages: An investigation of the discrepancy in attitudes and performance between boys and girls in modern languages. *Gender and Education*, 7(3), 315-325.

Clark, J. L. D. (1981). Language. In T. S. Barrows (Ed.), *College students' knowledge and beliefs: A survey of global understanding* (pp. 87-100). New Rochelle, NY: Change Magazine Press.

Clauser, B., & Linacre, J. M. (1999). Relating Cronbach and Rasch reliabilities. *Rasch Measurement Transactions*, 13(2), 696. Retrieved July 8, 2010, from <http://www.rasch.org/rmt/rmt132i.htm>

Clément, R. (1980). Ethnicity, contact, and communicative competence in a second language. In H. Giles, W. P. Robinson, & P. M. Smith (Eds.), *Language: Social psychological perspectives* (pp. 147-154). Oxford: Pergamon.

Clément, R., Baker, S. C., & MacIntyre, P. D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Psychology*, 22(2), 190-209.

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, 44, 417-448.

- Clément, R., Gardner, R. C., & Smythe, P. C. (1977). Motivational variables in second language acquisition: A study of Francophones learning English. *Canadian Journal of Behavioral Science, 9*, 123-133.
- Clément, R., & Kruidenier, B. (1983). Orientations in second language acquisition: I. The effects of ethnicity, milieu, and target language on their emergence. *Language Learning, 33*, 273-291.
- Clément, R., & Kruidenier, B. (1985). Aptitude, attitude, and motivation in second language proficiency: A test of Clément's model. *Journal of Language and Social Psychology, 4*, 21-37.
- Cohen, Y., & Norst, M. J. (1989). Fear, dependence, and loss of self-esteem: Affective barriers in second language learning among adults. *RELC Journal, 20*, 61-77.
- Concern over use of Japanese language in English conversation classes at high schools. (2010, December 15). *Zenkoku Ippan Tokyo General Union*. Retrieved February 1, 2011, from <http://tokyogeneralunion.org/2010/12/15/concern-over-use-of-japanese-language-in-english-conversation-classes-at-high-schools/>
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*, 187-276.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences, 13*(6), 653-665.
- Council of Europe. (2004). *Common European Framework of Reference for languages: Learning, teaching, and assessing*. Cambridge: Cambridge University Press.
- Cross, S. E., & Markus, H. R. (1999). The cultural constitution of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.) (pp. 378-396). New York: Guilford.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147-167.
- Davies, R. J., & Ikeno, O. (2002). *The Japanese mind: Understanding contemporary Japanese culture*. Boston: Tuttle.
- De Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. New York: Guilford.
- De Fruyt, F., McCrae, R. R., Szirmak, Z., & Nagy, J. (2004). The Five-Factor Personality Inventory as a measure of the Five-Factor Model: Belgian, American, and Hungarian comparisons with the NEO-PI-R. *Assessment, 11*(3), 207-215.
- De Raad, B., & Perugini, M. (2002). Big Five factor assessment: An introduction. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 1-26). Göttingen, Germany: Hogrefe & Huber.

- Di Blas, L., & Forzi, M. (1999). Refining a descriptive structure of personality attributes in the Italian language: The abridged Big Three circumplex structure. *Journal of Personality and Social Psychology*, 76(3), 451-481.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417.
- Digman, J. M. (1996). The curious history of the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five-Factor Model of personality: Theoretical perspectives* (pp. 1-20). New York: The Guilford Press.
- Ehrman, M. E. (1990). Owls and doves: Cognition, personality, and learning success. In J. E. Alatis (Ed.), *Linguistics, language teaching, and acquisition: The interdependence of theory, practice, and research* (pp. 413-437). Washington, DC: Georgetown University Press.
- Ehrman, M. E. (2008). Personality and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 61-72). Cambridge: Cambridge University Press.
- Ehrman, M. E., & Oxford, R. L. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal*, 74, 311-327.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- English classes mostly in Japanese. (2011, January 22). *The Japan Times*. Retrieved February 1, 2011, from <http://search.japantimes.co.jp/cgi-bin/nn20110122b2.html>

- Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). London: Methuen.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences*, *12*(8), 773-790.
- Eysenck, H. J. (1992). Four ways five factors are *not* basic. *Personality and Individual Differences*, *13*(6), 667-673.
- Eysenck, H. J. (1994). The Big Five or Giant Three: Criteria for a paradigm. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 37-52). Hillsdale, NJ: Erlbaum.
- Eysenck, H. J. (2000). The new look in personality and intelligence research: The advent of the paradigm. In J. Mohan (Ed.), *Personality across cultures: Recent developments and debates* (pp. 1-14). New Delhi: Oxford University Press.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. New York: Plenum Press.
- Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder and Stoughton.
- Fewer Japanese students studying abroad a cause for concern. (2010, July 7). *The Mainichi Daily News*. Retrieved October 22, 2010, from <http://mdn.mainichi.jp/features/archive/news/2010/06/20100607p2a00m0na018000c.html>

Ford, P. (2010, May 19). For study abroad, more Japanese prefer Chinese university over US one. *The Christian Science Monitor*. Retrieved October 22, 2010, from <http://www.csmonitor.com/World/Asia-Pacific/2010/0519/For-study-abroad-more-Japanese-prefer-Chinese-university-over-US-one>

Galton, F. (1884). Measurement of character. *Fortnightly Review*, 36, 179-185.

Gardner, R. C. (1988). The socio-educational model of second language learning: Assumptions, findings, and issues. *Language Learning*, 38, 101-126.

Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43, 157-194.

Gardner, R. C., Smythe, P. C., Clément, R., & Glikman, L. (1976). Second language acquisition: A social psychological perspective. *Canadian Modern Language Review*, 32, 198-213.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141-165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, 59(6), 1216-1229.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26-42.

- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26-34.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84-96.
- Gorsuch, G. J. (1998). *Yakudoku* EFL instruction in two Japanese high school classroom: An exploratory study. *JALT Journal*, 20(1), 6-32.
- Gorsuch, G. J. (1999). Exploring the relationship between educational policy and instruction in Japanese high school EFL classrooms. (Unpublished doctoral dissertation.) Temple University, Tokyo, Japan.
- Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34(4), 675-710.
- Gow, A. J., Whiteman, M. C., Pattie, A., & Deary, I. J. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39, 317-329.
- Graziano, W. G. (1994). The development of Agreeableness as a dimension of personality. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 339-354). Hillsdale, NJ: Erlbaum.
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827-838.

- Griffiths, R. (1991). Personality and second-language learning: Theory, research and practice. In E. Sadtano (Ed.), *Language acquisition and the second/foreign language classroom* (pp. 103-135). Singapore: SEAMEO Regional Language Centre.
- Guthrie, G. M., & Bennett, A. B. (1971). Cultural differences in implicit personality theory. *International Journal of Psychology*, 6, 305-312.
- Hahn, D.-W., Lee, K., & Ashton, M. C. (1999). A factor analysis of the most frequently used Korean personality trait adjectives. *European Journal of Personality*, 13(4), 261-282.
- Heine, S. J., Kitayama, S., & Lehman, D. R. (2001). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. *Journal of Cross-Cultural Psychology*, 32(4), 434-443.
- Heine, S. J., Kitayama, S., Lehman, D. R., Takata, T., Ide, E., Leung, C., & Matsumoto, H. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality & Social Psychology*, 81(4), 599-615.
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert-scale scales?: The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903-918.
- Heine, S. J., Takemoto, T., Moskaleiko, S., Lasaleta, J., & Heinrich, J. (2008). Mirrors in the head: Cultural variation in objective self-awareness. *Personality and Social Psychology Bulletin*, 34(7), 879-887.

- Hendry, J. (2003). *Understanding Japanese society* (3rd ed.). New York: Routledge.
- Hino, N. (1988). Yakudoku: Japan's dominant tradition in foreign language learning. *JALT Journal*, 10, 45-55.
- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Hofstede, G. (2007). A European in Asia. *Asian Journal of Social Psychology*, 10(1), 16-21.
- Hogan, R. T. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *Nebraska symposium of motivation: Personality-current theory and research* (pp. 58-89). Lincoln, NE: University of Nebraska Press.
- Hogan, R. T. (1986). *Hogan Personality Inventory*. Minneapolis, MN: National Computer Systems.
- Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly*, 20(3), 559-562.
- Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage.
- Hsu, F. L. K. (1971). Psychosocial homeostasis and jen: Conceptual tools for advancing psychological anthropology. *American Anthropologist*, 73, 23-44.

- Huang, D. C., Church, T. A., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross Cultural Psychology, 28*(2), 197-218.
- Isaka, H. (1990). Factor analysis of trait terms in everyday Japanese language. *Personality and Individual Differences, 11*(2), 115-124.
- Iwawaki, S., Eysenck, S. B. G., & Eysenck, H. J. (1980). Japanese and English personality structure. *Psychologia, 23*, 195-205.
- Jessor, R., & Hammond, K. R. (1957). Construct validity and the Taylor Anxiety Scale. *Psychological Bulletin, 54*(3), 161-170.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.) (pp. 102-138). New York: Guilford.
- Kagan, J., Reznick, J. S., & Snidman, N. (1988). Biological bases of childhood shyness. *Science, 240*, 167-171.
- Kagan, J., Snidman, N., Julia-Sellers, M., & Johnson, M. O. (1991). Temperament and allergic symptoms. *Psychosomatic Medicine, 53*, 332-340.
- Kano, Y. (2001). Structural equation modeling with experimental data. In R. Cudek, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 381-402). Lincolnwood, IL: Scientific Software International.

- Kaplan, D., & Wenger, R. N. (1993). Asymptomatic independence and separability in covariance structure models: Implications for specification error, power, and model modification. *Multivariate Behavioral Research*, 28(4), 467-482.
- Kashiwagi, S. (1999). The trait theoretic evaluation of the TEG from the view of the Five-Factor Model. *The Japanese Journal of Psychology*, 69, 468-477.
- Kashiwagi, S. (2002). Japanese adjective list for the Big Five. In B. De Raad & M. Perugini (Eds.), *Big Five Assessment* (pp. 305-326). Seattle: Hogrefe & Huber.
- Kashiwagi, S., & Wada, S. (1996). A study on the concurrent validity of personality inventory from the view of the Five-Factor Model concerning personality traits. *The Japanese Journal of Personality*, 67, 300-307.
- Kashiwagi, S., Wada, S., & Aoki, T. (1993). The Big Five and the oblique primary pattern for the items of the ACL Japan version. *The Japanese Journal of Personality*, 64, 153-159.
- Kashiwagi, S., & Yamada, K. (1995). Evaluation of the Uchida-Kraepelin test based on The Five Factor Model of personality traits. *The Japanese Journal of Personality*, 66, 24-32.
- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.
- Kenny, D. A. (2010). Measuring model fit. Retrieved July 26, 2010, from <http://davidakenny.net/cm/fit.html>
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.

- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed.) (pp. 233-265). Boston: McGraw-Hill.
- Kim, J.-H. (2000). Foreign language listening anxiety: A study of Korean students learning English. (Unpublished doctoral dissertation.) The University of Texas, Austin.
- Kim, M. R. (2005). The Relationship of social anxiety with identity, public self-consciousness, and interdependent self: A comparative study of Korean and Japanese university students. *The Japanese Journal of Personality*, 14(1), 42-53.
- Kimura, H. (2008). Foreign language listening anxiety: Its dimensionality and group differences. *JALT Journal*, 30(2), 173-196.
- Kitano, K. (2001). Anxiety in the college Japanese language classroom. *Modern Language Journal*, 85(4), 549-566.
- Kitayama, S. (2000). Collective construction of the self and social relationships: A rejoinder and some extensions. *Child Development*, 71(5), 1143-1146.
- Kitayama, S., & Markus, H. R. (1999). Yin and Yang of the Japanese self: The cultural psychology of personality coherence. In D. Cervone & Y. Shoda (Eds.), *The coherence of personality: Social-cognitive bases of consistency, variability, and organization* (pp. 242-302). New York: Guilford Press.

- Kitayama, S., Markus, H. R., & Lieberman, C. (1995). The collective construction of self esteem: Implications for culture, self, and emotion. In J. A. Russell & J.-M. Fernandez-Dols (Eds.), *Everyday conceptions of emotion: An introduction to the psychology, anthropology, and linguistics of emotion* (pp. 523-550). New York: Kluwer Academic/Plenum.
- Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any “free” choice? Self and dissonance in two cultures. *Psychological Science*, *15*(8), 527-533.
- Kitayama, S., & Uchida, Y. (2003). Explicit self-criticism and implicit self-regard: Evaluating self and friend in two cultures. *Journal of Experimental Social Psychology*, *39*, 476-482.
- Klonsky, B. G., Dutton, D. L., & Libel, C. N. (1990). Developmental antecedents of private self-consciousness, public self-consciousness, and social anxiety. *Genetic, Social, and General Psychology Monographs*, *116*, 273-297.
- Kubota, R. (1999). Japanese culture constructed by discourses: Implications for applied linguistics research and ELT. *TESOL Quarterly*, *33*(1), 9-35.
- Lalonde, R. N., & Gardner, R. C. (1984). Investigating a causal model of second language acquisition: Where does personality fit in? *Canadian Journal of Behavioral Science*, *16*(4), 224-237.
- Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin*, *9*, 371-375.
- Leary, M. R. (1990). Responses to social exclusion: Social anxiety, jealousy, loneliness, depression, and low self-esteem. *Journal of Social and Clinical Psychology*, *9*, 221-229.

- Leary, M. R., & Kowalski, R. M. (1995). *Social anxiety*. New York: Guilford.
- Leary, M. R., Tambor, E. S., Terdal, S. K., & Downs, D. L. (1995). Self-esteem as an interpersonal monitor: The sociometer hypothesis. *Journal of Personality and Social Psychology*, *68*, 518-530.
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-types work best? *Journal of Outcome Measurement*, *2*, 266-283.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3*(2), 193-122.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85-106.
- Linacre, J. M. (2006a). Data variance explained by Rasch measures. *Rasch Measurement Transactions*, *20*(1), 1045.
- Linacre, J. M. (2006b). WINSTEPS Rasch measurement computer program. [Computer software]. Chicago.
- Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago: MESA.
- Lun, V. M.-C., & Bond, M. H. (2006). Achieving relationship harmony in groups and its consequence for group performance. *Asian Journal of Social Psychology*, *9*(3), 195-202.
- MacCallum, R. C. (1995). Model specification: Procedures, strategies, and related issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 16-36). Thousand Oaks, CA: Sage.

- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*, 490-504.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*(1), 185-199.
- MacIntyre, P. D. (1994). Variables underlying willingness to communicate: A causal analysis. *Communication Research Reports, 11*, 135-142.
- MacIntyre, P. D. (1995). On seeing the forest for the trees: A rejoinder to Sparks and Ganschow. *Modern Language Journal, 79*(2), 245-248.
- MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). Boston: McGraw-Hill.
- MacIntyre, P. D., Babin, P. A., & Clément, R. (1999). Willingness to communicate: Antecedents & consequences. *Communication Quarterly, 47*(2), 215-229.
- MacIntyre, P. D., Baker, S. C., Clement, R., & Donovan, L. A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high French immersion students. *Language Learning, 52*(3), 537-564.

- MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language & Social Psychology, 15*(1), 3-26.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal, 82*, 545-562.
- MacIntyre, P. D., & Gardner, R. C. (1994). The effects of induced anxiety on cognitive processing in second language learning. *Studies in Second Language Acquisition, 16*(1), 1-17.
- MacIntyre, P. D., MacMaster, K., & Baker, S. C. (2001). The convergence of multiple models of motivation for second language learning: Gardner, Pintrich, Kuhl, and McCroskey. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language acquisition (Technical Report #23)* (pp. 461-492). Honolulu: University of Hawai'i.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.
- Markus, H. R., & Kitayama, S. (1998). The cultural psychology of personality. *Journal of Cross-Cultural Psychology, 29*(1), 63-87.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System, 32*(1), 21-36.

- Matsuoka, R. (2005). Japanese college students' willingness to communicate in English. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses. (UMI No. 3233454)
- McCrae, R. R. (2000). Trait psychology and the revival of personality and culture Studies. *American Behavioral Scientist, 44*(1), 10-31.
- McCrae, R. R., & Costa, P. T., Jr. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49*, 710-721.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the Five-Factor Model or personality across instruments and observers. *Journal of Personality and Social Psychology, 52*(1), 81-90.
- McCroskey, J. C. (1970). Measures of communication-bound anxiety. *Speech Monographs, 37*(4), 269-277.
- McCroskey, J. C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication Research, 4*, 78-96.
- McCroskey, J. C. (1978). Validity of the PRCA as an index of oral communication apprehension. *Communication Monographs, 45*, 192-203.
- McCroskey, J. C., & Beatty, M. J. (1998). Communication apprehension. In J. C. McCroskey, J. A. Daly, M. M. Martin, & M. J. Beatty (Eds.), *Communication and personality trait perspectives* (pp. 215-232). Cresskill, NJ: Hampton.

- McCroskey, J. C., & Richmond, V. P. (1987). Willingness to communicate. In J. McCroskey & J. A. Daly (Eds.), *Personality and interpersonal communication* (pp. 129-156). Newbury Park: Sage.
- McCroskey, J. C., & Richmond, V. P. (1991). Willingness to communicate: A cognitive view. In M. Booth-Butterfield (Ed.), *Communication, cognition, and anxiety* (pp. 19-37). Newbury Park, CA: Sage.
- McDonald, R. P., & Ho, M-H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64-82.
- MEXT. (2003). Regarding the establishment of an action plan to cultivate "Japanese with English abilities." Retrieved May 5, 2008, from <http://www.mext.go.jp/english/topics/03072801.htm>
- MEXT. (2007). *Eigo kyouiku kaizen jisshi jyoukyou chousa kekka gaiyo (koutou gakkou)* [Summary results of the survey regarding the current status of the implementation of English education reforms (high school)]. Retrieved February 1, 2011, from http://www.mext.go.jp/a_menu/kokusai/gaikokugo/1261353.htm
- Mlačić, B., & Goldberg, L. R. (2007). An analysis of a cross-cultural personality inventory: The IPIP Big-Five Factor Markers in Croatia. *Journal of Personality Assessment*, 88(2), 168-177.
- Mori, R. (2002). Entrance examinations and remedial education in Japanese higher education. *Higher Education*, 43(1), 27-42.
- Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five factor model. *Journal of Research in Personality*, 41, 1213-1233.

- Naiman, N., Frolich, M., & Stern, H. H. (1975). *The good language learner*. Toronto: The Ontario Institute for Studies in Education.
- Nakano, C. (1970). *Japanese society*. Tokyo: Tuttle.
- Nakayama, M., Yamamoto, H., & Santiago, R. (2006). Investigating the impact of learner characteristics on blended learning among Japanese students. In D. Remenyi (Ed.), *International conference on e-learning* (pp. 361-370). London: Academic Conferences.
- Neff, P., & Apple, M. (2008, August). Validating a Five-factor Model personality instrument for a Japanese population. Paper presented at the Pacific Rim Objective Measurement Symposium. Ochanomizu Women's University, Tokyo, Japan.
- Nishino, T. (2009). Communicative language teaching in Japanese high schools: Teachers' beliefs and classroom practices. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3371994)
- Noels, K. A., Pon, G., & Clement, R. (1996). Language, identity, and adjustment: The role of linguistic self-confidence in the acculturation process. *Journal of Language and Social Psychology, 15*(3), 246-264.
- Noguchi, K. (2007). Examination of the content of individualism/collectivism scales in cultural comparisons of the USA and Japan. *Asian Journal of Social Psychology, 10*(3), 131-144.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology, 66*, 574-583.

Only 20% of English conversation teachers give classes in English. (2011, January 22). *The Mainichi Daily News*. Retrieved February 1, 2011, from <http://mdn.mainichi.jp/mdnnews/news/20110122p2g00m0dm015000c.html>

Ortiz, F. A., Church, T. A., de Jesus Vargas-Flores, J., Ibanez-Reyes, J., Flores-Galaz, M., Iuit-Briceno, J. I., & Escamilla, J. M. (2007). Are indigenous personality dimensions culture-specific? Mexican inventories and the Five-Factor Model. *Journal of Research in Personality, 41*(3), 618-649.

Oxford, R. L., & Ehrman, M. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13*, 188-205.

Ozer, D. J. (1999). Four principles for personality assessment. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed.) (pp. 671-686). New York: Guilford.

Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology, 83*(4), 983-997.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57*, 552-567.

Pedhazur, E.J., & Pedhazur-Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Piedmont, R. L., & Chae, J-H. (1997). Cross-cultural generalizability of the Five-Factor Model of personality: Development and validation of the NEO PI-R for Koreans. *Journal of Cross-cultural Psychology, 28*(2), 131-155.

Population of Japan 2005. (n.d). Ministry of Internal Affairs and Communications: Statistics Bureau, Director-General for Policy Planning & Statistical Research and Training Institute. Retrieved September 15, 2008, from <http://www.stat.go.jp/english/data/kokusei/2005/poj/pdf/2005pofj.pdf>

Price, M. L. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 101-108). Englewood Cliffs, NJ: Prentice-Hall.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago.

Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in *Psychology and Aging*: Some proposed guidelines. *Psychology and Aging*, 6(4), 499-503.

Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, 83(2), 202-218.

Saucier, G., Hampton, S. E., & Goldberg, L. R. (2000). The language of personality: Lexical perspectives on the Five-Factor Model. In J. S. Wiggins (Ed.), *The Five-Factor Model of personality: Theoretical perspectives* (pp. 21-50). New York: Guilford.

Scarr, S., & Salapatek, P. (1970). Patterns of fear development during infancy. *Merrill-Palmer Quarterly*, 16, 53-90.

Schrieber, J. B., Stage, F. K., King, J., Nora, A., Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, 99(6), 323-337.

- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.
- Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning, 28*, 129-142.
- Shimizu, K. (1999). A survey of expectations of and characteristics attributed to native-speaker English teachers and Japanese English teachers by Japanese university students. *Journal of Intercultural Communication, 3*, 53-74.
- Sick, J. (2007). *The learner's contribution: Individual differences in language learning in a Japanese high school*. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3255147)
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120.
- Smith, E. V. (2000). Metric development and score reporting in Rasch measurement. *Journal of Applied Measurement, 1*(3), 303-326.
- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205-231.
- Smith, E. V., & Smith, R. M. (Eds.). (2004). *Introduction to Rasch measurement*. Maple Grove, MN: JAM.
- Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling, 3*(1), 25-40.

- Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement, 1*(2), 199-218.
- Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Hillsdale, NJ: Erlbaum.
- Sugawara, M. (1989). *Nihongo: A Japanese approach to Japanese*. Tokyo: The East.
- Sugimoto, Y. (2010). *An introduction to Japanese society* (3rd ed.). Cambridge: Cambridge University.
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.
- Takano, Y., & Osaka, E. (1999). An unsupported common view: Comparing Japan and the U.S. on individualism/collectivism. *Asian Journal of Social Psychology, 2*(3), 311-341.
- Taylor, J. A. (1953). A personality scale of manifest anxiety. *The Journal of Abnormal and Social Psychology, 48*(2), 285-290.
- Thurstone, L. L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology, 26*, 249-269.
- Thurstone, L. L. (1934). The vectors of the mind. *Psychological Review, 41*, 1-32.
- Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics, 23*(3), 361-374.

- Wada, S. (1996). Construction of the Big Five Scales of personality trait items and concurrent validity with NPI. *Japanese Journal of Personality*, 67, 61-67.
- Wang, D., Cui, H., & Zhou, F. (2005). Measuring the personality of Chinese: QZPS versus NEO PI-R. *Asian Journal of Social Psychology*, 8(1), 97-122.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13, 318-333.
- Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, 6(1), 80-99.
- Weaver, J. B. (1998). Personality and self-perceptions about communication. In J. C. McCroskey, J. A. Daly, M. M. Martin, & M. J. Beatty (Eds.), *Communication and personality: Trait perspectives* (pp. 95-118). Cresskill, NJ: Hampton.
- Wierzbicka, A. (1997). *Understanding cultures through their key words: English, Russian, Polish, German, and Japanese*. New York: Oxford University Press.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.
- Wolfe, E. W., & Smith, J., E. V. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I—instrument development tools. *Journal of Applied Measurement*, 8(1), 97-123.

- Wolfe, E. W., & Smith, J., E. V. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement, 8*(2), 204-234.
- Wright, B. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling, 3*(1), 3-24.
- Wright, B. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 65-104). Mahwah, NJ: Erlbaum.
- Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *Modern Language Journal, 86*(1), 54-66.
- Yashima, T., Zenuk-Nishide, L., & Kazuaki Shimizu, L. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning, 54*(1), 119-152.
- Yik, M. S. M., Russell, J. A., Ahn, C.-K., Dols, J. M. F., & Suzuki, N. (2002). Relating the five-factor model of personality to a circumplex model of affect. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 74-104). New York: Kluwer Academic.
- Yoon, K., Schmidt, F., & Ilies, R. (2002). Cross-cultural construct validity of the Five-Factor Model of personality among Korean employees. *Journal of Cross-Cultural Psychology, 33*(3), 217-235.
- Young, D. J. (1991). Creating a low anxiety classroom environment: What does language anxiety research suggest? *Modern Language Journal, 74*(4), 426-439.

- Zhang, J., & Bond, M. H. (1998). Personality and filial piety among college students in two Chinese societies. *Journal of Cross-Cultural Psychology*, 29(3), 402-417.
- Zuckerman, M. (1994). An alternative five-factor model for personality. In C. F. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality from infancy to adulthood* (pp. 53-68). Hillsdale, NJ: Erlbaum.
- Zuckerman, M., Juhlman, D. M., & Camac, C. (1988). What lies beyond E and N? Factor analyses of scales believed to measure basic dimensions of personality. *Journal of Personality and Social Psychology*, 54(1), 96-107.

APPENDIX A
ORAL SCRIPT OF INFORMED CONSENT AND RIGHT NOT TO
PARTICIPATE IN THE STUDY

“This is a research project being done by a colleague of mine. Participation is completely voluntary, and you don’t have to complete the questionnaire if you don’t want to.

“All of your answers will be anonymous, and will never be revealed to anyone. The questionnaire will not affect your grade for this course in any way.

“Please be sure to check whether you are a man or a woman, and please write your age also.”

APPENDIX B
INFORMED CONSENT AND RIGHT NOT TO PARTICIPATE IN THE
STUDY PROVIDED IN WRITING TO PARTICIPANTS

The Revised Big Five Factor Marker Questionnaire

Please use the below rating scale from 1 to 6 to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you see yourself, in comparison to other people of the same sex and age as you are. Your responses will remain anonymous and will have no effect on your course grades. Please answer as honestly as you can.

それぞれの記述をよく読んで、自分自身がどの程度当てはまるかを、1～6の尺度値を用いて回答してください。回答は現在の自分自身についてで、将来の希望についてではありません。同性で同年代の誰かと比べて、自分がどうであるかを、教えてください。回答は全くの匿名資料として使用され、個人の成績に影響を与えるものではありませんので、正直に教えてください。

APPENDIX B (continued)

The Foreign Language Classroom Speaking Confidence Questionnaire

This is a questionnaire about learning English. Read the statements and choose from the below scale of 1 to 6 to describe how accurately they describe yourself. Your responses will remain anonymous and will have no effect on your course grades. Individual student responses will not be made publicly available. Please answer as honestly as possible.

これは英語学習に関するアンケートです。それぞれの項目について、あなた自身がどの程度当てはまるかを、下記の 1 ～ 6 の尺度値を用いて回答してください。あなたの回答は全くの匿名資料として使用され、個人の成績に影響を与えるものではありません。個人が特定される形で公開することは決してありませんので、正直に答えてください。

APPENDIX C
REVISED BIG FIVE FACTOR MARKERS (ENGLISH VERSION)

Items marked with an asterisk (*) appear in the CFA and SEM.

Extraversion/Introversion

- EI1 Am the life of the party
- EI2 Like to talk to a lot of people
- EI3 Start conversations*
- EI4 Like to draw attention to myself
- EI5 Like to be the center of attention
- EI6 Talk a lot*
- EI7 Feel at ease in a large group of people
- EI8 Am skilled at dealing with people*

Agreeableness

- AGR1 Inquire about others' well-being
- AGR2 Feel concern for others
- AGR3 Am interested in other people's problems
- AGR4 Take time out for others
- AGR5 Am on good terms with nearly everyone*
- AGR6 Make people feel at ease
- AGR7 Have a good word for everyone
- AGR8 Know how to cheer people up

APPENDIX C (Continued)

Emotional Stability

- EMO1 Often feel blue*
- EMO2 Get stressed out easily*
- EMO3 Worry about things*
- EMO4 Have frequent mood swings
- EMO5 Panic easily
- EMO6 Feel threatened easily*
- EMO7 Get angry easily

Conscientiousness

- CON1 Am prepared*
- CON2 Put things back in their proper places
- CON3 Like rules
- CON4 Follow a daily schedule
- CON5 Do things according to a plan*
- CON6 Like to clean up things
- CON7 Continue my work until it's perfect*

Intellect/Imagination

- INT1 Have a vivid imagination*
- INT2 Am full of ideas*
- INT3 Have a rich vocabulary in my native language
- INT4 Am quick to understand things
- INT5 Love to read challenging materials
- INT6 Love to think of new ways to do things*
- INT7 Can handle a lot of information

APPENDIX D
REVISED BIG FIVE FACTOR MARKERS (JAPANESE VERSION)

Extraversion/Introversion

- EI1 盛り上げ役である。
- EI2 いろいろな人と話すのが好きだ。
- EI3 人に自分から話し掛ける。
- EI4 人の注意を引くのが好きだ。
- EI5 注目の的になるのが好きだ。
- EI6 よくしゃべる。
- EI7 たくさんの人の中になると落ち着く。
- EI8 人と関わるのが上手だ。

Agreeableness

- AGR1 周りの人の様子を尋ねる。
- AGR2 周りの人への関心がある。
- AGR3 周りの人の問題に興味がある。
- AGR4 周りの人のために時間を割く。
- AGR5 ほぼ誰とでも仲良くなれる。
- AGR6 周りの人を安心させる。
- AGR7 誰についても良いことを言える。
- AGR8 周りの人を元気づけることができる。

APPENDIX D (Continued)

Emotional Stability

- EM01 落ち込むことが多い。
- EM02 すぐにストレスがたまる。
- EM03 心配性である。
- EM04 気分が変わりやすい。
- EM05 慌てやすい。
- EM06 びくびくしやすい。
- EM07 怒りやすい。

Conscientiousness

- CON1 用意周到である。
- CON2 身の回りを整理整頓する。
- CON3 規則正しいのが好きだ。
- CON4 毎日の予定に従う。
- CON5 計画通りにする。
- CON6 片付けるのが好きだ。
- CON7 物事を完璧にする。

Intellect/Imagination

- INT1 想像力が豊かだ。
- INT2 発想が豊かだ。
- INT3 日本語の語彙力が豊富だ。
- INT4 ものわりが良い。
- INT5 難しい書物を読むのが好きだ。
- INT6 新しいやり方を試すのが好きだ。
- INT7 たくさんの情報に対応できる。

APPENDIX E
FOREIGN LANGUAGE CLASSROOM SPEAKING CONFIDENCE
(ENGLISH VERSION)

Items marked with an asterisk (*) appear in the CFA and SEM.

Foreign Language Classroom Speaking Anxiety

- FLCSA1 I'm worried that other students in class speak English better than I do.
- FLCSA2 I feel nervous speaking English in front of the entire class.
- FLCSA3 I'm worried that my partner speaks better English than I do.*
- FLCSA4 I start to panic when I speak in English with a classmate in a pair.
- FLCSA5 I'm worried about making mistakes while speaking English.*
- FLCSA6 I feel nervous when I can't express my opinion in English.*
- FLCSA7 I'm afraid my partner will laugh when I speak English with a classmate in a pair.
- FLCSA8 I'm worried about making mistakes when I speak English with a partner.*
- FLCSA9 I feel tense when I have to speak English with a classmate in a pair.*
- FLCSA10 I'm afraid that others in a group discussion will laugh if I speak English.*
- FLCSA11 I can feel my heart pounding when it's my turn to speak English in a group.*

APPENDIX E (Continued)

Perceived Foreign Language Speaking Self-Competence

- PFLSS1 I can order food in English in a restaurant.
- PFLSS2 I can introduce myself in English during pair work to a classmate.*
- PFLSS3 I can talk about my hobbies in English during pair work with a classmate.*
- PFLSS4 I can give an English speech in a group of my classmates.*
- PFLSS5 I can give an English presentation as part of a group in front of my class.*
- PFLSS6 I can easily join in a conversation in a group of native English speakers.
- PFLSS7 I can give street directions in English to a foreigner.*
- PFLSS8 I can give an individual speech in English in front of my class.*

Desire to Speak English

- DSE1 I want to find friends from English-speaking countries.*
- DSE2 I want to become a good speaker of English.
- DSE3 I want to speak English with people from different countries.*
- DSE4 I want to be able to express my opinions and ideas in English.
- DSE5 I want to visit English-speaking countries.*
- DSE6 I want to speak English as much as possible in my English class.
- DSE7 I want to find opportunities to speak English outside class.*
- DSE8 I want to speak English with foreign exchange students at my university.*
- DSE9 I want to speak in English to my English teacher.

APPENDIX E (Continued)

Past English Classroom Experiences

- PECE1 I felt encouraged to speak English in my high school English classes.*
- PECE2 I enjoyed doing pair work in my high school English classes.
- PECE3 I enjoyed speaking with classmates in my high school English classes.*
- PECE4 I spoke English frequently with classmates during my high school English classes.*
- PECE5 I did enjoyable and interesting activities in my high school English classes.
- PECE6 I liked my high school English teachers.
- PECE7 High school English classes had a comfortable atmosphere.*
- PECE8 I got along well with classmates in my high school English classes.

Current English Classroom Perception

- CECP1 I feel comfortable expressing my opinions in Japanese in my current English class.
- CECP2 I feel comfortable in my current English class.*
- CECP3 The current English class atmosphere is good.*
- CECP4 There are many opportunities to speak with classmates in my current English class.
- CECP5 I get along with my current English classmates.*
- CECP6 My current English class makes me feel relaxed.
- CECP7 I enjoy speaking in English with classmates in my current English class.*
- CECP8 I enjoy doing pair work in my current English class.*

APPENDIX E (Continued)

Perceived Social Value of Speaking English

- PSV1 My parents hope I become a good speaker of English.
- PSV2 My friends think that it's cool to speak English.
- PSV3 My classmates want to be able to speak in English.
- PSV4 Speaking English will help me become a valued member of society.*
- PSV5 My English teachers hope that I improve my English speaking ability.
- PSV6 Japanese companies think highly of workers who have English speaking ability.*
- PSV7 Speaking English will help me become an internationally-minded person.*
- PSV8 My classmates want to speak in English during class.

APPENDIX F
FOREIGN LANGUAGE CLASSROOM SPEAKING CONFIDENCE
(JAPANESE VERSION)

Foreign Language Classroom Speaking Anxiety

- FLCSA1 クラスの他の学生のほうが自分よりも英語を話すのがうまいのではないかと心配になる。
- FLCSA2 クラス全体の前で英語を話すとあがってしまう。
- FLCSA3 自分のパートナーが自分よりも英語を話すのがうまいのではと心配になる。
- FLCSA4 クラスメイトとペアになって英語で話をするとき、うろたえ始める。
- FLCSA5 英語を話すときに間違いを犯すのではないかと心配になる。
- FLCSA6 自分の意見を英語で表現できないと、あがってしまう。
- FLCSA7 クラスメイトとペアになって英語を話すときパートナーが笑うのではないかと不安に思う。
- FLCSA8 パートナーと英語を話すとき間違いを犯すのではないかと心配になる。
- FLCSA9 クラスメイトとペアになって英語を話さなければならないとき、緊張する。
- FLCSA10 グループ・ディスカッションで英語を話すとき他の人が笑うのではないかと不安に思う。
- FLCSA11 グループの中で、自分が英語を話す番が回ってくるとドキドキする。

APPENDIX F (Continued)

Perceived Foreign Language Speaking Self-Competence

- PFLSS1 レストランでの料理の注文を、英語ですることができる。
- PFLSS2 ペア・ワークの際、クラスメートに英語で自己紹介ができる。
- PFLSS3 クラスメートとのペア・ワークの際、自分の趣味について英語で話すことができる。
- PFLSS4 クラスメートとのグループの中で、英語のスピーチをすることができる。
- PFLSS5 クラス全体の前で、グループ発表を英語ですることができる。
- PFLSS6 英語の母語話者と一緒のグループの中で、気軽に会話をするすることができる。
- PFLSS7 外国の人に英語で道案内を示すことができる。
- PFLSS8 クラス全体の前で、一人で英語のスピーチをすることができる。

Desire to Speak English

- DSE1 英語を話す国出身の友達を作りたい。
- DSE2 英語が上手に話せるようになりたい。
- DSE3 いろいろな国から来た人と英語で話をしたい。
- DSE4 英語で自分の意見や考えを表現できるようになりたい。
- DSE5 英語を話す国に行きたい。
- DSE6 英語の授業でできるだけ英語を話したい。
- DSE7 授業以外で英語を話す機会を見つけたい。
- DSE8 大学の留学生と英語で話をしたい。
- DSE9 英語の先生と英語で話をしたい。

APPENDIX F (Continued)

Past English Classroom Experiences

- PECE1 高校の英語の授業で英語を話そうという気持ちを持たされた。
- PECE2 高校の英語の授業でペア・ワークをするのは楽しかった。
- PECE3 高校の英語の授業でクラスメートと英語で話すのは楽しかった。
- PECE4 高校の英語の授業でクラスメートとよく英語で話した。
- PECE5 高校の英語の授業で楽しく面白い活動をした。
- PECE6 高校の英語の先生が好きだった。
- PECE7 高校の英語の授業は雰囲気よかった。
- PECE8 高校の英語の授業でクラスメートと仲が良かった。

Current English Classroom Perception

- CECP1 今の英語の授業で日本語で自分の意見を言うとき、落ち着いている。
- CECP2 今の英語の授業で、居心地はよい。
- CECP3 今の英語の授業の雰囲気はよい。
- CECP4 今の英語の授業で、クラスメートと話す機会がたくさんある。
- CECP5 今の英語の授業でクラスメートと仲が良い。
- CECP6 今の英語の授業で、自分はリラックスしている。
- CECP7 今の授業でクラスメートと英語で話すのは楽しかった。
- CECP8 今の英語の授業でペア・ワークをするのは楽しかった。

Perceived Social Value of Speaking English

- PSV1 親は、私が英語が話せるようになってほしいと望んでいる。
- PSV2 友達は英語が話せるのは格好いいと思っている。
- PSV3 クラスメートは英語が話せるようになりたがっている。
- PSV4 英語が話せたら、良い社会人になるのに役立つ。
- PSV5 英語の先生は、私が英語のスピーキング力が高まるのを望んでいる。
- PSV6 日本の会社は英語が話せる人を重要視している。
- PSV7 英語が話せたら、国際感覚のある人になるのに役立つ。
- PSV8 クラスメートは授業中に英語を話したがっている。

APPENDIX G
LIST OF ABBREVIATIONS AND ACRONYMS

This paper refers to psychological instruments, statistical analysis methods, and English proficiency examinations using abbreviations or acronyms listed here in alphabetical order according to the following categories: personality trait-related concepts/instruments, second language acquisition or communication-related examinations/concepts, and statistical techniques/concepts.

Personality trait-related concept/instrument	Acronym
Agreeableness	AGR
Chinese Personality Assessment Inventory	CPAI
Chinese Personality Scale	QZPS
Conscientiousness	CON
Emotional Stability / Instability	EMO
Extraversion	EI
Extraversion-Introversion	EI
Eysenck Personality Questionnaire	EPQ
Five Factor model	FFM
Hogan Personality Inventory	HPI

APPENDIX G (Continued)

Personality trait-related concept/instrument (Continued)

Imagination	IMA
Intellect/Imagination	INT
Minnesota Multiphasic Personality Inventory	MMPI
Myers-Briggs Type Indicator	M/BTI
Neuroticism	N
Neuroticism-Extraversion-Openness to New Experiences Five Factor Inventory	NEO-FFI
Neuroticism-Extraversion-Openness to New Experiences Personality Inventory Revised	NEO-PI-R
Norman Peer Rating Scales	NPRS
Psychoticism-Extraversion-Neuroticism model	PEN
Revised Big Five Factor Markers	RB5FM
Yawate-Guilford Personality Inventory	Y/GPI

APPENDIX G (Continued)

Second language acquisition or communication-related examination/concept	Acronym
Common European Frame of Reference	CEFR
Communication Apprehension	CA
Current English Classroom Perception	CECP
Desire to Speak English	DSE
English as a Second/Foreign Language	ESL/EFL
First Language	L1
Foreign Language Classroom Anxiety Scale	FLCAS
Foreign Language Classroom Speaking Anxiety	FLCSA
Foreign Language Classroom Speaking Confidence	FLCSC
Foreign Language Reading Anxiety Scale	FLRAS
Ministry of Education, Culture, Sports, Science, and Technology	MEXT
Perceived Value of Speaking English	PVS
Perceived Foreign Language Speaking Self-Competence	PFLSS
Personal Report of Communication Apprehension	PRCA
Past English Classroom Experiences	PECE

APPENDIX G (Continued)

Second language acquisition or communication-related examination/concept (Continued)

Second Language (or Foreign Language)	L2
Second Language Acquisition	SLA
State-Trait Anxiety Inventory	STAI
Super English high school	SELHi
Test of English for International Communication	TOEIC
Test of English as a Foreign Language	TOEFL
Trait Anxiety Scale	TAS
Willingness to Communicate	WTC

Statistical technique/concept Acronym

Adjusted goodness of fit	AGFI
Analysis of variances	ANOVA
Analysis of covariances	ANCOVA
Chi-squared	χ^2

APPENDIX G (Continued)

Statistical technique/concept (Continued)

Comparative fit index	CFI
Confirmatory factor analysis	CFA
Goodness of Fit Index	GFI
Mean squares/squared	MNSQ
Multiple analysis of variances	MANOVA
Non-normed fit index	NNFI
Principle components analysis	PCA
Root mean square error of approximation	RMSEA
Standardized root mean square residual	SRMR
Standardized <i>z</i> -scores	ZSTD
Structural equation modeling	SEM
Structural regression modeling	SRM
Variance accounted for	VAF