ENRICHING STRUCTURAL MODELS OF L2 WILLINGNESS TO COMMUNICATE: THE ROLE OF PERSONALITY, EGO PERMEABILITY, AND PERCEIVED DISTANCE

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Education

By
James A. Elwood
May, 2011

Examining Committee Members

David Beglar, Advisory Chair, CITE/TESOL
Marshall Childs, CITE/TESOL
Mark Sawyer, External Member, Kwansei Gakuin University
James Sick, External Member, J. R. Oberlin University
Edward Schaefer, External Member, Ochanomizu University

©

Copyright

2011

by

James A. Elwood

ABSTRACT

Willingness to communicate (WTC) in a second language (L2) is crucial to the development of communicative speaking skills. This study is a cross-sectional investigation of the role in models of second language (L2) willingness to communicate of three personality variables hitherto underresearched in the L2 field: extroversion, ego permeability (one's capacity to tolerate ambiguity), and perceived distance from one's core persona. A sample of 252 Japanese university students responded to a set of instruments used to measure individual difference variables and personality variables; the instruments were drawn from the fields of L2 acquisition and psychology as well as a 5-item instrument designed to measure perceived distance in a series of participatory L2 speaking activities.

Confirmatory factor analysis, Rasch analysis, and structural equation modeling were utilized to validate the respective instruments. The International Posture instrument was best represented by a two-factor configuration consisting of Intergroup Approach-Avoidance Tendency and Intercultural Friendship Orientation, while the L2 Communicative Confidence was altered to consist of three factors (L2 Anxiety, Perceived L2 Communicative Competence, and Extroversion). The hypothesized additions of Ego Permeability and Perceived Distance failed to improve the measurement models, and the original Ego Permeability variable functioned poorly in this context.

The MacIntyre and Charos (1996) model had marginal fit to the data even after undergoing considerable respecification. The models of Yashima (2002) and

Yashima, Zenuk-Nishide, and Shimizu (2004) were found to have good fit as originally conceptualized, but the addition of Extroversion and paths from International Posture and L2 Communicative Anxiety improved the fit of both models.

Collectively, the results indicate that extroversion plays an important role in models of L2 WTC and that the basic models of Yashima and colleagues are robust. These findings provide crucial insights into the process of L2 WTC, an important factor in the students' acquisition of communicative competence.

ACKNOWLEDGMENTS

I wish to acknowledge the contributions of a large number of people that allowed this project to come to fruition. First, with love I humbly acknowledge the three teachers in the Elwood clan: my parents, Warren and Donna, and my uncle, Henry, all of whom blazed the teaching trail for me. To David Beglar, my primary adviser, committee chair, and baseball confidante, I thank you for your many, many hours of wise and patient guidance. To Mark Sawyer, Marshall Childs, Ed Schaefer, Jim Sick, and David Beglar: gentlemen, for your guidance as my committee members, I tip my hat. To the many excellent teachers I have been privileged to learn from at Temple University Japan and elsewhere, I offer my appreciation for your excellent classes. Next, to my classmates in the 7th Tokyo Cohort at Temple University Japan—I thank you for your love, inspiration, and humor. To Shinichi Nagata, who so carefully coded a large part of the data, I offer my heartfelt appreciation. To Mike Hood, I offer my thanks for allowing me to include his students in this study. I wish to thank Kumiko Fushino and Mika Falout for their assistance with the translations in this study.

Closer to home (and the heart) are the members of my family. To my loving wife, Yuko, and our children (Naoki, Isaac, Chelsea, and Dawes), I thank you for the many years you all have been a source of love, inspiration, and laughter.

Finally, I must recognize the hundreds of nameless participants, without whom this study would not have been possible.

Needless to say, this is a group effort, and I humbly acknowledge the contributions of these fine people above. Thank you.

This dissertation is dedicated with love to my parents, Warren and Donna.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The Higgs boson—this elusive particle, often referred to as the "God particle" for its hypothesized ability to confer mass on other particles (Lederman & Teresi, 1993), represents the holy grail at the center of elaborate research in Europe and the United States. To the best of researchers' knowledge, this unconfirmed particle is one member of the family of elementary particles that constitute all matter. Young students learn about electrons, protons, and neutrons, and some students continue on to more exotic and elementary particles such as the 6-member family of quarks (up, down, top, bottom, charm, and strange are the six varieties) and its alter-ego family of anti-quarks. Such particles interact in a bewildering number of ways, with the result that something much larger and more complex (e.g., a carbon-12 atom) can exist. The quest to understand the workings of atomic and subatomic particles that form larger, more complex molecules has continued for many years, with new discoveries and insights periodically enhancing scientists' understanding of the fundamental structure of matter.

The process of second language (L2) communication is similar: Small "particles" such as a particular type of anxiety, some situational element, or a propensity toward extroversion[1] interact to produce L2 willingness to communicate or perhaps even a speech act. Similar, too, is the quest to understand the mechanisms of L2 communication. While the various "particles" such as L2

---

[1] Although both spellings (extr<u>a</u>version and extr<u>o</u>version) are correct, I prefer the parallelism of spelling extr<u>o</u>version with an O and intr<u>o</u>version with its O. Extroversion is spelled as such throughout the current manuscript.

anxiety and motivation have been investigated in isolation or in small configurations, the overall configuration and the interactions of its constituents have also been the foci of research, and periodic discoveries and insights have informed that development. This investigation of university English as a Foreign Language (EFL) students in Japan is situated at the intersection of several issues that I wanted to investigate: individual difference variables implicated in L2 communication (e.g., motivation, L2 anxiety, L2 WTC), personality variables such as extroversion, ego permeability (the degree of one's cognitive flexibility), and perceived distance from one's core first language (L1) self while engaged in L2 interaction. The ultimate issue concerns the agglutination of those variables in models of L2 willingness to communicate. In this chapter I present the background of the issue, the purpose and significance of the study, and the delimitations of the study.

The Background of the Issue

Although I could not have known it at the time, this research project began many years ago while I was an undergraduate student at Montana State University. I spent a modest amount of time studying (physics among other things) and considerably more time participating in a musical group known as New Genesis, which was loosely affiliated with the Christian campus ministry. Many of our performances were in Christian churches of various denominations, but when time permitted we also performed in schools and nursing homes with inspirational messages. I joined that group as a singer, but over my six years in the group I taught myself guitar and learned the rudiments of puppetry.

The puppets were Sesame Street characters, and we used them in short sketches to teach, discuss issues, and simply entertain. Usually we would invite children to come forward and sit close to the puppet stage, and the children were generally spellbound. However, I often felt that some adults were at least as entranced as the young people, if not more so. Furthermore, the reactions to the puppets were often unexpected and very much unpredictable: I remember sketches falling flat with audiences that initially seemed to be responsive, and some sketches succeeded brilliantly in circumstances that had appeared less than promising. The same was true with individuals who were asked to participate: Some seemingly shy people would converse with the puppets easily, while some outgoing people became very ill at ease while speaking with Oscar, Grover, or Cookie Monster.

Some years hence I resurrected my puppet skills when I began teaching EFL in universities and doing workshops on using puppets in teaching, and the same phenomenon occurred: Some shy participants would become quite loquacious when interacting with or animating a puppet, while a few garrulous participants suddenly would become self-conscious. Why?

At that point I was coming to more fully understand the plethora of variables that influenced L2 acquisition. Motivation, willingness to communicate, attitudes, and so forth—all are important qualities, but they did not seem to address the question of why the character of some individuals seemed to change so drastically. L2 anxiety seemed a logical candidate to begin to explain such variation, but because the situation included unusual elements—non-human interlocutors, for example—the explanation might be more complex, I surmised. I began to suspect that personality variables were playing important roles: Perhaps the degree of a person's flexibility played a role? Or perhaps some people more readily felt some resonance with the puppets, or felt free of inhibitions while animating a hand puppet?

In my own experiences, I had often felt that when speaking Japanese, I somehow became a different person and behaved differently than when speaking English, and my conversants have commented on how much I change. In doing puppetry, too, I have always found myself slipping easily into various roles depending on the situation. Although I have never known the origin of that ability, it seemed that the "possible selves" evoked in the work of Helen Markus and colleagues (e.g., Markus & Kitayama, 1991; Markus & Nurius, 1986) could be a starting point. Because another self would be somewhat distant from one's core self or personality and therefore free of the inhibitions that my core self has, that other self could be beneficial in acquiring another language. Moreover, given that one's self could play an important role, I wondered if such other personality variables as extroversion and ego permeability might also be implicated in this phenomenon.

At that point I began searching the literature for studies dealing with distance. In psychology and in education the notion of perceived distance has been accorded considerable attention, but it has appeared only sporadically in the second and foreign language literature. One example was Brown (1980), who wrote that L2 acquisition would be facilitated by “an optimal distance,” which referred to a critical stage based on cultural distance in which L2 learners have an optimal chance to acquire the target language. The concept, however, failed to gain momentum in L2 research. Originally proposed by Guiora (e.g., Guiora, 1972; Guiora, Brannon, & Dull, 1972), ego permeability has received considerably more attention, primarily in the work of Madeline Ehrman and colleagues (e.g., Ehrman, 1996; Ehrman & Oxford, 1996). Ehrman (1999) noted that it is essentially “the degree to which people tend to compartmentalize their experiences” (p. 68), which is manifest in such areas as receptivity to one’s own intuitive insights, to outside influences, and in particular to tolerance of ambiguity. However, it has remained largely absent from models of L2 communication. The situation of personality variables such as extroversion was similar, having been accorded some attention (e.g., Dewaele, 2005; Dewaele & Furnham, 1999) yet remaining mostly absent from models of L2 communication.

Statement of the Problem

The lack of attention on personality variables is, as part of the process of second language acquisition (SLA), an important issue for all foreign language (FL) learners. Unlike areas such as anxiety and motivation to which considerable research has been devoted, distancing has received little attention from EFL and English as a second language (ESL) researchers. As one aspect of SLA and to better understand the process of SLA, distancing and other personality variables including extroversion and ego permeability deserve a fuller accounting than they have received to date.

In this study, I investigate several aspects of SLA that have to date received minimal attention from SLA researchers. Personality, extroversion, perceived distance, and ego permeability have been explored in detail in other contexts, and extroversion (e.g., Furnham & Dewaele, 1999), ego permeability (Ehrman & Oxford, 1996) and aspects of personality (e.g., MacIntyre & Charos, 1996) have been shown to be of importance in SLA. Perceived distance, however, has received scant attention in the ESL literature, and such distance-inducing techniques as drama and roleplay are virtually absent in the EFL literature.

The Purposes and Significance of the Study

Beyond satisfying my own curiosity, this study is significant because it adds to the literature in the fields of ESL and EFL. The data in this study were collected using an eclectic set of instruments from various fields and locations (and one created for this study), so the first purpose was the validation of the instruments using Rasch analysis and structural equation modeling. The validation of instruments, although a prudent action, is seldom done in EFL studies. The second purpose is thus the modification of instruments to make them appropriate for Japanese EFL contexts. It is hoped that such instruments will be of benefit in further research on personality variables in L2 communication models and in improving the models themselves.

The third purpose of this study is to examine two models of L2 willingness to communicate: the MacIntyre and Charos (1996) model and the 2002 and 2004 variants of Yashima's model (Yashima, 2002; Yashima, Zenuk-Nishide, & Shimizu, 2004). In the initial step, the structure and dimensionality of the L2 Communicative Confidence measurement model and two multi-dimensional constructs (International Posture and Ego Permeability) are investigated. Next, the original three models are replicated with minor revisions based on the Rasch analysis results of the instruments. Finally, the three models with the hypothesized changes are investigated.

Fourth, this study will enhance our understanding of perceived distance, which is manifested in such common techniques as roleplay and drama. This understanding could be pedagogically significant for L2 instructors who already find or could find such techniques useful. Furthermore, knowing more about the degree of students' flexibility should allow teachers to use roleplay and drama as effectively as possible.

This study also has methodological significance because of the use of Rasch analysis and structural equation modeling to investigate the instruments and models. In addition, the extension of the criteria for adequate category function in Rasch analysis is a small but potentially useful step that should prove beneficial to researchers investigating category function in Rasch analysis.

The intended audience for this study is broad. Researchers should find useful findings in the instruments validated and the models examined, and the statistical techniques should prove useful in further research on personality variables and L2 models of communication. Moreover, I hope that the insights gleaned from this study will inform EFL teaching, too, to the benefit of students learning a second language and the teachers instructing them.

Delimitations

In the interest of transparency, certain delimitations are briefly covered in this section. The first limitation concerns the limits of the sample used in the present study. The participants were university-level EFL learners in Japan, and those learners were from six universities in eastern Japan (although three of the universities are very competitive and attract students from throughout Japan). Furthermore, most of the participants were first-year students.

The second limitation concerns the design of the study: The current study was cross-sectional and therefore provided only a snapshot of these learners. As MacIntyre (2007) suggested, SLA should be viewed as an ongoing process, not one frozen at a particular moment in time. The specification of directional influences often presupposes that those influences are not instantaneous, so interpreting results involving the concurrent measurement of such variables should be done with caution. However, Gollob and Reichardt (1987) suggested that restricting models based on a priori hypotheses and salient research can overcome the drawback of a cross-sectional design addressing sequential variables.

MacCallum and Austin (2000) pointed out that inferring causality or directional influences based on cross-sectional studies requires making one of two possible assumptions. The first possible assumption is that causal variables do not change substantially over a time interval under study (i.e., those have a slope of zero), which thus renders moot the necessity of considering a time interval for the causal influence to take effect. The second assumption, which they assert might not be unusual, is that causal effects are essentially instantaneous, so the lack of a time interval for any causal effects to take effect is not problematic.

In defense of using a cross-sectional design to evaluate what might be longitudinal processes, we might consider the theoretical process in mathematics that leads from a discernible change in a function over a two-dimensional interval (denoted by delta) to the derivative of a function, which is the slope at a given *point* (i.e., a one-dimensional “interval”) instead of over a two-dimensional interval. Because a non-linear function results in any measurement of slope other than a derivative being an inexact approximation, the smallest possible interval produces the most exact measurement. In the previous paragraph, the first assumption (“any causal variables do not change substantially over a time interval”) is analogous to a linear function, whereas the second assumption (“causal effects are essentially instantaneous”) is analogous to a derivative as an exact measurement of the slope at a single point.

Third, the category structure of all the instruments in this study could be construed as problematic: All of the questionnaires originally employed an odd number of response categories, at the center of which is a neutral midpoint. The existence of that midpoint allows participants to avoid providing a positive or negative answer, an avoidance that has been shown to be a propensity of Japanese (Chen, Lee, & Stevenson, 1995). On the other hand, the neutral midpoint allows participants to give such responses if those represent their actual feelings.

Fourth, some of the instruments and subscales consisted of few items. The foremost example in the current study is the Sensitivity subscale of the Ego Permeability instruments; two items are insufficient to measure a construct adequately. In addition, more items are needed to measure Attitudes (just two items in the MacIntyre and Charos study and four items in the current study), Frequency of L2 Communication (three and five, respectively), and Sensitivity (two items in both studies).

The Organization of This Study

In this introductory chapter, I have presented the genesis and evolution of this study. What began as curiosity about a recurring phenomenon expanded into the present study, which investigates the roles of personality variables in three models of L2 willingness to communicate. The purposes and significance were outlined next; it is hoped that those purposes contribute to the literature and inform future research. Finally, the delimitations of the current study were noted.

In Chapter 2, the literature relevant to the current study is introduced in five major sections: Models of L2 Communication, Improvements to SLA Communication Models through the Addition of Personality Variables, The Human Actor and the Self, Distance, and Ego Boundaries. At the end of this chapter, the hypotheses and research questions are presented. In Chapter 3, Methods, I describe the participants, the instrumentation, the procedures by which the data were gathered, and review the analytical approaches used in this study. Chapter 4 concerns the validation of the two proficiency instruments used in this study. In Chapter 5, I present the validation results for the seven individual difference instruments that were used in studies by MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004). Chapter 6 details the validation of the variables added to the respective models: Perceived Distance, Personality and its five subscales, and Extroversion. In Chapter 7 I discuss the psychometric properties of the instruments covered in Chapters 4-6. In Chapter 8, the primary results of the study are presented in five major sections: Structural Equation Modeling Assumptions, Measurement Models, Path Analysis of Models Based on MacIntyre and Charos (1996), SEM Assessment of Models Based on Yashima (2002), and SEM Assessment of Models Based on Yashima et al. (2004). Chapter 9 includes a discussion of the findings about the three research questions, theoretical implications of the study, and methodological innovations undertaken. Finally, Chapter 10 includes limitations, directions for future research, and my final remarks.

CHAPTER 2

REVIEW OF THE LITERATURE

Four major bodies of literature are reviewed in this chapter. First, the development of a series of models of second language acquisition (SLA) is outlined. Next, on the heels of the suggestions by MacIntyre (1994) and Yashima, Zenuk-Nishide, and Shimizu (2004) that additional factors might enhance L2 communication models, extroversion, perceived distance, and ego permeability are explained and posited to strengthen the three models of L2 communication. The notion of the human being as an actor is explained and then the concept of distancing is defined and located in several early models of SLA. Finally, gaps in the literature are identified, the purposes of this study are described, and the specific research questions for this study are outlined.

Models of L2 Communication

The basis of the current study is models of L2 communication, which have developed over time. Early models focused on a limited number of factors such as foreign language (FL) aptitude as predictors of L2 acquisition; factors such as motivation and anxiety played no part. However, that paucity of variables began to be remedied with the work of Robert Gardner and colleagues (e.g., Gardner & Lambert, 1959, 1972; Gardner & MacIntyre, 1991; MacIntyre & Gardner, 1991b, 1994a, 1994b). The story then continues with pieces added to the growing model from different fields (e.g., psychology and communication studies) and undergoing more sophisticated evaluation as improved instruments and analyses became available. The strand of research pursued by Gardner and colleagues (e.g., Gardner, 1985; Gardner & Lambert, 1959, 1972) culminated in the socio-educational model, which roughly coincided with the emergence of the Clément's (1980) social context model. Incorporating those two models, Peter MacIntyre (Gardner & MacIntyre, 1994) then expanded the model with the addition of two elements: the notion of willingness to communicate from the work of McCroskey and associates (e.g., McCroskey, 1992; McCroskey & Richmond, 1991) and the personality categories specified in the Big 5 model of personality (Goldberg, 1992, 1993). The final step came with the model hypothesized and empirically assessed by Yashima and colleagues (Yashima, 2002; Yashima et al., 2004).

In the following pages I explore the journey that has resulted in the models evaluated in this study.

Gardner's (1985) Socioeducational Model

Expanding on his earlier work and the research outlined above, Gardner (1985) produced the socioeducational model. In investigating Canadian contexts, the socioeducational model posits two basic attitudes, integrativeness, which Gardner (2001) himself noted was "used in different ways by different individuals" (p. 1) and attitude toward the learning situation, both of which contribute to the learner's level of motivation to acquire the L2. The subsequent motivational level then directly influences language learning outcomes such as proficiency and fluency. Figure 1 shows a portion of Gardner's (1985) model.

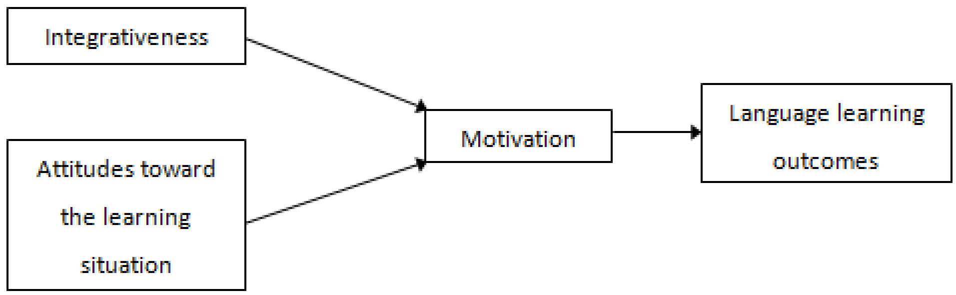


Figure 1. Portion of Gardner's (1985) model of L2 Communicative Competence. From "Personality, attitude, and affect as predictors of second language communication," by P. D. MacIntyre & C. Charos, (1996), *Journal of Language and Social Psychology, 15*(1), p. 5. Copyright 1996 by the *Journal of Language and Social Psychology.* Reprinted by permission.

Although the socioeducational model has proven very useful in helping researchers and educators to better understand SLA and in providing a theoretical basis for further research, it was developed in and from a specific context, the Canadian milieu, yet a general model needs to have support from various contexts (Dörnyei, 2003; Dörnyei & Csizér, 2002). Canada is officially bilingual, and its immigrant population is typically in an L2 situation vis-à-vis either English or French (or conceivably both). In that situation, the notion of integrativeness involves actual or metaphorical integration into a community. That notion is appropriate for that context, in which there is a clear need for non-English speaking immigrants to repeatedly use at least one L2 in order to function in daily life.

On the other hand, as in much of Asia outside of Hong Kong and Singapore, many Japanese learners of English are not seeking to integrate into an L2 community, as their primary goal is simply to communicate with the target language group or even to address some non-linguistic goal such as passing an English proficiency examination (an instrumental motivation). In Asia the chances are substantial that L2 speakers of English speak with other non-native speakers of English more frequently than with native speakers, suggesting that any integrative component is minimal. For this reason, any L2 model appropriate for most Asian contexts needs to account for non-proximal spatial orientation, such as Japanese learners of English who are far removed from frequent contact with English-speaking communities. Pratt (1991) wrote of the “contact zone,” which has generally included proximal contact; in the case of groups far removed from spatial contact, however, physical proximity plays a reduced role while an orientation toward the target language plays a more crucial role.

Whether integration is the goal or not, contact and context do play crucial roles in L2 communication. Echoing the findings of Clément, Dörnyei, and Noels (1994), Kormos and Csizér (2007) found that encounters with foreigners helped reduce L2 communicative anxiety, increase perceived L2 competence, and both change and sustain attitudes toward target language speakers. Interestingly, even contact perceived as negative can, if frequent enough, “help develop the individual’s expectations regarding the capacity to face successfully second language usage in [aversive] situations” (Labrie & Clément, 1986, p. 279).

Regarding actual frequency of L2 communication, researchers have found that context is an important determinant of L2 use (e.g., Clément, Baker, &

MacIntyre, 2003). As MacIntyre and Charos (1996) noted: “Having more opportunities for interaction in [the L2] may lead to an increase in perceived competence, a greater willingness to communicate in [the L2], and more frequent communication” (p. 17). Because English is learned in Japan as a foreign language, L2 learners use English primarily in the classroom setting and, to a much smaller extent, when opportunities for using English arise. Some of the possible contexts include travel abroad, study abroad, homestay abroad, residence abroad, attendance at an English conversation school, and the compulsory English education in secondary school. In the current study these various opportunities were grouped into a variable labeled English Experience, which replaces the context variable of the MacIntyre and Charos study. Furthermore, longer experience that should equate with a greater number of opportunities to use the target language was scored more highly.

Having suggested that context is crucial, I now address a slightly earlier model that posited exactly that point.

Clément’s (1980) Social Context Model

The work of Clément and colleagues (Clément, 1980; Clément & Kruidenier, 1985) makes two important points for the current study. First, Clément hypothesized that “frequent and pleasant contact with the L2 group will ultimately lead to variations in L2 confidence” (p. 192). Such frequent and pleasant contact should result in gains in L2 confidence.

The second important point concerned two predictors of WTC: communication anxiety and perceived communicative competence. While the importance of those two variables in predicting L2 WTC has been shown empirically (Baker & MacIntyre, 2000; MacIntyre, Clément, Baker, & Conrod, 2001; McCroskey & Richmond, 1991), Clément and Kruidenier (1985) took the innovative step of hypothesizing that perceived competence and anxiety constituted a single exogenous variable, L2 communicative confidence. This variable was in turn posited to predict L2 WTC, which then directly influences the frequency of L2 communication.

The two models outlined above have been evaluated and found to be empirically sound, but shortcomings were also pointed out. The applicability of the socioeducational model to foreign language contexts has been questioned, for research findings have shown that instrumental motivation is equally or more important in various foreign language learning contexts (Clément, Dörnyei, & Noels, 1994; Dörnyei, 1990; Samimy & Tabuse, 1992). As Dörnyei (1990) pointed out, in foreign language learning situations, “affective predispositions toward the target language community are unlikely to explain a great proportion of the variance in language attainment” (p. 49). Clément and Kruidenier (1985) emphasized the need to define operationally the integrative and other orientations that are relevant to a particular context.

Regardless of the shortcomings, the Gardner and Clément models were and have remained useful. The next step in this journey came with the addition of willingness to communicate to the Gardner model by Peter MacIntyre in 1994. The MacIntyre (1994) model is the topic of the following section.

MacIntyre's (1994) Willingness to Communicate Model

In his 1994 study, MacIntyre advanced a model whose terminus was L2 willingness to communicate, which was hypothesized to predict actual speech acts. In the model tested (Figure 2), introversion underpinned both perceived competence and communication anxiety, while self-esteem predicted communication anxiety only. Anxiety influenced perceived competence, and both perceived competence and communication anxiety significantly predicted L2 WTC. The overall model had good fit to the data with χ^2 (21) = 13.4, *p* = n.s., GFI = .99, and AGFI = .96.[2]

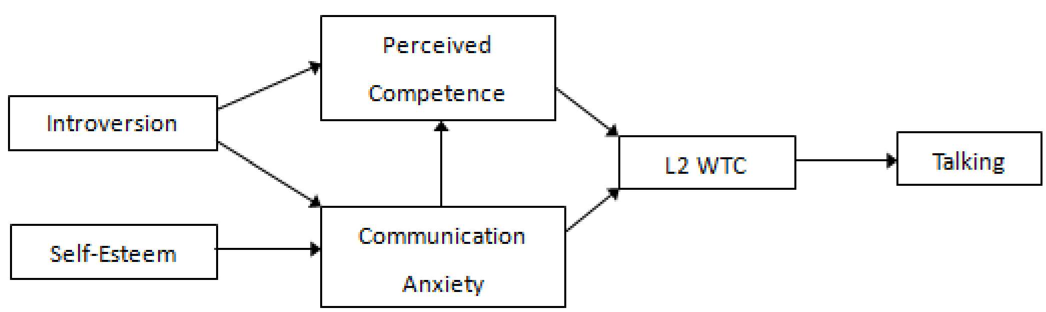


Figure 2. Portion of MacIntyre's (1994) willingness to communicate model. From P. D. MacIntyre & C. Charos, (1996), "Conceptualizing willingness to communicate in a L2: A situated model of confidence and affiliation." *Journal of Language and Social Psychology*, *15*(1), p. 8. Copyright 1996 by the *Journal of Language and Social Psychology*. Reprinted by permission.

[2] GFI is an asymptotic goodness-of-fit index and Adjusted GFI corrects for model complexity (Jöreskog & Sörbom, 1984). Values greater than .90 indicate good fit, but because both are insufficiently and inconsistently sensitive to model misspecification (Byrne, 2006) and strongly influenced by sample size (Marsh, Balla, & Hau, 1988), Hu and Bentler (1998) have advised against using them. Fit statistics are discussed in detail in Chapter 3.

The results offered empirical support for the predictive roles of communication anxiety and perceived competence. Furthermore, MacIntyre suggested that “[communication anxiety] has its roots in broader personality variables, such as introversion and self-esteem … [that] by no means exhaust the range of personality variables” (p. 139). This implication was investigated more fully in his study with Catherine Charos (MacIntyre & Charos, 1996) and comprises part of the rationale for the current study.

With MacIntyre having taken a small step toward integrating different models and orientations, MacIntyre and Charos (1996) then merged those into a larger model that offered a more comprehensive account of L2 communication. That model is the topic of the following section.

The WTC Model of MacIntyre and Charos (1996)

The MacIntyre and Charos (1996) model incorporated Gardner’s socioeducational model and the social context model of Clément (Clément, 1980, 1986) in addition to adding L2 WTC and the Big 5 personality subscales of Goldberg (1992, 1993). Building on MacIntyre’s (1994) study in which introversion and self-esteem were found to influence perceived L2 competence and communication anxiety, MacIntyre and Charos hypothesized that the five factors included in the Big 5 personality configuration would underpin the models, acting as lower-level variables akin to those in the Clément, Gardner, and MacIntyre models; those personality elements and context were posited to influence (primarily indirectly) L2 WTC and frequency of communication.

Earlier work on the role of personality variables in L2 communication models yielded mixed results. Lalonde and Gardner (1984), for example, included 18 personality variables in their study, but the results indicated that those variables had very few correlations with language achievement, aptitude, or perceived L2 competence. However, the personality variables were grouped into two groups based on a factor analysis; those two groups, analytic orientation and seriousness, did correlate with achievement, aptitude, and perceived L2 competence, which implies that more fundamental personality traits are present. This search for fundamental underlying personality traits was similar to the discussion on the optimal number of factors (e.g., Digman, 1990; McCrae & Costa, 1989) that culminated in the well-known Big 5 model of personality traits (Goldberg, 1992, 1993).

The combined model is shown in Figure 3. This model includes the elements in Figure 1 (attitudes, integrativeness, motivation, and learning outcomes [here, frequency of L2 communication]). The leftmost column is composed of context and the five subscales of the Big 5 personality construct: intellect, extroversion, agreeableness, emotional stability, and conscientiousness. The path analysis of the model yielded adequate fit to the data: χ^2 (45) = 55.75, p < n.s., GFI = .92, AGFI = .84, and RMSR = .067. However, the model underwent considerable respecification with four paths added to the hypothesized configuration and three paths deleted.

Building on this model, MacIntyre, Clément, Dörnyei, and Noels (1998) introduced the pyramid model (Figure 4), a conceptualization to account for individual differences in initiating communication in a L2 context. The pyramid model is outlined in the following section.

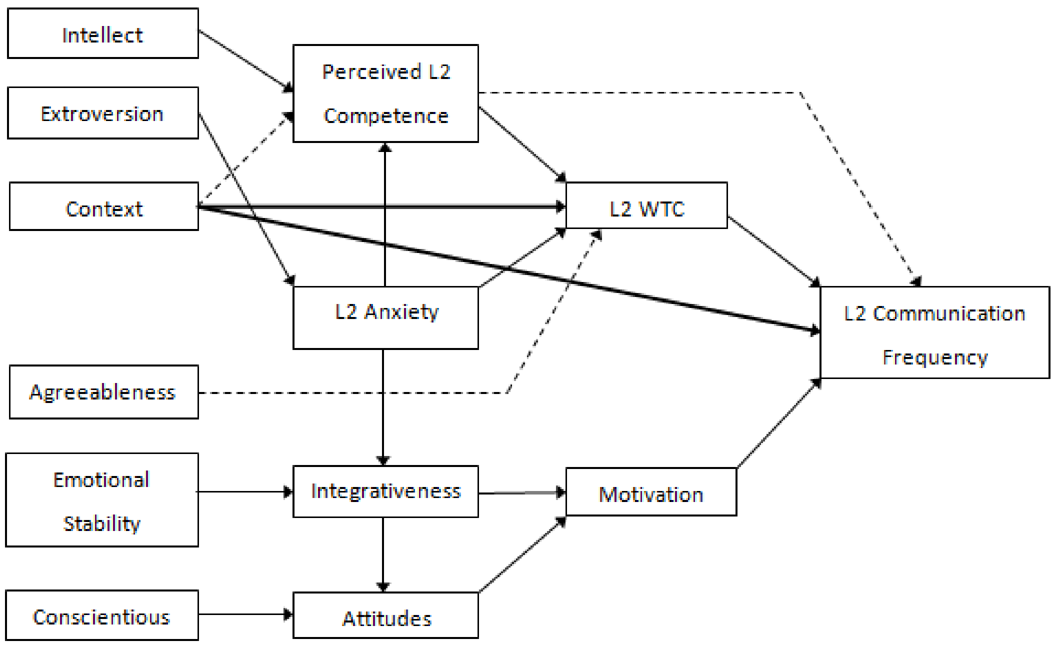


Figure 3. Final MacIntyre and Charos (1996) model of L2 willingness to communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 18. Copyright 1996 by *Journal of Language and Social Psychology*. Reprinted with permission.

The Pyramid Model of MacIntyre and Colleagues

MacIntyre et al. (1998) introduced the well-known pyramid model (Figure 4) to account for individual differences in initiating communication in a foreign language. The pyramid conceptualization is composed of six layers. The lower three strata, Layers IV-VI, are enduring, situation-independent elements. The bottom layer, Societal and Individual Context, includes personality and intergroup climate, both of which are general tendencies. These directly feed into the more specific components in the layer above, Affective and Cognitive Context. This group includes intergroup attitudes (which embrace Gardner's integrative orientation), the social situation or context, and communicative competence. Of note here, as McCroskey and Richmond (1991) pointed out, is that communicative competence is more heavily centered on *perceived* competence rather than on an objective measure of competence. In Japan, L2 learners commonly focus on self-criticism (Heine, Takata, & Lehman, 2000) and thus understate their own competence, exhibiting a culturally inculcated "modesty bias" that requires avoiding extreme responses (Chen, Lee, & Stevenson, 1995).

The third layer from the bottom is concerned with Motivational Propensities. Included therein is interpersonal motivation, which comprises both aspects of control and affiliation. The middle box is intergroup motivation, which is often manifested for L2 learners in the desire to make friends with speakers of the target language. The final box is L2 self-confidence, which reflects the relationship between the learner and the L2.

Dependent on social interaction for context, social distance includes both linguistic features and such non-linguistic elements as gestures and facial expressions. On a larger scale, proximal distance is important in current conceptualizations of integrative motivation and international posture. This type of distance refers to physical distance between or among groups and individuals and is important in today's world, in which groups learning English or other foreign languages are often separated physically from the target groups. This physical separation is the case with most Japanese learners of English, whose country is an archipelago. Of course, modern transportation and media have reduced this distance, but, inasmuch as many Japanese university students lack travel experience (Elwood, 2005), it remains an important factor.

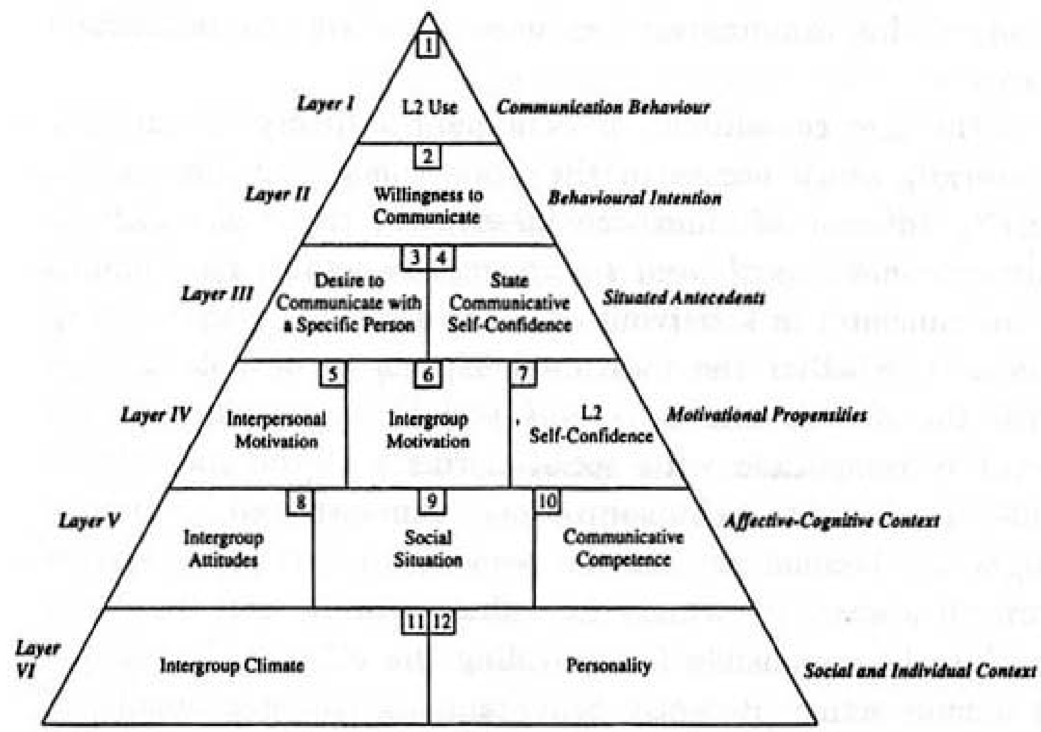


Figure 4. Schematic representation of the variables influencing L2 WTC. From "Conceptualizing Willingness to Communicate in a L2: A Situated Model of Confidence and Affiliation," by P. D. MacIntyre, R. Clément, Z. Dörnyei, and K. Noels, 1998, *Modern Language Journal, 82*, p. 547. Copyright 1998 by *The Modern Language Journal.* Reprinted with permission.

Whereas the lower three strata are enduring, situation-independent elements in the conceptualization of MacIntyre and Charos, the top three layers are concerned with situational factors. The third layer from the top includes the desire to communicate with a specific person and state communicative self-confidence, of which perceived competence and lack of anxiety are components. Given confidence and desire, the next layer, Willingness to Communicate, is the culmination of lower strata and represents “the readiness to enter into discourse at a particular time with a specific person or persons, using the L2” (MacIntyre et al., 1998, p. 547). This is the *intention* to initiate a communicative event, whether or not the event actually occurs. Finally, the apex represents a speech act in which the learner engages in a communicative event).

Social distance figures prominently in several recent models of L2 communication. In their pyramid conceptualization of willingness to communicate MacIntyre et al. (1998) proposed a number of factors that address social interactions, including intergroup actions, group motivation, and group climate. Occurring on the lower three levels, which underpin the entire model, these factors highlight the ubiquity of group activity, which naturally includes group boundaries and distance between and among groups (see Figure 4).

However, having been developed and evaluated in ESL contexts, the MacIntyre et al. (1998) model has features that might not be appropriate for EFL contexts. The work of Yashima and colleagues in extending this basic model to Japanese contexts is explored in the following section.

Yashima and EFL Contexts

Although Brown (1973) noted that his proposed model of optimal distance applied only to ESL contexts, it has been extended. For EFL contexts, Yashima and colleagues (Yashima, 2000, 2002; Yashima & Zenuk-Nishide, 2008; Yashima et al., 2004) have advanced the notion of *international posture*, which appears to satisfactorily supplant the concepts of acculturation and integrative orientation in Gardner's (1985) socioeducational model. Gardner (2001) defined integrativeness as "a genuine interest in learning the second language in order to come closer psychologically to the other language community" (p. 7). Specifically addressing the situation in Japan in which integrative orientation is of minor importance, Yashima (2002) noted that some learners "are more interested in or have more favorable attitudes toward what English symbolizes than other learners" (p. 57). This orientation can thus include "interest in foreign or international affairs, willingness to go overseas to stay or work, readiness to interact with intercultural partners, and, one hopes, openness or a non-ethnocentric attitude toward different cultures, among others" (p. 57). Furthermore, Yashima and colleagues have demonstrated that international posture does play an important role in models of SLA (Yashima et al., 2004).

The core model of L2 communication shown in Figures 5 and 6 (rotated 180 degrees about the Y-axis from the original figure) illustrates the importance of international posture. In this conceptualization, international posture directly influences frequency of L2 communication, willingness to communicate in the L2, and motivation. Motivation in turn influences L2 Communication Confidence with Proficiency playing some indeterminate, mediating role (the role of proficiency in the model was not specified in the original study). L2 communicative confidence directly influences L2 WTC, which together with Motivation determines the frequency of L2 communication.

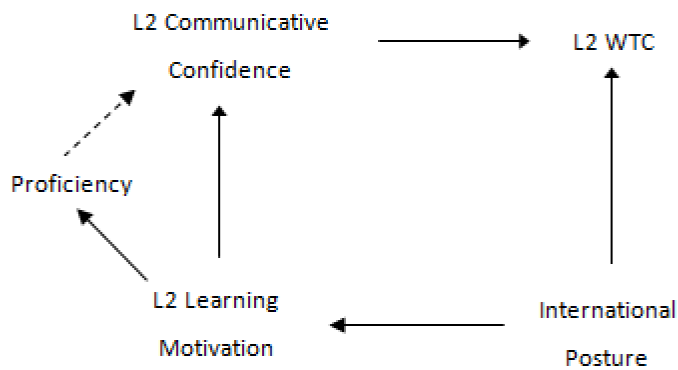


Figure 5. L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal*, *86*(1), 61. Copyright 2002 by *The Modern Language Journal*. Reprinted with permission. Note that the dashed path was hypothesized but found to be non-significant.

For both the 2002 and 2004 models, the results indicated a good fit of the respective models to the data. For the 2002 model shown in Figure 5, the fit statistics included χ^2 (49) = 62.63, (p = n.s.), CFI = .99, GFI = .97, adjusted GFI = .95, RMSEA = .031. Results for the 2004 model also indicated good fit of the model to the data with χ^2 (48) = 74.48 ($p < .01$), GFI = .93, CFI = .96, and RMSEA = .060.

As the reader will note, the models differ slightly. In the 2002 model, Frequency of L2 Communication was not included, whereas L2 Proficiency was. However, the hypothesized path from L2 Proficiency to L2 Communication Confidence was not significant. In the 2004 model the role of L2 Proficiency was implied with its inclusion parenthetically, but it was not included in the analysis. On the other hand, in the 2004 model, Frequency of L2 Communication was included in the model, which had very good fit to the data as noted above.

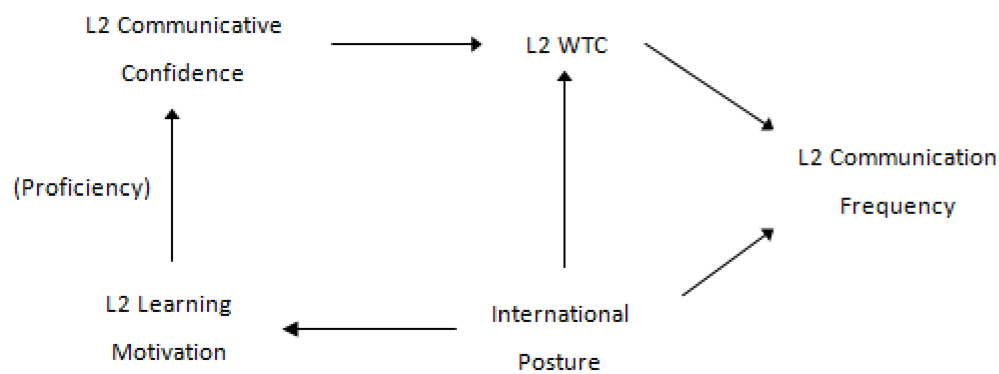


Figure 6. L2 communication model (minus proficiency). From "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Acquisition," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning*, *54*(1), p. 127. Copyright 2004 by *Language Learning*. Reprinted by permission.

Additions to such models have been investigated to some extent. In his (1994) study, MacIntyre suggested that although communication anxiety was underpinned by such personality variables as introversion and self-esteem, "[that] by no means exhausts the range of personality variables" (p. 139). Clément et al. (2003) examined the effects of adding ethnic identity and subjective L2 norms to the MacIntyre and Charos model, with results indicating that both played statistically significant roles (although path coefficients were not strong). Similarly, in commenting on the results of her 2002 study, Yashima noted that “factors outside language competence that were not included in the study (such as gender, personality, and communication tendency in the L1) might influence L2 communication confidence” (p. 62). While investigating the effects of gender and L1 communication tendency are beyond the scope of this study, alternative models in SEM are possible and might fit the data in question equally well. The current study thus focuses on (a) L2 Communication Confidence and possible alternative configurations with perceived distance, ego permeability, and extroversion, and (b) the role of the L2 Communicative Confidence variable in the three models of L2 communication.

Improvements to SLA Communication Models Through the Addition of Personality Variables

With MacIntyre’s (1994) suggestion and Yashima’s admonition in mind, I suggest that the MacIntyre and Charos (1996) model, the Yashima (2002) model, and the Yashima et al. (2004) model could be improved by the addition of personality variables. In the pyramid configuration (Figure 4), personality is included in the bottom tier, and its role in academic achievement is well established (e.g., Furnham, Chamorro-Premuzic, & McDougall, 2002; Lounsbury, Sundstrom, Loveland, & Gibson, 2003; Rothstein, Paunonem, Rush, & King, 1994). The above studies, however, dealt with L1 contexts, and the world of L2 and FL acquisition is somewhat different.

Personality

In L2 and FL contexts, the place of personality has been a matter of debate, with the results of previous research having been inconclusive. In early research on personality and language acquisition, Dunkel (1947) and Kawczynski (1951) were limited by their use of simple theoretical frameworks. However, the work of Gardner and colleagues heralded increasing interest in personality in SLA (e.g., Gardner & Lambert, 1972), which has continued to the present day. Interestingly, however, in 1984 Lalonde and Gardner reported that personality played no direct role in FL achievement. The authors investigated FL achievement in a general sense (i.e., not focusing on oral skills), but the performance-oriented nature of oral communication suggests that personality plays a direct role. Moreover, Onwuegbuzie, Bailey, and Daley (1999) suggested that shortcomings in the "situation-specific focus in some of the measures" might have made the Gardner and Lalonde conclusion premature (p. 5). This means that the measures did not adequately account for the role of context, which is an important aspect of any communication model. As noted above, early research on affective variables was inconclusive, a situation that has been remedied for the most part with more refined instruments and more extensive research. In much the same way, the inconclusive results concerning the role of personality in SLA necessitate further examination.

Conceptually, personality influences all aspects of the pyramid, but my assertion is that its role is stronger than just that of an underlying component. In the current study I group three variables under personality variables: extroversion, ego permeability, and perceived distance.

Personality, however, is a nebulous term that is generally understood to broadly mean one's character. In psychology, efforts to establish a taxonomy of personality have concluded that personality includes a number of dimensions, yet the exact nature of that paradigm (or paradigms) has been the focus of a lively debate, with various numbers of dimensions posited (Eysenck, 1991). However, several researchers (Costa & McCrae, 1992; Digman, 1989, 1990; Peabody & Goldberg, 1987) have settled on five dimensions, which have come to be known as the "Big 5" (Goldberg, 1992, 1993). These five dimensions are introversion-extroversion, agreeableness, conscientiousness (or diligence), emotional stability, and intellect or sophistication. Hereafter, the introversion-extroversion label is shortened to simply *extroversion*; this does not imply any difference in orientation or meaning, only an attempt to be slightly more parsimonious. The intellect or sophistication dimension is also known as *openness to experience*, the label that is used in this study.

In the current study, the Big 5 structure of personality is used as an umbrella term to indicate the character (or the *multiple* characters) of L2 learners. While personality seems to be comprised of these five factors, the focus of the current study is to explore the respective roles of those five constituents rather than the role of the overall personality construct. In the MacIntyre and Charos (1996) model, the five factors were explicitly used in the path analysis, and my additions to the Yashima et al. (2004) model include the extroversion subscale. Extroversion is the one dimension that has consistently shown statistical significance and at times surprising relationships with various measures of L2 and FL acquisition (see Dewaele & Furnham, 1999). Moreover, it has occupied a prominent and enduring place in personality configurations with various numbers of dimensions (Eysenck, 1991), and as Digman (1990) noted, “There is general agreement that Dimension I is Eysenck’s (1947) extraversion-introversion” (p. 422).

An important part of the continued development of these communicative models has been the concomitant growth of tools for investigating and quantifying the various parts of those models. Early efforts (in the 1970s) to investigate affective variables en route to a workable model met with mixed results (e.g., Chastain, 1975; Schumann, 1975). Later researchers (e.g., Young, 1991) have suggested that the theoretical underpinnings were not well developed, leading to results concerning affective variables that changed with the particular definition of the variable in question. Moreover, the instruments used in those studies were in a nascent stage, having undergone minimal usage and development.

By the 1980s, however, more reliable and valid instruments had begun to emerge. In the FL sphere, for example, the advent of the Foreign Language Classroom Anxiety Survey (FLCAS; Horwitz, Horwitz, & Cope, 1986) and the anxiety and motivation scales used in Gardner’s work (e.g., Gardner, 1985) marked the advent of increasing interest in and research on affective variables in L2 acquisition. In addition, the underlying conceptualizations were maturing, as in Gardner's (1985) socioeducational model and Clement's social context model, which were followed in the 1990s by numerous updates after the "reopening of the research agenda" (Dörnyei, 1994).

The following section begins with the notion of the human being as an actor, after which distance and ego permeability are explored. Thereafter, instances of distance in early conceptualizations of L2 communication are introduced.

The Human Actor and the Self

Acting is as universal as breathing, for, in Kipper's (1996) words, "Human beings are born actors" (p. 99) whose habitat is perfectly synchronized with that orientation—after all, as Shakespeare's Jaques succinctly noted, "All the world's a stage." From the moment of birth on, each person presents or projects an image (consciously or not) that reflects the person, his or her wishes, desires, and needs. Such an image comprises several components that include both linguistic and non-linguistic elements. Among the former would be such elements as appropriate use of extant registers in the language. In Japanese, for example, several levels of politeness are available and indeed necessary to function; failure to use polite Japanese (*keigo*) can result in social gaffes as well as miscommunication (Carroll, 2005).

On the other hand, non-verbal aspects of communication play important roles, too, perhaps more substantially so than the linguistic aspects: Birdwhistell (1970) suggested that “30-35% of the social meaning of a conversation or an interaction is carried by the words” (p. 158), and Berkowitz (1982) put the linguistic portion at a minuscule 7%. Such numbers point toward the fact that communication is not solely comprised of linguistic elements, yet in spite of the juxtaposition of both parts, over time people become increasingly sophisticated in using both verbal and non-verbal repertoires of their first language (L1) and culture (C1).

In the world of drama, this fusion of parts looms large as actors and actresses assume roles in as authentic a fashion as possible. For much of the last two centuries, acting has centered on two quite different approaches, those of Bertold Brecht and Constantin Stanislavski (Cohen, 2004). Brecht espoused the technique of *verfremdung* [artful detachment, estrangement, or even alienation], by which actors and actresses consciously separate from their roles, analyzing and performing with a palpable distance between themselves and the role. In the process of maturation, children learn this, too, as they gain the ability to think about their activities in a meta-awareness fashion instead of simply engaging in those activities. In SLA, however, the order is reversed as L2 learners are naturally distant from their L2. Over time they might draw closer to the L2 and C2 as proficiency increases, and eventually that detachment might diminish or even disappear.

The opposite pole is Stanislavski's *perezhivanie,* in which the actor comes to experience or actually live the part in a deep, emotionally invested manner. This corresponds to the status of small children less than three years of age who lack the ability to "step back" and consider what they do. As such, they are completely invested in and "living" the moment. As a child approaches five years of age, however, he or she begins developing meta-awareness, the ability to separate from and consider the activity, which approaches Brecht's *verfremdung*.

The world of drama or roleplaying does not exist independently from the actual world, as Bowman (2004) notes: "If the world is itself an elaborate game, and each of us struggle [*sic*] throughout our lives to learn the rules and find ways to succeed, roleplaying can be viewed as a microcosm of that process" (p. 13). Both worlds center on how people present themselves socially, which is clearly enunciated in the theory of self-presentation outlined below.

Self-Presentation Theory

In terms of self-presentation theory (Baumeister & Leary, 1995; Leary, 2004; Markus & Wurf, 1987; Tedeschi, 1981), each person uses various techniques (e.g., linguistic and kinesiological) to present a particular image to others, or, in Leary's words, "to manage a public identity" (2004, p. 205). This echoes a basic assumption of social identity theory (Tajfel, 1982), in which each individual possesses a repertoire of identities. The use of a particular identity depends on the social context, and the resultant social behavior varies from purely interpersonal to completely intergroup. That image—the social behavior—can be a consciously assembled entity in the Brechtian mold or simply the one projected while functioning socially, more in the tradition of Stanislavski.

In representing one's self, however, intervening factors can pose obstacles to appropriate self-representation. Such impediments include physiological factors such as fatigue, which in Japan is a pervasive concern as L2 learners face test-oriented education (Gorsuch, 2000) and the added burden of attending extracurricular schools (the ubiquitous cram schools, which include *yobiko* and *juku*). Social factors are also important as learners must contend with expectations of successful language acquisition set against a pervasive attitude that Japanese are poor L2 learners. Moreover, L2 learners also struggle with affective variables such as anxiety, of which foreign language anxiety is a recognized and significant obstacle for many learners (Horwitz, Horwitz, & Cope, 1986; MacIntyre & Gardner, 1991; Young, 1991). In addition, linguistic factors such as limited proficiency (e.g., an undesirable accent or limited eloquence) can limit self-presentation. Indeed, a Pandora's box of factors springs forth.

Moreover, to this daily task of self-presentation comes the additional burden of doing so in a second language (L2) and culture (C2). As proficient as speakers become in their first language (L1) and culture (C1), new languages and cultures confound the issue: An educated, literate person's ability to function often is reduced in an L2/C2 environment, and this poses an immense threat to identity for many people. The "vulnerable, inhibited central self that fears making mistakes"

(Heath, 1993, p. 673) enters a milieu in which, lacking competence, proficiency, or both, mistakes are inevitable and the carefully crafted and polished L1 and C1 image of one's self comes under siege (e.g., Kanno, 2003). The resolution of this conflict often takes considerable time, as Kanno's four informants noted. In addition, the struggle to deal with identity threats can lead to volitional choices to shield that vulnerable self (Koole, 2004).

Goffman (1959) noted that a "real life performer can utilize actions which convey some disdainful detachment … from the role he [*sic*] is performing" (p. 110). In interactions among different cultures and languages, this is common to mark one's station as "non-L2/C2" by, for example, the retention of a foreign accent even after many years of immersion in the L2/C2 (Jones, 2001). In presenting oneself, one constant is the need and desire to maintain one's own value or sense of worth.

Protecting the "Vulnerable Self"

In psychology, self-protection is viewed as imperative for students' well-being. In Covington's (1992) self-worth theory of motivation, the necessity of protecting one's self-worth is rooted in the fear of failure and its implications for one's sense of ability and subsequent self-worth. One strategy to address this fear is self-handicapping (Deppe & Harackiewicz, 1996), in which failure is deflected or attributed onto something other than one's own competence: Inadequate preparation, for example, provides a ready excuse for poor performance (Tice &

Baumeister, 1990). A second method with negative long-term consequences is defensive pessimism (Norem & Cantor, 1990), in which one sets unrealistically low expectations and reflects on various possible outcomes. Both these strategies, while offering protection, do not enhance progress and might impede it.

All is not lost, however, for there might be ways around such vulnerability. For education in general, Thompson (1994) pointed to the effective use of praise that is informational rather than directive of future performance, the importance of minimizing uncertainty and situations of evaluative threat that give rise to self-worth protection, and attribution retraining programs that encourage students to assume due credit for their successes. These means can help reduce self-worth protective behaviors in the classroom that are counterproductive.

In the L2 world, the recognition of methods to reduce classroom anxiety (Young, 1991) and increased awareness of social factors in SLA have helped make the SLA journey a bit smoother. Some of the methods to reduce anxiety include talking directly about anxiety and learners' sources thereof. Elaine Horwitz (1988) suggested that instructors "discuss with their students reasonable commitments for successful language learning and the value of some language ability if it is less than fluent" (p. 286).

The realistic assessment of their own English ability is important for L2 learners (Bandura, 1997; Pintrich, 2003; Schlenker & Trudeau, 1990). Overestimating one's ability, for example, can lead learners to devote insufficient time and effort to acquiring further proficiency (MacIntyre, 1994), while underestimating can result in learners viewing L2 acquisition as an insurmountable mountain, thus reducing motivation, a crucial shortcoming as motivation is an important factor in second language acquisition (Dörnyei, 2003). Other methods include the use of pair work and group work instead of individual tasks. Another possibility is to make the message optimally interesting by, for example, using games or unusual techniques. These all might help reduce anxiety, which should in turn facilitate L2 acquisition.

Another approach that can be used to address vulnerability is using the buffer of distance. Explored in such fields as psychology (e.g., Thompson, 1994), but as yet given little attention in the ESL/EFL field, distance likely plays important roles for L2 learners. Although not a panacea for the obstacles that EFL students face, distance-inducing activities might nonetheless prove useful in ESL/EFL education for some and perhaps many L2 learners.

Distance

What exactly is distance? In its most tangible sense, it refers to a physically measurable distance like a meter or a yard. A second sense of crucial importance in SL/FL acquisition is social distance, which invokes the world of groups and social dynamics (MacIntyre et al., 1998). Third is psychological distance, something that is perceived intrapsychically and is less dependent on social factors. Goffman (1959, 1961) noted that distancing of this type is a skill that small children acquire as they learn to inject a *cordon sanitaire* into their experiences, such that the signifier (e.g., the description or perception of riding a merry-go-round horse) separates or becomes distant from the signified, the actual act of riding a merry-go-round horse. This perceived distance contrasts with social distance, which Schumann (1975) used in the sense of spatially measurable distance (an etic sense). Brown (1980) then reformulated social distance as being an internally perceived distance (an emic usage): Social distance "refers to the cognitive and affective proximity of two cultures which come into contact *within an individual*" (p. 159; italics added). For the purpose of the present study, I follow Brown's formulation, treating distance as the learner's internal perception of the cognitive and affective proximity of two cultures.

Physical (i.e., spatial) distance is of relatively minor importance as the learner's perceived psychological distance is paramount. However, psychological distance is not a fixed entity, as it can change depending on the context. In a similar way, temporal distance is also subject to intrapsychic alteration. Ross and Wilson (2002) found that people often perceive unpleasant or unsuccessful incidents as more distant in time than was actually the case, whereas pleasant or successful things are often moved temporally closer to the present. Moreover, the fading affect bias (Ritchie, Skowronski, Wood, Walker, Vogl, & Gibbons, 2006) concerns how negative feelings associated with an event fade from memory faster than positive emotions, a phenomenon that likely underpins the "rose-colored glasses" phenomenon.

History of Distance in Psychology

In the field of psychology, distance has a long history, particularly in the form of roleplay (Kipper, 1996). Biddle and Thomas (1966) charted the origins of role theory in its modern connotation from the 1930s, but they noted that the term is much older, dating from antiquity in the sense of paper roll sheets or lists. Since that time various therapeutic methods have appeared using dramaturgical roles (Jones, 1996), including various forms of sensitivity training (Perls, 1969; Siroka, Siroka, & Schloss, 1971). In various guises it has been used as an intervention technique, generally for instances of pathology and in health care (Emunah, 1994; Lahad, 1999; Marsella, Johnson, & LaBore, 2003). Another use is in grief therapy, where children assuming drama roles in therapy were better able to successfully reach a resolution than children who had not done so (Curtis, 1999). A more recent use is in social and health education, especially concerning bullying issues: Characters in virtual environments were effective in increasing empathetic engagement in young children (Hall, Woods, Aylett, Newall, & Paiva, 2005).

A second form of roleplaying in psychology is training in group dynamics (Sogunro, 2004; van Ments, 1999). This includes areas such as leadership issues, group management, intragroup conflict, cooperation, and the formation of accurate perceptions of self and others within the group. The primary focus is self-improvement in such areas as business (Sogunro, 2004), military (Bowman, 2007; Dovey & Kennedy, 2006), and education (van Ments, 1999). Sensitivity training has been used in a similar way in SLA as a retraining strategy to overcome affective variables that inhibit L2 acquisition (e.g., Foss & Reitzel, 1988).

Acting and Distance

Rather less subtle is the approach in acting, in which actors or actresses consciously seek to empathize with a character to present a meaningful, convincing performance. Cohen (2004) noted that two schools dominate the acting profession: Adherents of the Stanislavski style hold that the dramatist must actually assume the role, while followers of the Brecht tradition espouse the importance of identifying and empathizing with the role and then recognizing a detachment or distance from the role to critically examine it (Rouse, 1989).

If used in a similar fashion in language teaching, such empathy, identification, and awareness of distancing in drama might help create the “right concatenation of natural psychological factors” (Stern, 1980) that should facilitate SLA. In ESL, drama has proven useful as a powerful motivating factor for inner-city L2 learners (Heath, 1993), and Makita-Discekici (1999) found drama to be effective in her FL classes. For language teachers in general, the workings of drama in various contexts were addressed in Stern’s (1980, 1983, 1993) studies.

All of these various scenarios, different as they are, rely on the essence of acting, which is to empathize with a character (another *self*) and present a believable facsimile to an audience, which could be in a contrived situation such as on stage or in an everyday social interaction.

Possible Selves

Hazel Markus and colleagues (Markus & Kitayama, 1991; Markus & Nurius, 1986; Markus & Wurf, 1987) explored how one element of self-knowledge is that of “possible selves.” These include not only “ideal selves we would very much like to become [but] also the selves we could become, and the selves we are afraid of becoming” (1986, p. 954). On a conceptual level, the use of distancing can allow L2 learners to accomplish several “self-related” tasks: explore possible selves, add additional possible selves, and establish distance from undesirable selves. As Ehrman (1999) noted,

> Individuals can have a variety of sub-personalities that are related to different roles they play. Most have some amount of consistency across roles and a set of stable ‘selves’ based on firm beliefs, attitudes and values. However, in certain social situations, they might well undertake sharply differing approaches and have a variety of transient ‘selves’ or repertoire of social identities. (p. 70)

Of particular importance is the idea that possible selves can be very liberating (Markus & Nurius, 1986). Although external factors (e.g., peer pressure and societal norms) can exert an influence, the contents of possible selves are constructed internally; not being openly available for scrutiny by others, these selves are thus psychologically safe. Rather like many religious doctrines, they cannot be disproved: Only the individual can ultimately ascertain what is possible, probable, or challenging. Because the self is mutable, such possible selves can free the person because the current self is not “set in concrete,” instead having the possibility of change. In other words, the locus of control is ultimately internal and thus as malleable as the person's mind can be. In the ESL/EFL sphere, this malleability points to the possibility that unsuccessful learners can construct such scenarios as ones in which they are successful language learners or in which they have overcome difficulties such as debilitating anxiety.

Such possible selves are extant in both the learners' L1 and L2, and they are part and parcel of identities and Jung's (1969) anima. Considerable literature has been devoted to the development of identity in a second language (e.g., Kanno, 2003; Mantaro, 2006; Pavlenko, 2002; Pavlenko & Blackledge, 2004; Pennycook, 1998; Vandrick, 1998), and one important conclusion is that the development of an L2 identity is a necessary and simultaneous step in acquiring a second language (Norton & Toohey, 2002).

Whereas the development of an L2 identity is certainly internally centered, external situations can affect the distance that learners and their L2 personae perceive from their own selves. In many L2 classrooms, such activities as drama, puppetry, pantomime, and roleplay require that students assume roles outside of their normal persona, what Bowman (2007) termed "identity alteration." Although the various media allow different degrees of freedom, in all such cases, the learners become someone or something different to an extent from their own, core persona (Jung's anima).

In drama, for example, an actor assumes a stage role in which the context and actual dialogue are decided (imagine, for example, the roles of Romeo and

Juliet). Each person brings a personal touch to the role through gestures and intonation (the Brechtian tradition), but the basic role itself is predetermined. In roleplay, however, the actor can also have the freedom to create original dialogue when given only the general situation. In mime the non-verbal elements assume the lead, and the non-spoken ‘dialogue’ can again be free or predetermined. Finally, in puppetry, dialogue and other elements can be free or predetermined, but puppetry often adds more distance in the form of non-human personae and in tangible objects that replace the actor’s human form. Here one can imagine a talking bear with human characteristics, which is simultaneously human-like in its ability to speak in human languages yet bear-like in its personality and needs.

The ability to act, to assume different roles easily or successfully, is not a universal ability among adults, and neither is L2 acquisition universal. In the work of Madeline Ehrman and Rebecca Oxford (e.g., Ehrman, 1993, 1996, 1999; Ehrman & Oxford, 1990, 1996), L2 learners of differing abilities faced different obstacles to L2 acquisition and preferred to deal with them in different ways. Very analytical people prefer more structured classes and tasks, while less analytical students often do better with open-ended and more creative tasks. This parallels the age-dependent stages in Goffman’s (1959) treatise, in which the need for structure and the development of meta-awareness come with increasing age.

As Cohen (2004) noted, “[O]ne doesn’t easily reverse (or accelerate) one’s lifelong psychobiological process of maturation. … [S]kills needed to move freely along the role embracement/distance continuum [are] … not merely a theatrical technique but a survival tool" (p. 6). This movement is intrinsically related to the notion of ego boundaries, which Ehrman found to be an important part of the different abilities exhibited by her L2 learners.

Ego Boundaries

Originating as a psychoanalytic concept, ego boundaries received some attention in SLA literature in the 1970s in the work of Alexander Z. Guiora and associates (e.g., Guiora & Acton, 1979; Guiora, Beit-Hallahmi, Brannon, Dull, & Scovel, 1972; Guiora, Brannon, & Dull, 1972; Guiora, Paluszney, Beit-Hallahmi, Catford, Cooley, & Dull, 1975) and John H. Schumann (1975, 1976). Much of Guiora's work concerned empathy and an individual's ability to pronounce a second language, and from this he went on to posit a theoretical model in which one's empathetic capacity equates with the permeability of ego boundaries (Guiora, 1972). While noting several shortcomings in Guiora's work, Schumann (1976) commented on the intuitive appeal of flexible ego boundaries, suggesting that "empathic capacity or ego flexibility, particularly as operationalized under the concept of "lowering of inhibitions," is best regarded as an essential factor in the overall ability to acquire a second language rather than simply in the ability to acquire an authentic pronunciation" (p. 226). However, with the notable exception of the work of Madeline Ehrman and colleagues, ego permeability has received little attention in subsequent years. At this point, I explore ego permeability in more detail.

Ego permeability begins with having a “well-defined, secure, integrated ego or sense of self” (Guiora & Acton, 1979, p. 199). In Ehrman’s (1999) words, “[B]y ego is meant a system of mental operations, cognitive and affective, that constitute an individual’s sense of self” (p. 69). The notion was originally used in explaining a band of psychopathological phenomena such as schizophrenia and bipolar disorders. However, current interpretations have challenged that conceptualization, positing it as a normal albeit neglected facet of personality (e.g., Stephen, 2004). Given that this notion centers on an ego (i.e., a system), boundaries are relatively unambiguous and become more clearly defined and perhaps less permeable with maturation. In psychology, more permeable boundaries correlate with a larger number of memories and increased attention to emotional events and subjective meanings (Stephen, 2004), which might correspond to the attention to linguistic and non-linguistic factors necessary to function in a second or a foreign language.

Regardless of age, a high level of ego permeability relates to fluidity of mental categories, including those concerned with one’s identity, with social relations, and with different ways of perceiving other cultures and languages (Ehrman, 1999). In the Ego Permeability instrument, this is measured by five subscales: Unusual Experiences (Item 1, *In my dreams, people sometimes merge into each other or become other people*), Need for Order (Item 13, *There is a place for everything and everything should be in its place*) Perceived Competence (Item 32, *I keep my desk and worktable neat and well organized*), Childlikeness (Item 40, *I think a good teacher must remain in part a child*), and Sensitivity (Item 46, *I am a very sensitive person*).[3]

Such fluidity is not just a passive awareness, but rather an ability to move back and forth between languages and personalities. Moreover, it might be possible to intervene to facilitate that movement: In two early studies (Guiora & Acton, 1979; Guiora, Brannon, & Dull, 1972) participants showed improved pronunciation after psychopharmacological intervention with either alcohol or valium. In both studies, the researchers suggested that improved linguistic performance resulted from lowered inhibitions, which might be analogous to more permeable ego boundaries.

L2 learners with thicker ego boundaries, who were more inclined to learn in an organized, analytical fashion, were less able to distance and found themselves somewhat handicapped in activities such as roleplays, which called for flexibility to assume new roles. On the other hand, L2 learners with thinner ego boundaries, who were more amiable to ambiguity and flexibility, more readily assumed new roles, yet they experienced problems when analysis and organization were more appropriate (e.g., when addressing grammar questions explicitly).

If these external roles are helpful in overcoming negative affective variables such as anxiety, then they are potentially useful in L2 education. For example, when students are asked to perform a roleplay, they must assume the persona of a

[3] The notion of ego permeability is similar to that of self-concept flexibility. Choi and Choi (2002) found that East Asian individuals tend to hold inconsistent and at times contradictory beliefs simultaneously, even at the most fundamental level of one's self-concept. This flexibility contrasts with the Hegelian dialectic of thesis-antithesis-synthesis in which one of a pair of inconsistent or contradictory views tends to be rejected.

character that is not in their present situation. Imagine, for example, an L2 learner being asked to assume the persona of an irate neighbor or a dog owner in order to address the problem of a noisy barking dog that disturbs neighbors in the middle of the night (which is a situation encountered outside the students' language classroom context).

This example is not so distant from L2 learners' reality because such roles and situations can occur. In drama, however, learners might have to assume a very unlikely or even impossible role—in playing the role of Romeo, for example, the learners cannot return to the past. While being lovestruck is a condition common to all eras, a lovestruck young man of Shakespeare's time likely thought and acted in different ways than now, for in modern life, "children of the scientific age" (Goffman, 1961, p. 204) are equipped with far more ways of thinking and acting (which might, in the words of Markus and Nurius (1986), constitute more possible selves). In a similar way, learners doing puppetry might have to assume a personality that might be highly unlikely (e.g., becoming a king) or not possible (becoming a talking bear).

In entering an unlikely or impossible role or situation, learners might perceive distance from their own personality. One possible outcome is that the individual's L1 personality as well as various obstacles to L2 acquisition recede and play diminished roles. If, for example, assuming the persona of a talking bear and thus assuming some distance allows learners to forget about their L2 difficulties or “helps people to maintain self-regard” (Ross & Wilson, 2002, p. 801), then this technique might facilitate SLA.

In that same context, however, the risks in becoming someone else might inhibit some learners, for whom becoming a talking bear is uncomfortable. This might correspond with less permeable ego boundaries, in which moving outside of one’s normal persona is difficult. However, the willingness to assume a different persona might or might not correspond with a person’s degree of loquacity—indeed, the cover might not match the contents of the book/person, a situation analogous to one memorable character created by Robert Louis Stevenson.

Dr. Jekyll and Mr. Hyde

In the course of assuming a new identity, an interesting phenomenon sometimes occurs among L2 learners. In my work with puppets, I often find a Jekyll-and-Hyde transformation in which normally shy L2 learners suddenly become quite loquacious with a puppet in hand while some extraverts become very quiet. While this transformation is not a universal occurrence, I have consistently observed it in my classes and workshops, and it certainly begs the question of why. One possibility is that some combination of personality factors reaches critical mass: Imagine, for example, that an introverted student with high ego permeability is quite amenable to distancing, in which case the opportunity to jump to another persona (and therefore be free of inhibitions) suddenly yields a talkative alter-ego. On the other hand, a student who is normally talkative but who has low ego permeability might be ill at ease and thus less talkative when asked or forced to move substantially outside his or her regular personality.

As noted above, various factors can mediate L2 learners' progress toward acquiring proficiency in the target language. Ego permeability likely plays a role, and the notion of distancing also is important. In the following section I attempt to locate distance in the context of models of SLA.

Education and Distance

L1 Education. Although long encouraged in first language education, the theoretical underpinnings of distance have been less rigorously investigated in that field than in psychology. Perhaps one reason for this is a fundamental difference in perception of what drama is or can be in education. Drama can be something taught for its own sake, to create skilled actors. A second view, widely endorsed in education, is that drama and other distancing techniques such as roleplay can be learning tools useful for teaching other skills or material. In the holistic development of elementary students, drama (Anderson, 2004; Stern, 1980, 1993) and fantasy (Bettelheim, 1975; Bowman, 2007; Combs, 1988) are important, for through such activities learners can vicariously experience and thus address various situations with no risk of detrimental consequences. The vicarious aspect has recently assumed a larger role through computers, which allow learners to interact in cyberspace, a medium absent physical risk (Sutton, 2001).

Adults certainly engage in distancing and psychological roleplay. This can be linked to narrative (i.e., storytelling), which humans use to understand the past and present as well as prepare for the future. Roleplaying then furthers this by allowing active engagement with the developing narrative (Bowman, 2007). In addition, people design and run mental simulations by constructing various scenarios about future events (Taylor, Pham, Rivkin, & Armor, 1998). The difficulty or potential realization of these scenarios varies widely, but they play important roles:

> What matters is not the ease with which these possibilities can be simulated, or their actual potential for being realized. What is important is that they exist as enduring elements that can be activated as part of a working self-concept and that can function as referents or standards by which the now self is evaluated and interpreted. (Markus & Nurius, 1986, p. 963)

Much like computer files sprinkled throughout today's expansive hard drives, it is possible to retrieve and use these scenarios later, as the present self or current situation requires. In EFL education, if students or teachers can activate these possibilities, then perhaps the resultant flexible self-concept can accept new versions of one's self.

L2 Education. Activities that utilize distance have had an even shakier footing in L2 education. Based on a common perception that they are both useful and effective (Spolin, 1986), drama and roleplay are recommended (MEXT, 2003) and used, yet systematic investigation of their use is lacking. While one might argue convincingly for the inclusion of distance-inducing activities in L2 education, in many cases the study is only illustrative, a post hoc look at a program that used distancing successfully (e.g., Heath, 1993; Makita-Discekici, 1999). Although convincing as an anecdotal account and probably successful in calling attention to a particular technique, empirical rigor has often been minimal.

A concern here is the extent that learning English by proxy transfers to real-world situations. As evidenced by many textbooks, how-to books, and guidelines from the Japanese Ministry of Education Sports, Science, Culture, and Technology (MEXT, 2003) the dominant paradigm is that classroom study transfers to real-life situations (Ratey, 2002). Of course, the extent of transfer depends on a number of factors, including the similarity of the situations, the strategies invoked, physiological factors, and individual differences. These are valid concerns and some evidence exists that domain-specific knowledge transfers poorly if the area of knowledge and the issue at hand are substantially different (Kimball & Holyoak, 2000). For so-called "adaptive experts" this is less of a problem, as they possess the flexibility to engage knowledge in novel tasks both within and outside their particular areas of expertise (Hatano & Inagaki, 1986). However, for both L1 and L2 students and in all subject areas, this seems at most a minor point: Education in general involves transfer from a learning context, which is generally a classroom or a practice venue (e.g., a laboratory), to any number of real-world situations.

A second consideration is flexibility, in which a large variety of experiences enhances transfer (Kimball & Holyoak, 2000). In addition, representativeness and variability of training examples helps optimize transfer from training to execution. In the case of English by proxy (e.g., roleplay, drama, or puppetry), learners experience a variety of roles, voices, and situations, all of which should prove helpful in transferring practiced skills to the inevitable variety of real-world situations with myriad roles, voices, and situations.

In spite of this somewhat slippery background, there might be an optimal distance at which "successful language learners see themselves as maintaining some distance between themselves and both cultures" (Brown, 1980, p. 161). Similar to the Brechtian technique in drama (Cohen, 2004), optimal distance refers to the extent to which L2 learners feel separation either from the target language and culture or from their own self. Both types of separation are important, yet the directions are different: In the former, L2 learners perceive distance between the target language and culture and their own language and culture. This is the default starting mode for FL learners as they embark on a language journey that takes them into distant lands and cultures, certainly figuratively and perhaps also literally. Over time that perceived distance can decrease if they become more proficient in the L2 and C2, and some learners might even go as far as to immerse themselves in the target language and culture by studying or living abroad.

In the latter case, however, L2 learners might distance themselves from their own person or personality, which naturally developed in their native language. This scenario, however, is an intrapsychic one, located in the learner’s mind. This can include any activity in which learners are not themselves to some extent, and this certainly applies to speaking a foreign language: As Kazin (1951) put it, “To speak a foreign language is to depart from yourself” (p. 127). This self resembles that posited by Carl Jung (1969), who outlined two opposing parts of one’s personality: the *anima* (the core, central self) and the *persona* (the mask worn for the outside world). That outer mask is the same one addressed by self-presentation theory: For many that mask is also a flexible one, depending on the role one assumes in a given situation. Is the core anima, however, a fixed entity, or is it also flexible, developing, and changing over time? While the answer likely varies by person, Srivastava, John, Gosling, and Potter (2003) found that the factors in the five-factor model of personality changed in adults, but gender and a variety of developmental influences determined the person’s degree of flexibility. Anecdotally, Kazin’s words suggest that the anima in learners’ L1 and C1 does not necessarily correspond to the developing anima in their L2 and C2, which can become additional personalities or perhaps *selves*.

In L2 education there is a dearth of systematic investigation of potentially useful distance-inducing methods. From a pedagogical viewpoint, this area might offer a rich trove of methods that are viewed positively by many educators and learners and that might be useful in lowering the debilitating effects of affective variables such as anxiety and thereby facilitating SLA.

Given that the notion of distance might be a useful pedagogical tool, where does it fit in L2 communication models?

Distance in Early Models of SLA

Distance in SLA is not a new idea, as it was included in several early models of L2 communication. An important early development was the work of Wallace Lambert, Robert C. Gardner, and others in the 1950s and 1960s. This research pointed to the paramount roles of two factors in SLA, aptitude and motivation. Subsumed in motivation are integrative motivation and instrumental orientation, both of which are related to the concept of distance. Integrative orientation is more transparently related, for it addresses learners that seek to “meet and communicate with valued members of the target language community” (Schumann, 1975, pp. 214). Gardner also addressed how the social milieu plays a fundamental role. This social aspect includes such factors as attitudes toward the target language group, familial support, and language vitality.

Expanding on the tenets of Gardner’s work, in Schumann’s (1975) Acculturation Model a fundamental premise was that:

> …second language acquisition is just one aspect of acculturation and the degree to which a learner *acculturates* to the target language group will control the degree to which he acquires the second language. (italics added; p. 34)

Acculturation and SLA are negatively affected by the extent to which the learner experiences social and psychological distance from the target language culture; in other words, decreasing distance correlates with increased levels of acculturation and SLA. Social factors take precedence over the psychological ones although the latter play important roles when the social factors are not clearly positive or negative. Some of the social factors include the relative status of the two groups (equal or not), the need for assimilation, and the similarities between the two groups.

A second model that addressed the notion of distancing was Giles and Byrnes' (1982) Accommodation Model, in which psychological distance depends on perception and thus is dependent on the situation. Drawing from Gardner's work in Canada, a complex and officially bilingual milieu in which myriad L1 groups must cope with not one but two L2s (i.e., French and English), the Accommodation Model is concerned with intergroup dynamics. These are naturally not static as both the learner and the target group negotiate identity and roles. As had Gardner (1979), Giles concluded that motivation plays a primary role in SLA. This is related to how learners define themselves in ethnic terms and is moderated by such variables as learner identification with the target group, comparison of different groups, L1 vitality (Clément, Baker, & MacIntyre, 2003), perceived boundaries between groups, and identification with subgroups within the target L2 community (e.g., occupational or school groups; Holliday, 1999). Again, the notions of distance and boundaries play important roles as L2 learners attempt to cross boundaries and bridge differences between groups.

Following Schumann, Ellis (1985) offered a more discrete formulation by suggesting that psychological distance includes language shock, culture shock, motivation, and ego boundaries. Language shock is negative self-perception, in which “students feel they cannot function properly within the community since they have been deprived of their real personality and are embarrassed to display a self that is fundamentally incompetent” (Hilleson, 1996). Because incompetence is generally to be avoided, a self so represented is not desirable. Similar to language shock and perhaps subsuming it is culture shock, a reaction to a different culture in which a person has or perceives difficulty functioning in a competent manner.

Motivation is, as noted above, a crucial part of SLA, and it certainly underpins psychological distance. For example, given a learner strongly motivated to acquire the target language or to communicate with a person or group in the target language, the distance should be perceived as smaller or even non-existent when compared with a less motivated learner. In efforts to communicate, learners naturally have to assume a new persona, that of a person speaking in the target language. If learners’ ego boundaries are permeable to the extent that assuming the new persona is not threatening or anxiety inducing, then the perceived distance is again minimal or non-existent.

In conclusion, the earlier models included such variables as distance to an extent, but the three models investigated in this study do not. The following section addresses gaps in the literature, the purposes of the current study, the hypothesized relationships, and the research questions.

Gaps in the Literature

In the current study I address several gaps in the SLA literature. First, most of the instruments used in this study have not been validated using Rasch analysis. This represents a prudent step in SLA research because the psychometric behavior of instruments has generally received limited attention.

Second, the models of L2 WTC in the current study have not been investigated in this particular Japanese context. To the best of my knowledge, the MacIntyre and Charos study has not been investigated with any Japanese group, and the variations of the Yashima model have not been replicated.

Finally, the specific inclusion of the personality variables in the models is a new step and has therefore not been investigated previously. Both MacIntyre (1994) and Yashima (2002) suggested that the inclusion of additional factors might strengthen the respective models, and this recommendation also has not been investigated.

Purposes of the Study

In the current study I propose to refine the models of L2 communication and ascertain whether the addition of personality dimensions, ego permeability, and distancing enhance them. Whereas motivation, anxiety, and WTC have been researched extensively, other affective variables such as perceived distance, ego permeability, extroversion, and personality dimensions play important roles and should be explicitly investigated and included in L2 communication models. In this study I address these absences by positing the following relationships and investigating the subsequent questions.

Hypothesized Relationships

Within the configurations of the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004), the following relationships are hypothesized concerning extroversion, perceived distance, and ego permeability.

1. In the MacIntyre and Charos (1996) model, ego permeability directly influences perceived distance.
2. Perceived distance is influenced by English Experience, and it directly influences L2 communicative anxiety and L2 WTC.
3. In the Yashima models, extroversion, perceived distance, and ego permeability underpin the exogenous L2 Communicative Confidence variable (in addition to the original L2 anxiety and perceived L2 competence variables).

Based on these hypothesized relationships, the following research questions are addressed in this study.

Research Questions

1. To what extent are the instruments used in this study reliable and valid in the Japanese university EFL contexts in this study?

2. To what degree will the 2-factor structure of the L2 Communicative Confidence factor be replicated in this university EFL context?

3. How much will the additional personality variables enhance the L2 Communicative Confidence factor?

4. To what extent will the L2 communication models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) be replicated in this university EFL context?

5. To what degree will data-driven additions improve the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004)?

6. To what extent will the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) benefit from the addition of the personality variables of distancing, ego permeability, and extroversion?

Summary

In this chapter, literature related to the current study was discussed. The lineage of the Yashima models was described, beginning with the socioeducational model of Gardner and colleagues and the social context model of Clement; those models were then integrated into the L2 WTC model of MacIntyre (1994), which was broadened in the model of MacIntyre and Charos (1996). This was then adapted to FL contexts with the advent of international posture in the models of Yashima and colleagues (Yashima, 2002; Yashima et al., 2004).

The second half of the chapter included an overview of the concept of distance and the related concept of ego permeability: The former has received attention outside of the L2/FL field, but very little attention within the field. The concept of ego permeability, while researched in the L2 sphere in the work of Ehrman and colleagues, is related to perceived distance and is hypothesized to improve the models investigated in this study.

Finally, gaps in the literature were identified, the purposes of this study were described, and the working hypotheses and research questions necessary to replicate and extend the respective models of L2 communication were presented.

The methods used to investigate the research questions of this study are the topic of Chapter 3.

CHAPTER 3

METHODS

The methods used in the current study are described in this chapter. First, an overview of the participants is provided. Second, the procedures and instruments are covered in detail. Third, the places of instruments added to the respective models are explained, after which the two preliminary studies are introduced. Next, the Rasch procedure used for data analysis is explained. Finally, the basics of structural equation modeling (SEM) are outlined.

Participants

The participants were Japanese university students taking EFL courses at six universities in the Tokyo area. Most of the participants were first-year students (n = 175, 69.44%), with second-year students (n = 40, 15.87%), third-year students (n = 19, 7.54%), and fourth-year students (n = 10, 3.97%) comprising successively smaller groups; eight participants (3.17%) did not provide this information. Of the 252 participants, female participants (n = 145, 57.54%) outnumbered male participants (n = 102, 40.48%), and five (1.98%) were of unknown gender. As Russell (2002) pointed out, many psychological studies utilize rather small sample sizes, and with the structural equation modeling analyses a larger size was prudent.

The selection of the six universities was carried out using convenience sampling. A variety of university students was included in the study: Those at B University were highly motivated, conscientious non-language majors, while the C

University students were highly-motivated foreign language majors. The D University group was composed of economics majors with relatively low English motivation, and their proficiency was similar to the F University students, who were English literature majors. The G University group was studying in evening English classes and was made up of students at various proficiency levels, and the H university group consisted of false beginners with low proficiency and limited interest in English.

The respective number of hours of classroom English instruction is indicated in Table 1, but the extent of English in the participants' surroundings could also be of importance. Three universities (B, C, and H) are located in cities of some 250,000 people in the Tokyo metropolitan area, while D, F, and G are located in Tokyo itself. Moreover, B and H are located in a city developed specifically for scientific research, meaning that a large contingent of foreigners lives and works there and that English is seen in everyday life. The other suburban university, C, is primarily a foreign language university with the largest number of hours of English instruction per week among these six universities. At the three downtown universities, the students are presumably exposed to English to some extent, and in the three suburban universities the special composition of the local population suggests that English is similarly present. It is thus assumed that all the participants experience roughly the same incidental exposure to English in their respective communities.

Table 1 details the six sub-samples with their respective number of participants, hours of English study per week, majors, and proficiency levels.

Table 1
Participants' Majors and English Proficiency Levels

Source	*n*	Major	Proficiency	Hrs / week
B University	137	Various	Mixed	4.5
C University	21	Foreign languages	High	6
D University	42	Economics	Low-intermediate	4.5
F University	14	English	Low-intermediate	4.5
G University	30	Various	Mixed	4.5
H University	8	Media and Information Studies	Low	3
Total	252			

Note. University names are pseudonyms. Hrs / week is the number of hours of English instruction per week.

A demographic questionnaire (Appendices A and B[4]) asked about the participants' school year, major, gender, English test scores (e.g., TOEIC), age at the onset of their English education, and experiences in the following areas: traveling abroad, living abroad, studying abroad, homestay abroad, and studying in extracurricular schools in Japan. As indicated in Table 2, 50 (19.84%) of the participants had studied abroad, while 61 (24.21%) had done a homestay abroad. Thirty-two had lived abroad with a total of 14 different countries represented, 116 (43.61%) had experience traveling abroad, 76 (30.16%) had attended an English conversation school, and 181 (71.83%) had attended cram schools. Finally, the age at which students began studying English varied from two years of age to 13 with a mean of 11.47.

[4] When two appendices are noted, the first appendix is the questionnaire that the participants answered, while the second is an English translation thereof.

Table 2
Demographic Information: Overseas and English Experience

Experience	*n*	Range	Mean	*SD*
Study abroad				
Short-term[a]	30	3 – 30 days	21.97 days	8.58
Long-term[a]	20	60 days – 3.5 years	456.85 days	305.09
Total	50			
Homestay				
Short-term[a]	47	1 – 30 days	17.23 days	8.78
Long-term[a]	14	90 days – 2 years	372.29 days	221.28
Total	61			
Live abroad	32	.5 – 18 years	3.85 years	4.72
Travel abroad	116	3 – 150 days	16.33 days	19.26
English school	76	2 months – 9 years	2.45 years	2.09
Cram school	181	2 months – 13 years	3.41 years	2.84
English start	249	2 – 13 years	11.47 years	2.41

Note. [a] Short-term refers to a duration of 30 or fewer days, while long-term refers to a duration of greater than 30 days. For participants with homestay experience, five did not provide the duration.

Instruments

A series of instruments was used to gather data in this study. Many had been piloted, published, and utilized in different contexts, but a pilot study using all of the instruments was conducted prior to administering them to the main group of participants.

The selection of instruments reflected the need to balance depth with parsimony. Although several of the instruments consisted of only five or six items, this number of items can have acceptable convergent and predictive validity (Gardner & MacIntyre, 1993). Moreover, to keep administration time to a minimum, I attempted to assemble a package of instruments that would be sufficiently reliable and valid while consuming only one class period and some out-of-class time on the participants' part. In order of description, the instruments included Breadth of Vocabulary Knowledge, Listening Comprehension, L2 Willingness to Communicate, Frequency of L2 Communication, Perceived Competence in English (L2 Communicative Competence), Perceived Distance, L2 Communicative Anxiety, Communication Anxiety in English, Foreign Language Classroom Anxiety Survey, Motivation to Learn English, International Posture, Personality, the Introversion-Extroversion subscale of the Personality instrument, Ego Permeability, English Experience (which replaced Context), and Attitude Toward the Learning Situation.

English Proficiency

Breadth of vocabulary knowledge. Two proficiency measures were used in this study. The first was based on a 72-item productive vocabulary test (Nation, 2001; Nation & Laufer, 1999) with a cued fill-in-the-blank format. In each sentence the correct response is cued by the initial letters of the missing vocabulary item. An example item is: “The pirates buried the trea_____ on a desert island.” (correct answer: treasure). In the preliminary study the test included 18 items each from the 2,000-word level, 3,000-word level, 5,000-word level, and the University Word List, respectively. The 72 items were analyzed with WINSTEPS 3.63.0 (Linacre, 2006), and misfitting and redundant items were culled to reduce the total number of items to 10 items per level for a total of 40 items (Appendix C). For the Breadth of Vocabulary Knowledge instrument and all other measures (except the English Experience instrument), the WINSTEPS software package was used to convert responses to Rasch measures, which are described in greater detail in the analyses section.

Listening comprehension. The second instrument was a 15-minute listening section similar to those used on numerous university entrance exams in Japan. It included two sections, the first of which had four short dialogues (4-12 lines) between two native speakers of English, and the second of which was a longer passage (198 words); comprehension questions (*k* = 16) followed the passages and dialogue (Appendix D). The first dialogue in Section 1 was as follows:

Dialogue 1. (Train directions for a foreigner)

> A: Excuse me, but you seem to be wondering about something.
> B: Well, now that you ask, yes, I am. This is the first time I've ridden the subways here, and I really don't know how to get where I want to go.
> A: Which is…?
> B: This station called Kōrakuen—I want to see a Giants' game!
> A: OK. First take the Yamanote Line (it's a JR train line) to Ikebukuro and then take the Marunouchi Line (a subway line marked in red on the signs) to Kōrakuen.
> B: Thank you so much.

1. Why is Person A looking confused?
   He has been drinking.
   He is using the train system for the first time.
   He doesn't know which team to cheer for.
   He wants to walk to the stadium.

The dialogues and passage for the listening comprehension measure were recorded on MD by two native speakers of English; the MD was played in the respective research venues.

Individual Difference Variables

First, the seven individual difference variables employed in the original models are presented: Frequency of L2 Communication, L2 Willingness to Communicate (WTC), Perceived L2 Communicative Competence, Communication Anxiety in English, the Foreign Language Classroom Anxiety Survey, Motivation, and International Posture. Thereafter the variables added to the MacIntyre and Charos (1996) model are explained: English Experience, Intercultural Friendship Orientation (Integrativeness), and Attitudes toward the Learning Situation. Finally, the Perceived Distance, Personality (and Extroversion) and Ego Permeability instruments are presented.

Frequency of L2 Communication. Frequency of L2 (English) Communication is defined as the frequency at which an individual engages in speech acts in his or her second language (English). This can differ from the intent or willingness to do so, which might or might not culminate in an actual speech act, but the intent to communicate and frequency thereof likely show a high correlation (MacIntyre et al., 1998). In Yashima (2002), a 5-item scale was used, but two items dealt with pairwork in the classroom. Because the frequency of communication in classroom pairwork depends on the extent to which the teacher utilizes such activities and not on the learner’s volition, those two items were not included in the questionnaire, resulting in a series of three self-report items designed to investigate frequency of L2 communication. However, in the current study the entire 5-item scale is used for two reasons. First, even when done at the behest of the teacher, classroom pairwork is interaction in English. Second, when facing the task of classroom pairwork, the manner in which the student engages in that pairwork (e.g., enthusiastically or grudgingly) remains the province of the student.

Although Yashima et al. (2004) used 10-point Likert scales, percentages were used in the current study (Appendices E and F). For example, participants who would nearly always not participate in a particular situation might write 10%, whereas participants that felt they would very likely participate might respond with 90%. MacIntyre and Charos (1996) used items that were much more heavily focused on the surrounding Anglophone-Francophone community, which reflects Canada being an officially bilingual context. Inasmuch as that is of less importance in Japan, Yashima's items with their emphasis on the foreign language classroom were used. Examples include "I volunteered to answer or ask questions in class," and "I asked teachers questions or talked to them outside the class period."

L2 Willingness to Communicate. In this study, L2 willingness to communicate (L2 WTC) is defined as the intention to initiate communication given the opportunity. McCroskey's (1992) WTC scale, the PRCA-20 (Appendices G and H), was used to measure the participants' willingness to communicate in English. This instrument is made up of the four situations and three audience groups mentioned above; in addition to the 12 permutations in the original scale, a further eight filler items yield a total of 20 items. Two examples are, "I would be willing to present a talk to a group of friends," and "I would be willing to talk in a small group of acquaintances." In each of the 12 permutations, the participants indicated a percentage of how willing they would be to communicate in that particular situation.

Perceived Competence in English (L2 Communicative Competence). Perceived Competence in English is defined as how competent the participants felt that they would be when communicating in a given situation. This was operationalized using the Perceived Competence in English instrument (Appendices I and J; MacIntyre & Charos, 1996), which is based on McCroskey's WTC instrument, the PRCA-20 (McCroskey, 1992).[5] The participants' responses were gathered for 12 situations based on four communication situations (public speaking, speaking in a large meeting, speaking in a small group, and speaking in pairs) with three audiences (strangers, acquaintances, and friends). An example item is, "I feel competent speaking in a small group of strangers." In each of the 12 permutations, the participants indicated a percentage (0-100%) of how competent they would feel about communicating in that situation. Participants who felt competent in a particular situation might write 90%, while participants who did not feel competent might write 10%; thus, a higher percentage indicates a larger degree of perceived competence in English.

[5]The original instrument includes eight filler items that are not analyzed, hence the number 20 in the instrument name (12 situation permutations plus eight filler items).

Communication Anxiety in English. L2 communicative anxiety is defined as the extent to which participants feel anxiety when engaged in communicative activities in their second language, English. In this study two anxiety questionnaires were used to measure it; the first was the Communication Anxiety in English questionnaire, which was used in both the MacIntyre and Charos (1996) study and the Yashima et al. (2004) study. Dealing with anxiety engendered by various situations, it was included to allow replication of those studies. In addition, the Foreign Language Classroom Anxiety Scale (FLCAS; Horowitz et al., 1986), which deals more with learner-internal elements such as emotions, was included as it was hypothesized that the construct of foreign language anxiety consists of both situational anxiety and internal anxiety. Elwood (2005) found a moderate correlation between the two (r = .35), suggesting that they measure different facets of anxiety, situation-specific anxiety and situation-independent internal anxiety (i.e., a trait anxiety; MacIntyre, 2007; MacIntyre et al., 1998). Thus, results from the two instruments entered the structural equation models separately.

The first anxiety instrument was the 12-item Communication Anxiety in English questionnaire (Appendices K and L; MacIntyre & Charos, 1996), which was also used in Yashima et al.'s (2004) study. The participants indicated the percentage of time that they would feel anxious engaging in a particular activity; the anchors were 0% (I would never feel nervous) and 100% (I would always feel nervous). The instrument includes the same 12 permutations (four situations, three receiver groups) introduced in McCroskey's (1992) PRCA-20; an example item is, “I would be anxious about speaking to a stranger while waiting in line.” A higher score thus indicates a higher degree of communicative anxiety in English.

Foreign language classroom anxiety survey. The Foreign Language Classroom Anxiety Survey (FLCAS; Horwitz et al., 1986) was the second anxiety scale used (Appendices M and N). Sample FLCAS items include, “I never feel quite sure of myself when I am speaking in my English class,” and “I worry about the consequences of failing my English class.” In the current study, a 7-point Likert scale was used; the scale was anchored by “This does not describe me at all” (1) and “This describes me very well” (7). Seven items were reverse coded so that a higher score indicates a higher degree of foreign language classroom anxiety.

Motivation. Motivation was defined as a person’s desire to learn English and any activities that reflect that desire (Appendices O and P). In this study, motivation to learn English was operationalized with two 6-item scales, the Desire to Learn English subscale and the Motivational Intensity subscale (Gardner & Lambert, 1972; Appendix M) that were used in Yashima’s (2002) study. A sample item from the former is “I would like the number of English classes in school increased,” and from the latter, two representative items are “Compared to my classmates, I think I study English relatively hard,” and “After I graduate from college, I will continue to study English and try to improve.” In the current study, a 7-point Likert scale was used; the scale was anchored by “This does not describe me at all" (1) and "This describes me very well" (7). A higher score thus indicates a higher degree of motivation.

International Posture. In this study I utilize Yashima's (2002) definition of international posture, in which some learners "are more interested in or have more favorable attitudes toward what English symbolizes than other learners" (p. 57). This orientation can thus include "interest in foreign or international affairs, willingness to go overseas to stay or work, readiness to interact with intercultural partners, and, one hopes, openness or a non-ethnocentric attitude toward different cultures, among others" (p. 57). Those four aspects underpin the four scales in the International Posture instrument (Appendices Q and R). Intergroup Approach-Avoidance Tendency (k = 7; Items Ipos1-Ipos7) includes "I want to make friends with international students studying in Japan," and "I try to avoid talking with foreigners if I can." Interest in International Vocations or Activities (k = 6; Items Ipos8-Ipos13) includes "I want to live in a foreign country," and "I want to work in an international organization such as the United Nations." Interest in International Affairs (k = 5) originally included only two items (Ipos14-Ipos15), "I often read and watch news about foreign countries," and "I am interested in international news." To strengthen this scale, three items were added in the pilot phase (Items Ipos16-Ipos18), for example, "International news is more important than local news." Finally, the Intercultural Friendship Orientation subscale (k = 8; Items Ipos19-Ipos26) includes "[A reason to study English is that] it will allow me to meet and converse with more and varied people." In the current study, a 7-point Likert scale was used; the scale was anchored by "This does not describe me at all" (1) and "This describes me very well" (7). Nine items were reverse coded so the valence of all 33 items matched; thus, a larger value indicated a higher level of international posture.

Additional Variables from MacIntyre & Charos (1996)

The MacIntyre and Charos (1996) model included Context, Attitudes toward the Learning Situation, and Integrativeness, which were changed for the current study. This section addresses the changes instigated to make variables appropriate for the contexts in the current study.

English Experience. The context under investigation was a foreign language context in which English is used almost exclusively in academic contexts, unlike the officially bilingual context of the MacIntyre and Charos (1996) study in which English is much more widely used. That study asked the extent to which English was used in the workplace and in the home, both of which are nearly moot points in Japan. For that reason, the Context variable was replaced with English Experience, a variable constructed for this study. English Experience is derived from the demographic information provided by informants and includes seven experiences in which English could have been encountered: study abroad experience, homestay experience, experience living overseas, overseas travel experience, experience in English conversation schools, the age at which English study was begun, and compulsory English education in secondary school. The length and richness of the experience constituted the score: Experience in an English-speaking country, for example, was scored higher than that in an ESL situation, which was in turn scored higher than experience in an EFL situation. A longer tenure was similarly scored more highly than a shorter time: Having lived abroad for three years or longer was scored higher than an interval of fewer than three years. The English Experience score was the sum of the respective categories and is detailed further in Table 3.

Table 3
Composition and Scoring Criteria of the English Experience Instrument

	Score			
Category	4	3	2	1
Live abroad	E, >3 yrs	E, ≤ 3 yrs		
		ESL, > 3 yrs	ESL, ≤ 3 yrs	
			EFL, > 3 yrs	EFL, ≤ 3 yrs
Study abroad			> 30 days	≤ 30 days
Homestay			> 30 days	≤ 30 days
Conversation school			> 3 yrs	≤ 3 yrs
Starting age			< 9 yrs	9 – 12 yrs
Travel				(yes) [a]
Compulsory education				everyone

Note. E = a country in which English is spoken as a first language; ESL = an ESL country; an EFL = EFL country; yrs = years. [a]Travel abroad was further subdivided into three categories: travel to an English L1 country was .5, travel to an ESL country was .25, and travel to an EFL country was just .1.

Intercultural Friendship Orientation (Integrativeness). Integrativeness in this study refers to “the desire to learn a L2 in order to meet and communicate with members of the L2 community" (Yashima, 2002, p. 56). Participants with a higher level of integrativeness should interact more with a L2 language group than those with a lower level. Integrativeness as conceptualized in the earlier sense of joining a target language community (Gardner & Lambert, 1972) is of minimal importance in this study's context because most of the informants grew up in Japan, were living in Japan at the time the data were gathered, and likely will always live in Japan. However, this is a necessary change to define the integrative orientation appropriate to this specific context (Clement & Kruidenier, 1983). In the current study the Intercultural Friendship Orientation subscale of the International Posture instrument is substituted for Integrativeness in the analyses of the original MacIntyre and Charos (1996) model and its revised form.

Attitudes Toward the Learning Situation. Finally, Attitudes toward the Learning Situation in the MacIntyre and Charos (1996) study included attitudes toward the teacher and the classroom, and the current study mirrored that orientation. Attitudes toward the Learning Situation was operationalized by grouping four items, two that deal with the classroom situation and two that concern interacting with teachers that are native speakers of English. The classroom items were Motivation Item 5 (*I believe absolutely English should be taught at school*) and FLCAS Item 5 (*It wouldn't bother me at all to take more English language classes*), and the native speaker items were FLCAS Item 14 (*I would not be nervous speaking English with native speakers*) and FLCAS Item 32

(*I would probably feel comfortable around native speakers of English*’). These four items were removed from their original scales. A higher value for attitude signified agreement with the four items, thus indicating a more positive attitude toward the English learning situation. The original scale in the MacIntyre and Charos (1996) included just two items, but the new 4-item instrument should better define the Attitudes toward the Learning Situation construct.

Personality Variables

Perceived Distance. In the present study, perceived distance is defined as the learner’s internal perception of the cognitive and affective proximity of two cultures and their respective languages. More specifically, this is the extent to which L2 learners feel separation either from the target language and culture or from their own self. Developed for this study, the Perceived Distance Questionnaire measured perceived distance by asking the participants to assign a percentage to the extent that they perceive distance from their own self while engaged in participatory exercises in which the entire class experienced a distance-inducing activity such as doing a roleplay (see Appendices S and T for the participants’ script and Appendices U and V for the researcher’s script). As the participants might have experienced roleplay in their junior high or high school English classes, this participatory approach involved all students in the same activity and thus tried to avoid students basing their responses on an experience from their past English courses or activities. The instrument measures perceived distance in six areas, beginning with casual conversation in the learners' L1 (Japanese) as a warm-up exercise. This was followed by five L2 (English) sections: casual conversation, formal public speaking (i.e., a speech), roleplay, drama, and puppetry.

The original wording asked about how much distance a speaker perceives while engaging in the various activities (in Japanese: Tsugi no katsudō wo suru baai ni, jibun no seikatsu kara dore gurai kyori wo kanjimasu ka?). However, after consulting with several native speakers of Japanese, the Japanese wording was changed to read, "How much do you change? (Dore gurai jibun jishin ga kawaru ka?) The native Japanese speakers felt that emphasizing the change would be more easily understood by the participants, and this was verified in the pilot stage. The scale was anchored by 0% change (My character doesn't change—I stay the same as usual [Seikaku ga kawaranai—futsū no jibun no mama]) and 100% change (I completely change and become like a different person [Seikaku ga kanzen ni kawaru—tanin ni naru]). In response to the question, "When doing [drama/puppetry/…], how much do you change?" participants answered with a percentage. For example, a participant who felt little change in perceived distance might write 10% while one who felt a great deal of change in perceived distance might write 90%. Omitting the "chatting in Japanese" item, the distancing measures were the Rasch person ability estimates for the five remaining Perceived Distance items.

The order of the activities reflected my expectation of increasing distance. The related study showed that mean values of the perceived distance for the respective activities increased in the expected order: English conversation, public speaking, roleplay, drama, and puppetry. In the preliminary study, the mean values of the first three were closely grouped, and the mean values for drama and puppetry were close but significantly different than those for the first three activities. As such, the preliminary study suggested that the above ordering corresponds with increasing perceived distance.

Personality. In the current study, personality is defined as a person's character and was operationalized with a shortened version (MacIntyre & Charos, 1996) of the Bipolar Scale of Global Personality Traits (Goldberg, 1992) to assess the Big 5 global personality traits (Appendices W and X). This scale consists of 35 pairs of adjectives, to which participants responded on a 7-point semantic differentiation scale anchored by "This is completely different than me" (1) and "This applies to me perfectly" (7). On all of the subscales, a higher score indicated a higher degree of the subscale focus (e.g., a high score on the first subscale would indicate the respondent has a high degree of extroversion). The current study investigated the composition of the five respective subscales; the five personality traits and sample pairs are as follows:

1. Extroversion (silent—talkative),
2. Agreeableness (cooperative—uncooperative),
3. Diligence (disorganized—organized),
4. Emotional stability (relaxed—tense), and

5. Openness to experience[6] (creative—uncreative).

The scale is termed transparent because Goldberg (1992) found that grouping similar items in a so-called transparent format yielded slightly better results than a format with items randomized. Moreover, although a reasonably short scale, Goldberg (1992) and MacIntyre and Charos (1996) maintained that this scale is an acceptable substitute for longer personality measures such as the NEO Personality Inventory-Revised (Psychological Assessment Resources, Inc., 2005), a 300-item questionnaire. In the interest of not burdening the participants more than necessary, a shorter scale was chosen.[7]

A 7-point Likert scale was used in the current study and was anchored by "This does not describe me at all" (1) and "This describes me very well" (7). Fifteen items were reverse coded so the valence would be the same for all items.

Ego permeability. Ego permeability is defined as fluidity of mental categories, including those concerned with one's identity, social relations, and different ways of perceiving other cultures and languages. This construct was operationalized using a shortened version (Rawlings, 2001) of the Hartmann Boundary Questionnaire (HBQ; Hartmann, 1991), a widely used measurement of ego permeability (Appendices Y and Z). The original HBQ was developed in connection with research on dreams (specifically, nightmares) and is designed to

---

[6] The fifth factor has at various times also been labeled Intellect, Sophistication, or Culture (MacIntyre et al., 1998).
[7] Two additional, freely available instruments (50-item and 100-item) for assessing the Big 5 factor markers are accessible at http://ipip.ori.org

measure the degree to which people separate aspects of their mental, interpersonal, and external experience through "thick" or "thin" psychological boundaries. The HBQ contains 146 items in 12 subscales: sleep, dreams, and wakefulness; unusual experiences; boundaries among thoughts, feelings, and moods; impressions of childhood, adolescence, and adulthood; interpersonal distance, openness, and closeness; physical and emotional sensitivity; preference for neatness; preference for clear lines; opinions about children, adolescents, and adults; opinions about lines of authority; opinions about boundaries among groups, peoples, and nations; opinions about abstract concepts; plus a total score covering all twelve of the subscales. Hartmann found that women and younger people consistently reported thinner boundaries than men and older people (Hartmann, 1991).

The HBQ has proven a reliable instrument for measuring thickness of personality boundaries (Ehrman, 1999; Ehrman & Oxford, 1996; Rawlings, 2001). In the interest of parsimony, however, a shortened version (Rawlings, 2001) containing 46 of the original 145 items was used to operationalize ego permeability. Rawlings found that six factors made significant contributions to the construct of ego boundaries, of which five are used to compile the ego permeability score. The five subscales and number of items, respectively, are as follows:

> 2. Unusual Experiences (*k* = 12). Example items are "In my daydreams, people kind of merge into one another or one person turns into another," and "I wake from one dream into another."
>
> 3. Need for Order (*k* = 12). Example items are "There is a place for everything and everything should be in its place," and "I think children need strict discipline."

4. Perceived Time-Money Competence[8] (*k* = 6). Example items are "I get to appointments right on time," and "I keep my desk and worktable neat and well organized."

5. Childlikeness (*k* = 5). Example items are "I think a good teacher must remain in part a child," and "A good parent has to be a bit of a child, too."

6. Sensitivity (*k* = 2). The two items are "I am easily hurt," and "I am a very sensitive person."

Ego permeability was measured in this study with a 7-point Likert scale anchored by "This does not describe me at all" (1) and "This describes me very well" (7). The respective constructs, instruments, and sources are shown in Table 4.

All instruments except the two proficiency measures (Breadth of Vocabulary Knowledge and Listening Comprehension) were translated and presented in Japanese. The directions on the proficiency measures were in Japanese, but the items were written in English. Some of the instruments are available in Japanese, whereas my additions were translated and then back-translated into English to check for accuracy. Two bilingual native speakers of Japanese provided the translation services. One holds a doctorate in TESOL, and the second has published several papers in the field and worked as a freelance translator.

In the following section I detail which of these variables were added to the two models of L2 communication, the model of MacIntyre and Charos (1996) and the International Posture models in Yashima (2002) and Yashima et al. (2004).

[8] This subscale was originally titled simply Perceived Competence, but in light of the items' content focusing heavily on time and money and in order to distinguish this subscale from the linguistically-oriented Perceived Competence in English subscale, in the current study it is referred to as Perceived Time-Money Competence.

Table 4
Summary of Instruments

Construct	Instrument (Appendix)	Source
Proficiency	Breadth of Vocabulary Knowledge (C)	Nation (2001); Nation & Laufer (1999)
	Listening Comprehension (D)	Author
Frequency of L2 Communication	Frequency of L2 Communication (E, F)	Yashima (2002)
L2 WTC	L2 Willingness to Communicate (G, H)	McCroskey (1992)
L2 Communicative Competence	Perceived Competence in English (I, J)	MacIntyre & Charos (1996)
L2 Communicative Anxiety	Communication Anxiety in English (K, L)	MacIntyre & Charos (1996)
	FLCAS (M, N)	Horwitz et al. (1986)
Motivation	Desire to Learn English (O, P)	Yashima (2002)
	Motivational Intensity (O, P)	Yashima (2002)
International Posture	Interest in International Vocation (Q, R)	Yashima (2002)
	Interest in International Affairs (Q, R)	Yashima (2002); Author
	Intergroup Avoidance-Acceptance Tendency (Q, R)	Yashima (2002)
	Intercultural Friendship Orientation (Q, R)	Yashima (2002)
Perceived distance	Perceived Distance Questionnaire (S, T)	Author
Personality	Bipolar scale of global personality traits (shortened form) (W, X)	Goldberg (1992); MacIntyre & Charos (1996)
Ego permeability	Hartmann Boundary Questionnaire (shortened form) (Y, Z)	Hartmann (1992); Rawlings (2001)
English experience	(from demographic data)	Author
Attitudes	(composite)	Author
Demographic information	(A, B)	Author

Note. The complete questionnaire included 274 items: 17 demographic items, 201 survey items, and 56 proficiency measure items.

Changes to the MacIntyre and Charos (1996) Model

MacIntyre and Charos' (1996) model (Figure 3) utilized Goldberg's (1992, 1993) Big 5 factors of personality: intellect, extroversion, agreeableness, conscientiousness, and emotional stability. In addition, the authors added context, a logical addition that markedly influenced several facets. Their results supported the posited influence of context on L2 WTC and L2 communicative frequency (the heavy, lines in Figure 3). Moreover, those results prompted the addition of the data-driven path from context to perceived L2 competence (shown by a dashed line in

Figure 4). Data also supported the dashed paths from perceived competence to L2 communicative frequency and from agreeableness to L2 WTC.

English Experience. Regarding context, MacIntyre and Charos (1996) noted that “the sociolinguistic context plays a potentially important role in providing the opportunity for frequent and/or pleasant L2 contact … because the number of opportunities to communicate in the second language should influence the frequency of doing so” (p. 16). In that study the extent of L2 usage at home and at work comprised the Context variable, but in the current study, the venues under investigation were foreign language contexts in which English was used almost exclusively in academic contexts rather than in social situations. For that reason, Context was replaced with English Experience, a variable constructed for this study. English Experience was derived from the demographic information provided by the informants and included seven experiences in which English could have been encountered: study abroad experience (DeKeyser, 2007), homestay experience, experience living overseas (Coleman, 1997), overseas travel experience, time in English conversation schools, the age at which English study was begun (Larson-Hall, 2008), and compulsory English education in secondary school. Both the length and richness of the experience contributed to the score, with experience in an English-speaking country scored higher than in an ESL situation, which was in turn scored higher than experience in an EFL situation. A longer tenure was similarly scored higher than a shorter tenure; the cutoff points were 30 days for travel and three years for residence abroad. The English Experience score was the sum of the respective categories and is detailed further in the Methods chapter (see Table 4).

Cultural Friendship Orientation (Integrativeness). The second substantial change to the MacIntyre and Charos model was substituting the Intercultural Friendship Orientation subscale from the International Posture instrument for Integrativeness.

Attitudes about the Learning Situation. MacIntyre and Charos (1996) operationalized attitudes with just two items, one asking about attitude toward the L2 teacher and the second asking about attitude toward the L2 course. In the current study the attitudes instrument consisted of four items, two about the teacher and two about the L2 class: One item is from the Motivation subscale (Item 5, *I believe absolutely English should be taught at school*), and three are from the FLCAS: Item 5 (*It wouldn't bother me at all to take more English language classes*), Item 14 (*I would not be nervous speaking English with native speakers*), and Item 32 (*I would probably feel comfortable around native speakers of English*).

Ego permeability. The first addition is ego permeability, which is fundamentally linked with context: depending on a given situation, the individual's ego permeability mediates the extent to which that person can assume new roles and thereby function adequately.

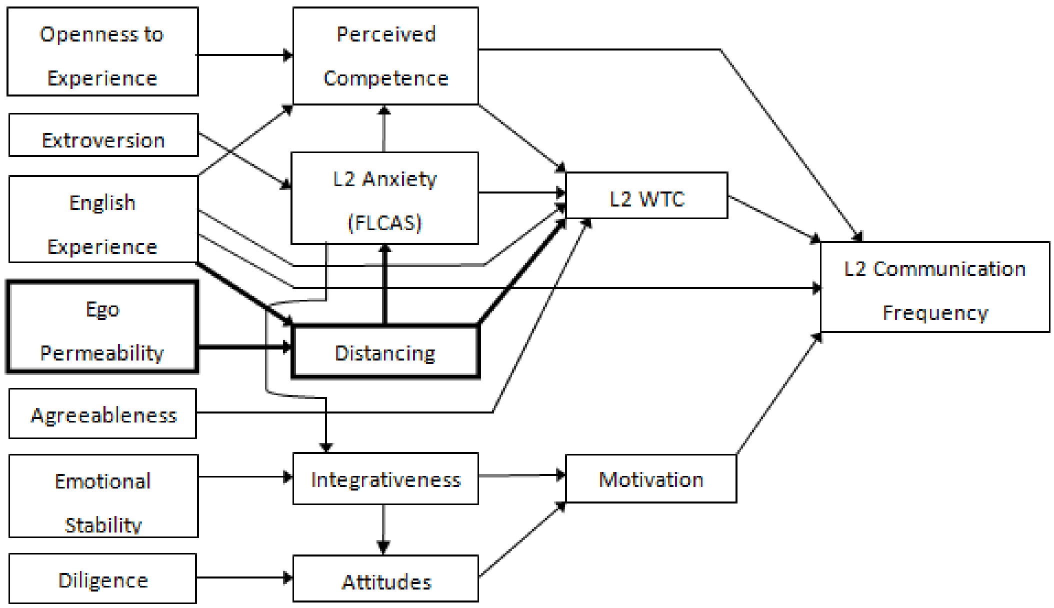


Figure 7. Proposed Model of L2 Willingness to Communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 12. Copyright 1996 by *Journal of Language and Social Psychology*.

Perceived Distance. Perceived Distance was then added as a higher-level construct. Ego permeability should affect distancing, for a low degree of ego permeability inhibits a learner's assuming or perceiving any degree of psychological distance; in short, such learners are limited mainly to their own persona. On the other hand, learners with a high degree of ego permeability might be able to assume and perceive larger degrees of distance as they adopt different personae. Similarly, context also influences distancing because different situations require different personae. Imagine, for example, individuals in a monolingual, L1 and first culture (C1) environment, which would allow them to remain comfortably in their own skin, so to speak. Given the need to assume another role or language in another context, however, the persons need to adapt and thereby might perceive some psychological distance from the usual L1/C1 persona. Ego permeability then directly influences perceived distance: If L2 learners can allow themselves to change to a new role, then they might perceive increased distance from their original persona.

Moreover, the role of distance in the second column from the left is closely and inversely related to L2 anxiety, for increasing distance should lower anxiety and decreasing distance should increase anxiety. An increase in anxiety (e.g., as a result of being in a new and therefore unfamiliar context) should make distancing less possible. In addition, distance also directly affects L2 WTC, although the relationship is not linear: Increasing distance initially facilitates WTC as, for example, speakers move from personal topics to more distant and thus safer topics. Thereafter, however, further increases in distance inhibit WTC as speakers lose interest in very distant topics in which they have little knowledge or interest. This curve resembles that posited by Csikszentmihalyi (2000) in which “flow” occurs in a middle ground between boredom and attention. Another similar relationship is outlined by the Yerkes-Dodson Law (Smith, Sarason, & Sarason, 1982), which describes a curvilinear relationship between anxiety and performance as a function of task difficulty. If we substitute distance for anxiety in the original figure, we see that tasks perceived as relatively close to the learner (i.e., of personal interest or perceived importance) are performed to a small degree, but with increasing distance the performance level increases. After reaching a maximum level (the apex of the curve), increasing distance corresponds with decreasing performance, which might be explained by decreasing interest or importance.

Additions to Yashima's (2002) Model

Because distancing and ego permeability might underpin several of the affective variables in the MacIntyre and Charos model above, they are potentially prudent additions to Yashima's (2002) model. These affective variables do not exist in a vacuum, however, for factors such as personality certainly influence them. Brown (1973) maintained that egocentric factors such as imitation, ego, and inhibition play important roles, for "a person is forced to take on a new identity if he [*sic*] is to become competent in a second language" (p. 233).

Building on the models proposed by Yashima et al. (2004) and MacIntyre and Charos (1996), one purpose of the current study is to better explain L2 Communicative Confidence through the addition of variables measuring personality dimensions, ego permeability, and perceived distance to the original scales measuring perceived L2 competence and L2 communicative anxiety (see Figure 8 below). Second, to assess L2 proficiency, TOEFL scores have been replaced by two measures, a Breadth of Vocabulary Knowledge test based on Nation's (2001) Productive Vocabulary Test and a listening comprehension examination. In Yashima's model, L2 communicative confidence and L2 proficiency are identified by two indicators, respectively, and L2 WTC was split into two parcels that functioned as indicators. However, Kline (2005) suggested that at least three indicators (variables) be used to identify latent variables, and Kenny (1979) put it this way: "Two *might* be fine, three is better, four is best, and anything more is gravy" (p. 143; emphasis in original).

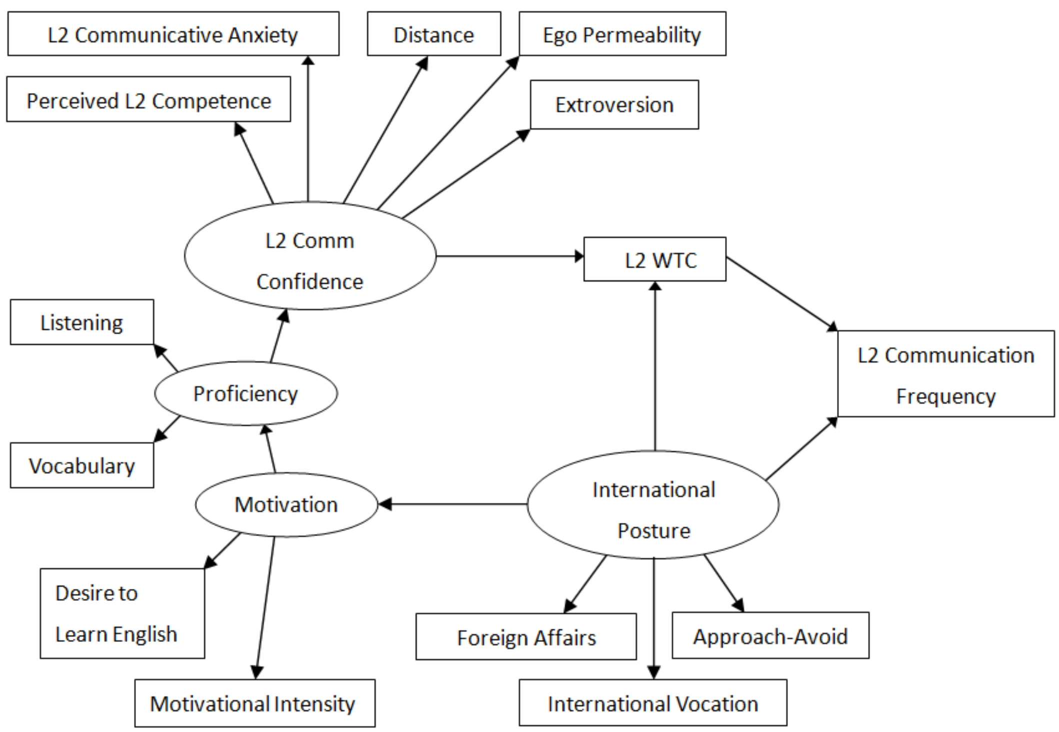


Figure 8. Proposed L2 Communication Model based on Yashima et al. (2004).

Although the models are similar, MacIntyre's included context and personality (the Big 5 personality factors) as an underlying layer. Yashima, however, addressed the L2 side more heavily by including International Posture (analogous to the integrative motivation of earlier models), which moderates motivation and L2 frequency. In addition, L2 Communicative Confidence was posited to include L2 anxiety, perceived L2 competence, and the three personality variables added in the current study.

The second half of this chapter is focused on perceived distance, ego permeability, and extroversion, which are hypothesized to enter the MacIntyre and Charos model and the L2 Communicative Confidence factor in the Yashima models. The third addition to the Yashima models, extroversion, was given a cursory treatment above; being a much more familiar element, extroversion is not addressed further here (for an extensive treatment of extroversion, see Dewaele and Furnham, 1999, and Dewaele, 2005).

Preliminary Studies

Two earlier studies contributed to the present study. The first was an in-depth investigation of the FLCAS in which data from a large group (N = 1,038) at B University were gathered and analyzed (Elwood, 2005). The results indicated that the FLCAS was unidimensional. Moreover, the results indicated a fair amount of redundancy in the 33 items, suggesting that the FLCAS could be shortened with no significant loss of reliability.

The second study was conducted in May of 2007 with two groups of students (N = 143). The first group was from the Economics Department at C University; these students had low English proficiency. The second group was from the Medical Department at B University. This group had a high level of English proficiency, and the students were quite serious and studious.

In the second study the entire package of instruments used in the current study was administered. The first purpose was to identify typographical errors and passages that were difficult to understand; a total of seven minor changes were instituted. A second purpose was to trim the number of items in the vocabulary proficiency questionnaire from 72 to 40 using output from WINSTEPS.

Procedure

During the 90-minute treatment session I first introduced myself and then explained the purpose of the research and the notion of distancing (Appendix AA), which took about five minutes. All participants received a 4-page handout with a written explanation of the research, the consent form, and the instructions and scripts for the activities (Appendix AB). Thereafter I walked the participants through the distancing activities in order to elicit a reaction to those particular activities, not activities that the participants might have experienced in junior high or high school English classes. First were the three speaking-only components (casual chatting in the L1, casual chatting in the L2, and public speaking); the participants engaged in the activity for 3-4 minutes with a partner, after which each participant noted the degree of distance that he or she perceived as a percentage from 0% (no perceived change) to 100% (a complete change of character). The students were allowed to speak about any topic in the chatting sections. In the public speaking section a portion of Martin Luther King, Jr.'s "I Have a Dream" speech was used. I demonstrated a rather dramatic style for the students to emulate, after which the speaker in each pair stood and all those individuals gave the speech simultaneously; when the first speaker had finished, the partners traded roles and the second speaker began giving the speech. As such, half the class was standing and speaking at any given moment. This format was used in all the distancing activities to ameliorate any anxiety that the speakers might have felt when speaking alone.

The participants next engaged in three activities that incorporated increasing degrees of psychological distance: a roleplay, a short drama, and finally a puppetry activity. The roleplay consisted of resolving a dispute between a person with a noisy, barking dog and a neighbor that was trying to sleep. I selected a gregarious pair of participants to illustrate the location of actors, produced one of my dog puppets who proceeded to "bark" noisily, modeled the two human parts, divided the participants into pairs, and then checked that each participant had a part. The entire class then began the activity, so all groups were speaking simultaneously in what became a noisy classroom. After 3–4 minutes, the participants were again asked to indicate as a percentage (0-100%) the degree of distance that they perceived while engaging in the roleplay.

A short excerpt from Shakespeare's Romeo and Juliet that the author revised into modern English was used in the drama section. Again, the dialogue was modeled in a dramatic fashion, after which a part was assigned to each participant and the activity was conducted.

In the final activity, hand puppets were used in a simplified version of the Bremen Town Musicians, a folktale in which a dog, a cat, a donkey, and a rooster are traveling to the city of Bremen to become musicians. I first introduced the four puppets (hand puppets in which one hand is inserted into the puppet's body to animate the puppet's mouth) and invited the students to conjure up voices for the animals. After a couple of minutes, I demonstrated the voices that I use for the respective animals and invited the students to provide a voice for each puppet; in addition, I asked the students to mimic the hand motion even if they did not have a puppet in hand. Because the story has four parts, the students worked in groups of four; I distributed the four puppets and one additional dog puppet, and while the story was in progress I circulated around the classroom and redistributed the puppets so as many students as possible could experience using a puppet. At the conclusion of each activity, the students marked the degree of distance they perceived while engaging in the activity as a percentage from zero to 100.

I had planned to explain and conduct the treatment in English, but in classes with a lower English level I had to explain the activities in Japanese so that the participants could understand and complete the activity successfully and in a timely fashion. The treatment lasted for 90 minutes, which is the length of a standard university class in Japan. At the conclusion of the distancing section (about 60 minutes), the participants completed the listening proficiency test and then began the productive vocabulary test. Finally, the remainder of the questionnaire was distributed and assigned as homework; the entire questionnaire was returned to the regular course instructor in the following week's class. The treatment schematic is shown in Table 5.

Table 5
Treatment Schematic

Duration (minutes)	Activity
5	Introduction, explanation of research
2-3	Distribution of questionnaires
2-3	Casual chatting in Japanese
4-5	Casual chatting in English
5	Public speaking (excerpt from Martin Luther King's "I Have a Dream" speech)
7	Roleplay (barking dog)
7	Drama (Romeo and Juliet)
10	Puppetry (Bremen Musicians)
15	Listening proficiency measures

Note. In most cases, the participants were able to begin the vocabulary section after completing the two proficiency instruments.

Ethical Considerations

Pursuant to the policies of the Graduate School of Temple University, in this study efforts were made to adhere to the accepted ethical practices in Japan for research utilizing human participants. Participants received a consent form with the distancing activities handout; on the form they had the choice of opting out of the research activities with no deleterious effect on their grade. Time borrowed from their regular class time was limited to just one 90-minute class period, and the participants received a small number of bonus points toward their regular class grade.

The Rasch Model

The Rasch measurement model (Rasch, 1960) is a mathematical probability model that permits investigating dimensionality and the ordering of items and persons on a continuum. It offers a simple yet elegant way to construct and analyze linear item and person measures of Rasch calibrations from ordinal measurements (Wright & Stone, 1979). Using Rasch modeling, the assessment of fit for items and for persons allows items and persons to be measured on a common interval scale.

The Rasch rating scale model estimates the probability that a respondent will choose a certain response category for a particular item with the following formula:

$$\ln \{ P_{nij} / P_{ni(j-1)} \} = B_n - D_i - F_j, \text{ where}$$

ln is a natural logarithm, P_{nij} is the probability of respondent n scoring in category j for item i, $P_{ni(j-1)}$ is the probability of scoring in category (j-1), B_n is the person measure of respondent n, D_i is the difficulty of item i, and F_j is the difficulty of category step j (the threshold at which there is a 50-50 chance of scoring in category j and category $j - 1$). The person's likely score is defined by the interaction between the person's measure, the criterion's (i.e., item's) difficulty, and the score's category threshold. Rasch analysis places persons (B_n) and items (D_i) on the same measurement scale where the unit of measurement is the logit (logarithm of odds unit).

The Rasch model reduces complex data matrices to a unidimensional variable regardless of the dimensionality of the original observational data. In other words, all systematic variation in the data is explained by only one latent variable while any residuals would represent random noise when the data fit the Rasch model. If a well-constructed instrument does not produce a single latent variable and instead yields unexpected residuals, then an examination of the residuals via a principal components analysis (PCA) can shed light on which items and persons poorly fit the construction of the measurement dimension. Patterns indicative of relevant second variables can emerge from a principal components analysis of standardized residuals. However, when data fit the model for a single latent variable, then residuals are independent and elements of the inter-item residual correlation matrix are zero.

Using the Rasch model, analyses included the following steps.

Rating Scale Functioning

Prior to conducting other analyses, an examination of the rating scale was conducted to ascertain whether the participants' response patterns represented the lack of use or inconsistent use of rating scale categories. Linacre (1997, 1999, 2002) suggested six criteria for evaluating rating scale effectiveness. First, each category should have a minimum of ten observations. Second, the probability curves should be peaked, thus forming a series of hills. Third, the average rating scale measure should increase with the rating scale category (i.e., the categories should be properly ordered with, for example, the second category being more difficult than the first, the third more difficult than the second, and so forth). Fourth, the outfit mean statistics should be less than 2.0. Fifth, threshold calibrations should increase with the rating scale category. Finally, category thresholds should be separated as indicated in Table 6.

Table 6
Category Separation Series

Number of Likert scale categories (*j*)	Series				
	ln(*j*)	*ln*(*j*) – *ln*(*j*-1)	Natural log series	Min sep (logits)	Min sep (CHIPS)
2	.69	.69	2.20	2.20	10.01
3	1.10	.41	1.51[a]	1.4*	6.37*
4	1.39	.29	1.10	1.1*	5.00*
5	1.61	.22	.81	.81*	3.69*
6	1.79	.18	.59	.59	2.68
7	1.95	.16	.41	.41	1.87

Note. Values with an asterisk are from Wolfe and Smith (2007, p. 210), and the corresponding CHIPS values are calculated from those data. [a] If the value from the natural log series is used here in 3-category row, the minimum separation would be 1.51 logits (6.85 CHIPS) instead of 1.4 logits (6.37 CHIPS).

To the best of my knowledge, separation criteria are only available for 3-, 4-, and 5-category scales (Wolfe & Smith, 2007, p. 210). Inasmuch as the current study included 6- and 7-category scales, a more complete set of guidelines was necessary. Because the three values (i.e., 1.4, 1.1, and .81) are *similar* to a logarithmic sequence, it was posited that the separation intervals could be explained as differences in natural logarithm values. If we assume a rounding or typographical error that yielded a value of 1.4 instead of 1.51, then we can extrapolate values for 6-point and 7-point scales from that sequence using the following equation:

$$\text{minimum separation } (j) = \text{minimum separation } (j{-}1) - [ln(j{-}1) - ln(j{-}2)].$$

For example, for six categories ($j = 6$), the minimum separation is .81 – [1.61 – 1.39], which simplifies to .81 – .22 and thus .59 logits (= 2.68 CHIPS). Similarly, for seven categories the values are .41 logits or 1.87 CHIPS. If we peruse differences in the minimum separation values, we find a steadily decreasing series of values (i.e., differences are .29, .22. .18, and .16). The appropriate value for the number of categories was then used to investigate the minimum separation of each scale.

The reader should be aware that this extrapolation represents at best an approximation, yet it is a useful approximation. If the value for a 3-category scale in fact follows the natural logarithmic sequence, then the corresponding separation values would be 1.51 logits or 6.85 CHIPS.

Item-Person Map

In addition to considering the category functioning, it is important to perform a visual inspection of the item-person map, also called the Wright variable map. This is a visual representation of Rasch analysis, with person and item measures on a common scale that shows the hierarchy and location of persons and items relative to one another. Typically both are displayed vertically, with the top of the person side indicating “more ability” of the construct and the top of the item side indicating “more difficult to endorse.” In this study the Rasch logit measures were delineated in units called CHIPS, a more user-friendly scaling unit than logits.

The fundamental makeup of the CHIPS scaling is that 4.55 CHIPS equal one logit and the standard errors tend to be about one CHIP in size (Linacre, 2007, p. 352); thus, one CHIP is a smaller unit, equaling .22 logits (put another way, if we imagine one CHIP being one Fahrenheit degree, then one logit—a larger unit—is analogous to a Celsius degree). In the current study, items are scaled using CHIPS with the mean set at 50.

On both sides of the vertical CHIPS scale line in the middle of the map are the letters M, S, and T, which indicate the mean, one standard deviation, and two (T = "two") standard deviations, respectively. The item mean and person mean should be reasonably close, ideally less than two measurement errors apart (which corresponds to a .05 significance level). A greater separation of means indicates that the instrument might not adequately target the particular sample. Moreover, the range of the two scales should be similar, as markedly different ranges can indicate that the scale is not adequately assessing the construct.

Where space permits (i.e., where there are few items in a scale), the regular item-person map has been replaced with a map showing Rasch-Thurstone item thresholds instead of item measures. An item threshold map more accurately shows the range of coverage than does simply showing the means of respective item measures.

The items should be distributed widely enough so that all levels of person abilities can be assessed; these generally form a normal distribution. Gaps in the item distribution can indicate insufficient mapping of the logit scale unless those gaps are small (less than .30 logits or 1.36 CHIPS; Reeve & Fayers, 2005). Moreover, the item map shows redundancy, allowing for the deletion of items with no loss in reliability or validity.

Rasch Fit Statistics

Rasch analysis works on the basis of fitting data to the Rasch model and checks data for perturbations caused by data failing to fit the model. Rasch analysis provides fit statistics for both persons and items to assess assumptions of fundamental measurement. Fundamental to this is the requirement that all data measure the same trait (i.e., that the measured construct is unidimensional).

The indicator of item functioning from classical test theory is corrected item-total correlations. These correlations are inspected for obvious off-dimension behavior; values less than or near zero indicate potential problems. Tables in the current study also include point-measure correlations (labeled Pt-M correl), which are appropriate when data are missing. This correlation between item responses and person raw scores is vital for assessing whether the measurement scheme and person responses yield results in which higher observations correspond to more of the latent variable and vice-versa. Although less informative than fit statistics, negative and zero values show when response strings contravene the variable in question.

The Rasch model provides two indicators of misfit, infit and outfit, which take two forms, chi-square statistics divided by their respective degrees of freedom and a standardized form that presents item fit in terms of a *z*-distribution. The basic difference in the two fit indicators is that the infit statistic is sensitive to unexpected response behavior near the person's ability level, while outfit is particularly sensitive to the effect of outliers (i.e., responses far from the person's ability level). The WINSTEPS User's Guide (2006b) suggested a range of .50 to 1.50 for the standardized fit statistics; that range is used in the current study.

In the event that the infit value fell outside the .50-1.50 range, unexpected responses were scrutinized. Ideally, the number of unexpected responses would be under 5% of the total, which is analogous to the .05 level of significance for statistical analyses (thus, with 252 respondents, fewer than 5% would be 13 responses). WINSTEPS for that particular scale was repeated after deleting a maximum of 5% of problematic responses; if the outfit value decreased to an acceptable value, then the item was considered to be functioning adequately, but if the outfit value did not meet the criterion, then the item was considered a candidate for deletion.

Mean square fit statistics are defined so that the model-specified anchor of randomness is 1.0 and the standardized *z*-statistic (Zstd) provides a significance test for which values greater than 2.0 are generally regarded as statistically significant. Although standardized *z*-statistic values in excess of 2.0 can occur with adequate fit statistics, Linacre (2006b, p. 308) suggested ignoring the standardized *z*-statistics, and in this study I follow that admonition.[9] Person fit indicates the degree to which

[9] Linacre (2006b, p. 308) also noted that with sample sizes greater than 300, the test can be too sensitive, resulting in a situation in which "everything misfits." The sample size in this study (*N*

a person's performance is consistent with how other participants responded. Item fit, on the other hand, shows whether use of a certain item is consistent when compared with how participants responded to other items. When potential misfitting items were identified, the content and quality of the item were considered in making a final decision about whether to retain the item.

Rasch Separation, Reliability, and Strata

WINSTEPS provides three statistics concerning person and item reliability. A Rasch person reliability estimate (R_p) is analogous in interpretation to the traditional alpha value of internal consistency (Smith, 2001). Person reliability is defined as the ratio of the sample variance adjusted for measurement error to the total observed variance; it represents the proportion of variance that is not from measurement error. In other words, it indicates the consistency of person ordering as measured by the measurement scale, which can also be conceptualized as the reliability of the persons being separated by the measurement scale.

Person reliability is bounded by values of zero and one and is non-linear. It can be mathematically transformed, however, to person separation (G_p) using the formula, $(G_p) = \{R_p / (1- R_p)\}^{1/2}$. Separation values can range from zero to infinity, and larger values indicate higher Rasch reliability.

Moreover, person separation (G_p) can be transformed into a strata statistic, H_p, using the formula, $H_p = (4G_p + 1) / 3$. The 3 in this formula is the basic unit of

---

= 252) is conceivably large enough that the tests could be overly sensitive.

comparison: Using “three standard errors apart” as the fundamental unit, H_p represents the number of strata of person measures that are statistically distinct on the measurement continuum (this arises because the statistically significant difference at $p < .05$ is 2.79, so the next largest integer was used).

Item separation, reliability, and strata are interpreted similarly. Item separation indicates the adequacy of scale measure in assessing the increase in the construct measured. Item reliability is analogous to the Cronbach reliability statistic. The spread of the item calibration as shown by the item strata statistic indicates the comprehensiveness of coverage of the construct by the items (Smith, 2001). The person strata value indicates the extent to which participants can distinguish statistically distinct regions of the construct in question.

Rasch Principal Components Analysis of Item Residuals

The third analysis involved checking the dimensionality of items designed to measure the same construct. The initial step was an exploratory factor analysis using SPSS; these results were treated as guidelines with Rasch principal components analysis of residuals as a more definitive method to detect the presence of other dimensions. The Rasch model extracts the primary dimension that explains the most variance, and the remaining variance is then inspected for the presence of any further dimensions; ideally, the remainder is only noise. The most common method (Linacre, 1998; Smith, 2002; Wright, 1996) is to plot factor loadings of the first factor against Rasch calibrations; these should be randomly distributed for both items and participants. Next is a perusal of the amplitude of the first remaining construct, for which Linacre (n.d.) offers the *general* guideline that the size should be relatively small, ideally less than 5% and comprising fewer than 3.0 localized units. Finally, items are divided into two subsets (i.e., positive loadings and negative loadings) based on the item residual factor loadings in the principal components analysis of residuals, after which separate analyses were performed on each half to measure each person on each subset of items. A plot of the two measures should be linear and the disattenuated Pearson correlation near unity if the two sets of items measure the same construct. Because “near unity” is not a helpful criterion and because no specific criterion based on empirical evidence exists, in the current study .80 is used as a rule of thumb.

Treatment of Misfitting Items

Misfitting items were first scrutinized to ascertain whether a small number of unexpected responses was the cause or whether the item was actually performing poorly. Second, the degree of misfit was checked using an Infit Mean Square (MNSQ) criterion of 1.5. Third, the extent of change in WINSTEPS reliability and separation when the item was deleted was examined, as was the correlation of the instrument with the item to the instrument without the item. Fourth, the composition of the PCA residual components with and without the item was examined. Finally, the theoretical necessity of the item was reviewed. In general, absent a strong statistical or theoretical reason to delete an item, it was included in the instrument.

Data Analysis

Initial Data Screening

Data were first carefully scrutinized for missing data. In addition, the data were screened for outliers, both univariate and multivariate. *Z*-scores were used to check for univariate outliers, with the criterion for potential outliers being $z > 3.29$ ($p < .001$, two-tailed test). The data were then examined for multivariate outliers using the Mahalanobis distance.

Data were also screened for normality, linearity, and homoscedasticity. Thereafter individual scales were examined for dimensionality, first by checking inter-item correlation and then by conducting an initial confirmatory factor analysis using SPSS.

The Rasch logit measures are scaled using CHIPS, a more user-friendly scaling unit than logits. The fundamental makeup of the CHIPS scaling is that 4.55 CHIPS equal one logit and the standard errors tend to be about one CHIP in size (Linacre, 2007, p. 352). One CHIP is thus a smaller unit, equaling .22 logits. In the current study, items are scaled using CHIPS with the mean set at 50 to minimize the possibility of obtaining negative CHIP values.

Scale Conversion from Percentages to Likert Scaling

Percentages were used for four of the instruments: Perceived Distance, L2 WTC, Perceived L2 Competence, and Communication Anxiety in English, but for three reasons these data were subsequently transformed to Likert-style data. One reason was to produce scales with a uniform number of response categories (7-point) inasmuch as the remaining scales in this study used that number. Second, with a smaller number of categories, category function can be carefully investigated; with percentage data this is not possible unless all 101 categories are represented with a minimum of 10 responses (Linacre, 1997, 1999, 2002), an extremely remote possibility since most responses occur as a multiple of 10 (e.g., 10%, 20%, 30%). Finally, research from psychology indicates that people have limited capacity to process information. Miller's seminal 1956 article memorably put that capacity at "the magical number seven, plus or minus two" (p. 81), and subsequent research has corroborated Miller's basic premise (Baddeley, 1994; Banks, 2003).

Based on older item response theory models, the most reliable scales have 7 to 10 response categories (Cicchetti, Shoinralter, & Tyrer, 1985; Oaster, 1989), and those with 6 or more categories having the highest levels of validity and discriminatory power (Chang, 1994; Hancock & Klockars, 1991). On the other hand, instruments with more response categories (10, 11, or even 101) are viewed as allowing greater precision for informants to express their feelings (Preston & Colman, 2000). However, scales with more than five categories are seldom useful, and scales with six or more categories generally fail to meet Linacre's guidelines (E. W. Wolfe, personal communication, August 3, 2009).

Thus, in the current study the percentage scales were transformed to scales with fewer response categories. The data were parsed to produce symmetrical intervals about the midpoint. The transformation to an 11-point scale reflects the original percentage data, while the second transformation was to two slightly different 7-point scales. The primary consideration in parsing was the number of multiples of 10 (i.e., 0%, 10%, 20%) in each category, for these are the responses that are most commonly used. As shown in Table 7, the "narrow center" 7-point scale has just one multiple of 10 (50%) in the neutral, #4 category, while the "wide center" 7-point scale has 3 multiples of 10 (40%, 50%, and 60%) in the neutral category.

Table 7
Scaling Parsing for Percentage-to-Likert Data Transformation

Likert scale category	Scale Parsing: 7-point scales: Narrow center	Scale Parsing: 7-point scales: Wide center	Scale Parsing: 11-point scale
1	0-9% (1)	0-9% (1)	0-4% (1)
2	10-24% (2)	10-19% (1)	5-14% (1)
3	25-40% (2)	20-39% (2)	15-24% (1)
4	41-59% (1)	40-60% (3)	25-34% (1)
5	60-75% (2)	61-80% (2)	35-44% (1)
6	76-90% (2)	81-90% (1)	45-55% (1)
7	91-100% (1)	91-100% (1)	56-65% (1)
8			65-75% (1)
9			75-85% (1)
10			85-95% (1)
11			96-100% (1)

Note. Parenthetical numbers indicate the number of multiples of 10 in each category. Thus, for example, Category 1 of the 11-point scale includes just one multiple of 10, namely, 0% (= 0 x 10).

These data sets were then examined with Rasch analysis to check for adequate fit statistics and category function, and the choice of a 7-point or 11-point scale was based on the descriptive statistics and results of the Rasch analyses. Depending on the category function results, a further reduction in the number of categories was investigated for the respective scales. As explained in detail in Chapter 5, when the optimal number of categories was found, the original percentage data were parsed into that number of categories. Results are reported in the Preliminary Results chapters.

Dimensionality

The next task was to investigate the dimensionality of the instruments. First, a factor analysis was conducted (principal axis factoring) to confirm the number of dimensions. In so doing, the scree plots, eigenvalues, and factor loadings were checked; basic criteria included eigenvalues in excess of unity, factors with at least three loadings in excess of .40, and preferably a marker variable loading in excess of .70. Second, inter-item correlations, internal reliability (Cronbach's alpha), and dimensionality for each hypothesized subscale were examined (Briggs & Cheek, 1986; Piedmont & Hyland, 1993). An average correlation in excess of .30 and internal reliability greater than .70 were the general criteria. Items that had a low inter-item correlation or significantly reduced the internal reliability of respective scales and subscales were deleted. Dimensionality was also investigated by examining inter-item correlation frequency distributions (Piedmont & Hyland, 1993); a unidimensional data set has a roughly normal distribution, but a multidimensional data set produces other distributions (e.g., bimodal) depending on the number of dimensions.

Instruments were then scrutinized in detail using Rasch analysis (WINSTEPS, 2006). The various aspects of the Rasch analyses are explained in the following section.

Structural Equation Modeling

Structural equation modeling (SEM) is a collection of statistical techniques that allow examination of sets of relationships between single or multiple independent variables (IVs) and single or multiple dependent variables (DVs). Both IVs and DVs can be factors (latent variables) or measured variables. Such versatility is matched by other advantages of SEM. Relationships among factors are free of measurement error because that error is estimated and removed, leaving only covariance. SEM is also known as causal modeling and analysis of covariance structures; factor analysis, path analysis, and regression all represent special, limited cases of SEM. In the current study, the models based on the MacIntyre and Charos (1996) model are investigated using path analysis and the models based on the Yashima models (Yashima, 2002; Yashima et al., 2004) are investigated using full-fledged SEM. All confirmatory factor analyses and structural models were tested using EQS 6.1 for Windows, Build 94 (Bentler, 2006a).

SEM is largely confirmatory rather than exploratory. That is, researchers are more likely to use SEM to determine whether the data fit a hypothesized model, rather than to build a model using SEM. However, the respecification of models to examine potentially better-fitting models does involve a certain exploratory element.

In SEM, interest usually focuses on latent constructs—abstract psychological variables such as intelligence or motivation—rather than on the measured variables used to investigate those constructs. By explicitly modeling measurement error, SEM users seek to derive unbiased estimates of the relationships between such latent constructs, and SEM specifically allows multiple measures to be associated with a single latent construct.

A structural equation model implies a structure of the covariance matrix of the measures (hence an alternative name for this field noted above, “analysis of covariance structures”). Once the model’s parameters have been estimated, the resulting model-implied covariance matrix can then be compared to an empirical covariance matrix. If the two matrices are consistent with one another, then the structural equation model can be considered a plausible explanation for the relations among the measures.

Applying SEM involves the following steps. These are summarized from Kline (2005, pp. 63-64):

1. Model specification, which means the researcher’s hypotheses are expressed in terms of structural equations;

2. Model identification, which means it is theoretically possible to derive a unique estimate for every model parameter;

3. Model estimation using computer software;

4. Evaluation of model fit, by which the researcher determines how well the model as a whole fits the data; and

5. Model respecification, in which the fit of a revised model is compared with the same data.[10]

The structural models in the present study follow conventions used in the social sciences. These include the following:

1. Rectangles show measured variables. Each observed variable has an error component with a mean of zero and a fixed loading of 1.0 (not shown in the figures).

2. Ovals represent latent variables, which cannot be measured. Each latent variable has one fixed loading of 1.0 on an observed variable, and the means of all latent variables are fixed to zero.

3. E terms represent the error terms or the residuals from the measured variables. As noted in #1, each error term has a mean of zero and a fixed loading of 1.0.

---

[10] Although model respecification can capitalize on chance characteristics of the data (MacCallum, Roznowski, & Necowitz, 1992), it does provide a useful technique to discover relationships that the research might not have considered a priori. Bearing the risks in mind, models in the current study will be respecified if statistically and theoretically prudent with the understanding that replication would address the possibility of any respecification being based on a chance occurrence.

4. D terms are disturbances, which are error terms for the latent (unmeasured) variables. Each disturbance has a fixed loading of 1.0 (disturbances are also not shown in the figures).

5. Straight, single-headed arrows show the direction of influence. After analysis the values that appear can be interpreted as beta weights (from multiple regression) or as factor loadings (from factor analysis).

6. Curved, double-headed arrows are covariances and can be interpreted as correlations.

Error and disturbance terms are not shown in SEM diagrams in the Results chapter in order to avoid excessively cluttering the diagrams. The complete standardized solutions with error terms, disturbance terms, and standard errors are presented in Appendices AB-AD.

Rasch Data Screening and SEM Assumptions

In the unlikely event that instruments have perfect fit to the Rasch model, true interval measures can be constructed from the raw scores, which are ordinal data. Although the instruments used in the current study did not meet the strict criterion of fitting the Rasch model perfectly, it is assumed that the Rasch measures approximate true interval scales better than the raw scores from which they are derived. Pursuant to this, Rasch measures should be screened for patterns of distribution and covariance necessary to conduct a particular statistical analysis. The steps taken in screening the Rasch measures to meet the assumptions of SEM are presented in the following pages.

Sample size and missing data. Kline asserted (2005) that a sample size in excess of 200 is necessary to obtain trustworthy results. With a sample size of 252, the current study satisfies this criterion. Kline also suggested a ratio of 20 respondents per freely estimated parameter is ideal, while 10:1 is more practical (p. 111). In the current study the path analyses have a ratio of 5.2:1, and the SEM analyses have a much better ratio of 9.7:1. Although a larger sample size would have been better for evaluating the rather complex path analysis models, the fairly large sample size was deemed acceptable.

In the structural equation models, Rasch person ability estimates (CHIPS) were used; these estimates correct for missing data, so the data were complete.

Multivariate normality. One assumption of SEM is multivariate normality. SEM can tolerate a certain degree of non-normality, with robust methods able to handle egregious cases of non-normality. An examination of the significance of skewness and kurtosis indicates non-normality for small samples, yet for large samples minor perturbations in the data can yield statistically significant skewness and kurtosis. Tabachnick and Fidell (2004, p. 714) suggested perusal of distribution plots for samples of 200 or more, so histograms for the 22 variables were produced and examined using SPSS.

Outliers. An outlier is a person with an extreme value on one variable (a univariate outlier) or an unusual combination on multiple variables (a multivariate outlier). Either case for parametric analyses is problematic because outliers exert an undue influence that threatens the generalizability of the results. Diagnosing outliers can be done by examining *z*-scores and checking distribution plots. *Z*-scores with an absolute value in excess of 3.29 are indicative of univariate outliers, and scores that are isolated from the distribution are also suggestive of outliers.

Linearity. To investigate linearity, bivariate scatterplots were examined. Examining all 231 possible permutations of the 22 variables was an impractical task, but several potentially problematic combinations were examined (Tabachnick & Fidell, p. 79).

Homoscedasticity. Homoscedasticity was also examined with scatterplots. In a bivariate distribution, scedasticity refers to the extent that the variance in one variable is the same at all values of the second variable. Homoscedasticity refers to variance that is the same, while heteroscedasticity denotes variance that is not the same. Violations of homoscedasticity are investigated by examining scatterplots; an oval shape is indicative of homoscedasticity, whereas something like a rounded triangle is indicative of skewness in one of the variables and thus of heteroscedasticity. The scatterplots examined exhibited no indication of heteroscedasticity.

Multicollinearity and singularity. Multicollinearity refers to excessively high correlation of variables, a situation that makes matrix inversion unstable due to excessively small determinants. To investigate multicollinearity, the correlation matrix is examined and values in excess of the .90 criterion are indicative of multicollinearity (Tabachnick & Fidell, 2001, p. 83). Although examining bivariate scatterplots is also prudent, with 22 variables and 231 possible permutations, that task becomes impractical. However, several potentially problematic combinations were examined (Tabachnick & Fidell, p. 79), and in particular, the distancing–extroversion and the distancing–ego permeability permutations were carefully scrutinized. In all cases, scatterplots were not indicative of any particular problems.

Singularity refers to a situation in which variables are redundant, which prohibits matrix inversion. Although an assumption of SEM, the lack of singularity is confirmed post ipso facto. In short, if the model converges when the SEM analysis is conducted, then no singularity was present.

Residuals. Basically, the residuals should be small and symmetrically distributed around the mean. This is addressed by examining the distribution of the residuals of the covariances.

Methodology

When the assumptions above have been satisfactorily met, the actual path analysis or structural equation model is conducted. In the current study, all model estimation was conducted using the covariance matrix and maximum likelihood (ML) estimation. All analyses were done using EQS 6.1 for Windows, Build 94 (Bentler, 2006a).

Model fit is traditionally checked with chi-square statistics, yet the applicability of chi-squared tests for model fit has been questioned (Mulaik, James, Van Alstine, Bennett, Link, & Stilwell, 1989). Kline (2005) suggested assessing fit with four fit indices: the model chi-square, the Steiger-Lind root mean square error of approximation (RMSEA; Steiger, 1990) the Bentler comparative-fit index (CFI; Bentler, 1990), and the standardized root mean square residual (SRMR). Bearing in mind that the discussion on which of the many adjunct fit indices to use continues (and even whether to use them; see Barrett, 2007 and Bentler, 2006), five fit indices are reported in the current study: the model chi-square, RMSEA, CFI, SRMR, and the Incremental Fit Index (IFI). Generally accepted levels of significance for the five are reported and used, but the reader should be cognizant that “the jury is still out as to whether .90, .95, or any rule-of-thumb cutoff is appropriate” (Lance, Butts, & Michels, 2006, pp. 204-205). Families of fit indices and the indices used in the current study are briefly described below.

Absolute fit indices are based on comparing the observed covariance matrix to the one estimated on the assumption that the model is true. As such, they do not posit or subsequently compare with an alternative model. Among these tests is the venerable model chi-square test, for which a non-significant value is desired (i.e., $p > .05$). However, the chi-square statistic is sensitive to sample size, model size, and violations of multivariate normality (Newsom, 2007), so statistically significant results should be viewed with caution. A second absolute-fit index included in the present study is the standardized root mean square residual (SRMR), which is among those suggested by Kline (2005). Newsom (2008) suggested that SRMR is a good choice to report inasmuch as, although related to chi-square, it is less susceptible to the problems that the chi-square statistic faces.

The family of parsimonious fit indices penalizes models for complexity because lack of parsimony (i.e., more complexity) generally leads to better fit. Among these is the Steiger-Lind root mean square error of approximation (RMSEA; Steiger, 1990), which is based on the non-centrality parameter. The formula for RMSEA is $\chi^2/df - 1/(N - 1)$, in which df is the degrees of freedom and N is the sample size. Values of .05 or less indicate models that fit well, while values less than .06 (Hu & Bentler, 1999) or .08 (Brown & Cudeck, 1993; Garson, 2007) indicate adequate fit. Values over .10 indicate models with poor fit. In addition, following the recommendation of MacCallum, Browne, & Sugahara (1996), the 90% confidence interval is reported for RMSEA. As Byrne (2006, p. 100) notes, the use of confidence intervals provides information on the precision of the estimate of model fit.

The comparative-fit index (CFI; Bentler, 1990) is one member of the family of incremental fit indices, which are also based on the non-centrality parameter. These are based on comparisons of the proposed model with another model, which is generally the null model. Values greater than .95 are considered indicative of good-fitting models (Hu & Bentler, 1999).

Among indices based on residuals, the standardized root mean square residual (SRMR) is the standardized difference between observed variance and covariance and predicted variance and covariance, meaning it estimates the lack of fit in a model compared to a perfect saturated model. Values of less than .08 indicate a good fit (Browne & Cudeck, 1993; Garson, 2007).

A slightly different parsing of these families (e.g., Maruyama, 1998; Tanaka, 1993) combines incremental fit indices and residual-based indices under noncentrality-based indices. The rationale for this class emerges from the notion that structural modeling does not seek to reject the null hypothesis (which is based on the usual χ^2 distribution being a 'central' distribution), so testing should aim to reject an alternative hypothesis, which would be evaluated on an alternate, 'non-central' χ^2 distribution.

Table 8 shows the fit indices reported in the current study.

Table 8
Summary of Fit Indices

Index Family	Index	Significance Level
Absolute Fit	Model χ^2	p > .05 good
Relative Fit	Incremental Fit Index (IFI)	> .90 good, > .95 better
Parsimonious Fit	Root mean square error of approximation (RMSEA) with 90% confidence intervals	< .05 good, > .10 poor
Incremental Fit	Comparative Fit Index (CFI)	> .95 good
Residual-based Fit	Standardized Root Mean Square Residual (SRMR)	< .08 good, < .05 better

Methods Used to Address the Research Questions

The first research question addresses the psychometric properties of the instruments used in this study: *To what extent are the instruments used in this study reliable and valid in the university EFL contexts in this study?* This is assessed using WINSTEPS to validate the instruments and to compare the results obtained with the participants in the current study with previously published data.

The second and third research questions address one measurement model crucial to the Yashima (2002) and Yashima et al. (2004) studies: *To what degree will the 2-factor structure of the L2 Communicative Confidence factor be replicated in this university EFL context? How much will the additional personality variables enhance the L2 Communicative Confidence factor?* This question is evaluated by confirmatory factor analyses of the various permutations of the measurement model.

The fourth research question concerns the replicability of the three models: *To what extent will the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) be replicated in this university EFL context?* This is assessed using SEM to compare the results obtained with the participants in the current study with the previously published data.

The fifth research question concerns the respecification of the three models: *To what degree will data-driven additions improve the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004)?* This is assessed by using the Lagrange multiplier test results to add theoretically justified paths to the model; the fit of the respecified model is then evaluated.

The sixth research question addresses the extension of the three models with the addition of the three personality variables (Perceived Distance, Ego Permeability, and Extroversion) to the structural models: *To what extent will the above L2 communication models benefit from the addition of the personality variables of distancing, ego permeability, and extroversion?* This is assessed using SEM to check the fit indices of the revised models. The results are compared with the fit indices of the original models in this context in order to ascertain if the added variables improve the model fit.

Summary

In this chapter, the methods used in the current study were described. The first two sections examined the participants and the procedures and instruments. The two related studies were then introduced. Next, the Rasch procedure used for data analysis was explained, and the basics of structural equation modeling (SEM) were outlined.

In Chapters 4 through 7 the results of preliminary analyses are covered in detail. Because of the length of those results, they are presented in four separate chapters. Chapter 4 examines initial data screening and validation of the two proficiency instruments, Breadth of Vocabulary Knowledge and Listening Proficiency. In Chapter 5 results of preliminary analyses for the individual difference variables are presented; those variables include Motivation, L2 Communicative Anxiety (both the L2 Communicative Anxiety instrument the FLCAS), Frequency of L2 Communication, L2 Willingness to Communicate, and International Posture. In Chapter 6 validation of the four personality variables (Distancing, Extroversion, Ego Permeability, and Personality) is presented. Finally, Chapter 7 is a discussion of the preliminary analysis results presented in Chapters 4, 5, and 6.

CHAPTER 4

PRELIMINARY ANALYSES: PROFICIENCY INSTRUMENTS

In this chapter I begin to describe the initial analyses. First is an overview of the initial data screening. This is followed by an in-depth look at the two proficiency instruments (Breadth of Vocabulary Knowledge and the Listening Proficiency Test) using the Rasch procedures outlined in the Methods chapter: category function; item-person map; Rasch fit statistics; Rasch separation, reliability, and strata; Rasch principal components analysis of item residuals; and the treatment of misbehaving items.

Initial Data Screening

As detailed in the Methods sections, data from the five percentage-scale instruments were first converted to Likert-scale data. All data were then checked for missing information and improbable values. Of the 302 questionnaires returned, 37 were incomplete and thus deleted, leaving 265 participants. In the 265 surveys, 152 cells of missing data were found (of 53,998 possible responses). These data represented only .28% of the responses and exhibited no particular pattern; because the Rasch analyses account for missing data, no further action was necessary.

Analyses of Proficiency Scales

To identify misfitting items and persons, Rasch analysis (Rasch, 1960) was performed on the two proficiency instruments, Listening Proficiency and Breadth of Vocabulary Knowledge, which had dichotomous and partial-credit scaling, respectively.

Listening Proficiency

The Listening Proficiency instrument was created for this study and thus had not been used elsewhere. The results of a WINSTEPS analysis indicated that all 16 items had reasonable point-measure correlations and adequate fit (Table 9). The Listening Proficiency instrument had an item reliability estimate of .98, a person reliability estimate of .54, item separation of 7.63, person separation of 1.09, and thus a person strata statistic of 1.79. These results indicate that the person difficulty estimates were not well separated in relation to their standard errors.

Figure 9 shows the item-person map for the listening proficiency measure. The person ability estimates covered a range of 65.01 CHIPS, from 34.78 to 64.36, while the item difficulties ranged from 39.46 to 58.23, a span of 18.77 CHIPS. The difference between the means of the person ability and item difficulty estimates was small, just .84 CHIPS (50.00–49.16).

The item-person map shows four gaps between Items 2.3 and 4.4, 3.3 and 2.2, 2.2 and 1.2, and 2.1 and 1.1. The CHIPS gaps were 3.82, 3.05, 3.76, and 2.58, respectively, which indicates that more items would result in more precise person ability estimates. Moreover, some redundancy in terms of item difficulty was present (e.g., Items 23, 44, and 45). The item difficulties aligned generally as expected: The first item in each section was written to be fairly simple, and Items 11, 21, and 31 were the easiest (Item 11 is the first item in the first section, Item 21 is the first item in the second section, and so forth). Three of the five most difficult

Table 9

Listening Proficiency Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
3-2 aloe prior remark	58.23	.73	1.10	.9	1.11	.6	.21
4-4 ship new city name	54.41	.66	.99	-.2	.93	-.6	.38
4-5 ship material	54.41	.66	1.08	1.3	1.18	1.6	.26
2-3 farmer send crops	54.31	.66	.96	-.6	1.18	1.6	.38
4-2 ship name	53.49	.64	1.11	1.8	1.22	2.2	.24
3-5 aloe green thumb	52.79	.63	.92	-1.5	.89	-1.2	.45
3-4 aloe host offer	52.28	.62	.91	-1.7	.87	-1.6	.46
1-3 baseball # transfers	52.11	.62	1.04	.8	1.03	.4	.34
4-1 ship purpose	51.37	.61	1.06	1.2	1.03	.5	.33
4-3 ship original date	50.96	.61	.98	-.4	.97	-.4	.40
3-3 aloe good for burns	50.88	.61	1.04	.9	1.01	.2	.35
2-2 farmer kind of crop	47.83	.61	1.00	.0	.96	-.5	.39
1-2 baseball easy route	44.07	.68	1.00	.1	1.04	.4	.35
2-1 farmer topic	43.00	.71	.93	-.8	.78	-1.7	.44
1-1 baseball confused	40.42	.83	.95	-.4	.77	-1.2	.38
3-1 aloe talk location	39.46	.88	.93	-.5	.85	-.6	.37
M	50.00	.68	1.00	.1	.99	.0	
SD	5.30	.08	.06	1.0	.13	1.1	

Note. *N* = 252, *k* = 16; Pt-M Corr = point-measure correlation.

items (Items 42, 44, and 45) were from the long listening passage (198 words), which was designed to be more challenging than the three dialogues.

The plot of item residuals against the item calibrations indicated no problems. The items appeared randomly distributed, and 51.6% of variance was explained by the measures. The first residual component accounted for 1.6 units (4.8%) of the unexplained variance, figures that lie below the 3.0 (5%) values that suggest the presence of additional components (Linacre, n.d.). In addition, 1.6 eigenvalue units is only slightly above the value of 1.4 that Smith and Miao (1995) found using random data. When the items with positive and negative loadings from the PCA of item residuals were used to produce two sets of person ability estimates, the estimates had a disattenuated correlation of .91, suggesting that the Listening Proficiency instrument was fundamentally unidimensional.

Breadth of Vocabulary Knowledge

The Breadth of Vocabulary Knowledge instrument, based on the work of Nation and Laufer (1999), had four levels with 10 items measuring each level: the 2,000-word level (Items 1-10), the 3,000-word level (Items 11-20), the 5,000-word level (Items 21-30), and the University Word List level (Items 31-40). These data were analyzed using a Rasch partial credit model (i.e., 2 = correct, 1 = partially-correct, and 0 = incorrect). Items that received partial credit fell into two categories: incorrect word forms (*usually* instead of *usual*, the correct answer) or answers with egregious spelling mistakes (*phaz* instead of *phase*, the correct answer). For all 40

```
More proficient      +  More difficult items
listeners
                  .. -|
58                    +  aloe32
                      |
57                    +
                 ###  |
56                    +
                      |S
55              .###  +
                      |  farmer23    ship44      ship45
54                   S+
                      |  ship42
53          .#######  +  aloe35
                      |  aloe34
52           .######  +  baseball13
                      |  ship41
51                    +  aloe33      ship43
     .###########     |
50                    +M
                     M|
49         #########  +
                      |
48                    +  farmer22
        .#########    |
47                    +
                      |
46                    +
              .##### S|
45                    +
                      |S
44                    +  baseball12
               #####  |
43                    +  farmer21
                      |
42                    +
                  ##  |
41                   T+
                      |  baseball11
40                    +
                      |T aloe31
39                  . +
Less proficient         Less difficult items
listeners
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 9. Rasch item-person map of the Listening Proficiency measure.

items, infit values were satisfactory with a range of .62-1.41, which met the .50-1.50 criterion used in this study (Table 10). However, four items had outfit MNSQ values in excess of 1.5 (5k7-appliance, U3-project, 5k15-whole, and 2k8-motor). MNSQ outfit is sensitive to outliers, which indicate unexpected behavior on items far from the person's measure level. Two items (5k7-appliance and 5k15-whole) with slightly high outfit MNSQ values had outfit *t*-values less than the 1.96 criterion that indicates statistical significance (Linacre, 2006). For all four items, temporarily deleting the dozen most misfitting responses yielded outfit MNSQ values with a range of 1.07-1.18. This indicated that the items functioned satisfactorily, so these four items were thus retained.

Table 10
Breadth of Vocabulary Knowledge: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
5k13-devise	61.80	1.93	1.41	.8	.53	-.3	.11
5k7-appliance	55.80	.74	1.03	.2	*1.71	1.3	.25
5k5-stool	55.57	.71	1.25	1.1	1.16	.5	.22
5k11-bruises	55.25	.68	.90	-.5	.35	-1.6	.38
5k3-mess	55.05	.67	1.04	.2	.68	-.6	.34
U13-assess	54.86	.65	.98	.0	.46	-1.3	.37
3k12-whirling	54.59	.63	.91	-.5	1.23	.6	.32
U3-project	54.18	.60	1.10	.7	*2.63	2.7	.29
5k17-gloom	54.03	.59	1.05	.4	.96	.0	.32
5k15-whole	53.21	.53	1.27	1.8	*1.83	1.8	.28
3k9-veins	52.91	.52	1.04	1.0	.82	-.4	.37
3k4-chill	52.41	.49	1.02	.2	1.32	1.0	.35
5k2-phase	52.35	.49	1.21	1.6	.98	.1	.36
3k11-trim	51.90	.47	1.03	.3	.69	-1.0	.44
2k11-examined	51.76	.46	1.07	.7	.93	-.1	.37
2k7-tips	51.67	.46	.75	-2.5	.76	-.7	.45

Table 10 (continues)

Table 10 (continued)
Breadth of Vocabulary Knowledge: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
3k17-supreme	50.89	.43	1.10	1.0	1.09	.4	.42
U17-rational	50.81	.43	.81	-2.0	.64	-1.4	.56
5k12-hugging	50.57	.42	.87	-1.3	.82	-.6	.46
2k13-connects	50.34	.41	1.19	2.0	1.06	.6	.41
U1-vision	50.30	.41	1.15	1.5	1.05	.3	.44
U11-indicates	49.69	.40	.92	-.9	.84	-.7	.54
U5-democracy	49.59	.39	.97	-.3	.80	-.8	.53
2k18-brave	49.09	.38	1.09	1.1	1.02	.2	.49
U18-dynamic	48.71	.38	.97	-.3	.98	.0	.55
3k16-normal	48.58	.38	.93	-.8	.78	-1.2	.56
U4-sex	48.40	.37	1.04	.6	.89	-.6	.54
2k9-copy	48.16	.37	1.28	3.5	1.49	2.4	.41
U12-participate	48.07	.37	.82	-2.6	.70	-1.8	.63
5k6-trumpet	48.01	.37	.89	-1.5	.73	-1.6	.60
3k18-aware	47.83	.37	1.15	2.0	1.10	.6	.48
3k10-assisted	47.40	.36	.86	-2.0	.95	-.3	.55
2k5-skirts	46.27	.36	.62	-6.4	.88	-.1	.52
2k8-motor	44.56	.36	1.09	1.2	*1.54	3.3	.46
2k16-usual	43.91	.37	.84	-2.2	.87	-.8	.57
3k6-structure	43.14	.38	1.06	.8	1.40	2.2	.55
3k1-apartment	40.74	.44	1.18	1.6	1.37	1.4	.51
2k3-nurse	39.44	.50	.95	-.4	.96	.0	.55
2k6-justice	38.30	.55	1.01	.1	1.06	.3	.54
M	50.00	.50	1.03	.0	1.03	.1	
SD	4.66	.25	.16	1.7	.40	1.2	

Note. $N = 252$, $k = 40$; Pt-M Corr = point-measure correlation. 2K, 3K, 5K, and U denote the 2,000-word level, the 3,000-word level, the 5,000-word level, and the University Word List, respectively. Items marked with an asterisk indicate values greater than the 1.5 cutoff criterion used in this study.

The Breadth of Vocabulary Knowledge instrument exhibited good coverage of the range of person abilities as shown in Table 10 and Figure 12. A WINSTEPS analysis revealed that the vocabulary construct had an item reliability estimate of .98, item separation of 7.57, a person reliability estimate of .88, person separation of 2.75, and a person strata statistic of 4.00.

For the Breadth of Vocabulary Knowledge instrument, about 75% of the data were scored and input by a research assistant (a fourth-year undergraduate), so inter-rater reliability between the research assistant and the researcher was checked using Cohen's kappa (Cohen, 1968). Some controversy exists about the use of kappa statistics with interclass correlation statistics available as an alternative (Muller & Buttner, 1994), but in the present study the widely-used Cohen's kappa was chosen and yielded a value of .94, which indicated adequate inter-rater reliability.

As shown in Figure 10, the 40 items of the Vocabulary instrument exhibited various levels of difficulty. The most difficult item was Item 13 on the 5,000-word list (*We'll have to be inventive and de______ a scheme for earning more money*; correct answer = devise), and the least difficult item was Item 6 on the 2,000-word list (*Laws are based on the principle of jus_____*; correct answer = justice). Several items appeared to be misplaced, with, for example, Item 12 on the 3,000-word list (*People were whir_______ around on the dance floor*; correct answer = whirling) being more difficult than several 5,000-word list items, while Item 1 on the 3,000-word list (*I live in a small apa_______ on the second floor*; correct answer = apartment) was less difficult than eight of the 2,000-word list items. The latter lexeme, *apartment*, is a widely-known cognate in Japan, so its position in the Rasch results as an easy item was not surprising. In general, the item difficulties were as

```
Persons with larger          More difficult vocabulary words
vocabularies
62                   +  5k13devise
61                   +
60                   +
59                   +T
58                   +
57                   +
56              .    +  5k5stool           5k7appliance
55              .    +S 3k12whirling       5k11bruises       5k3mess
                        uwl14assess
54             .# T+    5k17gloom          uwl3project
53             .#    +  3k9veins           5k15whole
52            .##    +  2k11examined       2k7tips           3k11trim
                        3k4chill           5k2phase
51         .#####    +  3k17supreme        5k12hugging       uwl17rational
50          .#### S+M   2k13connects       uwl11indicates    uwl1vision
                        uwl5democracy      uwl9crisis
49       .#######    +  2k18brave          3k16normal        uwl18dynamic
48          #####    +  2k9copy            3k18aware         5k6trumpet
                        uwl12participate   uwl4sex
47     .#########    +  3k10assisted
46      ######### M+    2k5skirts
45      .########    +S 2k8motor
44     .#########    +  2k16usual
43          .####    +  3k6structure
42         ###### S+
41          #####    +T 3k1apartment
40              #    +
39            .##    +  2k3nurse
38             .#    +  2k6justice
37              . T+
36             ##    +
35                   +
34             .#    +
33              #    +
Persons with                 Less difficult vocabulary words
smaller vocabularies
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 10. Item-person map for the Breadth of Vocabulary Knowledge instrument.

expected with 5,000-word list items comprising the more difficult items, the 3,000-word list items and UWL items in the middle, and the 2,000-word list items being the easiest.

The range of the Breadth of Vocabulary Knowledge instrument was satisfactory with a span of 23.50 CHIPS (61.80-38.30). The person ability estimates exhibited a similar range of 21.95 CHIPS (33.86 to 55.81); this indicated that the instrument was broad enough to measure the person distribution adequately. However, the difference between item difficulty and person ability means was 4.44 CHIPS (50.00 – 45.56), which indicates that the items on this instrument were somewhat difficult for this sample. This is evident in a visual inspection of Figure 10, in which the item distribution is skewed toward the bottom of the figure.

To investigate the external validity of the two L2 proficiency instruments, the demographic section included a request for proficiency exam scores (e.g., TOEIC); these two instruments should have high correlations with such exams. The most widely reported score was for the TOEIC (n = 59), for which respondents had a mean of 580.03 (SD = 211.68). For this small subsample, the TOEIC and Listening Proficiency correlated at r = .61, TOEIC and Breadth of Vocabulary Knowledge correlated at r = .86, and Listening Proficiency and Breadth of Vocabulary Knowledge correlated at r = .68. However, due to the low reliability (.54) of the Listening Proficiency instrument, these reliabilities were attenuated; the three correlation coefficients corrected for attenuation were .83, .92, and .99, respectively. These corrected correlation coefficients suggest that the somewhat low values of .61 and .68 are due to measurement error and that the actual correlation values are higher (Schumacker & Muchinsky, 1996). These findings offer support for the external validity of the two L2 proficiency instruments.

Summary

In this chapter, the initial data screening was covered, after which psychometric properties of the two L2 proficiency instruments were presented. A Rasch analysis of the Listening Comprehension instrument indicated that the items had very good fit to the Rasch model and adequate reliability. The range of the item difficulties was adequate, yet more items would allow better coverage of the person ability estimates.

A Rasch analysis of the Breadth of Vocabulary Knowledge instrument indicated that the vocabulary items had satisfactory fit to the Rasch model and adequate reliability. The items also had adequate coverage of the person ability estimates although the items were somewhat difficult for the participants.

Having addressed the two L2 proficiency tests, the focus of Chapter 5 is the validation of the eight additional individual difference instruments that were used in the MacIntyre and Charos (1996) study, the Yashima (2002) study, and the Yashima et al. (2004) study.

CHAPTER 5

PRELIMINARY ANALYSES: INDIVIDUAL DIFFERENCE INSTRUMENTS USED IN PREVIOUS STUDIES

In this chapter I cover the initial analyses of the eight instruments employed in the MacIntyre and Charos (1996) study, the Yashima (2002) study, and the Yashima et al. (2004) study. Those eight instruments include L2 Willingness to Communicate (L2 WTC), Frequency of L2 Communication, Perceived Competence in English, Communication Anxiety in English, the Foreign Language Classroom Anxiety Survey (FLCAS), Motivation, International Posture with its four subscales, and Personality with its five subscales. Raw data were screened as outlined in Chapter 4. The first section of this chapter offers an in-depth look at each of the seven instruments with the procedures outlined in the Methods chapter: category function; item-person map; Rasch fit statistics; Rasch separation, reliability, and strata; Rasch principal components analysis of item residuals; and the treatment of misfitting items. In the second section, the results of the confirmatory factor analysis using EQS to investigate the dimensionality of the International Posture instrument are presented.

Analyses of Instruments Used in Previous Communication Models

Following are the seven instruments used in the studies by MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004): L2 Willingness to

Communicate (L2 WTC), Frequency of L2 Communication, Perceived Competence in English, Communication Anxiety in English, the Foreign Language Classroom Anxiety Scale (FLCAS), Motivation, and International Posture and its four subscales. Thereafter the Attitudes and English Experience instruments are presented and analyzed; these two instruments were substantially different than analogous instruments used in the two studies. Analyses include exploratory factor analysis results, the Rasch item statistics, the item-person map, and a summary table (for multi-dimensional instruments).

L2 Willingness to Communicate

The L2 Willingness to Communicate instrument was the PRCA-20 (McCroskey, 1992), a 20-item instrument that includes eight distracters. As mentioned above, the data were converted from percentages to Likert-scale data prior to conducting analyses. When examined with WINSTEPS, the 7-point Likert scale functioned poorly, as the thresholds were disordered. To remedy this situation, categories were combined into various configurations, As shown in Table 11, the 7-category statistics had two category thresholds separated by just .19 CHIPS, far below the necessary separation of 1.87 CHIPS (recall that for seven categories, the minimum acceptable separation is 1.87 CHIPS; see Table 6). Those two categories were thus combined and the Rasch analysis repeated for the 6-category instrument; results indicated that the minimum separation (1.65) was below the minimum of 2.68, so those two categories were combined and Rasch analysis conducted on the new 5-category instrument. This configuration also failed to yield adequate separation, so the closest categories were combined and the analysis repeated for the 4-category instrument. With a minimum threshold separation of 5.54, this configuration met the minimum criterion for a 4-category instrument (5.00).

Table 11
Steps in Combining Categories for L2 Willingness to Communicate

Separation categories	Category Threshold Separation: First iteration	Second iteration	Third iteration	Fourth iteration
Min separation	.19	1.65	3.41	5.54
Separation criterion[a]	1.87	2.68	3.69	5.00

Note. Min separation = minimum separation among the average measures for that iteration. [a]The respective separation criteria are from Table 6.

The raw (percentage) data were then transformed into a symmetrical 4-point scale with correct ordering of thresholds, good fit, and adequate separation (Table 12). The same procedure was employed for the other three instruments that originally used percentages (Perceived L2 Competence, Perceived Distance, and L2 Communicative Anxiety).

Table 12
Category Statistics for L2 Willingness to Communicate

Category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very unwilling	653 (21.53)	-9.77	-9.87	1.11	(none)	
Unwilling	897 (30.60)	-3.23	-2.96	.89	-7.66	.26
Slightly willing	898 (30.06)	2.94	2.66	.95	-.12	.23
Willing	549 (18.81)	8.50	8.65	1.11	7.78	.27

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

The dimensionality of the L2 WTC subscale was then checked. The average inter-item correlation for the 12-item instrument was adequate (r = .50), and internal reliability was high (Cronbach's α = .92). An initial EFA yielded both a 2-component solution that accounted for 67.56% of the variance and a single-component solution that covered 54.67% of the variance. Both solutions had strong component loadings and communalities. Again, however, addressing the question of dimensionality with a PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person measures derived from items with positive and negative residual loadings was .81, suggesting that the L2 WTC instrument was fundamentally unidimensional.

The *WTC* instrument yielded a Rasch item reliability estimate of .99, item separation of 10.07, a Rasch person reliability estimate of .88, person separation of 2.72, and a person strata statistic of 3.96. All 12 items exhibited satisfactory fit statistics and satisfactory point-measure correlations (Table 13). In addition, the PCA of item residuals indicated that the Rasch model explained 76.6% of the variance and that the unexplained variance in the first residual contrast accounted
variance and that the unexplained variance in the first residual contrast accounted
for 2.4 units (4.7%) of the total variance. These values are within the criteria that are suggestive of a single dimension (Linacre, n.d.). As shown in Figure 11, the 12 items were widely distributed. The participants indicated the most reluctance to communicate in the four contexts that involved interacting with strangers. Moreover, for all three groups, speaking in a large meeting corresponded with the

Table 13

L2 WTC Measure: Rasch Item Fit Statistics and Item-item Correlations

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
17-meeting strangers	57.82	.48	1.13	1.4	1.18	1.3	.64
12-line strangers	57.03	.47	1.27	2.8	1.21	1.6	.62
8-group strangers	54.18	.44	1.07	.8	1.10	1.0	.68
3-speech strangers	53.35	.43	1.20	2.2	1.18	1.7	.68
11-meeting acqnts	50.87	.42	.71	-3.8	.77	-2.6	.76
6-meeting friends	50.06	.42	.70	-3.9	.69	-3.8	.78
14-speech friends	49.65	.42	.74	-3.3	.73	-3.2	.77
20-speech acqnts	47.81	.42	.81	-2.3	.80	-2.3	.77
15-group acqnts	47.25	.43	.83	-2.1	.87	-1.5	.75
4-line acqnts	45.45	.44	1.34	3.6	1.45	4.1	.67
9-line friends	43.28	.45	1.16	1.7	1.30	2.6	.70
19-group friends	43.25	.46	1.04	.5	1.09	.8	.73
M	50.00	.35	1.00	-.2	1.03	.0	
SD	4.69	.02	.22	2.6	.24	2.4	

Note. N = 252; k = 12; Pt-M Corr = point-measure correlation; acqnts = acquaintances.

lowest level of WTC. The participants were most willing to communicate when speaking with a group of friends (Item 19) or while in line with a friend (Item 9). However, some unexpected results emerged: the participants were *more* willing to communicate when giving a speech to acquaintances (Item 20) than to friends (Item 14), and giving speeches was not the lowest-rated WTC activity when communicating with strangers or with acquaintances (speaking in a meeting of strangers and speaking in a line with a stranger had the lowest WTC levels, respectively). These results likely illustrate that for a planned activity such as a speech, people generally exhibit more willingness to communicate than for an impromptu (i.e., unplanned) communicative act, as when speaking to someone while waiting in line.

The breadth of the L2 WTC items was adequate as the difficulty estimates covered a span of 14.57 CHIPS (43.25–57.82). The person ability estimates, however, ranged from 29.37 to 67.27, a span of 37.90 CHIPS, so the coverage was considered adequate. The difference between the item difficulty and person ability means was 1.20 CHIPS (48.80 – 50.00), which indicates that the instrument was appropriate for this sample.

```
Persons more willing     Context—less willing to communicate
to communicate
60            ##  +
59            .#  +T
58           .##  +  17meeting strangers
57            .#  +  12line strangers
56            .# S+
55      ########  +S
54           .##  +  8group strangers
53           .##  +  3speech strangers
52          ####  +
51    .#########  +  11meeting acqnts
50         #####  +M 14speech friends     6meeting friends
49          #### M+
48      .#######  +  20speech acqnts
47           .##  +  15group acqnts
46           .##  +
45    .#########  +S 4line acqnts
44             #  +
43            .# S+  19group friends      9line friend
42            .#  +
41           .##  +T
40           .##  +
Persons less willing     Context—more willing to communicate
to communicate
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 11. Item-person map for the L2 WTC instrument.

Frequency of L2 Communication

On the 5-item Frequency of L2 Communication instrument, the 7-point Likert scale functioned poorly, as the thresholds were disordered in the first iteration. To remedy this situation, the categories were combined as shown in Table 14. In the second iteration, three categories (5, 6 and 7) were combined into a single category indicated by a box. However, separation of thresholds was too small for categories 3 and 4, so they were combined in the third iteration. The combined "3" category and 2 were not adequately separated, so they were combined in the fourth iteration, which yielded a 3-point scale resulted with correct ordering of thresholds, good fit, and good separation (Table 15). This procedure was used for all the instruments that originally had 7-point Likert scales.

Table 14
Steps in Combining Categories for Frequency of L2 Communication

Iteration	# of Categories	Categories
First	7	1 – 2 – 3 – 4 – 5 – 6 – 7
Second	5	1 – 2 – 3 – 4 – $\boxed{5 - 5 - 5}$
Third	4	1 – 2 – $\boxed{3 - 3}$ – $\boxed{5 - 5 - 5}$
Fourth	3	1 – $\boxed{3 - 3 - 3}$ – $\boxed{5 - 5 - 5}$

Table 15
Category Statistics for Advance of Step Difficulties for Frequency of L2 Communication

Frequency category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Very limited	234 (19.02)	-8.36	-8.23	1.01	(none)	
Limited	622 (50.57)	.63	.54	.96	-8.85	.41
Some	374 (30.41)	12.19	12.26	1.01	8.85	.39

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

The dimensionality of the Frequency of Communication instrument was investigated using both SPSS and WINSTEPS. The average inter-item correlation for the 3-item instrument was adequate (r = .44), but the internal reliability estimate was slightly low (Cronbach's α = .70), which is understandable given the small number of items. An exploratory factor analysis yielded a single component that accounted for 61.77% of the variance. An analysis of the PCA item residuals from WINSTEPS yielded a disattenuated correlation of .79, suggesting that this instrument was sufficiently unidimensional. In addition, the PCA of item residuals revealed that the Rasch model accounted for 74.0% of the variance. The unexplained variance in the first contrast accounted for 1.6 units (8.6%) of the total variance.

All five items displayed adequate fit (Table 16); the Rasch item reliability estimate of .99 and item separation of 10.66 were good, but the Rasch person

Table 16
Frequency of Communication in English Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
5-speak E outside class	57.45	.59	1.27	3.2	1.29	3.0	.67
1-volunteer in class	54.92	.61	.91	-1.1	.90	-1.2	.71
4-ask Q outside class	53.43	.58	.84	-1.9	.83	-2.0	.75
2-answer when called on	45.15	.61	.82	-2.1	.85	-1.4	.64
3-participate pairwork	39.04	.58	1.16	1.8	1.11	.6	.45
M	50.00	.61	1.00	.0	1.00	-.2	
SD	6.86	.04	.18	2.1	.13	1.8	

Note. N = 252, k = 5; Pt-M Corr = point-measure correlation.

```
68     .########  +
67               T+
66                +                            5-speak E outside class.3
65             #  +
64                +T                           1-volunteer in class   .3
63                +
62          .###  +                            4-ask Q outside class  .3
61                +
60               S+
59            .#  +
58    .#########  +
57                +S
56                +
55                +
54   .##########  +                            2-answer called on     .3
53             .  +
52               M+
51                +
50 .############  +M
49            .#  +  5-speak E outside class.2
48                +                            3-participate pairwork .3
47                +
46     .########  +  1-volunteer in class   .2
45               S+  4-ask Q outside class  .2
44             .  +
43       .#####   +S
42                +
41                +
40                +
39                +
38           .## T+
37                +  2-answer called on     .2
36                +T
35                +
34                +
33                +
32             .  +
31                +
30                +  3-participate pairwork .2
```
Figure 12. Item-person map with Rasch-Thurstone thresholds of the Frequency of L2 Communication instrument.

reliability estimate of .53, person separation of 1.07, and a person strata statistic of 1.76 were low. All five items exhibited adequate point-measure correlations.

The difficulty of the items measuring *Frequency of L2 Communication* covered a span of 3.44 CHIPS (48.50-51.94), but the Rasch-Thurstone thresholds ranged from 30 to 66 (36 CHIPS); while the person measures ranged from 39.59 to 67.85, a span of 28.26 CHIPS. The difference in means between item difficulty measures and person ability estimates was 5.63, which indicated that the participants found these items difficult to endorse. As shown in Figure 12, of the five items, speaking English outside of the classroom (Item 5) predictably yielded the lowest frequency of L2 communication, whereas participating in pairwork (Item 3) had the highest frequency.

Perceived Competence in English

The participants' assessment of their own English competence was investigated with the Perceived Competence in English instrument (Yashima et al., 2004). As mentioned above, the data were converted from percentages to Likert-scale data prior to conducting analyses. However, when examined with WINSTEPS, the 7-point Likert scale functioned poorly, with inadequate separation of structure measures. Combining categories in the same manner as outlined above yielded a 4-category scale with proper ordering, good fit, and adequate separation of thresholds (Table 17).

Next, the dimensionality of the Perceived Competence in English items was investigated. The average inter-item correlation for the 12-item instrument was adequate (r = .58), and the internal reliability estimate was high (Cronbach's α = .95). The initial EFA yielded both a 2-factor solution that accounted for 75.91%

Table 17
Category Function Statistics for Perceived Competence in English

Competence category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
None	599 (20.32)	-12.21	-12.30	1.14	(none)	
Very little	901 (30.56)	-4.10	-3.77	.86	-9.76	.29
Limited	900 (30.53)	3.98	3.71	.94	.01	.24
Good	548 (18.59)	11.47	11.57	1.03	9.76	.29

Note. *N* = 252; Avg Measure = average measure; Exp Measure = expected measure.

of the variance and a single-factor solution that accounted for 62.84% of the variance. Both solutions had strong factor loadings and communalities. The more definitive answer, however, came from an analysis of the PCA of item residuals from WINSTEPS: The disattenuated correlation of person ability estimates from items with positive and negative residual loadings was .95, suggesting that this instrument was strongly unidimensional.

When examined with WINSTEPS, all 12 items displayed adequate fit to the Rasch model and reasonable point-measure correlations. Items 10 and 11 were the easiest to endorse, while Items 1 and 9 were the most difficult (Table 18). A WINSTEPS analysis revealed that the Perceived Competence in English instrument had a Rasch item reliability estimate of .99, item separation of 8.22, a Rasch person reliability estimate of .80, person separation of 1.99, and thus a person strata statistic of 2.99.

The plot of the item residuals against the item calibrations showed a random distribution. In addition, the PCA of item residuals indicated that the Rasch model explained 65.3% of the variance, and the first residual contrast accounted for just

Table 18

Perceived Competence in English: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
9-meeting strangers	60.43	.53	1.27	2.6	1.51	2.9	.62
1-speech strangers	59.63	.52	1.41	3.9	1.52	3.1	.60
3-group strangers	54.53	.48	1.01	.1	.99	-.1	.72
8-line strangers	53.84	.47	1.27	2.9	1.22	2.1	.68
5-meeting acquaint	50.58	.47	.77	-2.8	.74	-3.0	.79
2-meeting friends	49.86	.46	.83	-2.1	.83	-2.0	.79
12-speech acquaint	49.72	.46	.74	-3.3	.73	-3.2	.80
6-speech friends	47.98	.46	.90	-1.2	.91	-.9	.79
7-group acquaint	45.38	.47	.84	-1.9	.83	-1.8	.81
4-line acquaint	43.74	.48	.88	-1.4	.90	-.9	.81
11-group friends	42.30	.50	1.08	.9	1.06	.6	.79
10-line friends	41.03	.51	1.09	.9	1.17	1.2	.79
M	50.00	.49	1.01	-.1	1.03	-.2	.76
SD	6.08	.02	.21	2.3	.26	2.1	

Note. *N* = 252, *k* = 12; Pt-M Corr = point-measure correlation. Acquaint = acquaintances.

3.1 units (8.8%) of the unexplained variance. The variance accounted for and the small number of localized units accounted for (3.1) were good, while the percentage is slightly above the suggested level of 5% (Linacre, n.d.). As shown in Figure 13, the 12 items on the Perceived Competence in English instrument covered the range of person ability estimates well. The four items dealing with interactions with strangers were predicted to be areas with lower perceived competence, which was borne out by the results. Speaking with a friend in line and speaking with a group of friends were viewed as contexts in which participants would have the highest levels of perceived competence.

```
Persons, more perceived    Contexts, less perceived competence
competence
 66            ## T+
 65                +
 64            .#  +
 63                +
 62            ##  +T
 61            ##  +
 60            .#  +  1-speech strangers    9-meeting strangers
 59           ###  +
 58        ###### S+
 57      .#######  +
 56         #####  +S
 55         .####  +  3-group strangers
 54         #####  +  8-line strangers
 53         #####  +
 52       .######  +  5-meeting acqnts
 51         #####  +
 50        .##### M+M 12-speech acqnts      2-meeting friends
 49       .######  +
 48      .#######  +  6-speech friends
 47           ###  +
 46          ####  +
 45         .####  +  7-group acqnts
 44          .###  +S 4-line acqnts
 43          .###  +
 42            .# S+  11-group friends
 41         #####  +  10-line friend
 40        ######  +
 39          ####  +
 38           ###  +T
 37                +
 36           .##  +
 35            .#  +
 34                +
 33            ## T+
Persons, less perceived    Contexts, more perceived competence
competence
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 13. Item-person map for the Perceived Competence in English instrument.

L2 Communicative Anxiety

The 12-item L2 Communicative Anxiety instrument (MacIntyre & Charos, 1996) was one of two instruments used to measure anxiety. As mentioned above, the data were converted from percentages to Likert-scale data prior to conducting the analyses. However, when examined with WINSTEPS, the 7-point Likert scale functioned poorly, with structure measures not adequately separated. Combining categories yielded a 4-category scale with proper ordering, good fit, and adequate separation of thresholds (Table 19).

Table 19
Category Function Statistics for L2 Communicative Anxiety

Anxiety category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
No anxiety	592 (20.09)	-7.87	-8.48	1.14	(none)	
Very little	958 (33.57)	-3.55	-2.82	.81	-7.71	.26
Limited	913 (31.55)	2.52	2.21	.86	-.09	.22
Some	477 (16.07)	7.80	7.70	1.06	7.80	.28

Note. $N = 252$; Avg Measure = average measure; Exp Measure = expected measure.

Next, the dimensionality of the L2 Communicative Anxiety items was investigated. The average inter-item correlation for the 12-item instrument was adequate ($r = .55$), and the internal reliability estimate was high (Cronbach's $\alpha = .93$). The initial confirmatory factor analysis yielded both a 2-factor solution that accounted for 78.07% of the variance and a 1-factor solution that accounted for 59.42% of the variance. Both solutions had strong factor loadings and communalities. An analysis of the PCA of item residuals from WINSTEPS indicated that the disattenuated correlation of person ability estimates from items with positive and negative residual loadings was .76, suggesting that this instrument was perhaps not unidimensional.

The plot of the item residuals against the item calibrations showed a random distribution. In addition, the PCA of item residuals indicated that the Rasch model explained 70.6% of the variance, and the first residual component accounted for 4.3 units (10.7%) of the unexplained variance. Based on the disattenuated correlation and the values for the first residual contrast that were in excess of the respective criteria the cutoff value used in this study, the composition of the components from the PCA of residuals was examined.

The content of the respective components of the positive and negative loadings is suggestive of different dimensions (Table 20). The salient

Table 20
Item Loadings from the Rasch PCA of Residuals for the L2 Communicative Anxiety Instrument

Item loadings	MNSQ	
	Infit	Outfit
Positive loadings		
9. Meeting with strangers.	1.65	1.49
1. Speech with strangers	1.61	1.45
3. Group strangers	.89	.83
8. Line with strangers	.85	.94
2. Meeting with friends	.77	.74
5. Meeting with acquaintances	.53	.52
Negative loadings		
11. Group friends	1.13	1.37
10. Line with friends	1.37	1.62
7. Group with acquaintances	.82	.99
4. Line with acquaintances	1.15	1.33
12. Speech with acquaintances	.57	.60
6. Speech with friends	.54	.53

characteristics of the items with positive loadings include anxiety when speaking with strangers and in meetings. For items with negative loadings, speaking with friends or acquaintances and in two informal settings (in a group or in line) were the primary defining points. This arrangement coincided with the 2-factor solution from the confirmatory factor analysis above.

The two subscales were then examined with WINSTEP, and all items on the respective L2 Communicative Anxiety subscales showed good fit to the model (Tables 21 and 22). Although the original fit statistics for the 1-dimension configuration indicated six items were misfitting, in the 2-dimension all 12 items had adequate fit statistics, which indicates the separate subscales better represent the structure of the L2 Communicate Anxiety variable.

A WINSTEPS analysis revealed that the Friend / Acquaintance Anxiety subscale instrument had a Rasch item reliability estimate of .95, item separation of 4.24, a Rasch person reliability estimate of .85, person separation of 2.38, and thus a strata statistic of 4.51. In addition, the PCA of item residuals indicated that the Rasch model explained 70.7% of the variance, and the first residual contrast accounted for 2.5 units (12.1%) of the unexplained variance. The variance accounted for and the small number of localized units accounted for (3.1) were good, while the percentage is slightly above the suggested level of 5% (Linacre, n.d.). The disattenuated correlation of person ability estimates from items with positive and negative residual loadings was .91, suggesting that this instrument was strongly unidimensional.

Table 21
L2 Communicative Anxiety Measure, Friend / Acquaintance Anxiety Subscale: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
10-line friends	53.38	.59	1.05	.6	1.01	.1	.83
11-group friends	52.10	.58	.77	-2.7	.76	-2.6	.87
4-line acquaint	51.51	.58	1.11	1.1	1.08	.8	.83
7-group acquaint	49.60	.58	.66	-4.1	.65	-4.1	.89
12-speech acquaint	47.45	.57	1.16	1.7	1.17	1.8	.81
6-speech friends	45.95	.57	1.21	2.2	1.20	2.1	.80
M	50.00	.53	.99	-.2	1.00	-.1	
SD	2.55	.01	.11	1.2	.14	1.5	

Note. N = 252, k = 12; Pt-M Corr = point-measure correlation; Misfitting values are indicated with an asterisk. Acquaint = acquaintances.

Of the six items, Item 6 (*I would feel anxious presenting a speech to a group of friends*) was the easiest to endorse, indicating that it was the most anxiety-inducing scenario, while Item 10 (*I would feel anxious talking with a friend while standing in line*) was the most difficult to endorse and thus the least anxiety-inducing situation.

The *Stranger Anxiety subscale* instrument had a Rasch item reliability estimate of .96, item separation of 4.61, a Rasch person reliability estimate of .85, person separation of 2.38, and thus a strata statistic of 4.51. In addition, the PCA of item residuals indicated that the Rasch model explained 75.9% of the variance, and the first residual contrast accounted for just 1.9 units (7.8%) of the unexplained variance. The variance accounted for and the small number of localized units accounted for (3.1) were good, while the percentage is slightly above the suggested level of 5% (Linacre, n.d.). The disattenuated correlation of person ability estimates from items with positive and negative residual loadings was .88, suggesting that this instrument was strongly unidimensional.

Table 22
L2 Communicative Anxiety Measure, Stranger Anxiety Subscale: Rasch Item Fit Statistics

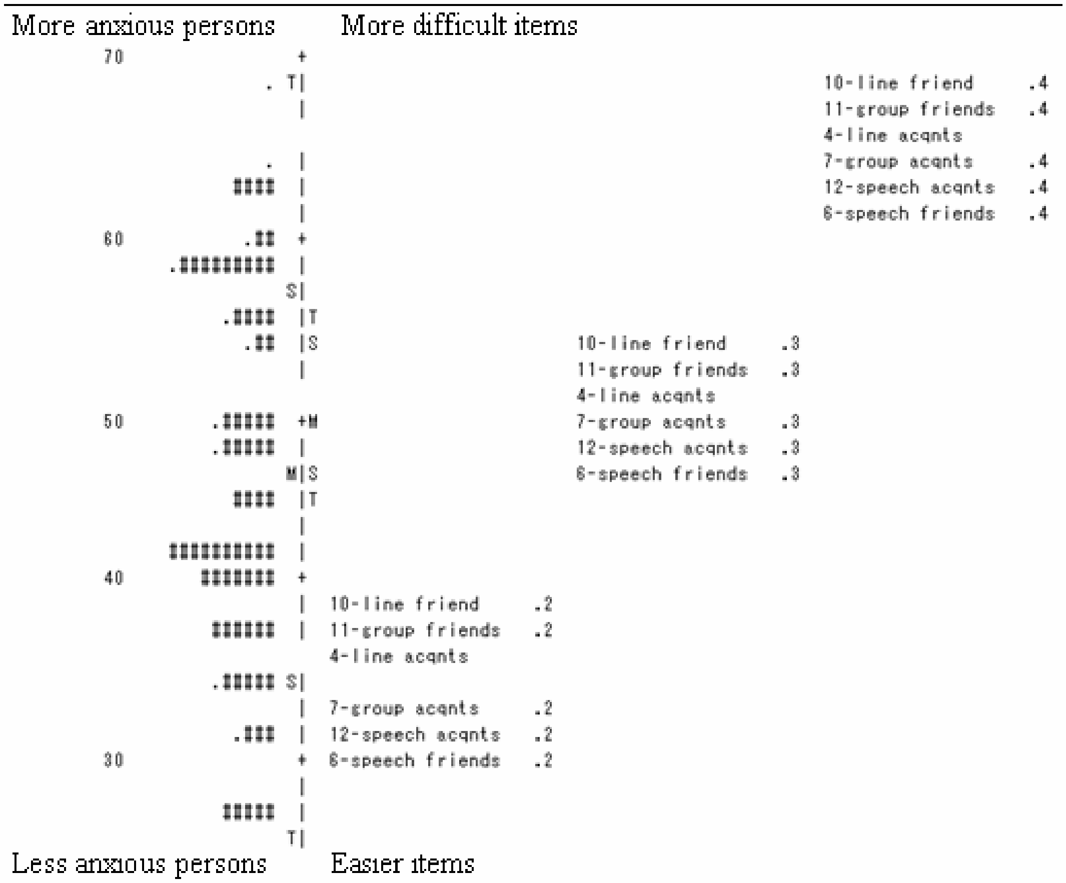
Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
8-line strangers	52.57	.52	1.06	-1.5	1.13	1.2	.83
5-meeting acquaint	52.21	.52	.91	-6.8	1.11	1.1	.84
2-meeting friends	51.61	.52	1.11	-2.6	1.16	1.6	.82
3-group strangers	50.58	.52	.78	-1.2	.75	-2.7	.88
9-meeting strangers	46.96	.54	1.02	6.1	.91	-.9	.87
1-speech strangers	46.10	.54	1.04	5.7	.93	-.6	.86
M	50.00	.53	.99	-.2	1.00	-.1	
SD	2.55	.01	.11	1.2	.14	1.5	

Note. N = 252, k = 12; Pt-M Corr = point-measure correlation; Misfitting values are indicated with an asterisk.

Of the six items, Item 1 (*I would feel anxious presenting a speech to a group of strangers*) was predictably the easiest to endorse, indicating that it was the most anxiety-inducing scenario, while Item 8 (*I would feel anxious talking with a stranger while standing in line*) was the most difficult to endorse and thus the least anxiety-inducing situation.

The existence of a second dimension in the L2 Communicative Anxiety variable is not entirely unexpected because the instrument focuses on two factors, the type of interaction (making a public speech, for example) and the nature of interlocutor (friend, acquaintance, or stranger). The importance of the interlocutor(s) was prominently displayed in Kang’s (2005) study, in which situational WTC was found to be affected by a host of interlocutor factors: the language of interlocutor, knowledge about the interlocutor's proficiency, the relative difference in language proficiency, relative familiarity with the person, the number of interlocutors, and the interest, attitudes and responses of the interlocutor(s). These findings were echoed by Cao and Philp (2006), who found that among the factors that L2 learners perceived as influencing their WTC behavior in class were familiarity with and participation by interlocutor(s). While these studies addressed the role of interlocutor(s) vis-à-vis L2 WTC, the same influences can be posited with respect to communicative anxiety, which underpins L2 WTC.

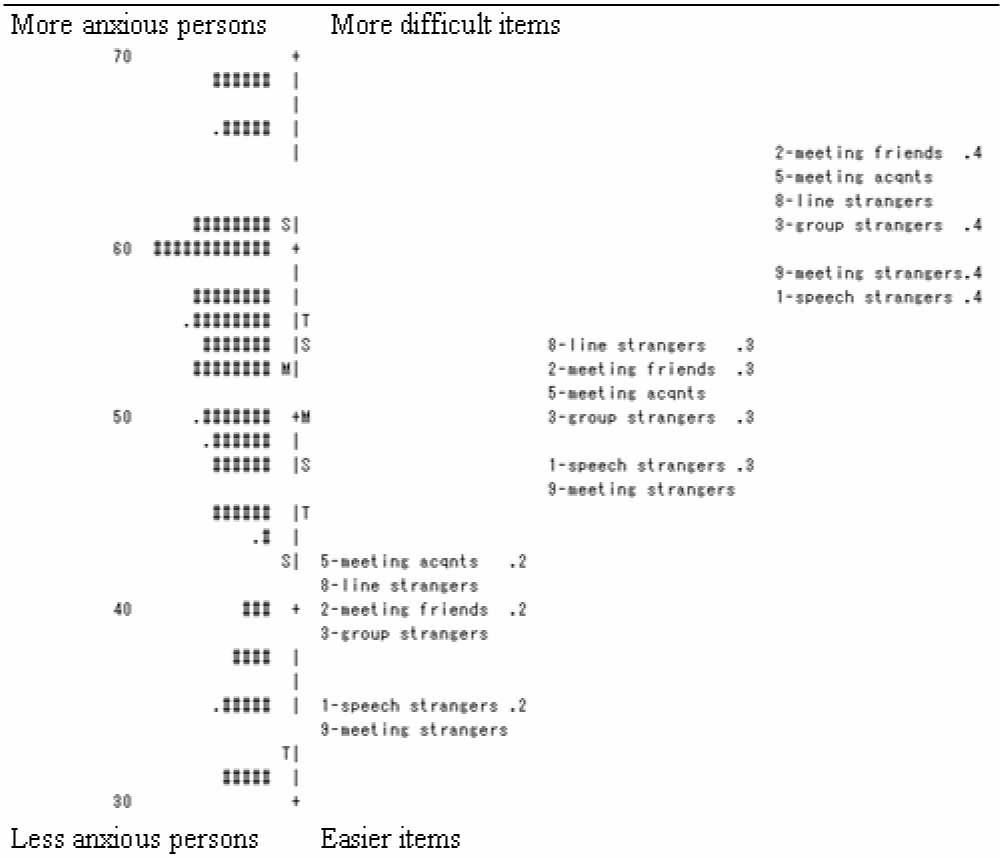
The item-person maps (Figures 14 and 15) indicated that although the range of the item means was somewhat limited when compared with the range of person ability estimates, the Rasch-Thurstone thresholds indicated adequate coverage. The difference between item difficulty and person ability means was only .73 CHIPS, which indicates that the instrument was appropriate for the participants in this study. As shown in Figure 14, interactions with friends and acquaintances were less anxiety-inducing than interactions with strangers. Giving a speech and speaking in a meeting induced nearly the same level of anxiety in each of the three groups, but doing so with friends (Items 2 and 6) was, oddly, more anxiety-inducing than doing so with acquaintances (Items 5 and 12). However, a certain distance and perhaps reticence (e.g., to express criticism) is likely more prominent in speaking with acquaintances than when speaking with friends; this might explain why less communicative anxiety was perceived in this scenario.



Note. M = mean; S = 1 *SD*; T = 2*SD*.

Figure 14. Item-person map for the L2 Communicative Anxiety, Friend / Acquaintance Anxiety subscale.

In subsequent analyses of the MacIntyre and Charos (1996) model, the estimates of Rasch person measures from the two subscales were averaged and the L2 Communicative Anxiety variable treated as a measured variable.



Note. M = mean; S = 1 *SD*; T = 2*SD*.

Figure 15. Item-person map for the L2 Communicative Anxiety, Stranger Anxiety subscale.

Foreign Language Classroom Anxiety Survey

The second anxiety instrument was the Foreign Language Classroom Anxiety Survey (FLCAS; Horwitz et al., 1986), a 33-item instrument that used a 7-point Likert scale. For the 33-item instrument, the average inter-item correlation was adequate (r = .35), and internal reliability was high (Cronbach's α = .95). However, Items 2, 5, 6, 7, and 8 correlated poorly with the scale with average inter-item correlations of .27, .27, .16, .23, and .25, respectively, so these five items were treated as candidates for deletion pending the results of the following analyses.

When examined with WINSTEPS, the 7-point Likert scale functioned poorly with disordered thresholds. To remedy this situation, categories were combined into various configurations. Ultimately, the 7-point scale was reduced to a 4-point scale that had correct ordering, good fit, and good separation (Table 23).

Table 23
Category Function Statistics for the FLCAS

Anxiety category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
No anxiety	1067 (12.21)	-4.58	-4.24	1.17	(none)	
Very little	2262 (25.88)	-.04	-.37	.97	-5.62	.17
Limited	3192 (36.51)	2.60	2.73	.92	-.35	.12
Some	2221 (25.41)	6.06	6.04	1.07	5.97	.13

Note. *N* = 252; Avg Measure = average measure; Exp Measure = expected measure.

An initial exploratory factor analysis using SPSS was conducted to investigate the dimensionality of the FLCAS items. A 2-factor solution accounted for 43.69% of the variance with factor loadings ranging from .29 to .77, and a 1-factor solution accounted for 34.13% of the variance with factor loadings ranging from .29 to .77. Item 6 (*During English class, I find myself thinking about things that have nothing to do with the course*), Item 7 (*I keep thinking that the other students are better at English than I am*), and Item 8 (*I am usually at ease during tests in my English class*) loaded below the .40 cutoff point and exhibited low communalities in both solutions, and because all three items also had low inter-item correlations, they were considered candidates for deletion. Both components exhibited good reliability and sufficient inter-item correlations.

Previous research (Elwood, 2005) has suggested that the FLCAS is unidimensional, and an analysis of the PCA of item residuals from WINSTEPS indicated that the disattenuated correlation of items with person measures from positive and negative residual loadings was .83, which indicates that this instrument was fundamentally unidimensional. The Rasch model explained 58.1% of the variance, and the first residual contrast accounted for 3.2 units (4.1%) of the unexplained variance; this first contrast would thus consist of just three items of the total of 30 items, which is too few to warrant further consideration.

A WINSTEPS analysis yielded an item reliability estimate of .97, item separation of 6.14, a person reliability estimate of .92, person separation of 3.51, and thus a person strata statistic of 5.01; moreover, all 30 items had reasonable point-measure correlations. However, two items were slightly misfitting. Item 6 (*I often think about other things in English class*) had an infit MNSQ value of 1.31 and an outfit MNSQ value of 1.67, so it was checked for the influence of unexpected responses. Twelve persons (4.5%) showed unusual responses; temporarily deleting those persons resulted in improved fit statistics with an infit MNSQ statistic of 1.05 and an outfit MNSQ statistic of 1.07. Item 7 (*I always feel that the other students are better at English than I am*) yielded an infit value of 1.54 and an outfit value of 2.10. It had 14 unexpected responses (5.3%), which when temporarily deleted yielded markedly improved fit statistics of infit MNSQ = 1.08 and outfit MNSQ = 1.01. As responses from a small group of persons appeared to be the cause of the misfit, Items 6 and 7 were retained. Rasch item fit statistics and inter-item correlations for the FLCAS items are shown in Table 24.

Table 24

FLCAS Measure: Rasch Item Fit Statistics and Inter-Item Correlation

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
21-study confused	55.44	.38	.98	-.2	.96	-.5	.56
19-afraid correct all mis	54.10	.37	.92	-1.0	1.03	.4	.56
31-others laugh at me	53.35	.36	.80	-2.7	.83	-2.2	.68
26-more nervous E class	52.42	.36	.76	-3.3	.81	-2.4	.69
25-class pace too fast	52.34	.36	.76	-3.3	.75	-3.3	.67
3-tremble called on	52.29	.36	.81	-2.6	.80	-2.5	.70
30-too many E rules	52.29	.36	1.05	.7	1.07	.9	.58
29-nerv not every word	52.19	.36	.97	-.3	1.04	.5	.59
16-even prepped, nervous	51.98	.36	.78	-3.0	.78	-2.9	.68
17-not go to English	51.79	.36	1.24	2.9	1.25	2.9	.55
27-confused in E class	51.15	.37	.70	-4.2	.70	-4.1	74
4-afraid not understand	51.07	.36	.77	-3.1	.78	-2.9	.67
10-conseq failing E	50.85	.37	1.44	5.0	1.39	4.2	.55
12-nerv, forget things	50.53	.37	.99	-.1	1.05	.6	.54
5-not OK more E classes	50.29	.37	1.35	4.0	1.36	4.0	.48
8-not at ease E tests	50.23	.37	1.37	4.2	1.49	5.2	.42
2-worry about mistakes	49.94	.37	1.22	2.5	1.31	3.5	.47
9-panic if no prep	49.86	.37	.84	-2.1	.82	-2.2	.69
6-think about other things	49.64	.37	1.31	3.6	*1.67	6.7	.32

Table 24 (continues)

Table 24 (continued)

FLCAS Measure: Rasch Item Fit Statistics and Inter-Item Correlation

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
14-native speaker not OK	49.20	.38	1.18	2.1	1.18	2.0	.58
22-feel pressure to prep	49.06	.38	1.18	2.2	1.19	2.2	.45
32-not comfy native spkr	48.63	.38	.99	-.1	.96	-.4	.63
15-upset not catch correct	48.35	.38	.95	-.6	.95	-.6	.55
24-self-conscious speak E	48.28	.38	.78	-2.9	.75	-3.1	.70
20-heart pounds call on	48.21	.38	.72	-3.7	.71	-3.7	.71
13-embarrass volunteer	48.13	.38	.87	-1.6	.84	-1.9	.65
33-nervous if no prep	47.92	.39	.85	-1.9	.81	-2.2	.70
28-not conf going to E	47.67	.39	.86	-1.8	.86	-1.6	.61
1-unsure in E class	47.33	.39	.94	-.7	.90	-1.1	.68
23-others speak better	46.99	.40	1.23	2.6	1.35	3.5	.41
7-other students better	46.86	.40	*1.54	5.4	*2.10	9.1	.37
28-not conf going to E	47.67	.39	.86	-1.8	.86	-1.6	.61
1-unsure in E class	47.33	.39	.94	-.7	.90	-1.1	.68
23-others speak better	46.99	.40	1.23	2.6	1.35	3.5	.41
7-other students better	46.86	.40	*1.54	5.4	*2.10	9.1	.37
18-not conf in E class	46.01	.41	.87	-1.5	.86	-1.5	.60
11-why others upset	45.60	.42	1.18	2.0	1.17	1.6	.47
M	50.00	.38	1.01	-.1	1.05	.2	
SD	2.35	.01	.23	2.7	.31	3.2	

Note. N = 252, k = 30; Pt-M Corr = point-measure correlation; acqnts = acquaintances; nerv = nervous; mis = mistakes; prep = preparation; conf = confident.

The breadth of the FLCAS item difficulties was 10.28 CHIPS (45.30 to 55.58). The person ability estimates, however, ranged from 35.65 to 67.16, a span of 21.51 CHIPS, meaning that the instrument did not adequately measure the tails of the distribution. In addition, considerable redundancy in terms of item difficulty estimates was present in the 30 items. The difference between item difficulty and person ability means was 1.94 CHIPS, which indicates that the instrument was somewhat easy to endorse for this sample and that participants exhibited some anxiety.

As shown in Figure 15, the majority of items were relatively easy to endorse, thus indicating a substantial degree of anxiety in the foreign language classroom. The items easiest to endorse dealt with limited *personal* confidence (e.g., Items, 1, 18, and 28) and the feeling that other students were better (Items 7 and 23). Interestingly, the participants expressed little anxiety about being laughed at (Item 31), which suggests that group cohesion plays an important role. When prepared for English class, the participants indicated lower levels of anxiety (Item 16), but with inadequate preparation they felt nervous (Item 33). One surprising result was that the participants did not strongly agree that “[They] feel overwhelmed by the number of rules you have to learn to speak English” (Item 30). In light of the considerable attention devoted to grammar minutiae in English instruction in Japan, it was expected that students would strongly endorse this item, yet that was not the case. Although somewhat puzzling, it might reflect the common use of grammar rules as test questions rather than as tools to be used while speaking English.

```
More anxious         Items showing more classroom anxiety
persons

61            . T+
              .  |
60           .#  +
              .  |
59         .###  +
              .  |
58            .  +
             ##  |
57           .#  +
       .##### S|
56          ###  +
        #######  |  21-study confused
55           ##  +T
             ##  |
54    .########  +  19-afraid correct all mis
      .########  |  31-others laugh at me
53    .########  +
        .#####   |S 25-E class pace too fast       26-more nervous in E class
                    3-tremble called on            30-too many E rules
52       ###### M+  16-even prepped, nervous       17-not go to English
                    29-nervous not every word
        .#####   |
51      #######  +  10-conseq failing E            27-confused in E class
                    4-afraid not understand T
          .###   |  12-nervous, forget things
50        ####   +M 2-worry about mistakes         8-not at ease E tests
                    9-panic no prep
      .######    |  6-think other things
49    .######    +  22-pressure to prep
     ########    |  15-upset not catch correction
48          ##   +  13-embarrass re volunteer      20-heart pounding called on
                    24-self-conscious speaking E   33-nervous no prep
           .##   |S 28-not confident going to E
47          .# S+   1-unsure in E class            23-others speak better E
            .#   |  7-other students better
46           .   +  18-not confident in E class
             #   |  11-understand why others upset
45           .   +T
           .##   |
44           #   +

Less anxious         Items showing less classroom anxiety
persons
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 16. Item-person response map for the FLCAS.

Motivation

The Motivation instrument is from Yashima's (2002) study and uses items originally from Gardner and Lambert's (1972) study. It consists of 12 items in two 6-item subscales, Desire to Learn English (Items 1-6), and Motivational Intensity (Items 7-12). For the 12-item instrument, the average inter-item correlation was adequate (r = .43), and the internal reliability estimate was high (Cronbach's α = .95). However, Items 1 and 5 correlated poorly with the scale, as indicated by their inter-item correlations of .22 and .28, respectively; Item 1 was treated as a candidate for deletion pending the results of the following analyses. Recall that Item 5 (*I absolutely believe English should be taught at school*) was removed and used in the Attitudes about the Learning Situation scale, but its low correlation indicates that it adds little to the Motivation scale.

When examined with WINSTEPS, the 7-point Likert scale was problematic with category 3 being underutilized; this caused the thresholds to be disordered. Combining categories yielded a 4-category alignment with proper ordering, good fit, and adequate spacing; Item 1 was removed as explained below, yielding adequate category function statistics for the 10-item Motivation instrument (Table 25).

In the revised, 4-category Motivation instrument, all 11 items had reasonable point-measure correlations, but Item 1 (*When I have assignments to do in English, I try to do them immediately*) had an infit value of 1.71 and an outfit value of 1.91, so it was checked for the influence of unexpected responses. 20

Table 25
Category Function Statistics for Motivation

Motivation category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very weak	264 (10.40)	-8.33	-9.12	1.20	(none)	
Weak	677 (26.65)	-2.00	-1.51	.89	-9.23	.38
Low-medium	966 (38.03)	4.29	4.26	.95	-.24	.25
Medium	633 (24.92)	11.44	11.28	1.00	9.47	.27

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

persons (7.5%) had unusual responses, and temporarily deleting responses from 13 persons (5%) resulted in slightly improved fit statistics with an infit MNSQ = 1.54 and outfit MNSQ = 1.75, which are still misfitting. As Item 1 appeared to be poorly fitting, it was deleted from further analysis. Rasch item fit statistics and inter-item correlations for the 10-item, 4-category *Motivation* instrument are shown in Table 26.

Next, the dimensionality of the Motivation instrument was investigated. The average inter-item correlation for the 10-item instrument was adequate (r = .52), and the estimate of internal reliability was quite high (Cronbach's α = .90). A 2-factor solution accounted for 63.15% of the variance with factor loadings ranging from .39 to .97, and a 1-factor solution accounted for 52.96% of the variance with factor loadings ranging from .63 to .77. Components in both solutions exhibited good reliability and adequate inter-item correlations. An analysis of the PCA of item residuals from WINSTEPS indicated that the disattenuated correlation of person ability estimates from items with positive and negative residual loadings was .84, suggesting that this instrument was fundamentally unidimensional

Table 26
Motivation Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
2-read outside class	54.63	.45	1.45	4.0	1.33	2.9	.72
7-study E harder	54.21	.44	.95	-.6	.96	-.4	.72
8-think about E learned	52.36	.44	.78	-2.8	.75	-2.9	.76
10-long hours studying E	52.02	.44	.90	-1.2	.88	-1.3	.73
4-want more E classes	52.02	.44	1.23	2.6	1.23	2.4	.68
3-concentrate in E class	49.43	.45	1.01	.1	1.17	1.8	.66
6-E most interesting	46.97	.47	.87	-1.6	.87	-1.3	.72
11-try hard to learn E	46.00	.48	.84	-1.9	.78	-2.2	.73
9-self-study if no E class	45.85	.48	.99	.0	.88	-1.1	.72
12-after uni continue E	43.31	.51	1.07	.8	.94	-.4	.70
M	50.00	.46	1.00	-.1	.98	-.3	
SD	4.25	.02	.18	2.0	.19	1.9	

Note. *N* = 252, *k* = 10; Pt-M Corr = point-measure correlation; subj = subject; uni = university.

although it was originally posited as separate subscales, Desire to Learn English and Motivational Intensity.

As shown in Table 27, the variance explained by the model (76.8%), the number of localized units (2.1) in the first contrast, and the percentage of variance explained by the first contrast (4.8%) are all within acceptable ranges (Linacre, n.d.).

Table 27
PCA of Residuals for Motivation

Index Family	Localized Units	Percentage
Total variance	43.0	100.0%
Variance explained	33.0	76.8%
Unexplained variance	10.0	23.2%
First contrast	2.1	4.8%
Suggested criteria[a]	3.0	5.0%

The suggested criteria for the variance explained and the values for the first criteria are from Linacre (n.d.).

Moreover, when the content of the respective components of the positive and negative loadings are examined, the three strongest loadings from each are not indicative of different dimensions (Table 28). Incidentally, these loadings (Items 7-12) are all from items included in the original Motivational Intensity subscale, which suggests that the original subscale did not represent a dimension distinct from the Desire to Learn English subscale.

Table 28
Three Strongest Item Loadings from the Rasch PCA of Residuals for Motivation

Index family
Positive loadings
12. After university, I plan to continue studying English.
11. I try hard to study English.
9. I would study by myself if there were no English classes.
Negative loadings
7. I study English harder than my classmates.
10. I spend long hours studying English.
8. I think about things that I learned in my English class(es).

Because this would represent a fundamental change in the configurations tested via SEM, this was further investigated with a confirmatory factor analysis using EQS. The results of that confirmatory factor analysis indicated that the 1-

factor model and the 2-factor model had nearly identical fit statistics; those for the 2-factor model were χ^2 (32) = 115.262 (p < .01), CFI = .928, IFI = 929, SRMR = .053, RMSEA = .102, and 90% C.I. = .082 - .122. These numbers are suggestive of barely adequate fit of both the 1-factor and 2-factor configurations to the data, which does not definitively answer the question of dimensionality.

However, in looking at the content of the items, I'm not convinced that two distinct subscales are present. For example, Item 12 (the easiest item to endorse), which was originally in the Motivational Intensity subscale, deals with continuing to learn English after finishing college; however, it could just as easily fall under the Desire to Learn English subscale, and my inclination is that Item 12 is more indicative of 'desire' than 'motivational intensity'.

Thus, based on (a) the strong disattenuated correlation of .84 and the adequate statistics from the first contrast of the PCA of item residuals, (b) the ambiguous finding that both configurations had reasonable fit statistics, (c) ambiguous theoretical footing for two separate subscales based on the content of the items, and (d) a more parsimonious configuration with one factor instead of two, the Motivation instrument was treated as a single dimension in this study.

For the 10-item Motivation instrument, the PCA of item residuals indicated that the Rasch model explained 76.8% of the variance, and the unexplained variance in the first residual component accounted for 2.1 units, which was 4.8% of the total unexplained variance. The 10-item Motivation instrument yielded an item reliability estimate of .99, item separation of 8.86, a person reliability estimate of .86, person separation of 2.45, and thus a person strata statistic of 3.60.

As shown in Figure 17, the Motivation instrument exhibited reasonable coverage of the persons. Items were generally positioned as expected with several items

```
More motivated          Items indicating less motivation (difficult to endorse)
persons
67           .##  +
66                +
65           .##  +
64                +
63           ###  +
62                +
61           .##  +
60           .## S+
59           ###  +
58                +T 2-read E outside class
57         #####  +
56       .######  +
55           ###  +
54      .#######  +S 7-study E harder than classmates
53        .##### M+
52    .#########  +  10-long hours studying E         4-want more E classes
                     8-think about E learned in class
51         .####  +
50        .#####  +M
49          .###  +  3-concentrate in E class
48          ####  +
47            .#  +  6-E most interesting subject
46            ## S+S 11-try hard to learn E           9-self-study if no E class
45           ###  +
44             .  +
43            .#  +  12-after uni continue E
42                +T
41             .  +
40             #  +
39             # T+
Less motivated          Items indicating more motivation (easy to endorse)
persons
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 17. Item-person response map for the Motivation instrument.

indicating motivated behavior in class (e.g., Items 3, 6, and 11) and a strong propensity toward future study of English (Item 12). However, behavior outside class (e.g., Item 2, reading English materials outside class) was endorsed less, which likely reflects how busy the students are (or, unfortunately, that perhaps they don't read much).

The breadth of the range of Rasch-Thurston thresholds of the Motivation instrument was 27.85 CHIPS, while the range of person ability estimates was from 35.49 CHIPS indicating that the instrument covered the distribution adequately. The difference between item difficulty and person ability means was 3.08 CHIPS (53.08 - 50.00), which indicates that the items on the instrument were somewhat easy to endorse for these participants.

International Posture

The International Posture instrument was from Yashima's (2002) study. It originally consisted of four subscales with a total of 23 items, and in the current study three items were added to the two original items of the Interest in Foreign Affairs subscale. A 7-point Likert scale was used in the current study.

First, to investigate the dimensionality of the instrument, an exploratory factor analysis was conducted using SPSS. A principal components analysis with orthogonal rotation and then with oblique rotation was requested; the best solution had four components that accounted for 42.96% of the variance. Five items (12, 17, 18, 23, and 26) failed to achieve the cutoff loading point of .40, with Item 18 loading at just -.16. Moreover, the four subscales emerged with several changes in their respective configurations. The first factor, the International Approach-Avoidance Tendency subscale, originally included Items 1-7 but gained Item 11 (*I'm interested in volunteer activities in developing countries such as participating in Youth International Development Assistance*) and Item 12 (*I don't think what's happening overseas has much to do with my daily life*). The fourth factor, the Interest in International Vocations/Activities subscale, originally was made up of six items (8-13), but Items 11 and 12 loaded on the International Approach-Avoidance Tendency subscale, leaving four items (8, 9, 10, and 13). The third factor, the Interest in International News subscale, originally consisted of Items 14-18, from which Items 17 and 18 were deleted. Finally, Item 17 (*International news makes interesting, useful content for school classes*) loaded on the second factor, the Intercultural Friendship Orientation subscale (originally Items 19-26). The exploratory factor analysis indicated that Item 18 (*International news is too difficult to understand*) did not load on any of the four subscales, and it was deleted from further analyses. All four components exhibited adequate internal reliability estimates (Cronbach's alpha) with the two shorter subscales (Interest in International Vocations/Activities and Interest in Interest in Foreign Affairs) having slightly lower reliability (Table 29).

Table 29

26-Item International Posture Measure Rotated Pattern Matrix

Item	Approach-Avoid	Cultural Friendship	Foreign Affairs	Vocation	h^2
Ipos1	.76				.69
Ipos3	.72				.49
Ipos5	.69				.59
Ipos7	.62				.46
Ipos4	.58				.38
Ipos6	.57				.42
Ipos2	.47				.30
Ipos11*	.44				.39
Ipos12*	*.30				.27
Ipos20		.59			.34
Ipos19		.58		.	.63
Ipos22		.57			.65
Ipos25		.53			.51
Ipos21		.52			.27
Ipos24		.50			.67
Ipos23		*.37			.17
Ipos26		*.36			.45
Ipos17*		*.33			.31
Ipos15			.76		.60
Ipos14			.63		.52
Ipos16			.43		.22
Ipos8				.63	.36
Ipos13				.62	.48
Ipos9				.50	.49
Ipos10				.40	.44
Ipos18*				*-.16	.07
% of var	29.07	5.68	4.53	3.68	
Eigenvalue	7.56	1.48	1.18	.96	
Reliability	.86	.80	.72	.64	
I-I correl	.40	.36	.39	.37	

Note. N = 252, k = 26; Extraction Method: Principal components analysis; Rotation Method: Oblimin with Kaiser Normalization. Items marked with an asterisk changed from their original subscales. Underlined loadings indicate the item did not achieve the cutoff value of .40 used in this study. % of var = percentage of variance accounted for.

The item performance of each subscale was then checked using WINSTEPS. The subscales are described in the order of the size of their respective eigenvalues.

Intergroup Approach-Avoidance Tendency Subscale. On the revised International Approach-Avoidance Tendency subscale, WINSTEPS yielded poor category function with improperly ordered structure measures and inadequate separation. However, combining the categories yielded a 4-category alignment with proper ordering, good fit, and adequate spacing (Table 30).

On the revised Intergroup Approach-Avoidance Tendency subscale, the results from the WINSTEPS analysis indicated that all nine items had very good fit statistics (Table 31). The subscale had person separation of 2.00, a Rasch person

Table 30
Category Function Statistics for the Revised Intergroup Approach-Avoidance Tendency Subscale

Approach-Avoidance category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Strongly avoid	399 (17.66)	-6.10	-5.98	1.05	(none)	
Avoid	756 (32.43)	-2.10	-2.21	.93	-6.95	.29
Weakly approach	791 (34.05)	1.72	1.77	.96	-.50	.23
Approach	391 (17.86)	6.96	6.95	1.12	7.45	.31

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

reliability of .80, item separation of 5.50, and a Rasch item reliability of .97. Item 1 (*I want to make friends with international students studying in Japan*) and Item 6 (*I would not feel somewhat uncomfortable if a foreigner moved in next door*) were the easiest to endorse, whereas Item 3 (*I would talk to an international student if there were one at school*) was the most difficult to endorse. This was rather surprising given that respondents quite readily endorsed the item concerning wanting to make friends with international students in Japan (Item 1), yet it might indicate that students would approach international students more readily if they themselves had chosen to do so (i.e., because they want to make friends). On the other hand, Item 3 might tap into student reluctance to engage in spontaneous conversation, which might well be unplanned and therefore anxiety-inducing.

Table 31
Intergroup Approach-Avoidance Tendency Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
3-talk intnl students	53.70	.42	.86	-1.6	.86	-1.6	.69
11-intnl volunteer	52.28	.41	1.29	3.2	1.30	3.2	.65
4-live w/ intnl students	51.34	.41	1.31	3.4	1.33	3.5	.62
7-help foreigner in store	51.01	.41	.96	-.4	.95	-.6	.67
5-volunteer foreigners	50.45	.41	.82	-2.3	.80	-2.4	.76
2-talk to foreigners	49.93	.41	1.15	1.8	1.23	2.6	.61
12-overseas related	47.13	.41	.93	-.8	.96	-.4	.59
1-friends intnl students	46.82	.42	.95	-.7	.95	-.6	.65
6-foreigner next door	46.72	.42	.73	-3.5	.74	-3.2	.74
M	50.00	.42	1.00	-.1	1.01	.1	
SD	2.43	.00	.19	2.3	.21	2.3	

Note. N = 252, k = 9; Pt-M Corr = point-measure correlation; intnl = international. The item descriptions for Items 2, 6, and 12 reflect the recoding of the items so all items had the same valence.

A second interesting pair of items was Item 4 (*I wouldn't mind sharing an apartment or room with an international student*), which students were reluctant to endorse, and Item 6 (*I would not feel somewhat uncomfortable if a foreigner moved in next door*), which they generally agreed with. This seems to reflect a propensity to allow foreigners to live in proximity (i.e., next door), but not *too* close.

The dimensionality of the Intergroup Approach-Avoidance Tendency subscale was then investigated. The average inter-item correlation for the 9-item instrument was adequate (r = .40), and the internal reliability estimate was high (Cronbach's α = .87). An exploratory factor analysis yielded a two-component solution that accounted for 55.70% of the variance. Loadings on both components were adequate and communalities ranged from .41 to .71. Addressing this question with a PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived using items with positive and negative residual loadings was .96, suggesting that this instrument was strongly unidimensional. The PCA of residuals indicated that the Rasch model accounted for 60.7% of the variance and the first contrast accounted for 1.9 localized units, which was 8.2% of the variance explained by the first contrast. Although the first two values were satisfactory, the 8.2% value is slightly high.

Thus, in lieu of (a) the hypothesized composition of the scale as a single dimension, (b) the strong results from the initial confirmatory factory analysis (eigenvalue = 7.57), (c) the strong disattenuated correlation result, and (d) the adequate results from the Rasch PCA of residuals, the Intergroup Approach-

Avoidance Tendency subscale was treated as a single dimension. The breadth of the Intergroup Approach-Avoidance Tendency subscale was 4.38 CHIPS (48.31-52.69), which is narrower than the range of person ability estimates (Figure 17). The items showed some redundancy, but the instrument was appropriate for this sample with a difference of only .39 CHIPS between the mean item difficulty and the mean of person ability estimates.

Intercultural Friendship Orientation Subscale. On the revised Intercultural Friendship Orientation subscale (C-Friend; Items 17, 19-26), category function was investigated and yielded a series of hills with properly ordered difficulty, yet separation was inadequate. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 32).

The dimensionality of the revised International Cultural Friendship Orientation subscale was investigated next. The average inter-item correlation for the 9-item instrument was adequate (r = .36), and the internal reliability estimate was high (Cronbach's α = .83). An exploratory factor analysis yielded a one-component solution that accounted for 44.70% of the variance with good loadings (.47 to .83) and communalities from .22 for Item 23 to .69 for Item 22. However, a 2-component solution accounted for 56.00% of the variance with stronger loadings and communalities. A PCA of residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived from items with

```
Persons more likely     Items indicating less likelihood of approaching
to approach
  62          .##  +
  61              T+
  60         .###  +
  59                +
  58          .##  +
  57           ##  +
  56          .## S+
  55         .###  +T
  54         ####  +  3-talk to intnl ss at school
  53         ####  +
  52        .####  +S 11-intnl vol activities        4-live with intnl ss
  51         ####  +  5-volunteer activity w/ forgnr  7-help forgnr in store
  50      .###### M+M 2-talk to forgnrs
  49 .#############  +
  48      .######  +S
  47     .#######  +  1-friends intnl ss in Japan     12-overseas related to life
                      6-comfy forgnr next door
  46         ####  +
  45         .###  +T
  44         #### S+
  43          .##  +
  42           .#  +
  41                +
  40            #  +
  39           .# T+
  38                +
Persons less likely     Items indicating more likelihood of approaching
to approach
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 18. Item-person map for the Intergroup Approach-Avoidance Tendency subscale.

Table 32

Category Function Statistics for the Revised Intercultural Friendship Orientation Subscale

Cultural Friendship Orientation	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Little interest	226 (9.96)	-4.28	-5.52	1.38	(none)	
Slight interest	637 (28.07)	-1.55	-1.10	.93	-8.00	.37
Some interest	888 (39.14)	3.07	3.36	.89	-.79	.24
Strong interest	518 (22.83)	9.41	8.85	.93	8.80	.27

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

positive and negative residual loadings was .82, which indicates the presence of a single dimension. To further investigate this question, a confirmatory factor analysis was conducted. As shown in Table 33, although the 1-dimension model had adequate fit, the 2-dimension model fit the data slightly better: $\chi^2 = 51.213$ ($p < .01$), CFI = .968, IFI = .968, SRMR = .030, RMSEA = .063, and 90% C.I. = .037-.087.

Table 33
Summary of Fit Indices for 1-Factor and 2-Factor Intercultural Friendship Orientation Models

	1-factor	2-factor
Reliability Coefficient (rho)	.849	.860
Multivariate Kurtosis		
Mardia's coefficient	17.904	17.904
Normalized estimate	10.039	10.039
Residuals		
Average absolute standardized residuals	.024	.028
Average off-diagonal absolute standardized residuals	.029	.035
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 36)	673.030	673.030
χ^2 (df = 27, 26)	58.184	43.320
Probability value for the χ^2 statistic	.000	.018
χ^2/df ratio	2.155	1.666
Fit Indices		
Comparative fit index (CFI)	.951	.973
Incremental fit index (IFI)	.952	.973
Standardized root mean square residual (SRMR)	(.050)	(.040)
Root mean-square error of approximation (RMSEA)	.068	.052
RMSEA 90% confidence interval	.044-.092	.022-.078

Although the analyses indicated that a 2-component configuration was plausible, the decision was reached to treat the Intercultural Friendship Orientation as a single dimension based on (a) its theoretical basis as a single dimension, (b) the disattenuated value of .82, and (c) the adequacy of both the 1-component and the 2-component configurations.

The Intercultural Friendship Orientation instrument was then examined with WINSTEPS. The scale had an item reliability estimate of .98, item separation of 6.53, a person reliability estimate of .77, person separation of 1.82, and thus a person strata statistic of 2.64. All nine items exhibited adequate fit and reasonable point-measure correlations for their respective subscales (Table 34). The easiest items to endorse dealt with getting to know various people (e.g., Item 19, *[English] will allow me to meet and converse with more and varied people*). The other end of

Table 34
Intercultural Friendship Orientation Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
21-necessary for Net	55.35	.42	1.30	3.5	1.34	3.7	.53
23-help tests Eiken	53.89	.41	1.35	3.9	1.41	4.5	.53
26-necessary future job	52.32	.41	1.23	2.6	1.22	2.6	.63
17-intnl news content	49.97	.42	1.01	.1	1.13	1.4	.57
24-join cultural active	49.03	.43	.74	-3.3	.75	-3.2	.75
25-info in English	47.64	.44	.81	-2.4	.78	-2.7	.72
20-get job in future	47.60	.44	1.02	.3	1.00	.0	.58
22-know culture’s people	47.23	.44	.69	-3.9	.66	-4.2	.77
19-meet various people	46.97	.44	.84	-1.9	.81	-2.1	.71
M	50.00	.43	.99	-.1	1.00	.0	
SD	2.95	.01	.11	1.3	.23	3.0	

Note. *N* = 252, *k* = 9; Pt-M Corr = point-measure correlation.

the spectrum was more concerned with instrumental motivation such as using the Internet (Item 21, *A reason to study English is that it is necessary for using the Internet*), and thus was only tangentially related to the notion of friendship. Based on my teaching experience, a reasonable explanation is that Japanese students have seldom used English on the Internet and avoid doing so unless absolutely necessary. The range of the Intercultural Friendship Orientation subscale was 8.38 CHIPS (46.97-55.35). The subscale showed some redundancy and did not cover the tails of

```
Persons with more       Items indicating less friendship orientation
friendship orientation
65           .##  +
64               T+
63        .#####  +
62                +
61          ####  +
60          ####  +
59                +
58       .###### S+
57         .####  +
56       .######  +T
55                +  21-necessary for Net
54 .############  +  23-help tests Eiken
53   ########### M+S
52 .############  +  26-necessary future job
51   .##########  +
50     .########  +M 17-intnl news class content
49      ########  +  24-join cultural activities
48     #########  +  20-get job in future        25-info in English
47           .## S+S 19-meet various people      22-know cultures people
46        ######  +
45          ####  +
44          ####  +T
43           .##  +
42            .#  +
41             # T+
Persons with less       Items indicating more friendship orientation
friendship orientation
```

Note. M = mean; S = one standard deviation; T = two standard deviations.
Figure 19. Item-person map for the Intercultural Friendship Orientation Subscale.

the distribution as well as could be hoped (Figure 18). The subscale was somewhat easy for these respondents to endorse with a difference between item difficulty and person ability in means of 2.57; this, however, would indicate a reasonably high degree of Intercultural Friendship Orientation (a desirable quality in our students!).

Interest in International Vocation/Activities Subscale. The revised Interest in International Vocation/Activities subscale (I-vocation; Items 8, 9, 10, and 13) was investigated using WINSTEPS. The scale yielded disordered category thresholds and inadequate separation. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 35).

Table 35
Category Function Statistics for the Revised Interest in International Vocation/Activities Subscale

Motivation category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Little interest	199 (20.73)	-6.05	-6.17	1.07	(none)	
Slight interest	324 (33.75)	-2.85	-2.72	.82	-6.66	.42
Some interest	311 (32.40)	1.02	1.00	1.05	-.73	.36
Strong interest	126 (13.13)	5.89	5.80	.99	7.39	.51

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

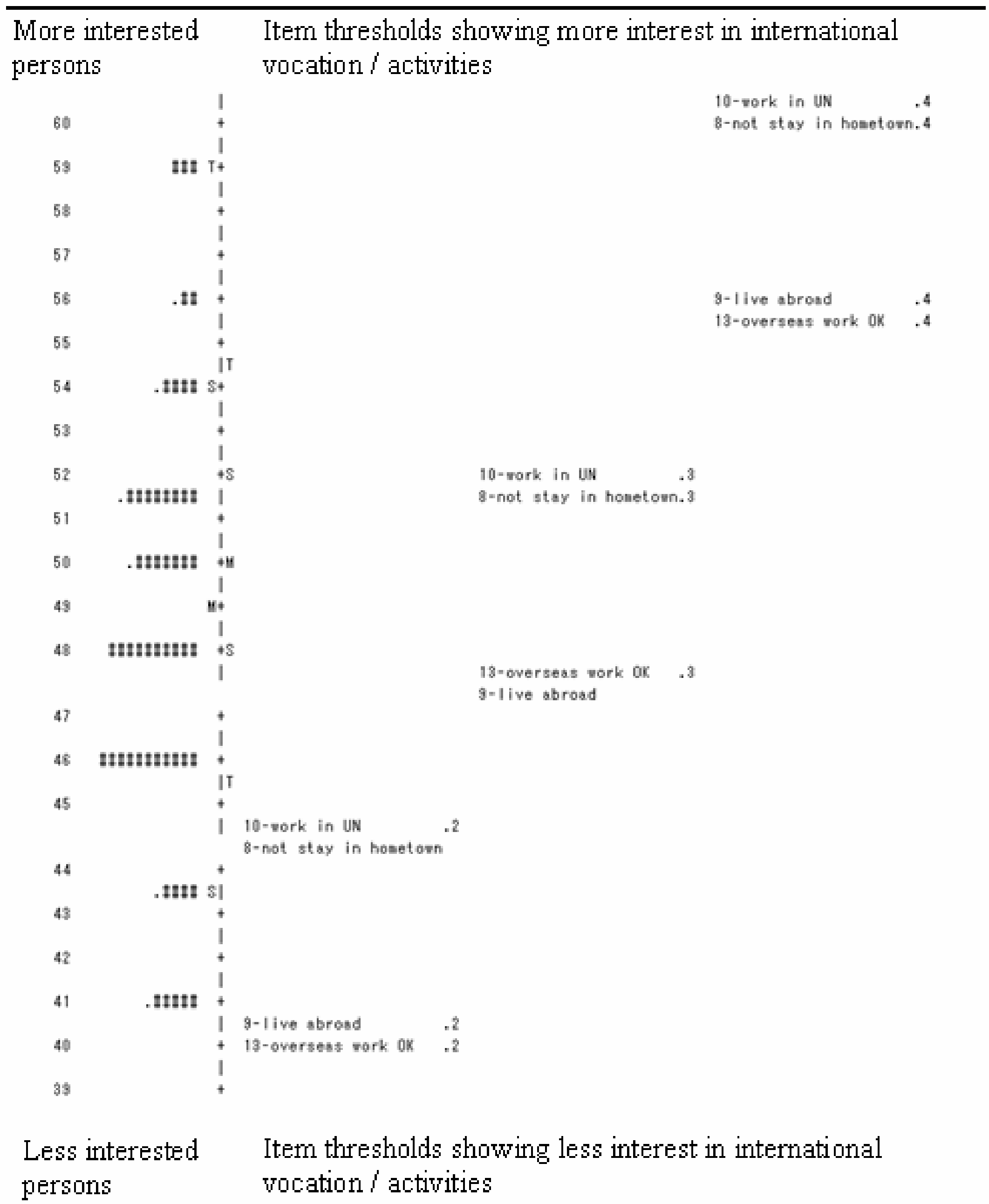
The dimensionality of this subscale was investigated. The average inter-item correlation for the 4-item instrument was adequate (r = .39), and the estimate of internal reliability was also satisfactory (Cronbach's α = .73). An initial EFA yielded a one-component solution that accounted for 55.18% of the variance. Loadings on the single component were strong (.67 to .80) and communalities were adequate, from .45 to .63. A PCA of item residuals showed that the disattenuated correlation of person measures derived from items with positive and negative residual loadings was .52, suggesting that this instrument was *not* unidimensional. However, splitting the subscale further would have resulted in two 2-item subscales; such small scales are at best minimally adequate for defining a construct, so the Interest in International Vocation/Activities subscale was treated as a single dimension.

Four of the values were satisfactory with a Rasch item reliability estimate of .96, item separation of 5.14, a Rasch person reliability estimate of .47, person separation of .95, and thus a person strata statistic of 1.60; however, the reliability was quite low, which would be problematic for SEM. As shown in Table 36, all four items exhibited adequate fit with reasonable point-measure and inter-item correlations. In addition, the PCA of item residuals indicated that the Rasch model

Table 36
Interest in International Vocation/Activities Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
10-work in UN	52.30	.41	1.10	1.2	1.08	.9	.68
8-not stay in hometown	52.11	.41	1.10	1.2	1.12	1.4	.67
9-live abroad	47.89	.41	.91	-1.1	.89	-1.3	.76
13-overseas work OK	47.71	.41	.90	-1.1	.90	-1.1	.72
M	50.00	.41	1.02	.2	.99	-.1	
SD	2.20	.00	.10	1.1	.10	1.2	

Note. N = 252, k = 4; Pt-M Corr = point-measure correlation.



Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 20. Item-person map for the Interest in International Vocation/Activities subscale.

accounted for 55.5% of the variance, and the unexplained variance in the first residual component accounted for 1.5 units (16.6%) of the total variance.

Of the four items, Items 9 and 13 (living overseas or frequently traveling overseas for work) were the easiest to endorse, while working for the United Nations or a similar organization (Item 10) was the most difficult to endorse. Item 8 (*I would rather [not] stay in my hometown*)[11] was surprisingly difficult to endorse, but in lieu of the ongoing movement of people from rural areas of Japan to urban areas, this result was not completely unexpected.

The breadth of the Interest in International Vocation/Activities subscale was limited as the item difficulty estimates covered a span of 5.19 CHIPS (47.71-52.90) and the category thresholds covered about 20 CHIPS. The person ability measures, however, ranged from 37.19 to 63.25, a span of 26.06 CHIPS, meaning that the instrument measured just the center of the distribution (Figure 20). The difference between item difficulty and person ability means was 1.83 CHIPS, which indicates that the items were slightly difficult to endorse for this sample.

Interest in Foreign Affairs Subscale. Finally, the revised Interest in Foreign Affairs subscale (Items 14-16) yielded disordered category thresholds and inadequate separation when examined with WINSTEPS. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 37).

[11] This item was reverse-coded so the valence matched the other items on the Interest in International Vocation / Activities subscale.

Table 37
Category Function Statistics for the Revised Interest in Foreign Affairs Subscale

Interest category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Little interest	142 (19.40)	-12.01	-12.30	1.15	(none)	
Slight interest	351 (47.95)	-5.37	-4.95	.90	-12.95	.53
Some interest	193 (26.37)	1.98	1.93	1.03	1.30	.46
Strong interest	46 (6.28)	8.39	8.39	.85	11.65	.82

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

As shown in Table 38, all three items exhibited adequate fit to the model, and all three items had reasonable point-measure correlations. Rasch statistics were satisfactory with a Rasch item reliability estimate of .97, item separation of 5.85, a Rasch person reliability estimate of .46, person separation of .92, and thus a person strata statistic of 1.56.

The dimensionality of the Interest in Foreign Affairs subscale was then checked. The average inter-item correlation for the 3-item instrument was adequate (r = .37), but internal reliability (Cronbach's α = .64) and the Rasch reliability (.46) were low, which was not unexpected given the small number of items. A confirmatory factor analysis indicated the presence of one factor with good loadings that accounted for 42.17% of the variance.

The PCA of item residuals indicated that the Rasch model accounted for 57.1% of the variance, and the unexplained variance in the first residual component accounted for 1.6 units (22.9%) of the total variance. In spite of its marginal reliability, it was included in the current study to allow replication of the Yashima et al. (2004) model.

Table 38

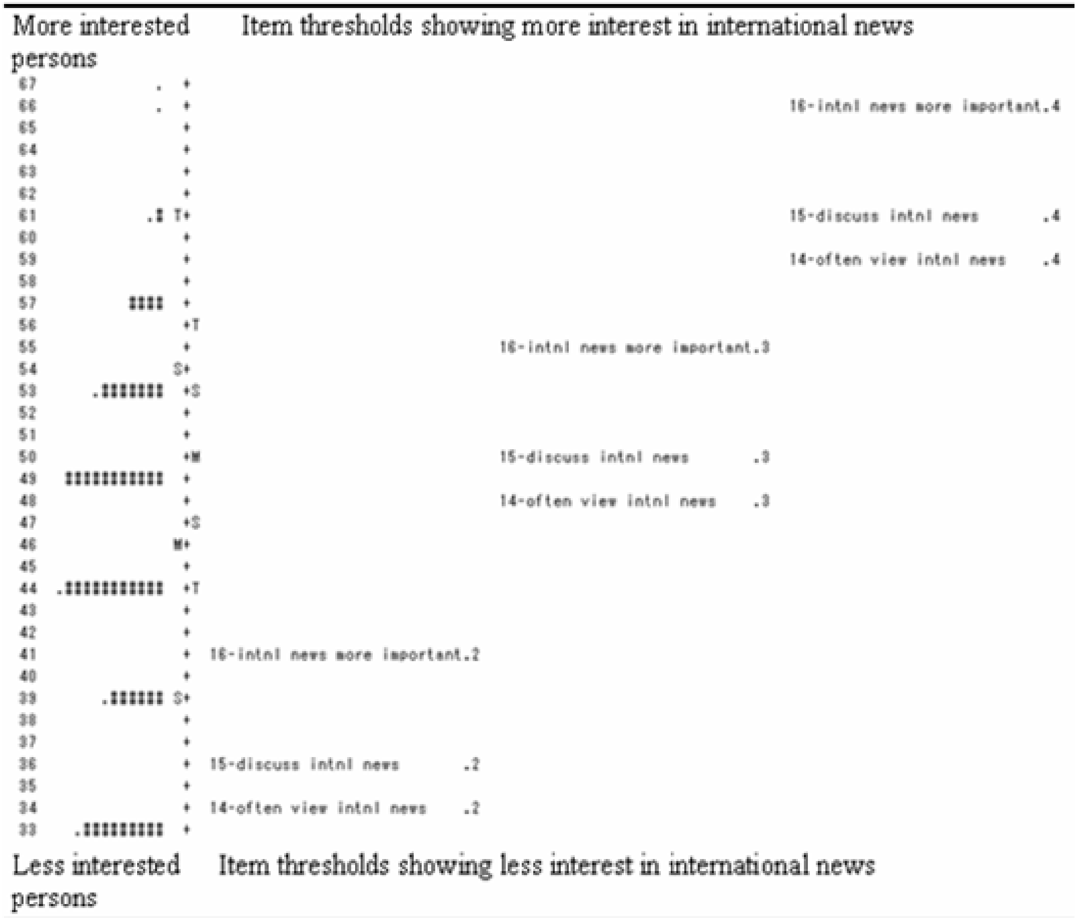
Interest in Foreign Affairs Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
16-intnl news important	54.22	.54	1.22	2.3	1.23	2.3	.62
15-discuss intnl news	49.15	.50	.89	-1.3	.90	-1.2	.81
14-often view intnl news	46.63	.49	.86	-1.7	.87	-1.6	.80
M	50.00	.51	.99	-.3	1.00	-.2	.74
SD	3.16	.02	.17	1.8	.17	1.8	

Note. *N* = 252, *k* = 3; Pt-M Corr = point-measure correlation. intnl = international.

As shown in Figure 20, the three items covered the person distribution reasonably well. Item 14 (*I often read and watch news about foreign countries*) was the most easily endorsed, with discussion of international news (Item 15) being somewhat less easy to endorse. The item most difficult to endorse was Item 16, *International news is more important than local news*.

Although the breadth of the item measure means of the Interest in Foreign Affairs subscale was limited with a span of 7.59 CHIPS (46.63-54.22), the range of the category thresholds was much larger at about 32 CHIPS. The person measures, however, ranged from 33.21 to 65.84, a larger span of 26.06 CHIPS (Figure 20). The difference between the mean item difficulty and the mean of the person ability estimates ability estimates was 5.10 CHIPS, which indicates that the items were somewhat difficult to endorse for this sample.



Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 21. Item-person map for the Interest in Foreign Affairs subscale.

A summary of the International Posture subscales and the overall International Posture instrument is shown in Table 39. The number of items decreased from 26 to 25 with the deletion of Item 18, and the analyses yielded reconfigurations of several subscales. Items 11 and 12 were moved from the Interest in International Vocation/Activities subscale to the Intergroup Approach-Avoidance subscale, and Item 17 was moved from the Interest in International News subscale to the Interest in International Vocation/Activities subscale.

Table 39
Subscale Correlation Coefficients and Rasch Reliability and Separation Statistics for the International Posture Subscales

Category	1	2	3	4
Number of items	9	4	3	9
Correlation				
1. Approach-avoidance tendency				
2. Interest in vocation/activities	.54			
3. Interest in foreign affairs	.40	.27		
4. Intercultural friendship orien	.66	.55	.39	
Item reliability	.97	.96	.98	.99
Item separation	5.34	5.17	6.54	4.63
Person reliability	.74	.47	.56	.75
Person separation	1.67	.95	1.13	1.73

Note. Orien = orientation.

With the four subscales adequately defined and all sufficiently unidimensional, the question at hand then became which of the four subscales to include in the International Posture instrument. In Yashima (2002), all four subscales were used, while in Yashima et al. (2004), the Intercultural Friendship Orientation was omitted based on item overlap with the other three subscales. If that were the case, then inter-item correlations should be excessively high. However, five of the eight items dealt with international things, while three dealt specifically with interacting with people in international contexts. The items dealing with interacting with foreigners (i.e., all the items of the Approach-Avoidance Tendency subscale and the three from the Intercultural Friendship Orientation) would, in a sense, overlap in that the basic action of all those items is interaction. However, the inter-item correlations were not excessively high with a maximum of .56.

This question of which of the four subscales to include in the International Posture instrument was addressed with a confirmatory factor analysis using EQS (this was an assessment of one of the measurement models for the SEM). The best model was the 2-factor configuration with Intergroup Approach-Avoidance Tendency and Intercultural Friendship Orientation; statistics indicated reasonable fit of the model to the data with χ^2 (32, N = 252) = 185.716 (p < .01), CFI = .935, IFI = .937, RMSEA = .066, and 90% C.I. = .052-.080.

In spite of that particular result, the earlier factor analysis yielded four factors, raising the question of why two factors did not enter the new configuration. One possibility is that both had relatively few items and were therefore not well defined. A second possibility is that the two shorter subscales were subsumed by the two strong factors. For example, Interest In Foreign Affairs could be a manifestation of an amicable orientation toward other cultures (which is conceptually close to Intercultural Friendship Orientation). Similarly, Interest in International Vocations/Activities would, if acted upon, necessarily involve approaching and interacting with foreigners. To explore this issue further, a confirmatory factor analysis was conducted using 24 of the original 26 items (Items 12 and 18 were deleted earlier). However, this model exhibited poor fit to the data with χ^2 (251, N = 252) = 625.912 (p < .01), CFI = .814, IFI = .816, RMSEA = .077, and 90% C.I. = .070-.085. As shown in Table 39, the four subscales had moderate correlations, and the individual items were not highly correlated, with a maximum correlation of .64.

SEM Analysis of the Dimensionality of the International Posture Scale

Because Rasch analysis of the International Posture scale yielded configurations different than originally posited, a confirmatory factor analysis using SEM was conducted to investigate further the dimensionality of the International Posture instrument.

The original configuration of International Posture consisted of four subscales, but as noted earlier, both the Interest in Foreign Affairs subscale and the Interest in International Vocation/Activities subscale included a small number of items and had suspect reliability. Thus, the configuration of the entire 4-factor instrument was investigated with a confirmatory factor analysis using EQS. The 4-factor model fit the data poorly, while the 2-factor model with Intergroup Approach-Avoidance Tendency and Intercultural Friendship Orientation displayed much better fit: χ^2 (32, N = 252) = 185.716 (p < .01), CFI = .935, IFI = .937, RMSEA = .066, and 90% C.I. = .052-.080. Statistics for the two models are presented in Table 40, and the standardized solution for the 2-factor model is shown in Figure 21.

Table 40

Summary of Fit Indices for 2-Factor and 4-Factor International Posture Models

	2-factor	4-factor
Reliability Coefficient (rho)	.900	.914
Multivariate Kurtosis		
Mardia's coefficient	37.206	70.347
Normalized estimate	11.573	15.520
Residuals		
Average absolute standardized residuals	.046	.053
Average off-diagonal absolute standardized residuals	.052	.057

Table 40 (continues)

Table 40 (continued)

Summary of Fit Indices for 2-Factor and 4-Factor International Posture Models

	2-factor	4-factor
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 136, 276)	1465.847	2288.288
Satorra-Bentler scaled χ^2 (df = 118, 248)	232.315	528.861
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	1.969	2.133
Fit Indices		
Comparative fit index (CFI)	.914	.860
Incremental fit index (IFI)	.915	.862
Standardized root mean square residual (SRMR)	(.062)	(.070)
Root mean-square error of approximation (RMSEA)	.062	.067
RMSEA 90% confidence interval	.050-.074	.059-.075

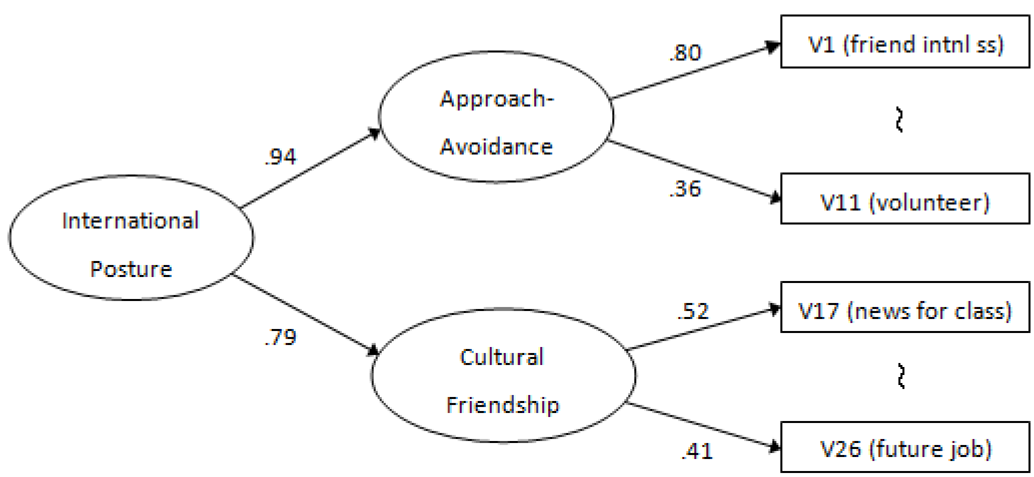


Figure 22. Standardized solution of the 2-factor International Posture instrument.

In addition, the 3-factor model of International Posture used in Yashima et al. (2004) was analyzed and yielded the following fit statistics: χ^2 (87, N = 252) = 281.236 (p < .01), CFI = .847, IFI = .849, RMSEA = .095, and 90% C.I. = .082-.107. These values indicate fit that is very similar to the 4-factor model and inferior to the 2-factor model.

In the subsequent analyses, International Posture thus consisted of two subscales instead of the original four subscales or the three subscales used in Yashima et al. (2004).

Personality

The five posited personality subscales were measured with the Bipolar Scale of Global Personality Traits (Goldberg, 1992). The participants indicated the extent to which a list of 35 pairs of adjectives matched their own personality. After an initial look at the configuration of the overall scale, the respective subscales were examined individually.

Overall Personality scale. The overall measure was developed under the aegis of the so-called Big 5 personality traits, so a confirmatory factor analysis was conducted to verify the overall structure of the Personality instrument and the composition of the respective subscales. As shown in Table 41, the factor analysis yielded a strong 5-factor solution with a single complex loading that accounted for 48.3% of the variance. Items 11, 14, and 21 had the lowest loadings and correspondingly low communalities.

The original instrument was composed of five 7-item subscales (1-7, 8-14, 15-21, 22-28, and 29-35), yet the factor analysis yielded a somewhat different alignment. The Extroversion subscale expanded with the addition of Items 12 (*pleasant*) and 21 (*wealthy, extravagant*) to include the following: *outgoing,* *energetic*, *talkative*, *bold*, *spunky–active*, *assertive*, and *pleasant–agreeable*). In the original English instrument, Item 6 was rendered as *active*, but the Japanese translation is closer to *spunky*. This group of adjectives fits together well and captures the essence of an extroverted person.

Table 41
35-Item Personality Measure Rotated Pattern Matrix

Item	Extro	Diligence	Emotional Stability	Agree	Open Exper	h^2
Pers5	.78					.64
Pers3	.76					.51
Pers6	.75					.51
Pers1	.73					.60
Pers2	.66					.48
Pers4	.58					.42
Pers7	.50					.49
Pers12*	.40					.37
Pers21*	<u>.26</u>					.05
Pers16		.59				.44
Pers29*		.58				.29
Pers31*		.57				.33
Pers30*		.56				.24
Pers19		.55				.31
Pers20		.54				.30
Pers15		.45				.33
Pers35*		.45				.28
Pers14*		<u>.27</u>				.13
Pers26			.74			.51
Pers22			.66			.45
Pers24			.64			.47
Pers28			.56			.35
Pers23		-.40	.53			.40
Pers18*			.47			.36
Pers25			.42			.22
Pers27			.41			.25

Table 41 (continues)

Table 41 (continued)
35-Item Personality Measure Rotated Pattern Matrix

Item	Extro	Diligence	Emotional Stability	Agree	Open Exper	h^2
Pers8				-.61		.48
Pers13				-.61		.39
Pers9				-.58		.30
Pers10				-.53		.39
Pers17*				-.53		.26
Pers33					-.83	.69
Pers34					-.77	.68
Pers32					-.63	.57
Pers11*					<u>.34</u>	.13
Variance	19.36	8.79	8.30	6.09	4.62	
Eigen	6.77	3.08	2.91	2.14	1.62	
Reliab	.87	.68	.72	.66	.82	
I-I correl	.45	.20	.25	.30	.59	

Note. $N = 252$, $k = 35$. Extraction method: principal axis factoring. Rotation method: oblique rotation with Kaiser normalization. Items marked with an asterisk changed to a different subscale than originally posited. Underlined values failed to achieve the cutoff loading value of .40. Eigen = eigenvalue and reliab = reliability (Cronbach's alpha). I-I correl = mean inter-item correlation.

The Diligence subscale (also labeled Conscientiousness) originally consisted of Items 15-22, but lost Items 17, 18, and 21 (*conscientious*, *practical*, and *simple–frugal*) and added Items 29, 30, 31, and 35 (*intelligent*, *analytical*, *reflective*, and *sophisticated*). The Diligence subscale thus included the following adjectives: *generous*, *organized*, *responsible*, *thorough*, *hardworking*, *intelligent*, *analytical*, *reflective*, and *sophisticated*.

The Agreeableness subscale (originally Items 8-14) gained Item 17 (*conscientious*) and lost Items 11, 12, and 14 (*not selfish*, *pleasant*, and *generous*). The resulting configuration included Items 8-10, 13, and 17 (*warm*, *kind*, *trustful*, *cooperative*, and *conscientious*).

The Emotional Stability subscale originally included Items 22-28: *calm*, *relaxed*, *at ease*, *not envious*, *stable*, *contented*, and *emotional*. Item 18 (*practical*) was added, which is a curious addition; however, subsequent analysis indicated that it did not fit the Rasch model well, and it was summarily omitted.

Finally, the Openness to Experience subscale (originally Items 28-35) gained Item 11 (*selfish*) and lost Items 28-31 and 35. Item 11 was deleted later (see below), but the three remaining items (*curious*, *imaginative*, and *creative*) effectively capture the idea of a person interested in the world and new experiences.

The five reconfigured subscales were then investigated individually using WINSTEPS.

Extroversion. The initial WINSTEPS analysis of the Extroversion subscale (Items 1-7, 12, and 21) yielded adequate category function with a series of hills with properly ordered difficulty and separation (Table 42). Item 21 (*simple–frugal*),

Table 42
Category Function Statistics for the Revised Extroversion Subscale

Category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very intro	81 (3.85)	-6.48	-6.72	1.17	(none)	
Introverted	251 (11.94)	-3.34	-3.06	.91	-9.67	.59
Slightly intro	342 (16.26)	-1.14	-1.30	1.05	-3.54	.32
Neutral	473 (22.49)	.27	.30	1.05	-1.98	.26
Slightly extro	441 (20.97)	2.13	2.06	.86	1.48	.25
Extroverted	370 (17.64)	4.18	4.22	1.03	3.89	.28
Very extro	144 (6.85)	6.94	6.97	1.10	9.82	.77

Note. N = 252; k = 7; Avg Measure = average measure; Exp Measure = expected measure.

however, underfit the model with an infit MNSQ value of 2.12 and an outfit MNSQ value of 2.64. Temporarily deleting 13 (5%) of the 31 unexpected responses slightly improved the MNSQ fit statistics to 1.78 and 2.20, respectively, so Item 21 was deleted and the initial WINSTEPS analysis was repeated. The second iteration yielded adequate category function and fit statistics. Incidentally, this was the sole instrument of the 22 used in this study to emerge with seven categories intact.

Next, the dimensionality of the Extroversion instrument was investigated. The average inter-item correlation for the 7-item instrument was adequate (r = .44), and internal reliability was good (Cronbach's α = .87). The initial factor analysis yielded a 1-component solution that accounted for 52.92% of the variance with the seven items having loadings from .56 to .82. This suggests the subscale is unidimensional; dimensionality was checked in more detail using WINSTEPS. The disattenuated correlation of person ability estimates derived using items with positive and negative residual loadings was .93, suggesting that this instrument was fundamentally unidimensional. In addition, the PCA of item residuals indicated that the Rasch model accounted for 65.5% of the variance, and the unexplained variance accounted for by the first residual component was 1.8 units (7.8%).

As shown in Table 43, all eight items exhibited adequate fit statistics and reasonable point-measure correlations. The Extroversion subscale yielded a Rasch item reliability estimate of .55, item separation of 1.10, a Rasch person reliability estimate of .84, person separation of 2.30, and thus a person strata statistic of 3.40.

Table 43

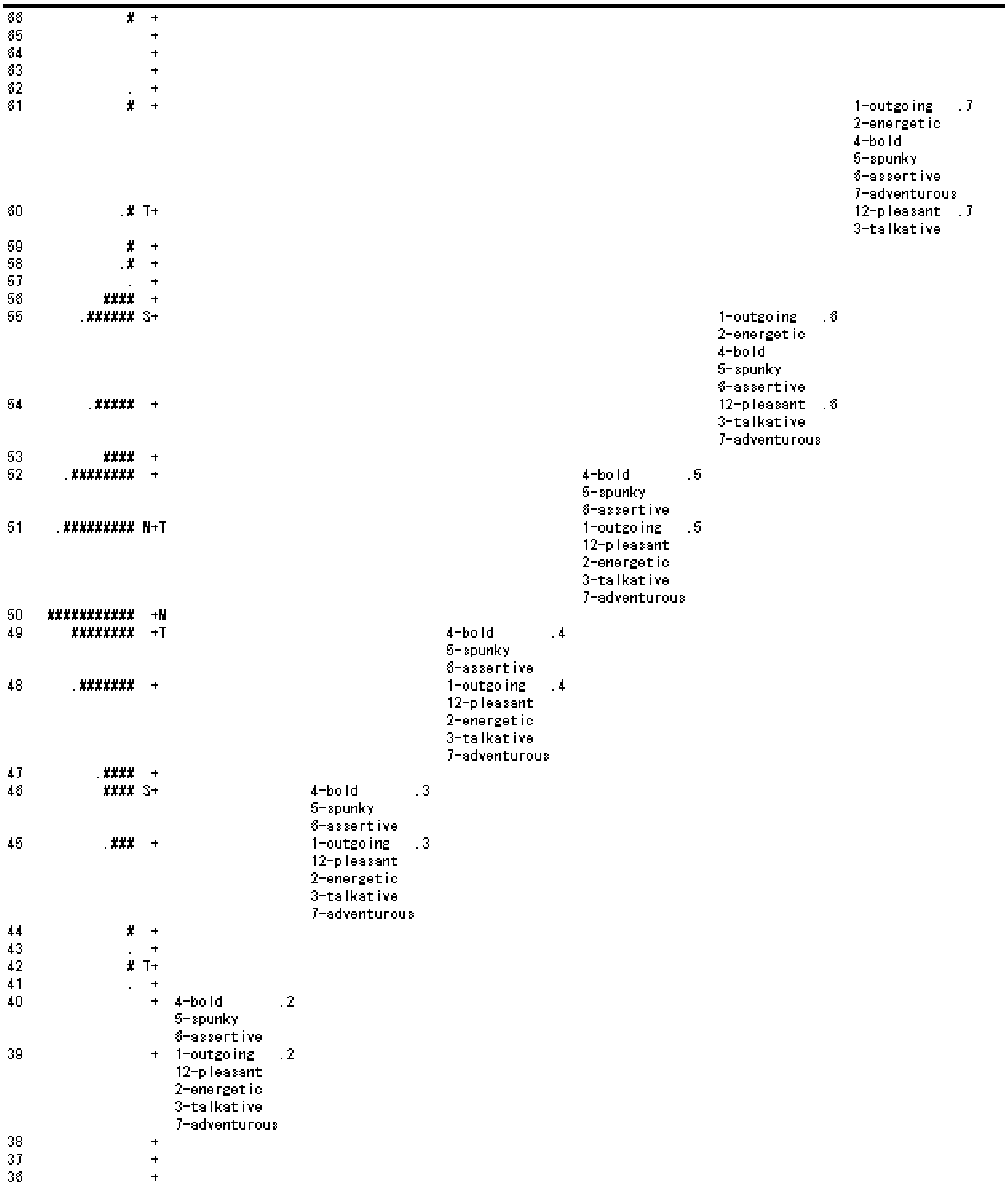
Extroversion Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
4-bold	50.45	.25	1.04	.1	1.06	.8	.67
6-assertive	50.43	.25	.90	-1.2	.93	-.8	.70
5-spunky (active)	50.40	.25	.77	-2.7	.77	-2.9	.76
1-outgoing	50.10	.25	.86	-1.8	.85	-1.8	.74
2-energetic	49.92	.25	.97	-.1	.99	-.1	.70
12-pleasant (agreeable)	49.69	.26	1.34	3.7	1.38	3.1	.56
3-talkative	49.33	.26	1.04	.1	1.01	.1	.68
7-adventurous	49.24	.27	1.10	.2	1.01	.1	.68
M	50.00	.25	.99	-.1	1.01	.0	
SD	.39	.00	.14	1.7	.17	20.	

Note. N = 252, k = 8; Pt-M Corr = point-measure correlation.

The breadth of the Extroversion subscale was just .82 CHIPS (49.63-50.45), indicating a large degree of redundancy in the item difficulties. However, the Rasch-Thurstone thresholds ranged from about 39 to 60 CHIPS, indicating that the items provided adequate coverage of the person abilities. The person ability measures ranged from 31.14 to 65.61, a span of 34.47 CHIPS. As shown in Figure 23, this instrument had considerable redundancy, yet the Rasch-Thurstone thresholds indicate adequate coverage of the person ability estimates. The difference between item difficulty and person ability means was just .97, which indicated that the items were appropriately centered on this sample.

Diligence. Next, the revised Diligence subscale (Items 14-16, 19, 20, 29-31, and 35) was investigated using WINSTEPS. The category function of the 7-category subscale was problematic with disordered category thresholds and inadequate separation. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 44).



Note. M = mean, S = one standard deviation, T = two standard deviations.
Figure 23. Item-person map with Rasch-Thurstone thresholds for the Extroversion subscale.

Table 44

Category Function Statistics for the Revised Diligence Subscale

Stability category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very unstable	137 (5.77)	-3.32	-3.04	.95	(none)	
Slightly unstable	647 (27.24)	.26	-.24	.98	-8.39	.43
Slightly stable	1149 (48.38)	3.30	3.23	.91	-.88	.23
Very stable	442 (18.61)	6.67	6.79	1.10	9.28	.27

Note. N = 252; k = 8; Avg Measure = average measure; Exp Measure = expected measure.

Next, the dimensionality of the Diligence instrument was investigated. The average inter-item correlation for the 8-item instrument was adequate with r = .44, and internal reliability was good (Cronbach's α = .87). The initial EFA yielded a 1-component solution that accounted for 52.92% of the variance with the seven items having loadings from .56 to .82. In WINSTEPS, the disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .93, suggesting that this instrument was strongly unidimensional. In addition, the PCA of residuals indicated that the Rasch model accounted for 65.5% of the variance. The unexplained variance accounted for by the first residual contrast was 1.8 units (7.8%).

Rasch statistics yielded a Rasch item reliability estimate of .96, item separation of 4.83, a Rasch person reliability estimate of .61, person separation of 1.24, and a person strata statistic of 1.99. As shown in Table 45, eight of the nine items exhibited adequate fit and reasonable point-measure correlations. Although Item 35 (*sophisticated*) was overfitting with infit and outfit MNSQ statistics of .55 and .56, respectively, it was retained as those values do not degrade the model.

Table 45
Diligence Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
19-thorough	53.42	.41	1.28	3.2	1.36	4.0	.41
15-organized	51.68	.41	1.15	1.8	1.16	1.9	.55
14-generous	51.15	.41	1.18	2.1	1.19	2.2	.38
29-intelligent	50.85	.41	.74	-3.4	.75	-3.3	.58
20-hardworking	50.13	.42	1.12	1.4	1.09	1.1	.55
30-analytical	49.94	.42	1.00	.1	.99	-.1	.56
35-sophisticated	49.58	.43	.55	-6.4	.56	-6.2	.52
31-reflective	47.81	.43	.86	-1.8	.86	-1.8	.56
16-responsible	45.44	.45	1.10	1.2	1.03	.4	.62
M	50.00	.42	1.01	-.2	1.00	-.2	
SD	2.17	.01	.21	2.9	.23	2.9	

Note. N = 252, k = 9; Pt-M Corr = point-measure correlation.

The breadth of the Diligence subscale was 7.98 CHIPS (45.42-53.44), and some redundancy was present in the item difficult estimates (Figure 24). The person measures ranged from 40.50 to 65.59, a span of 15.09 CHIPS. The difference in means of the item difficulty and person ability estimates was 2.61, which indicated that the items on this instrument were somewhat easy to endorse for this sample. In other words, the participants felt they were relatively diligent, which is a desirable quality in students.

```
Persons with more        Item indicating more diligence
diligence
  63             .   +
                     |
  62                 +
                .#   |
  61                T+
                     |
  60                 +
               ###   |
  59                 +
              .###   |
  58                 +
                     |
  57         .####  S+
                     |
  56             .   +
         .########   |
  55                 +
         .########   |T
  54             .   +
        ##########   |  19-thorough
  53             .   +
         ######### M |
  52                 +S
                     |  15-organized
  51 .############   +  14-generous       29-intelligent
                     |
  50         .####   +M 20-hardworking    30-analytical
                 .   |  35-sophisticated
  49       #######   +
                    S|
  48         .####   +S 31-reflective
                     |
  47          .###   +
                     |
  46             .   +
                     |T 16-responsible
  45            .#   +
                    T|
  44                 +
                     |
  43             #   +
                     |
  42             .   +
Persons with less        Item indicating less diligence
diligence
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 24. Item-person map for the Diligence subscale.

Emotional Stability. Next, the revised Emotional Stability subscale (Items 18, 22-28) category function was investigated using WINSTEPS; the initial results indicated disordered category thresholds and inadequate separation. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 46).

Table 46

Category Function Statistics for the Revised Emotional Stability Subscale

Distance category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Not stable	387 (18.28)	-7.66	-7.68	1.02	(none)	
Slightly unstable	889 (42.14)	-2.91	-2.79	.88	-8.80	.30
Rather stable	676 (32.17)	.73	.45	.90	.11	.24
Very stable	157 (7.42)	3.15	3.73	1.16	8.70	.41

Note. N = 252; k = 8; Avg Measure = average measure; Exp Measure = expected measure.

Rasch statistics produced an item reliability estimate of .96, item separation of 5.01, a person reliability estimate of .67, person separation of 1.43, and a person strata statistic of 2.24. As shown in Table 47, all eight items exhibited adequate fit and reasonable point-measure correlations. Item 25 (*not envious*) and Item 28 (*emotional*) were the most difficult to endorse, and the three items dealing with calmness were the easiest to endorse (*at ease*, *calm*, and *relaxed*).

Next, the dimensionality of the Emotional Stability instrument was investigated. The average inter-item correlation for the 8-item instrument was adequate with r = .44, and internal reliability was good (Cronbach's α = .87). The initial EFA yielded a 1-component solution that accounted for 52.9% of the

Table 47

Emotional Stability Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
25-not envious	54.09	.44	1.04	.5	1.10	1.1	.47
28-emotional	52.69	.42	1.17	1.9	1.22	2.4	.45
27-contented	50.27	.41	1.10	1.2	1.14	1.6	.53
26-stable	49.90	.41	.96	-.5	.95	-.6	.69
18-practical	49.12	.41	.94	-.8	.98	-.5	.52
24-at ease	48.42	.41	1.03	.3	1.02	.3	.64
22-calm	48.09	.41	.87	-1.7	.87	-1.7	.60
23-relaxed	47.41	.41	.88	-1.5	.90	-1.2	.60
M	50.00	.42	1.00	-.1	1.02	.2	
SD	2.17	.01	.10	1.2	.12	1.3	

Note. N = 252, k = 8; Pt-M Corr = point-measure correlation.

variance with the seven items having loadings from .56 to .82. This suggests that the subscale is unidimensional, which was checked in more detail using WINSTEPS. The disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .80, suggesting that this instrument was fundamentally unidimensional. In addition, the PCA of item residuals indicated that the Rasch model accounted for 47.2% of the variance. The unexplained variance accounted for by the first residual component was 1.6 units (10.5%).

As shown in Figure 25, with a range of 6.68 CHIPS (47.41-54.09), the Emotional Stability subscale covered the person distribution of 30.34 CHIPS (31.78-62.12) poorly, yet the Rasch-Thurstone item thresholds are indicative of adequate coverage. The difference between the means of the item difficulty and person ability estimates was 2.24 (47.76-50.00), which indicated that the Emotional Stability subscale was slightly difficult to endorse for these participants.

```
Persons with more        Item thresholds indicating more emotional stability
emotional stability
  63               +                                        25-not envious.4
  62             . +                                        28-emotional  .4
  61               +
  60             . +                                        27-contented  .4
  59             . +                                        26-stable     .4
  58              T+                                        18-practical  .4
                                                            24-at ease
  57             . +                                        22-calm       .4
                                                            23-relaxed
  56            .# +
  55            .# +
  54               +T                  25-not envious.3
  53         .#### S+                  28-emotional  .3
  52       .###### +S
  51      .####### +
  50    .######### +M                  26-stable     .3
                                       27-contented
  49  .########### +                   18-practical  .3
                                       24-at ease
  48    .######### M+S                 22-calm       .3
  47      .####### +                   23-relaxed    .3
  46               +T
  45    .######### +  25-not envious.2
  44         .#### +
  43           .## S+ 28-emotional  .2
  42               +
  41           .## +  26-stable     .2
                      27-contented
  40           .## +  18-practical  .2
  39               +  22-calm       .2
                      24-at ease
  38           .## T+ 23-relaxed    .2
  37               +
  36               +
  35             # +
Persons with less        Item thresholds indicating less emotional stability
emotional stability
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 25. Item-person map with Rasch-Thurstone thresholds for the Emotional Stability subscale.

Agreeableness. Next, category function was investigated for the revised Agreeableness subscale (Items 8-10, 13, and 17) using WINSTEPS; the initial results yielded disordered category thresholds and inadequate separation. Combining categories ultimately yielded four categories with proper ordering, good fit, and adequate separation (Table 48).

Table 48
Category Function Statistics for the Revised Agreeableness Subscale

Distance category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Disagreeable	109 (8.25)	-6.84	-8.09	1.27	(none)	
Slightly disagree	468 (35.43)	-3.44	-2.79	.83	-11.86	.52
Agreeable	633 (47.92)	2.64	2.33	.86	-1.72	.30
Very agreeable	111 (8.40)	9.23	9.49	1.08	13.58	.52

Note. N = 252; k = 5; Avg Measure = average measure; Exp Measure = expected measure.

Rasch statistics yielded an item reliability estimate of .56, item separation of 1.12, a person reliability estimate of .61, person separation of 1.24, and thus a person strata statistic of 1.99. As shown in Table 49, all five items exhibited adequate fit and reasonable point-measure correlations.

Next, the dimensionality of the Agreeableness instrument was investigated. The average inter-item correlation for the 8-item instrument was adequate with r = .44, and internal reliability was good (Cronbach's α = .87). The initial EFA yielded a 1-component solution that accounted for 52.9% of the variance with the seven items having loadings from .56 to .82. This suggests that the subscale is unidimensional, which was checked in more detail using WINSTEPS. The disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .86, suggesting that this instrument was fundamentally unidimensional. In addition, the PCA of item residuals indicated that the Rasch model accounted for 47.3% of the variance. The unexplained variance accounted for by the first residual contrast was 1.5 units (16.1%).

Table 49

Agreeableness Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
8-warm	51.12	.48	.70	-3.9	.70	-3.8	.70
9-kind	50.54	.48	.88	-1.4	.87	-1.5	.67
13-trustful	49.74	.48	1.36	3.8	1.39	4.0	.61
17-conscientious	49.67	.48	1.14	1.5	1.16	1.7	.60
10-cooperative	48.93	.48	.91	-1.0	.89	-1.3	.65
M	50.00	.48	1.00	-.2	1.00	-.2	
SD	.76	.00	.23	2.7	.24	2.7	

Note. N = 252, k = 5; Pt-M Corr = point-measure correlation.

As shown in Figure 26, with a range of 2.22 CHIPS (48.93-51.15), the Agreeableness subscale covered the person distribution of 38.44 CHIPS (31.55-69.99) somewhat poorly. The difference between item difficulty and person ability means was .35 (50.35-50.00), which indicated that the Agreeableness subscale was at an appropriate level for this sample.

```
Persons with more        Item thresholds indicating more agreeableness
agreeableness

 65                 +                                            8-warm          .4
 64                 +                                            9-kind          .4
 63                T+                                            10-cooperative  .4
                                                                 13-trustful
                                                                 17-conscientious
 62             ##  +
 61                 +
 60                 +
 59           .###  +
 58                 +
 57                S+
 56    ##########   +
 55                 +
 54                 +
 53      ########   +
 52                 +T
 51                 +S
 50   ########### M+M                     8-warm          .3
 49              .  +S                    9-kind          .3
 48    .#########   +T                    10-cooperative  .3
                                          13-trustful
                                          17-conscientious
 47                 +
 46              .  +
 45      ########   +
 44                S+
 43       .######   +
 42                 +
 41            ###  +
 40                 +
 39                 +  8-warm          .2
 38              .  +  9-kind          .2
 37                T+  10-cooperative  .2
                       13-trustful
                       17-conscientious
 36              .  +
Persons with less        Item thresholds indicating less agreeableness
agreeableness
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 26. Item-person map with Rasch-Thurstone thresholds for the Agreeableness subscale.

Openness to Experience. Finally, on the *Openness to Experience* subscale (Items 11, 32-34), WINSTEPS yielded adequate category function with a series of hills with properly ordered difficulty. However, Category 1 was rarely used, and a preliminary look at fit statistics indicated that Item 11 (*not selfish*) fit the model poorly with infit and outfit MNSQ values of 1.88 and 2.11, respectively. Of the 34 unexpected responses, temporarily omitting 13 (5%) improved the MNSQ fit statistics to 1.57 and 1.72, but as this was still misfitting, Item 11 was deleted. Combining Categories 1 and 2 yielded a 6-category, 3-item scale with a series of hills with properly-ordered difficulty, good fit, and adequate separation (Table 50).

Table 50
Category Function Statistics for the Revised Openness to Experience Subscale

Openness category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very closed	51 (7.17)	-11.32	-12.00	1.54	(none)	
Closed	89 (12.52)	-7.55	-7.08	.84	-12.00	.84
Neutral	132 (18.57)	-3.17	-2.61	.69	-6.62	.61
Slightly open	182 (25.60)	2.22	2.13	.82	-1.76	.52
Open	188 (26.44)	8.80	7.90	.74	4.77	.51
Very open	69 (9.70)	12.71	14.22	1.43	15.61	.71

Note. N = 252; k = 3; Avg Measure = average measure; Exp Measure = expected measure.

With just three items, the revised *Openness to Experience* subscale was treated as a single dimension. The subscale yielded a Rasch item reliability estimate of .98, item separation of 7.92, a Rasch person reliability estimate of .74, person separation of 1.70, and a person strata statistic of 2.60. As shown in Table 51, all three items exhibited satisfactory fit and reasonable point-measure correlations. These three items were located as expected, with Item 32 (*curiosity*) the easiest item to endorse. Being curious is a common innate characteristic, whereas creativity is a trait that exists in a much more limited way (e.g., in the world of music, interest and curiosity about music are common traits, and musical proficiency is common; however, musical creativity is much less common).

Table 51

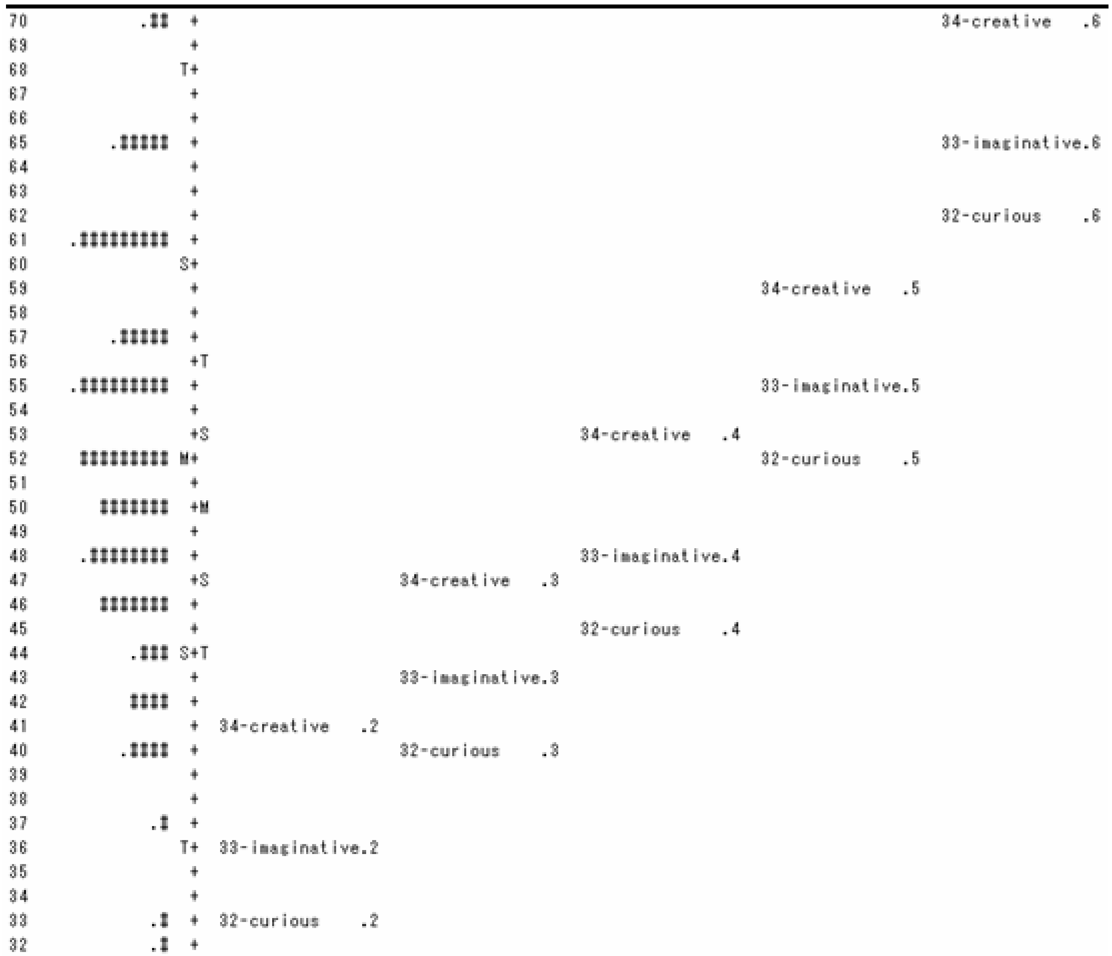
Openness to Experience Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
34-creative	54.18	.37	.76	-2.9	.74	-3.0	.88
33-imaginative	49.50	.38	.92	-.9	.88	-1.3	.84
32-curious	46.32	.40	1.31	3.0	1.25	2.6	.78
M	50.00	.38	1.00	-.3	.96	-.6	
SD	3.23	.01	.23	2.5	.22	2.3	

Note. N = 252, k = 3; Pt-M Corr = point-measure correlation.

The average inter-item correlation for the 3-item instrument was adequate with r = .59, and internal reliability was good (Cronbach's α = .82). The PCA of item residuals indicated that the variance explained by the Rasch model was 79.0%, and the first residual contrast had unexplained variance of 1.7 units (11.8%).

As shown in Figure 27, with a range of 7.86 CHIPS (46.32-54.18), the Openness to Experience subscale poorly covered the person distribution of 36.97 CHIPS (32.81-69.78), but the category thresholds were much more widely distributed. The difference between item difficulty and person ability means was 2.96, which indicates that the Openness to Experience subscale was rather easy to endorse for this sample. These figures must be viewed with caution, however, for with only three items this subscale is short for measuring a construct.



Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 27. Item-person map with Rasch-Thurstone thresholds for the Openness to Experience subscale.

A summary of the subscales is shown in Table 52. Although all five subscales originally had seven items, the results from these data indicated that the deletion of two items and the realignment of the items on the subscales were appropriate.

Table 52
Summary of Personality Subscales

Subscale	*k*	*j*	Item Rel	Item Sep	Per Rel	Per Sep	% of Var	I-I Corr
Extroversion	9	7	.55	1.10	.84	2.30	65.5	.45
Emotion Stability	8	4	.96	4.85	.61	1.24	41.9	.20
Diligence	8	4	.96	5.02	.67	1.43	47.2	.25
Agreeableness	5	4	.57	1.15	.61	1.99	47.3	.30
Openness to Experience	3	6	.98	7.92	.74	1.70	79.0	.59
Total	33							

Note. $N = 252$; k = number of items; j = number of response categories; Rel = reliability; sep = separation; per = person. % of variance is from WINSTEPS PCA of residuals. I-I Corr = average inter-item correlation.

Summary

In this chapter, the results of the preliminary analyses of the individual difference variables were presented; those variables include L2 Communicative Anxiety (both the L2 Communicative Anxiety instrument and the FLCAS), Frequency of L2 Communication, L2 Willingness to Communicate, Motivation, International Posture, and the Personality subscales. The first four instruments were found to be valid as originally configured. However, the Motivation instrument was found to consist of a single dimension rather than two subscales as originally hypothesized. Finally, the configuration of the respective International Posture subscales changed somewhat, and a confirmatory factor analysis using EQS indicated that a two-factor configuration made up of the Intergroup Approach-Avoidance Tendency subscale and the Intercultural Friendship Orientation subscale had the best fit to the model; the 2-factor model was used in subsequent analyses.

In this chapter, the respective individual difference variables were validated. With additional variables hypothesized to augment the original three models, the topic of Chapter 6 is the validation of the added personality variables: Distancing and Ego Permeability with its five subscales.

CHAPTER 6

PRELIMINARY ANALYSES: VARIABLES ADDED TO THE MODELS

In this chapter I cover the initial analyses of the variables added to the respective models: Perceived Distance and Ego Permeability with its five subscales (although the Extroversion subscale of the Personality instrument was added to the Yashima models, it was addressed in the previous chapter because the five subscales were included in the MacIntyre and Charos (1996) model). The first section is followed by an in-depth look at each of the instruments and subscales with the procedure outlined in the Methods chapter: category function; item-person map; Rasch fit statistics; Rasch separation, reliability, and strata; Rasch principal components analysis of item residuals; and the treatment of misbehaving items. In addition, a structural equation model was tested to investigate further the dimensionality of the Ego Permeability instruments. As detailed in Chapter 4, the data from the instruments were first carefully screened. In the second section of this chapter the results from confirmatory factor analyses are presented. The purpose of this analysis was to evaluate the dimensionality of the Ego Permeability with its five subscales.

Analyses of Instruments Added to the L2 Communication Models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004)

In this section I examine the instruments that were added to the two communication models. The instruments were the Perceived Distance Questionnaire and Ego Permeability with its five subscales.

Perceived Distance

The Perceived Distance instrument created for this study consists of five items that asked the participants about changes in perceived distance when they engaged in various second language tasks. As noted in Chapter 4, the data were converted from percentages to Likert-scale data prior to conducting the analyses. Category function was then investigated using WINSTEPS; the results indicated proper ordering yet inadequate separation of the thresholds. Combining categories yielded a 4-category alignment with proper ordering, good fit, and good separation (Table 53).

Table 53
Category Function Statistics for the Revised Perceived Distance Instrument

Distance category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Very little	224 (17.78)	-8.01	-8.68	1.22	(none)	
Little	299 (23.73)	-2.98	-2.28	.96	-6.69	.44
Neutral	416 (33.02)	3.33	3.43	1.00	-.92	.36
Considerable	321 (25.48)	9.90	9.59	.95	7.61	.38

Note. N = 252; Avg Measure = average measure; Exp Measure = expected measure.

Next, the dimensionality of the Perceived Distance subscale was investigated. The average inter-item correlation for the 5-item instrument was adequate (r = .59), and internal reliability was good (Cronbach's alpha = .85). An exploratory factor analysis yielded two possible solutions: a 1-factor solution accounting for 51.64% of the variance with factor loadings from .43 to .87 and communalities from .19 (Item 1) to .77, and a 2-component solution accounting for 59.00% of the variance with factor loadings from .40 to .75 and communalities from .30 to .80. However, addressing dimensionality with a PCA of residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived using items with positive and negative residual loadings was .85, indicating that the Perceived Distance instrument was fundamentally unidimensional.

All five items exhibited adequate fit and reasonable point-measure correlations, but Item 1 (*chatting in English*) showed barely adequate fit to the model with an infit MNSQ value of 1.34 and an outfit MNSQ value of 1.56. However, temporarily deleting the responses from 12 persons with unusual responses improved the outfit value to 1.17, indicating that the item functioned satisfactory; Item 1 was thus retained. In Table 54, the reader should be aware of the valence: Item 1 (*chatting in English*), was the most difficult item for the respondents to endorse, meaning they perceived less distance when chatting in English. However, Item 5 (*doing puppetry*) was the easiest item to endorse, meaning the respondents perceived the most distance when doing puppetry.

Table 54

Perceived Distance Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
1-chatting in English	56.23	.43	1.34	3.6	1.56	4.4	.63
3-roleplay	51.69	.42	.78	-2.8	.75	-3.0	.80
2-public speaking	51.54	.42	.85	-1.7	.85	-1.6	.77
4-drama	45.89	.45	.77	-2.7	.74	-2.8	.82
5-puppetry	44.65	.47	1.25	2.5	1.23	1.9	.71
M	50.00	.44	1.00	-.2	1.03	-.2	
SD	4.23	.02	.28	2.7	.32	2.9	

Note. N = 252, k = 5; Pt-M Corr = point-measure correlation.

The 5-item Perceived Distance instrument yielded an item reliability estimate of .99, item separation of 9.07, a person reliability estimate of .73, person separation of 1.64, and a person strata statistic of 2.93. The PCA of item residuals indicated that the Rasch model accounted for 76.8% the variance. The unexplained variance in the first residual contrast accounted for 2.2 units (10.3%) of the total variance.

Figure 28 shows the item-person map with the Rasch-Thurstone thresholds for the five items on the Perceived Distance instrument. The breadth of the means of the item difficulties was 7.71 CHIPS (46.40-54.11), yet the thresholds span 15.88 CHIPS (37.96-63.84). This indicates reasonable coverage of the person ability estimates, which ranged from 34.62 to 66.79, a span of 32.17 CHIPS. The difference between item difficulty and person ability means was just 1.32 CHIPS, which indicates that the items were appropriate for this sample.

```
Persons perceiving    Item thresholds more difficult to endorse (show less perceived distance)
more distance
66       .###### T+
65               +
64               +                                         chatting English.4
63               +
62               +
61          .### +
60               +                                         public speaking .4
                                                           roleplay
59      ######## S+
58               +T
57               +
56     .######## +                        chatting English.3
55               +
54   ########### +S                                        drama          .4
53               +                                         puppetry       .4
52     .######## +
51      .####### M+                       public speaking .3
                                          roleplay
50               +M
49     .######## +  chatting English.2
48               +
47          #### +
46               +S
45        ###### +                        drama           .3
44          #### S+  public speaking .2   puppetry        .3
                     roleplay
43               +
42               +T
41            ## +
40               +
39        ###### +
38               +  drama           .2
37               T+  puppetry       .2
36               +
35            .# +
34           ### +
Persons perceiving    Item thresholds less difficult to endorse (show more perceived distance)
less distance
```

Note. M = mean, S = one standard deviation, T = two standard deviations

Figure 28. Item-person Rasch-Thurstone threshold map of the Perceived Distance instrument.

Ego Permeability

The *Ego Permeability* instrument was a shortened form (BQ-SH; Rawlings, 2001) of the *Hartmann Boundary Questionnaire* (Hartmann, 1991). The shortened form consists of 40 statements culled from the original 146; participants indicate the extent to which they agree or disagree. The 40 items comprise five subscales: Unusual Experiences, Need for Order, Childlikeness, Perceived Time-Money Competence, and Sensitiveness. As noted above, the Perceived Time-Money Competence subscale was originally titled Perceived Competence, but because the items deal with skill in using time and money and to distinguish it more clearly from the Perceived L2 Competence scale, hereafter the label 'Perceived Time-Money Competence' is used.

The ego permeability construct was examined with a confirmatory factor analysis using principal axis factoring and oblique rotation (Table 55). The five factors that emerged correspond closely with the subscales hypothesized in the BQ-SH instrument; only Item 39 (*There are no sharp dividing lines between normal people, people with problems, and people who are considered psychotic or crazy*) was moved from the Perceived Time-Money Competence subscale to the Childlikeness subscale. At first glance this seems to be an odd change since Item 39 does not specifically concern children, but the items in the Childlikeness subscale all deal with how the division between groups such as children and adults is blurred; viewed in that light, the blurring of lines between crazy or psychotic people and normal people is similar to the blurring of divisions between other groups. Item 18 (*I cannot imagine living with or marrying a person of another race*) had the smallest loading at .32. The resulting 40-item scale accounted for 43.35% of the variance and had an overall internal reliability estimate of .71 (Cronbach's alpha). This analysis thus offered support for the underlying structure of the shortened Ego Permeability instrument.

Table 55
40-Item Ego Permeability Measure Rotated Pattern Matrix

Item	Unusual Exp	Need for Order	Childlike	Perceived T-M Comp	Sensitive	h^2
Ego6	.76					.43
Ego11	.73					.57
Ego12	.71					.50
Ego10	.69					.53
Ego4	.67					.44
Ego1	.64					.43
Ego9	.63					.48
Ego3	.63					.38
Ego7	.62					.43
Ego5	.59					.41
Ego8	.58					.44
Ego2	.58					.35
Ego16		.66				.41
Ego15		.60				.42
Ego23		.60				.38
Ego17		.55				.34
Ego20		.52				.40
Ego19		.49				.47
Ego14		.49				.30
Ego13		.48				.39
Ego22		.45				.36
Ego21		.45				.33
Ego24		.43				.22
Ego18*		*.38				.22

Table 55 (continues)

Table 55 (continued)
40-Item Ego Permeability Measure Rotated Pattern Matrix

Item	Unusual Exp	Need for Order	Childlike	Perceived T-M Comp	Sensitive	h^2
Ego41			.85			.73
Ego40			.84			.70
Ego42			.67			.46
Ego43			.49			.42
Ego39			-.47			.31
Ego44			.41			.39
Ego34				.73		.56
Ego32				.67		.45
Ego33				.59		.37
Ego35				.59		.39
Ego37				.57		.41
Ego31				.52		.45
Ego36				.49		.28
Ego38				.41		.26
Ego45					.85	.74
Ego46					.77	.64
Variance	43.62	29.25	45.43	51.53	86.77	
Eigen	5.23	3.51	2.73	2.58	1.74	
Reliab	.87	.77	.74	.81	.85	
I-I correl	.38	.23	.36	.57	.75	

Note. N = 252; k = 40; E-value = eigenvalue; Rel = reliability; I-I correl = inter-item correlation. Extraction method: principal axis factoring. Rotation method: Oblim rotation with Kaiser normalization. Item 18 (marked with an asterisk) fell beneath the .40 cutoff criterion but was retained. Exp = experiences; T-M Comp = time-money competence.

The five Ego Permeability subscales were then examined using WINSTEPS, and all performed adequately. The individual subscales were checked for dimensionality using WINSTEPS, and the Rasch CHIPs measures of person ego permeability estimates of the five subscales were used in subsequent analyses.

Unusual Experiences. On the Unusual Experiences subscale (Items 1-12), category function was investigated using WINSTEPS; the initial results showed disordered category thresholds and inadequate separation. Responses were positively skewed with Category 1 having the largest count. Combining categories ultimately yielded three categories with proper ordering, good fit, and adequate separation (Table 56).

Table 56
Category Function Statistics for the Unusual Experiences Subscale

Extent of experiences	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Very seldom	1238 (41.25)	-8.32	-8.03	.90	(none)	
Occasional	1185 (39.43)	-.81	-1.39	.92	-4.32	.21
Some	587 (19.40)	3.10	3.68	1.16	4.32	.25

Note. N = 252; k = 12; Avg Measure = average measure; Exp Measure = expected measure.

The dimensionality of the *Unusual Experiences* subscale was then checked. The average inter-item correlation for the 12-item instrument was adequate with r = .38, and internal reliability was good (Cronbach's α = .87). An initial EFA yielded a one-component solution that accounted for 43.62% of the variance. Loadings on the single component were strong (.56 to .76) and communalities ranged from .31 to .57. A PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .70, suggesting that this instrument was possibly multi-dimensional. However, the PCA of item residuals indicated that the variance explained by the Rasch model was a robust 62.7%; unexplained variance in the first residual component accounted for a mere 1.9 units (5.9%) of the total variance, which suggested that the *Unusual Experiences* subscale instrument was unidimensional.

Because of the low disattenuated correlation, a confirmatory factor analysis was conducted using EQS. Neither the 1-factor model nor the 2-factor model had good fit although the latter model was slightly better: χ^2 (52, N = 252) = 180.327 (p < .01), CFI = .864, IFI = .866, RMSEA = .099, and 90% C.I. = .083-.115. In lieu of the ambiguous results from both the Rasch analysis and the confirmatory factor analysis, the Unusual Experiences subscale was treated as a single dimension on theoretical grounds.

Rasch statistics included an item reliability estimate of .96, item separation of 4.74, a person reliability estimate of .77, person separation of 1.83, and a person strata statistic of 2.77. As shown in Table 57, all 12 items exhibited adequate fit statistics and reasonable point-measure correlations.

The items in the *Unusual Experiences* subscale were positioned as expected. Items 1, 6, and 10 dealt with people or things changing form, whereas Item 8 queried sensory convergence in which, for example, a person perceives a color to have sound. As expected, these items were difficult to endorse. Items that were easy to endorse were concerned with transitions between dreaming and being awake, which can be disorienting. As shown in Figure 29 and by the difference in means between the Rasch person ability and item difficulty estimates, many of the items on the Unusual Experiences subscale were difficult to endorse, indicating that many participants had seldom encountered such experiences.

Table 57

Ego Permeability, Unusual Experiences Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
8-senses converge	53.70	.54	.95	-.6	1.27	1.8	.57
1-daydreams ppl change	53.38	.53	.96	-.5	.88	-.9	.60
6-things change	53.04	.53	.66	-4.4	.60	-3.5	.68
10-own body changes	52.74	.53	.93	-.8	.90	-.7	.64
3-have daydreams	50.02	.50	1.12	1.4	1.28	2.5	.57
4-dreams people change	49.55	.50	1.06	.8	1.03	.3	.63
5-body injured	48.91	.50	1.17	2.0	1.15	1.5	.62
9-dreams vivid real	49.21	.50	.94	-.7	.99	.0	.64
12-real or not	48.94	.49	.91	-1.1	.91	-1.1	.68
7-scary to nightmares	47.97	.49	1.16	2.0	1.10	1.0	.63
11-called real not real	46.98	.49	.96	-.5	.90	-1.1	.70
2-dream to dream	45.54	.50	1.15	1.8	1.15	1.6	.61
M	50.00	.51	1.00	-.1	1.01	.1	
SD	2.55	.02	.14	1.7	.18	1.6	

Note. *N* = 252, *k* = 12; Pt-M Corr = point-measure correlation.

The breadth of the Unusual Experiences subscale was 8.16 CHIPS (45.54-53.70), and some redundancy in the items was present. The person measures, however, ranged from 34.05 to 65.96 CHIPS, a very broad span of 31.91 CHIPS, yet the Rasch-Thurstone thresholds indicated coverage of the person ability estimates was adequate (Figure 29). The difference between the means of the item difficulty and person ability estimates was 3.93 CHIPS (46.07-5,000), which indicated that the participants found the items on the instrument somewhat difficult to endorse and thus had had relatively few unusual experiences.

```
Persons with more        Item thresholds indicating more unusual experiences
experiences

60            . T+
59            .  +                              8-senses converge      .3
58            .  +                              1-daydreams ppl change .3
                                                10-own body changes
                                                6-things change
57           .#  +
56            .  +
55          .##  +T                             3-have daydreams       .3
54          ###  +                              12-real or not         .3
                                                4-dreams ppl change
                                                5-body injured
                                                9-dreams vivid real
53         .### S+S                             7-scary to nightmares  .3
52        .####  +                              11-called real not real.3
51      #######  +
50    #########  +M                             2-dream to dream       .3
49          .##  +  8-senses converge      .2
48       ######  +  1-daydreams ppl change .2
                    10-own body changes
                    6-things change
47        ##### M+S
46           ##  +
45       .#####  +T 3-have daydreams       .2
                    4-dreams ppl change
44          .##  +  12-real or not         .2
                    5-body injured
                    9-dreams vivid real
43         .###  +  7-scary to nightmares  .2
42               +  11-called real not real.2
41        #####  +  2-dream to dream       .2
40       .##### S+
39               +
38        #####  +
37               +

Persons with fewer       Item thresholds indicating fewer unusual experiences
experiences
```

Note. M = mean, S = one standard deviation, T = two standard deviations.

Figure 29. Item-person map with Rasch-Thurstone thresholds for the Unusual Experiences subscale.

Need for Order. On the *Need for Order* subscale (Items 13-24), WINSTEPS initially yielded disordered thresholds and inadequate separation of thresholds. Combining categories yielded a 4-category alignment with proper ordering, good fit, and good separation (Table 58).

Table 58

Category Function Statistics for the Need for Order Subscale

Need for Order category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Little need	396 (13.25)	-4.21	-4.14	1.01	(none)	
Slight need	919 (29.43)	-.91	-.80	.90	-6.19	.28
Some need	1258 (40.92)	2.29	2.06	.92	-0.79	.19
Strong need	590 (19.40)	4.81	5.06	1.07	6.98	.23

Note. N = 252; k = 12; Avg Measure = average measure; Exp Measure = expected measure.

The dimensionality of the Need for Order subscale was then investigated. The average inter-item correlation for the 12-item instrument was adequate with r = .40, and internal reliability was good (Cronbach's alpha = .85). An exploratory factor analysis yielded two possible solutions: a 1-factor solution, which accounted for 51.64% of the variance with factor loadings from .43 to .87 and communalities from .19 (Item 1) to .77, and a 2-component solution, which accounted for 59.00% of the variance with factor loadings from .40 to .75 and communalities from .30 to .80. However, addressing dimensionality with a PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .88, suggesting that the Need for Order instrument was strongly unidimensional.

The Rasch statistics yielded an item reliability estimate of .98, item separation of 6.87, a person reliability estimate of .74, person separation of 1.69, and a person strata statistic of 2.59. All 12 items exhibited adequate fit and reasonable point-measure correlations (Table 59).

Table 59
Ego Permeability, Need for Order Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
21-frames picture	53.88	.37	.98	-.2	.96	-.5	.53
20-good guys bad guys	53.24	.37	.93	-.9	.91	-1.1	.56
13-everything place	52.95	.37	1.00	.1	.99	-.1	.57
18-partner not diff race	52.02	.37	1.36	4.2	1.36	4.1	.46
16-m/f different	50.31	.37	.93	-.8	.92	-1.0	.54
23-def walls functions	49.76	.37	.82	-2.3	.82	-2.3	.57
14-strict discipline	49.54	.37	.91	-1.1	.90	-1.2	.50
17-stories definite parts	49.33	.38	.96	-.5	.97	-.4	.55
15-org definite roles	48.80	.38	.82	-2.3	.82	-2.3	.55
24-East is East	48.70	.38	1.40	4.4	1.37	4.1	.47
19-precise borders	46.92	.40	.86	-1.7	.86	-1.7	.58
22-neat dress important	44.56	.43	.98	-.2	.98	-.2	.45
M		.38	1.00	-.1	.99	-.2	
SD		.02	.18	2.1	.18	2.0	

Note. *N* = 252, *k* = 12; Pt-M Corr = point-measure correlation. Org = organization; m/f = male / female; diff = different; def = definite.

As shown in Figure 30, the easiest item to endorse was Item 22 (*dressing well*); this was not surprising given the widespread consciousness about fashion in Japan. Other frequently endorsed items dealt with things (e.g., borders in Items 19 and 24, stories in Item 17, and organizations in Item 15), whereas items dealing with people were generally more difficult to endorse (Items 14, 16, 18, and 20). An interesting dyad is also present with Item 21 (*Good solid frames are very important* *for a picture or a painting*) and Item 19 (*I like clear, precise borders*) being difficult and easy to endorse, respectively. At first glance this seemed to be contradictory, but it might reflect a specific example (the picture frame in the world

```
Persons with more        Item indicating more need for order
need for order
62                  +
61             #    +
60                  +
59             #   T+
58            ##    +
57            .#    +
56          .###    +
55            ##   S+T
54  ############    +  21-frames picture
53        .#####    +S 13-everything place       20-good guys bad guys
52  ############    +  18-partner not diff race
51     .########   M+
50    .#########    +M 14-strict discipline      16-m-f diff
                       23-def walls functions
49  ############    +  15-org def roles          17-stories def parts
                       24-East is East
48          .###    +
47        .#####   S+S 19-precise borders
46             .    +
45           .##    +T 22-neat dress important
44           .##    +
43            .#   T+
42             .    +
41                  +
Persons with less        Item indicating less need for order
need for order
```

Note. M = mean, S = one standard deviation, T = two standard deviations.
Figure 30. Item-person map for the Need for Order subscale.

of art, about which people might have no particular opinion) and a general tendency toward careful, detailed organization.

In addition, the PCA of item residuals indicated that the variance explained by the measures was 48.1%, and unexplained variance in the first contrast accounted for a mere 1.7 units (7.4%) of the total variance. With a range of 11.20 CHIPS (43.47-54.67), the Need for Order subscale covered the range of person ability estimates of 48.46 CHIPS (23.21-61.67) reasonably well. The difference between item difficulty and person ability means was 1.57 CHIPS, which indicated that the Need for Order subscale was appropriate for this sample (Figure 30).

Perceived Money-Time Competence. Next, on the Perceived Money-Time Competence subscale (Items 31-38) WINSTEPS initially yielded disordered thresholds and inadequate separation of thresholds. Combining categories yielded a 4-category alignment with proper ordering, good fit, and good separation (Table 60).

The dimensionality of the Perceived Money-Time Competence subscale was then investigated further. An initial exploratory factor analysis yielded two reasonable configurations, the first of which was a one-component solution that accounted for 35.96% of the variance. Loadings on the single component were strong (.39 to .77) and communalities ranged from .15 (Item 38) to .59. The second configuration was bifurcate, with two 4-item components consisting of Items 31-34 and 35-38, respectively. The two subscales accounted for 50.54% of the variance. However, a PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .81, suggesting that this instrument was fundamentally unidimensional.

Table 60
Category Function Statistics for the Perceived Money-Time Competence Subscale

Perceived competence category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Low	301 (14.25)	-6.37	-5.96	.94	(none)	
Slight	802 (38.43)	-1.37	-1.45	.88	-7.98	.33
Some	720 (34.92)	1.96	1.72	1.02	.67	.23
Good	298 (14.40)	4.55	5.03	1.21	7.31	.32

Note. N = 252; k = 5; Avg Measure = average measure; Exp Measure = expected measure.

When analyzed further with WINSTEPS all eight items of the Perceived Money-Time Competence subscale exhibited adequate fit and reasonable point-measure correlations (Table 61). The subscale yielded an item reliability estimate of .95, item separation of 4.23, a person reliability estimate of .70, person separation of 1.54, and a person strata statistic of 2.39, all of which are adequate. Moreover, the average inter-item correlation for the 4-item instrument was adequate (r = .37), and internal reliability was adequate (Cronbach's α = .70). The items dealing with psychotherapy and money were the most difficult to endorse (i.e., respondents perceived themselves to be less competent), which is not surprising: psychotherapy is likely a mysterious area for many, and managing money is challenging for many people. On the other hand, the time items (e.g., Item

31, *I get to appointments right on time*) were the easiest to endorse, as was expected; Japanese are generally meticulous about time.

Table 61
Ego Permeability, Perceived Money-Time Competence Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
38-good psychother	52.49	.40	1.29	3.3	1.38	4.1	.68
33-good with money	52.25	.40	1.20	2.3	1.22	2.5	.68
37-down to earth	49.95	.39	.89	-1.3	.89	1.4	.68
36-clear memory	49.59	.39	1.09	1.1	1.11	1.4	
32-desk neat	49.55	.39	.89	-1.4	.89	-1.3	.72
35-know safe areas	49.50	.39	.81	-2.6	.81	-2.4	
34- clear time sense	48.74	.39	.68	-4.6	.71	-4.0	.80
31-appointments on time	46.71	.40	1.16	1.8	1.14	1.7	.78
M	50.00	.39	1.00	-.2	1.02	.1	
SD	1.79	.00	.20	2.5	.22	2.6	

Note. N = 252, k = 8; Pt-M Corr = point-measure correlation. Psychother = psychotherapist.

As shown in Figure 31, the Perceived Money-Time Competence subscale covered the range of person ability estimates reasonably well: The range of Rasch-Thurstone thresholds was about 24 CHIPS, while the distribution of person ability estimates covered 34.04 CHIPS (32.70-66.74). Some redundancy was present in the instrument (e.g., Items 32, 35, and 36). The difference in the means of the person ability and item difficulty estimates was very small (0.10 CHIPS), which indicated that the Perceived Money-Time Competence subscale was at an appropriate level for this sample.

```
Persons with more       Item thresholds indicating more T-M competence
T-M competence
 62                +
 61           .    +
 60               T+                                                    38-good psychother .4
 59          .#    +                                                    33-good w money    .4
 58         .##    +                                                    37-down to earth   .4
                                                                        32-desk neat       .4
                                                                        35-safe areas
                                                                        36-clear memory
 57          .#    +                                                    34-clear time sense.4
 56           .    +
 55         .## S+                                                      31-appts on time   .4
 54    ########    +T
 53   .########    +                              33-good w money    .3
                                                  38-good psychother
 52       .####    +S                             37-down to earth   .3
 51     .######    +
 50 ########### M+M                               32-desk neat       .3
                                                  35-safe areas
                                                  36-clear memory
 49   #########    +                              34-clear time sense.3
 48   .########    +S
 47       .####    +                              31-appts on time   .3
 46       #####    +T
 45        .### S+
 44                +  33-good w money    .2
                      38-good psychother
 43         .##    +  37-down to earth   .2
 42          ##    +
 41          ##    +  32-desk neat       .2
                      35-safe areas
                      36-clear memory
 40               T+  34-clear time sense.2
 39         .##    +
 38                +  31-appts on time   .2
 37                +
Persons with less       Item thresholds indicating less T-M competence
T-M competence
```

Note. M = mean, S = one standard deviation, T = two standard deviations.
Figure 31. Item-person map with Rasch-Thurstone thresholds for the Perceived Money-Time Competence instrument.

Childlikeness. On the revised Childlikeness subscale (Items 39-44), WINSTEPS initially yielded disordered thresholds and inadequate separation of the thresholds. The data were negatively skewed, but combining the three disagree categories yielded a 5-category alignment with proper ordering, good fit, and good separation (Table 62).

Table 62
Category Function Statistics for the Childlikeness Subscale

Category	Count (%)	Avg Measure	Exp Measure	Outfit MNSQ	Structure Measure	*SE*
Not childlike	154 (12.25)	-4.22	-5.28	1.53	(none)	
Neutral	276 (22.43)	-3.68	-2.73	.72	-6.64	.45
Slightly	326 (26.00)	.01	-.02	.73	-2.16	.34
Childlike	300 (24.92)	3.49	3.15	.87	1.90	.34
Very childlike	193 (15.40)	6.64	6.72	1.07	6.90	.43

Note. N = 252; k = 5; Avg Measure = average measure; Exp Measure = expected measure.

The dimensionality of the Childlikeness subscale was then investigated. The average inter-item correlation for the 5-item instrument was adequate with r = .36, and internal reliability was adequate (Cronbach's α = .74). An initial EFA yielded a one-component solution that accounted for 51.53% of the variance. Loadings on the single component were strong (.52 to .87) and communalities ranged from .27 to .76. A PCA of item residuals in WINSTEPS showed that the disattenuated correlation of person ability estimates derived from items with positive and negative residual loadings was .92, indicating that this instrument was unidimensional. In addition, the PCA of item residuals indicated that the variance explained by the measures was 63.1%, and unexplained variance in the first residual contrast accounted for a mere 1.9 units (13.9%) of the total variance.

In the revised Childlikeness subscale, five items exhibited adequate fit and reasonable point-measure correlations, but Item 39 (*There are no sharp dividing lines between normal people, people with problems, and people who are considered psychotic or crazy*) was badly misfitting (infit MNSQ = 2.36, outfit MNSQ = 2.85). A perusal of misfitting responses found 31 persons (11%), which when deleted only improved the fit statistics to 2.07 and 2.53, respectively. Item 39 was thus deleted, and the Childlikeness subscale as originally postulated (Items 40-44) yielded satisfactory category function with a Rasch item reliability estimate of .91, item separation of 3.26, a Rasch person reliability estimate of .68, person separation of 1.47, and a person strata statistic of 2.29 (Table 63).

Table 63

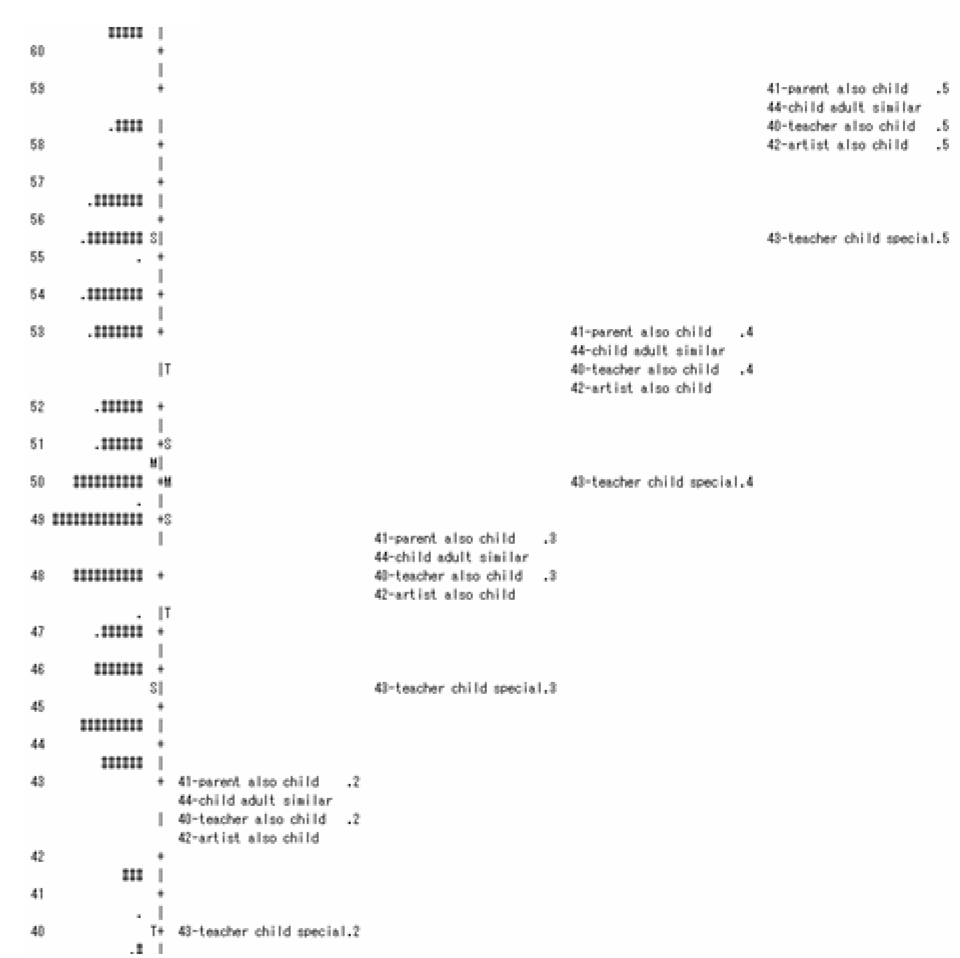
Ego Permeability, Childlikeness Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
41-parent also child	50.87	.33	.65	-4.7	.62	-4.6	.81
44-child adult similar	50.76	.33	1.37	-.5	1.42	4.4	.64
40-teacher also child	50.58	.33	.62	-5.3	.62	-5.1	.80
42-artist also child	50.13	.33	1.24	2.7	1.24	2.6	.70
43-teacher child special	47.66	.34	1.08	3.9	1.11	1.2	.67
M	50.00	.33	.99	-.5	1.01	-.3	
SD	1.20	.00	.31	3.8	.32	3.9	

Note. *N* = 252, *k* = 5. Item 39 was deleted. Pt-M Corr = point-measure correlation.

As shown in Figure 32, four of the five items were clustered around the mean. The exception was Item 43 (*A good teacher needs to help a child remain special*), which was easier to endorse. With a range of 2.42 CHIPS (48.22-50.64), the Childlikeness subscale covered a small portion of the person distribution of 27.88 CHIPS (36.14-64.02), yet the Rasch-Thurstone thresholds adequately covered the distribution of person ability estimates. The difference between the means of item difficulty and person ability estimates was 1.45 CHIPS, which indicated that the Childlikeness subscale was appropriate for this sample.

Persons with more childlikeness | Item thresholds indicating more childlikeness



Persons with less childlikeness | Item thresholds indicating less childlikeness

Note. M = mean, S = one standard deviation, T = two standard deviations.
Figure 32. Item-person map with Rasch-Thurstone thresholds for the Childlikeness subscale.

Sensitiveness. Finally, on the Sensitiveness subscale (Items 45 and 46), WINSTEPS initially yielded disordered thresholds and inadequate separation of thresholds. Combining categories yielded a 4-category alignment with proper ordering, good fit, and good separation (Table 64).

Table 64

Category Function Statistics for the Sensitiveness Subscale

Category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Not sensitive	109 (8.25)	-6.84	-8.09	1.27	(none)	
Somewhat	468 (35.43)	-3.44	-2.79	.83	-11.86	.52
Sensitive	633 (47.92)	2.64	2.33	.86	-1.72	.30
Very sensitive	111 (8.40)	9.23	9.49	1.08	13.58	.52

Note. N = 252; k = 5; Avg Measure = average measure; Exp Measure = expected measure.

The Rasch statistics yielded an item reliability estimate of .94, item separation of 3.93, a person reliability estimate of .78, person separation of 1.89, and a person strata statistic of 2.85. As shown in Table 65, both items exhibited excellent fit and reasonable point-measure correlations, and the Rasch-Thurstone thresholds indicated reasonable coverage of the person ability estimates (Figure 33).

Table 65

Ego Permeability, Sensitiveness Subscale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
46-very sensitive (self)	52.08	.51	.98	-.2	.95	-.5	.93
45-easily hurt	47.92	.51	.99	.0	.98	-.2	.93
M	50.05	.51	.98	-.1	.97	-.3	
SD	2.08	.00	.01	.1	.01	.1	

Note. N = 252, k = 2; Pt-M Corr = point-measure correlation.

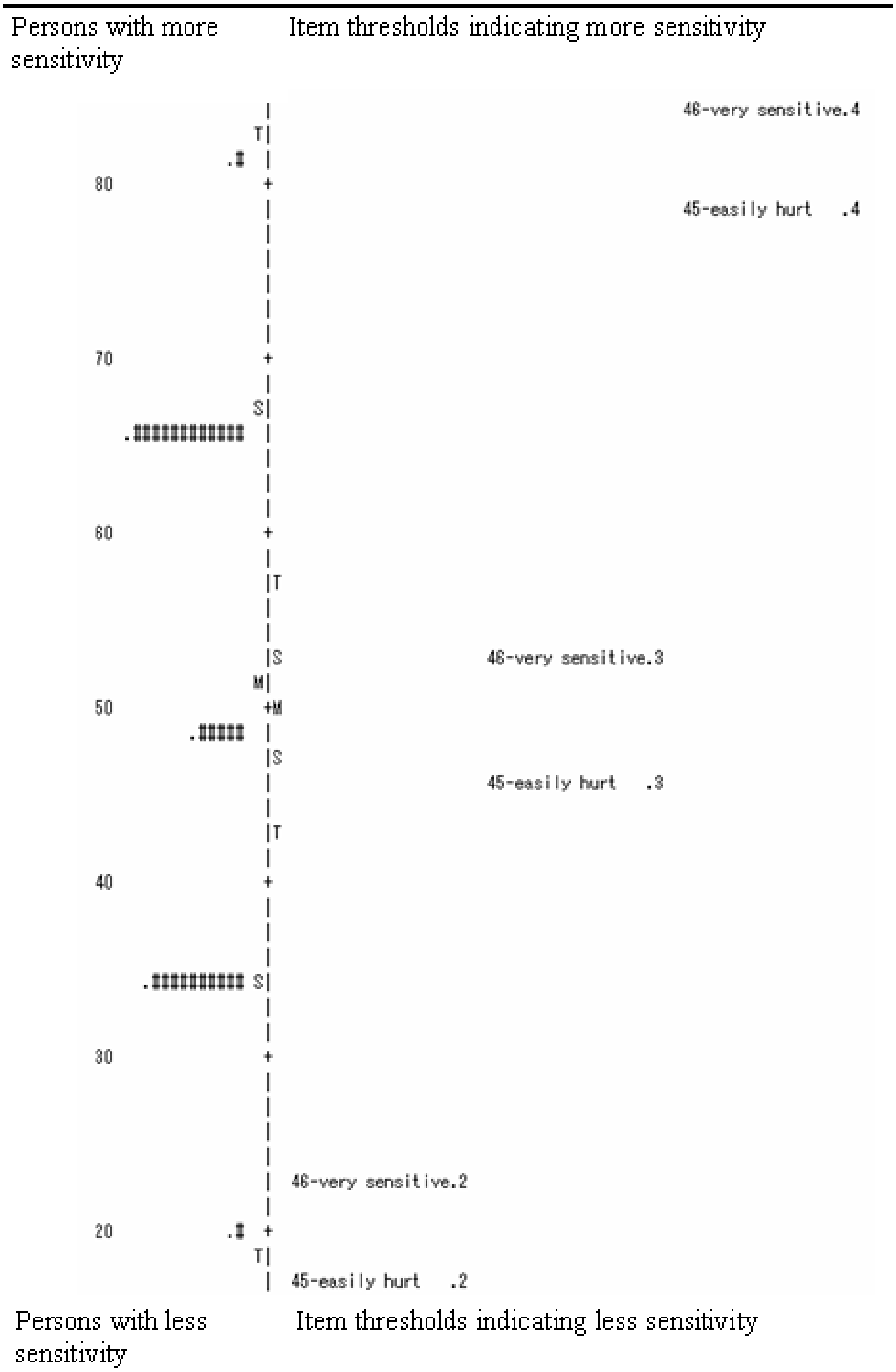


Figure 33. Item-person map with Rasch-Thurstone thresholds for the Sensitiveness subscale.

The dimensionality of the Sensitiveness subscale was then checked. The average inter-item correlation for the 2-item instrument was good (*r* = .75), and internal reliability was also good (Cronbach's α = .85). The PCA of item residuals indicated that the variance explained by the Rasch model was a strong 78.7%. With only two items, there was no unexplained variance in the first residual component, indicating that this subscale was strongly unidimensional.

As shown in Figure 33, the Rasch-Thurstone thresholds covered the range of person ability estimates, which ranged from 20 to more than 80 CHIPS. The difference in means was small (1.47 CHIPS), which indicates that the Sensitiveness subscale was at an appropriate level for these participants. These figures must be viewed with caution, however, for with only two items the subscale is poorly defined.

Summaries of the Ego Permeability subscales and the overall Ego Permeability instrument are shown in Table 66.

Table 66
Ego Permeability Subscale Summary

Subscale	*k*	*j*	Item Rel	Item Sep	Per Rel	Per Sep	% of Var	I-I Corr
Unusual Exper	12	3	.98	6.59	.76	1.80	44.97	.45
Need for Order	12	4	.98	6.87	.74	1.69	29.25	.20
Time-Money	4	5	.97	5.69	.65	1.36	52.97	.25
Childlikeness	5	4	.90	2.93	.65	1.37	51.53	.30
Sensitiveness	2	4	.93	3.66	.78	1.88	86.77	.59
total	39		.99	9.12	.81	2.05	-	

Note. N = 252; *k* = number of items; *j* = number of response categories; Rel = reliability; sep = separation; per = person; Exper = Experiences; Time-Money = Perceived Time-Money Competence. % of variance is from the WINSTEPS PCA of residuals for the respective subscales.

Ego Permeability Measurement Model

The original configuration of Ego Permeability consisted of five subscales: Unusual Experiences, Need for Order, Perceived Time-Money Competence, Childlikeness, and Sensitiveness. To confirm the 5-factor configuration, a confirmatory factor analysis was conducted, but the 5-factor model exhibited poor fit. The factor with the weakest path coefficient (Childlikeness) was then deleted, but the 4-factor model also had poor fit. The Sensitiveness subscale was removed, yet the resulting 3-factor model also had inadequate fit. Omitting the Unusual Experiences subscale yielded the best-fitting model, a 2-factor configuration with the Need for Order and Perceived Time-Money Competence subscales (Table 67).

Table 67
Step-by-Step Procedure for Revising the Ego Permeability Instrument

Model	χ^2	*df*	CFI	SRMR	RMSEA
5-factor model	1384.462	741	.732	.091	.063
4-factor model (delete Childlikeness)	1000.156	561	.763	.089	.061
3-factor model (delete Sensitiveness)	991.016	461	.706	.090	.068
2-factor model (delete Unusual Experiences)	322.436	168	.801	.065	.061

Note. CFI = Comparative fit index; SRMR = standardized root mean square residual. RMSEA = root mean square error of approximation.

However, the 2-factor configuration represents a somewhat different construct than Ego Permeabilty. The Need for Order subscale, with such items as *There is a place for everything and everything should be in its place* (Item 13), is concerned with acceptance of how the world and roles therein are ordered, and the

Perceived Time-Money Competence subscale (e.g., Item 32, *I keep my desk and worktable neat and well organized*) is concerned with how capably one imposes order on the world. As such, conceptualizing this latent factor as Imposition of Order would better represent the underlying concept.[12]

To further confirm that omitting the three subscales was prudent, a second confirmatory factor analysis using SEM was conducted (Figure 34). Because the three omitted subscales deal with cognitive, internally-perceived constructs, they

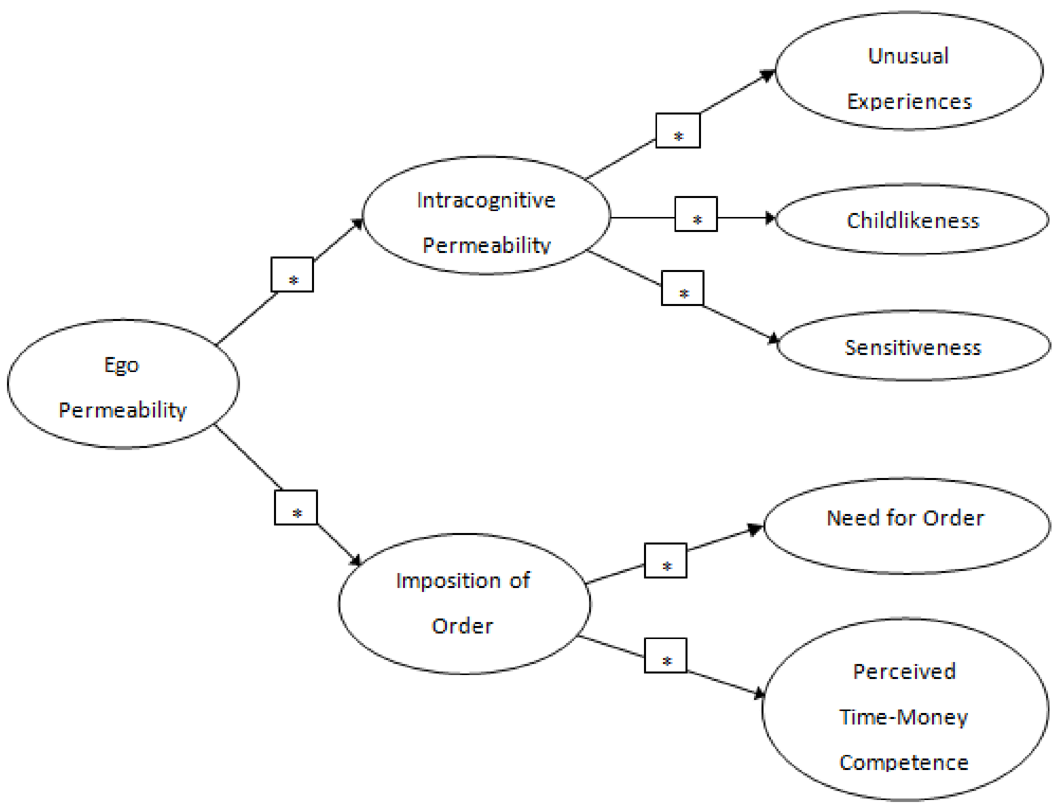


Figure 34. Hypothesized 2-factor model of Ego Permeability with Imposition of Order and Intracognitive Permeability.

[12] This could also be conceptualized as 'tolerance of ambiguity' (Budner, 1962; Ely, 1989; Furnham & Ribchester, 1995), but here I opt for Imposition of Order as it more transparently reflects the content of the items.

were posited to form a factor that was labeled Intracognitive Permeability. Thus, the model tested included two second-order factors, Imposition of Order and Intracognitive Permeability. Note that the valences of these two factors should be reversed: Imposition of Order should be negatively related to the notion of permeability, whereas Intracognitive Permeability would be positively related.

However, the SEM results indicated the model fit the data poorly: χ^2 = 1389.777 (p < .01), CFI = .729, RMSEA = .063, and 90% C.I. = .058-.068. More importantly, the path from Intracognitive Permeability to Ego Permeability was not significant, thus lending support to the 2-factor Imposition of Order configuration as the more appropriate model for these data. When assessed, the 2-factor model exhibited much better albeit moderately acceptable fit: χ^2 = 330.005 (p < .01), CFI = .827, RMSEA = .061, and 90% C.I. = .051-.070. Detailed results for both models are presented in Table 68.

Although Ego Permeability was posited to have five subscales, the results of a series of confirmatory factor analyses indicated that a 2-factor model with the Need for Order and Perceived Time-Money Competence subscales exhibited the best fit statistics of the four models tested. Moreover, with just those two subscales the configuration represents an Imposition of Order construct. Thus, in subsequent analyses the 5-factor Ego Permeability configuration is replaced by the 2-factor Imposition of Order construct.

Table 68

Summary of Fit Indices for 2-Factor and 5-Factor Ego Permeability Models

	2-factor	*5-factor*
Reliability Coefficient (rho)	.827	.833
Multivariate Kurtosis		
Mardia's coefficient	50.550	141.919
Normalized estimate	13.535	19.919
Residuals		
Average absolute standardized residuals	.047	.070
Average off-diagonal absolute standardized residuals	.052	.074
Model χ^2		
Model estimation method	ML	ML
Independence model χ^2 (*df* = 190, 741)	965.018	3307.450
χ^2 (*df* = 167, 693)	305.161	1389.777
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	1.827	2.005
Fit Indices		
Comparative fit index (CFI)	.822	.729
Incremental fit index (IFI)	.827	.733
Standardized root mean square residual (SRMR)	.063	.096
Root mean-square error of approximation (RMSEA)	.057	.063
RMSEA 90% confidence interval	.047-.067	.058-.068

Although Ego Permeability was posited to have five subscales, the results of a series of confirmatory factor analyses indicated that a 2-factor model with the Need for Order and Perceived Time-Money Competence subscales exhibited the best fit statistics of the four models tested. Moreover, with just those two subscales the configuration represents an Imposition of Order construct. Thus, in subsequent analyses the 5-factor Ego Permeability configuration is replaced by the 2-factor Imposition of Order construct.

Replacement Instruments Created for This Study

Attitudes about the Learning Situation

The Attitudes about the Learning Situation (hereafter Attitudes) instrument was a 4-item hybrid instrument constructed for this study. Specifically, it included two items which queried feelings about having more English classes in school and two items which asked about comfort levels when dealing with native speakers of English. The original data yielded poor separation and disordered thresholds, but combining categories into a 3-level scheme produced alignment with proper ordering, good fit, and good separation (Table 69).

Table 69
Category Function Statistics for the Attitudes about the Learning Situation Instrument

Attitudes category	Count (%)	Avg measure	Exp measure	Outfit MNSQ	Structure measure	*SE*
Negative	244 (8.25)	-5.16	-5.15	1.06	(none)	
Neutral	454 (35.43)	.10	.09	1.30	-5.38	.40
Positive	275 (47.92)	6.30	6.29	.89	5.38	.39

Note. N = 252; k = 5; Avg Measure = average measure; Exp Measure = expected measure.

Rasch statistics yielded an item reliability estimate of .97, item separation of 6.06, a person reliability estimate of .33, person separation of .70, and a person strata statistic of 2.85. As shown in Table 70, all items exhibited excellent fit and reasonable point-measure correlations. The PCA of item residual results indicated that unexplained variance in the first contrast was 1.8 (22.9%) and the total variance explained by the Rasch model was 7.9 units (100%).

Table 70
Attitudes about the Learning Situation Scale Measure: Rasch Item Fit Statistics

Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
4-comfy with NS	52.81	.51	.74	-3.4	.74	-3.4	.74
3-not nervous NS	52.37	.51	.80	-2.7	.84	-2.0	.72
2-OK more Eng	50.42	.51	1.10	1.2	1.07	.9	.64
1-absolutely Eng	44.40	.54	1.31	3.4	1.61	4.9	.48
M	50.00	.52	.98	-.4	1.07	-.3	
SD	3.36	.01	.23	2.8	.01	.1	

Note. N = 252, k = 4; Pt-M Corr = point-measure correlation.

As shown in Figure 35, the two items dealing with interacting with native speakers of English (Items 3 and 4) were predictably difficult to endorse, likely reflecting the participants' anxiety about engaging in English conversation.

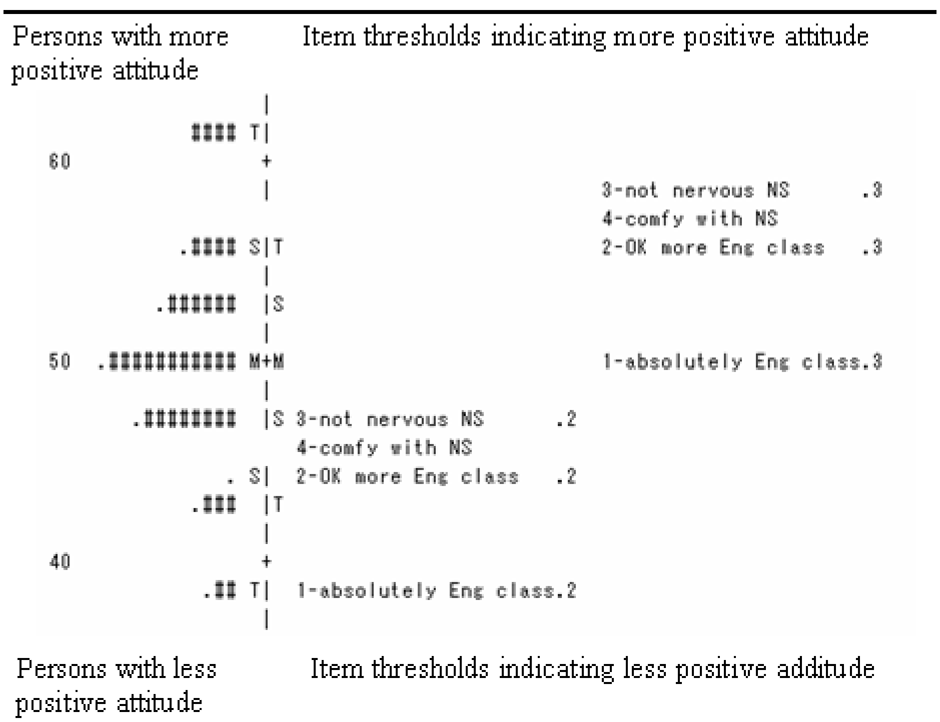


Figure 35. Item-person map with Rasch-Thurstone thresholds for the Attitudes about the Learning Situation instrument.

However, taking more English classes (Item 2) was slightly easier to endorse, and the general belief in the necessity of more English classes (Item 1) was much easier to endorse.

English Experience

The English Experience variable was a composite that quantifies seven experiences in which participants could have been in contact with English. This was used in lieu of the Context factor in the replication of the MacIntyre and Charos (1996) model. The seven experiences include living abroad, study abroad, a homestay in a foreign country, conversation school attendance, the age at which English study began, and compulsory English education (Table 71). Because English is a compulsory subject in secondary education in Japan, the default score for all Japanese participants was one; in addition, the one non-Japanese participant had also undergone compulsory English classes in his secondary education. The length and richness of any additional English experience counted for more points with, for example, having begun English at age six counting for an extra two points and between nine and 12 garnering one extra point.

As shown in Table 71, compulsory education constitutes the most common English experience. The second most common was travel abroad, yet just over half of the participants had done so: 128 (51.82%) of the 247 participants that responded. Interestingly, 81 respondents (32.79%) began English before the onset

Table 71
Composition and Scoring Criteria of the English Experience Instrument

	Score			
Category	4	3	2	1
Live abroad				
English (L1)	≥ 3 years (*n* = 9)	< 3 years (*n* = 2)		
ESL		≥ 3 years (*n* = 2)	< 3 yrs (*n* = 3)	
EFL			≥ 3 yrs (*n* = 4)	< 3 yrs (*n* = 7)
Study abroad			> 30 days (*n* = 20)	< 30 days (*n* = 30)
Homestay			> 30 days (*n* = 14)	< 30 days (*n* = 47)
Conversation school			≥ 3 yrs (*n* = 17)	< 3 yrs (*n* = 46)
Starting age			< 9 yrs (*n* = 27)	9–12 yrs (*n* = 54)
Travel				(yes)[a] (*n* = 128)
Compulsory education				everyone (*n* = 247[b])

Note. English (L1) = country in which English is spoken as a first language; ESL = ESL country; EFL = EFL country. [a]Travel abroad was further subdivided into three categories: travel to an English L1 country was .5, travel to an ESL country was .25, and travel to an EFL country was just .1. [b]Five respondents did not provide information.

of compulsory English education in junior high school. The English Experience measure was the sum of the various scores, and it ranged from one point for those whose only English experience was the compulsory English education in school to a maximum of 18.85.

Summary

In this chapter, the personality variables hypothesized to improve the three original models were validated. Using Rasch analyses, the reliability, validity, and appropriateness of the instruments were evaluated, and in cases some minor post-hoc adjustments allowed improvements to be made to the instruments by revising the number of category function steps or deleting misfitting items. In addition, the optimal configuration of the Ego Permeability instrument was found to consist of just two subscales, the Need for Order and Perceived Time-Money Competence subscales, which together constitute an Imposition of Order construct. In the primary analyses in Chapter 8, the 5-factor Ego Permeability configuration is replaced by the 2-factor Imposition of Order construct.

In Chapters 4 through 6 the results of preliminary analyses were covered in detail. Chapter 4 examined initial data screening and validation of the two proficiency instruments, Breadth of Vocabulary Knowledge and Listening Proficiency. In Chapter 5 the results of the preliminary analyses for the individual difference variables were presented; those variables include Motivation, L2 Communicative Anxiety (both the L2 Communicative Anxiety instrument the FLCAS), Frequency of L2 Communication, L2 Willingness to Communicate, and International Posture. In Chapter 6 the validation results of the four personality variables (Distancing, Extroversion, Ego Permeability, and Personality) was presented. Chapter 7 is a brief discussion of the preliminary analysis results presented in Chapters 4, 5, and 6, after which the primary results of this study are presented in Chapter 8.

CHAPTER 7

DISCUSSION OF THE PSYCHOMETRIC PROPERTIES OF THE INSTRUMENTS

In this chapter the psychometric properties of the instruments are summarized and discussed. Many of the results have been covered in the previous three chapters, and in this chapter my purpose is to more concisely present the various preliminary analysis results.

Psychometric Properties of the Instruments in this Study

The first research question, which concerned the psychometric behavior of the instruments utilized in the current study, asked, “To what extent are the instruments used in this study reliable and valid in the university EFL contexts in this study?” The instruments were found to be fundamentally sound and configured much as originally constructed. The current study is, to the best of my knowledge, the first time in which many of these instruments were validated using Rasch analysis or, when necessary, with structural equation modeling. The findings for the respective scales are discussed below.

Breadth of Vocabulary Knowledge

The Breadth of Vocabulary Knowledge instrument included 40 items drawn from the 2,000-, 3,000-, and 5,000-word levels of the Vocabulary Levels Test

(Nation, 1990) in addition to 10 items from the University Word List. Using Rasch analysis, these 40 items were culled from the original list of 72 items (18 items per frequency level), and in the current study they were evaluated using the partial-credit Rasch model. The analysis indicated that all 40 items functioned well and were, for the most part, ordered as expected in terms of difficulty.

Listening Proficiency

The Listening Proficiency instrument was created for the current study, but the format is familiar to Japanese students as it is commonly used on entrance examinations in Japan. It consisted of four short dialogues with three or four multiple-choice comprehension questions each and a longer passage of 198 words with five multiple-choice comprehension questions. The analysis indicated that all 16 items functioned well and were generally ordered as expected in terms of difficulty.

Motivation

In this study, the Motivation instrument was operationalized using the original bifurcate configuration of Motivational Intensity and Desire to Learn English (e.g., Gardner & Lambert, 1972); this configuration was also utilized in the studies by MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004). However, the disattenuated correlation value of .85 from the Rasch analysis was suggestive of a strong single dimension rather than separate subscales. When further investigated with a confirmatory factor analysis using EQS, the results were ambiguous with both a single dimension and the original configuration of two dimensions having similar fit statistics, yet based on the strong disattenuated correlation and theoretical considerations, a single dimension was deemed more appropriate.

The unidimensionality of the Motivation instrument could be due to two factors. First, English is generally treated first and foremost as a school subject rather than a tool of communication (Sick, 2006), so the notion of intensity might not be appropriate. Second, although many Japanese learners of English seem to have two types of motivation for learning English (Yashima et al., 2004), those two types might conflate because of the de facto role of English for many Japanese EFL learners. Specifically, Yashima et al. (2004) noted that the more pressing of the two motivational types is to pass the ubiquitous entrance examinations, while the second type of motivation is a rather vague notion that English will be useful in the future in some capacity not yet known. Based on these observations, the role of English is primarily instrumental: Passing entrance examinations is of paramount importance to one's subsequent education and thereafter to one's position in society, which is intrinsically linked to one's educational background.

Second, in examining the items, the delineation into two subscales seems questionable as some items could logically fit in either subscale. For example, the original Desire to Learn English subscale includes Item 1 (*When I have assignments to do in English, I try to do them immediately*), which seems to fit as well in the Motivational Intensity category. One's desire to learn English might lead to the immediate completion of homework, but motivational intensity should lead to the same result. Similarly, although Item 12 (*I intend to continue studying English after graduating from university*) is in the Motivational Intensity subscale, it encapsulates a strong desire to learn English. Neither item definitively belongs to one or the the other of the two subscales.

Moreover, the validity of one of the motivation items is suspect. Item 1 addresses when a student does homework, but this question could be confounding learning style (or study style) with motivational elements. I have taught students majoring in International Studies who ostensibly had substantial desire to learn English, yet some of those students were chronically late with homework and exhibited poor attendance in my English class. While a logical suppostion is that students will enthusiastically (read: immediately) devote time and effort to homework in classes in their major field of study, my experience is that study habits and learning styles are quite consistent across the spectrum of classes; thus, students that procrastinate generally do so regardless of the class, and punctual, well-organized students conduct themselves in that fashion in all their classes.

Thus, based on statistical support for unidimensionality, the poor discrimination of the two posited dimensions by some items, and suspect validity for some items, the Motivation instrument was treated as unidimensional.

International Posture

The International Posture instrument with its four subscales was a primary focus of the current study based on the central role that it plays in SLA. Because the original configuration included the four subscales, all four were examined using a confirmatory factor analysis and then a Rasch analysis. The basic structure of the four persisted, but with minor changes: Item 11 (*I'm interested in volunteer activities in developing countries such as participating in Youth International Development Assistance*) was added to the Approach-Avoid Tendency subscale, the Interest in International Vocation/Activities subscale lost Items 11 and 12, the Item 16, (*International news is more important than local news*) was added to the Interest in Foreign Affairs subscale, and Item 17 (*International news makes interesting, useful content for school classes*) was added to the Intercultural Friendship Orientation.

With the four subscales adequately defined and sufficiently unidimensional, the subsequent question was which of the four subscales to include in the International Posture instrument. In Yashima (2002), all four subscales were used, while in Yashima et al. (2004), the Intercultural Friendship Orientation was omitted based on item overlap with the other three subscales. However, neither the items nor the subscales exhibited overly high correlations that would have been indicative of overlap (the maximum correlation was .56). In addition, although the various items overlapped in the sense that they all dealt with international things or people, conceptually the four subscales address different aspects of an international orientation. The Intergroup Approach-Avoid Tendency subscale looks at the proclivity to interact with individual *persons* (e.g., individuals such as a neighbor or someone in need of assistance while shopping), whereas Intercultural Friendship Orientation is more focused on outcomes of activities with an international element such as taking an English test or interacting with *people* from another culture (i.e., the focus is on people in general and not individuals). The Interest in Foreign Affairs subscale is concerned with interest as manifested by the consumption and use of foreign news, while the Interest in International Vocation/Activities subscale measures an instrumental orientation concerned with living, working, and traveling abroad.

The original configuration of the International Posture instrument included these four subscales, but the Rasch analyses of the respective subscales indicated that some reconfiguration was necessary. A confirmatory factor analysis using EQS was conducted to examine the dimensionality of the International Posture instrument; the results indicated that both the original 4-factor model and a 3-factor model (Yashima et al., 2004) fit the data poorly. The model with the best fit was a 2-factor model with Intergroup Approach-Avoidance Tendency and Intercultural Friendship Orientation: χ^2 (32, N = 252) = 185.716 (p < .01), CFI = .935, IFI = .937, RMSEA = .066, and 90% C.I. = .052-.080. This 2-factor configuration was thus used in subsequent analyses.

L2 Anxiety

Two instruments were used to assess L2 anxiety. On the L2 Communicative Anxiety scale (MacIntyre & Charos, 1996; Yashima, 2002), results indicated that the presence of two dimensions, one dealing primarily with anxiety about interactions with strangers and the other dimension concerned with interactions with friends or acquaintances. When analyzed with Rasch analysis, the respective subscales displayed adequate fit to the Rasch model and satisfactory unidimensionalality. However, the question remains of whether this is an appropriate instrument for measuring anxiety in EFL contexts in which most L2 interactions are not with strangers or in such contexts as standing in line, but rather within the confines of L2 classrooms.

The FLCAS was the larger of the two scales used to measure anxiety in the current study. After removing three items for use in the Attitudes scale, the FLCAS consisted of 30 items addressing anxiety related to the foreign language classroom (rather than the extracurricular situations in which L2 speakers might encounter English). Of the 30 items, 28 items had good Rasch fit statistics and formed a single dimension.

Of the two scales, the FLCAS is the more logical one to assess foreign language anxiety in this context because it deals with more common anxiety-inducing elements than does the L2 Communicative Anxiety instrument. For example, the FLCAS includes Item 19 (*I am afraid that my English teacher is ready to correct every mistake I make*), which describes a common experience given the extensive grammar focus in secondary school English education. However, speaking with an acquaintance or a stranger in English while standing in line (Item 8 and Item 4, respectively, of the L2 Communicative Anxiety scale) are probably much less common occurrences for most Japanese EFL learners.

Perceived L2 Competence

This instrument is another based on the 3 x 4 WTC matrix of venues and speaker groups. The results of the Rasch analysis indicated that the items fit the model well and formed a single dimension.

However, in hindsight a more classroom-focused instument or at least several classroom-oriented questions would have made this instrument more appropriate for this EFL context. Much as anxiety was better operationalized using the FLCAS than the L2 Communicative Anxiety instrument, this instrument could have benefitted from the addition of items modeled after those on the Frequency of L2 Communication instrument such as *I feel competent volunteering answers in my English class(es) at school* or *I feel competent participating in English classroom activities such as pairwork*.

Frequency of L2 Comunication

The short Frequency of L2 Communication scale should have been longer, and ideally should have included an evaluation by the researcher of the participants' communication activities. A further point is that the proficiency and frequency scales represent a mismatch, with the former focused on receptive skills and knowledge while the latter is focused on productive activities. However, I would argue that the mismatch is not problematic, for two-way communication is of necessity an exercise in production, while proficiency—whether receptive or productive—includes vocabulary knowledge. In social interactions, moreover, listening is crucial to understanding the interlocutor's message and, more profitably, to responding appropriately. In the current study, frequency of communication included both volitional acts of communication outside the classroom context and compulsory communication in the language classroom (e.g., participating in pairwork activities). These both constitute communication, and even when made to communicate at the behest of the teacher in a classroom, the degree of effort expended in doing so reflects a certain type of volition on the part of the learner. For example, when students in my speaking classes are given speaking tasks, some engage briefly and grudgingly, while others enthusiastically speak at length. Recall that the scale for frequency of L2 communication was a 7-point Likert scale, which allows participants to express varying degrees of speaking frequency vis-à-vis individual items.

In the Yashima et al. (2004) study, Items 2 (*I answered when I was called upon by the teacher*) and Item 3 (*I participated in classroom activities such as pair work*) were omitted because communication in those situations was based not on the individual's volition but rather on the fashion in which the teacher conducted the class (p. 670, Note 4). In this study, however, I retained all five items because communication does not necessarily have to be a volitional act, and one can argue that most L2 communication for Japanese EFL learners is in the junior and senior high school English classroom, which is one facet of the compulsory curriculum in secondary education in Japan. While volition plays an important role in L2 communication outside the classroom, the reality remains that (a) communication is still communication, regardless of volition, and (b) the majority of L2 communication for most EFL learners in Japan takes place inside the L2 classroom.

Furthermore, the teacher's influence on frequency of L2 communication goes beyond calling on students (Item 2) and having students participate in classroom activities such as pairwork (Item 3). The atmosphere established by the teacher can influence the frequency of communication regardless of the location: I had teachers to whom I was loathe to speak, whether inside the classroom or outside, and some of my students resist talking with me outside the classroom.

Finally, Item 4 (*I asked teachers questions or talked to them outside the class period*) concerns extracurricular communication with a teacher, yet the teacher's identity from the student's viewpoint is intrinsically linked with the classroom. When speaking Japanese the notion of the teacher's identity as a teacher is overtly coded with the lexeme *sensei* (teacher), which is the appropriate form of address when conversing directly with a teacher. In American English the form of address when speaking with a teacher includes an everyday title of respect (e.g., mister), but the teacher relationship is not lexically coded. In class I address students as "Mr. Suzuki" or "Miss Tanaka", and I require students to reciprocate by addressing me as "Mr. Elwood." However, some of my students prefer to code-mix (which I allow) when speaking English by using the Japanese form of address preceding a sentence in English: "Sensei, I have a question." It seems that for my students the teacher is always intrinsically linked with the classroom regardless of the code. I suspect that this is true for Japanese students in general, and this implies that Item 4 is a classroom-oriented item and that four of the five items deal with the classroom context.

Ego Permeability

The ego permeability instrument with its five subscales was viewed as a prime candidate for reconfiguration, but the Rasch analysis indicated that the individual subscales were valid and reliable. One item misfit the Rasch model to the extent that it was deleted (Item 39, *There are no sharp dividing lines between normal people, people with problems, and people who are considered psychotic or crazy*). The remaining items loaded on their respective factors, which were sufficiently unidimensional. The one correction made on the overall scale was to rename the Perceived Competence subscale as the Perceived Time-Money Competence subscale based on the content of the items (and to distinguish it from the Perceived L2 Competence scale).

Although the subscales emerged nearly as originally hypothesized with five subscales, the results of a series of confirmatory factor analyses using EQS indicated that a 2-factor model with the Need for Order and Perceived Time-Money

Competence subscales was superior to configurations with more subscales. The subscale was dubbed Imposition of Order to reflect the content of the items, and in subsequent analyses it was used instead of the original 5-factor Ego Permeability instrument.

The Rasch analysis indicated that the individual subscales were fundamentally sound, but the SEM results showed that the optimal configuration consisted of just two subscales instead of the five originally posited. The reasons the ego permeability variable crumbled as it did in this context are unclear, but one possibility is that the underlying construct of ego permeability might be different in this context than in North American contexts. Choi and Choi (2002) examined this question, arriving at the conclusion that in East Asian contexts, one's self-concept generally consists of different co-existing parts; this is somewhat different than the North American identity in which the primary construction is a positive / not positive dyad. However, in many East Asian contexts one can, with no contradiction, include elements that North Americans would view as incompatible.

An analogy might serve to illustrate this: In the researcher's North American upbringing, the Hegelian dyad consists of a one-dimensional construct in which a quality changes in one direction or its diametric opposite (e.g., black or not black, which is white). However, in Asia such a change occurs in a two-dimensional construct (or, arguably, three) in which any change is not necessarily indicative of a change in a particular dimension. In other words, the color change could be the addition of degrees of red instead of a change in the black-white element.

A further result that calls into question the viability of Ego Permeability in this context was the unsuccessful Ego Permeability measurement model that posited two latent variables underpinning the variable: Intracognitive Permeability and Imposition of Order. As noted, the former is an intra-psychic factor, while the latter is very much concerned with interacting with the world—in short, those form an inner-outer Hegelian dyad, the measurement model of which did not adequately account for the data.

Personality

The Big 5 personality construct played a central role in the MacIntyre and Charos model, and the extroversion subscale also played a crucial role in my extension of the Yashima et al. (2004) model. In the current study a shortened version (MacIntyre & Charos, 1996) of the Bipolar Scale of Global Personality Traits (Goldberg, 1992) was used to assess the Big 5 global personality traits; this includes five of the original 12 subscales. Because three of the original subscales underwent considerable realignment, I address those changes next.

Extroversion. According to the initial WINSTEPS analysis, the Extroversion subscale (Items 1-7) gained Item 12 (*pleasant, agreeable*) and Item 21 (*simple, frugal*). However, Item 21 had poor fit to the Rasch model, and the wealthy – frugal dyad did not logically fit well with the other eight items; Item 21 was thus deleted. The Extroversion subscale included the following items: 1 (*outgoing*), 2 (*energetic*), 3 (*talkative*), 4 (*bold*), 5 (*spunky, active*), 6 (*assertive*), and 12 (*pleasant, agreeable*).

Of the five subscales, the Extroversion subscale (originally labeled the Introversion-Extroversion subscale) was robust, playing important roles in both the replication of the various models and the extension of the Yashima models. As noted by Dewaele (2005), this subscale is probably the most robust of the five, consistently appearing as the strongest subscale regardless of the number or composition of additional subscales.

Diligence. Next, the revised Diligence subscale (Items 14-16, 19, 20, 29-31, and 35) included four original descriptors, 15 (*organized*), 16 (*responsible*), 19 (*thorough*), 20 (*hardworking*) in addition to five new ones: 14 (*generous*), 29 (*intelligent*), 30 (*analytical*), 31 (*reflective*), and 35 (*sophisticated*). The addition of Items 29 (*intelligent*), 30 (*analytical*), and 31 (*reflective*) is a logical step, as these three qualities are commonly associated with diligence related to school and extracurricular activities. The addition of reflection is especially prudent in a Japanese context because *hansei* [reflection] in the form of a *hanseikai* [meeting for reflection] is a common addendum to an activity or in response to a misdeed, for which a student can be directed to write a *hanseibun* [self-reflection essay] to atone for the malfeasance. Similarly, being generous is also part of diligence in study or activities, for students commonly work collaboratively with other students.

Emotional Stability. Next, the revised Emotional Stability subscale (Items 18, 22-28) included Item 18 (*practical*) and the original seven items: 22 (*calm*), 23 (*relaxed*), 24 (*at ease*), 25 (*not envious*), 26 (*stable*), 27 (*contented*), and 28 (*not emotional*)[13]. The addition of Item 18 (*practical*) is a logical step, for the ability to act in a practical manner indicates a certain degree of objectivity (read, 'stability').

Agreeableness. Next, the revised Agreeableness subscale (Items 8-10, 13, and 17) included four original descriptors, Items 8 (*warm*), 9 (*kind*), 10 (*cooperative*), and 13 (*trustful*) in addition to Item 17 (*conscientious*). Three of the original items, 11 (*not selfish*), 12 (*pleasant, agreeable*), and 14 (*not stingy*) loaded on different factors.

As shown below, this subscale was the only one that was dropped from the models because of non-significant paths, which might have occurred because the Agreeableness construct was poorly defined; as noted, three of its original items loaded on other factors.

Openness to Experience. Finally, on the Openness to Experience subscale (Items 11, 32-34), just three of the original seven items remained: 32 (*curious*), 33

[13] Items 9, 11, 13, 25, and 28 were reverse-coded so the valence would match the other items.

(*imaginative*), and 34 (*creative*); recall that Item 11 (*selfish*) had poor fit statistics and was deleted. Items that were moved to other subscales included Items 29 (*intelligent*), 30 (*analytical*), 31 (*reflective*), and 35 (*sophisticated*). The three items (and the deleted fourth item, *selfish*) appear to define an impulsive, emotional construct, whereas the items that loaded elsewhere (e.g., *analytical*) are suggestive of a rational, considered approach to experiences. In this case, the subscale appears to be more accurately labeled as openness to experience rather than sophistication or intelligence. Finally, with only three items, this subscale would benefit from additional items to better define the construct and increase measurement precision.

Attitudes about the Learning Situation

This 4-item scale yielded good Rasch fit statistics, and it represents an improvement over the 2-item instrument used by MacIntyre and Charos (1996), which simply asked the extent to which the participants had a good attitude vis-à-vis the teacher and the classroom. “Straight from the horse’s mouth” might be the most direct method of getting information, but a self-adjudicated estimate is at risk of being subjective. In hindsight and with an eye toward future research, a more nuanced look at attitudes, something similar to the original Motivational Intensity subscale of Motivation (Gardner & Lambert, 1972) that asked about activities indicative of the degree of intensity, would probably assess attitudes more objectively.

English Experience

The composite English Experience instrument replaced the Context variable in the MacIntyre and Charos (1996) model, and it assigned higher values for experiences of longer duration or for those spent in English-speaking contexts. Essentially a demographic variable, it was the sole instrument not analyzed with Rasch analysis. As expected, it was heavily skewed toward the pole reflecting little English experience: Just over half of the participants (51.82%) had traveled abroad, and even fewer had experience living abroad (10.93%), studying abroad (20.24%), or doing a homestay abroad (24.70%). The scale represents a novel attempt to quantify English experience by including both the breadth and depth of the various activities.

Rating Scale Performance

In addition to the performance of the respective scales listed above, the number of categories bears mentioning. As noted in the Results chapter, for 16 of the 22 scales used in the current study, the results of the Rasch analyses indicated that four was the optimal number of categories. Of the four remaining scales, Extroversion had seven categories, Openness to Experience had six, Childlikeness had five categories, while three instruments, Frequency of L2 Communication, Unusual Experiences, and Attitudes, resolved into 3-category scales.

In the current study, longer scales were mostly unnecessary as the findings indicate that Likert scales of five or more categories can result in underutilized categories. Based on that fact, the category function of instruments with five or more categories should be carefully examined and such instruments used with caution. This finding agrees with results from Cowan's (2000) study, in which mental storage capacity was found to average four chunks of information. Moreover, it suggests that Miller's well-known (1956) admonition about the magic number of categories being "seven (plus or minus two)" was too generous. Although longer scales can be perceived to allow finer delineation of responses (Preston & Colman, 2000), constraints on the capacity of humans for processing information (Baddeley, 1994) limit the usefulness of such scales, as demonstrated in the current study.

Summary

In this chapter, the psychometric properties of the instruments used in this study were discussed. The Rasch analyses provided screening of the various instruments for validity, reliability, and appropriateness for the participants, and confirmatory analyses were used to arrive at the optimal configuration for several of the scales. Of particular note is that just one of the 22 instruments retained the full contingent of seven category steps, while 16 were revised to have just four category steps.

The Rasch person ability estimates represented an improvement over the raw data, but we must bear in mind Linacre's (2006) admonition that instruments and data never attain perfect fit to the Rasch model; if they did, the resultant data would truly be interval data. Failing to meet that high standard, however, it is assumed that these instruments and data are superior to the original data.

In Chapter 8, these data are first screened to investigate whether they meet the assumptions of structural equation modeling. Thereafter, the models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) are replicated and extended.

CHAPTER 8

RESULTS

In this chapter I present the results of the various analyses conducted in the present study. Research questions are presented and answered sequentially. The SEM questions were addressed in three stages. In the first stage, the results are presented from the assessment of the L2 Communicative Confidence measurement models. The second stage involves the path analyses of the MacIntyre and Charos model. Two versions of the original model are investigated, one using the original Communication Anxiety scale and the second using the FLCAS. Thereafter the revised MacIntyre and Charos model with the personality variables is tested. In the third step, results are presented for the replication and the hypothesized extension of the Yashima (2002) model and the Yashima et al. (2004) model. Both the original and extended configurations are evaluated, so results are presented for four models: the original Yashima (2002) model, the extended Yashima (2002) model, the original Yashima et al. (2004) model, and the extended Yashima et al. (2004) model.

Structural Equation Models

The following sections present results from analyses with structural equation modeling. First, screening of the Rasch data is presented, after which the assumptions necessary for SEM are examined. Next, the L2 Communicative Confidence measurement model included in the Yashima models is investigated.

Finally, the original model and variations are presented for the MacIntyre-Charos (1996) model, the Yashima (2002) model, and the Yashima et al. (2004) model.

Rasch Data Screening and SEM Assumptions

Given instruments that have perfect fit to the Rasch model, true interval measures can be constructed from the raw scores, which are ordinal data. Although the instruments used in the current study did not meet the strict criterion of perfectly fitting the Rasch model, it is assumed that the Rasch measures approximate true interval scales better than the raw scores from which they are derived. Pursuant to this, Rasch measures should be screened to guarantee they meet the assumptions necessary to conduct a specific statistical analysis. The steps taken in screening the Rasch measures to meet the assumptions of SEM are presented in the following pages.

Sample size and missing data. Kline suggested (2005) that a sample size in excess of 200 is advisable, and the current study with a sample size of 252 exceeds that value. Kline also suggested that a ratio of 20 respondents per freely estimated parameter is ideal, while a ratio of 10:1 is more practical. In the current study, the minimum ratio among the various path analysis models was 5.7:1, which is somewhat small, but the SEM models had a better minimum ratio of 9.7:1. Although a larger sample size would have been better for evaluating the rather complex path analysis models, the sample size was deemed appropriate.

In the structural equation models, Rasch person ability estimates were used. Because these estimates compensate for missing data, the data were complete.

Multivariate normality. One assumption of SEM is multivariate normality. SEM can tolerate a certain degree of non-normality, with robust methods able to handle modest departures from non-normality (Bentler, 2006). An examination of the significance of skewness and kurtosis indicates non-normality for small samples, yet for large samples minor perturbations in the data can yield statistically significant skewness and kurtosis. To check for normality, Tabachnick and Fidell (2004, p. 721) suggested perusing distribution plots for samples of 200 or more, so histograms for the 22 variables were produced and examined using SPSS. Some skewness and kurtosis was present for most variables, with L2 WTC, Intergroup Approach-Avoidance Tendency, and Need for Order having the highest levels (see Table 70). To reduce the levels of skewness and kurtosis, transformation of the variables was considered, but the results indicated little or no improvement. While excessive kurtosis can result in underestimates of variance, this problem disappears for sample sizes greater than 200 (Waternaux, 1976). Thus, the data were not transformed.

EQS output enables further investigation of the extent of multivariate kurtosis in a given model. Mardia’s coefficient and its standardized coefficient are provided as well as the five cases (persons) that make the largest contribution to the kurtosis. Byrne (2006, p. 199) notes that a case that is ‘strikingly different’ from other cases can be deleted, and in the models tested in this study several such cases were deleted and the analysis repeated. Regarding the size of the standardized Mardia's coefficient, Bentler (2006, p. 106) suggested, "In practice, values larger than three provide evidence of nontrivial positive kurtosis, though modeling statistics may not be affected until values are five, six, or beyond" (p. 106). Furthermore, Byrne (2006, p. 140) pointed out that a comparison of the uncorrected χ^2 and the Satorra-Bentler corrected χ^2 also sheds light on the extent to which data are non-normally distributed. In the current study, the discrepancy between the two values of χ^2 was quite small when the standardized Mardia's coefficient was less than 10. That discrepancy as well as the value of Mardia's coefficient were the general criteria used in deciding whether to use robust ML estimation.

Outliers. An outlier is a person with an extreme value on one variable (a univariate outlier) or an unusual combination of scores on multiple variables (a multivariate outlier). Either case is problematic for parametric analyses because outliers exert an undue influence that threatens the generalizability of the results. Diagnosing outliers can be done by examining *z*-scores and checking distribution plots. *Z*-scores with an absolute value in excess of 3.29 are indicative of univariate outliers, and scores that are separate from the distribution are also suggestive of outliers.

The initial perusal of *z*-scores yielded 18 scores from 14 respondents in excess of 3.29; of those 18, two persons accounted for three scores each. A series of three regression analyses using SPSS REGRESSION with a cutoff Mahalanobis value of $p < .001$ ($\chi^2 = 51.148$, $df = 21$) indicated the 13 persons were multivariate outliers.

Stepwise regression was then employed to discern the variables on which the multivariate outliers differed from the remaining 252 cases. The outliers differed on four variables: L2 Communicative Anxiety, L2 WTC, Frequency of L2 Communication, and Motivation. Although Frequency was virtually the same for the two groups, the outlier group exhibited lower Motivation, L2 WTC, and L2 Communicative Anxiety. This combination of low Motivation and L2 WTC should correspond with high levels of L2 Communicative Anxiety, but that was not the case for these 13 outliers. Upon looking at the characteristics of these 13 participants, three were of quite high proficiency (one with a reported score of 900 on the TOEIC), yet no clear characteristics of the group were in evidence. This suggested that the outliers were randomly distributed and thus pose a minimum threat to the generalizability of this study's results.

Deletion of the 13 multivariate outliers yielded a final sample size of 252, for which descriptive statistics are reported in Table 72. The correlation and covariance matrices for the models based on MacIntyre and Charos (1996) appear in Appendix AC, while the matrices for the Yashima models are in Appendix AD.

Linearity. To investigate linearity, bivariate scatterplots were examined. Examining all 231 possible permutations of the 22 variables was an impractical task, but several potentially problematic combinations were examined (Tabachnick & Fidell, p. 79). In all cases, plots were not indicative of collinearity or curvilinear relationships. In particular, the distancing and extroversion permutations were checked, but no special problems were evident.

Table 72
Descriptive Statistics of Screened Variables

Variable	Mean	*SD*	Min	Max	Skew	Kurt
Listening Proficiency	49.82	4.38	34.65	64.44	-.11	1.18
Vocabulary	45.83	4.09	33.85	55.81	-.38	.30
Distance	51.22	8.29	24.84	75.19	-.37	.63
L2 Comm Anxiety	49.58	9.16	24.84	55.81	.18	1.66
FLCAS	51.91	4.56	37.46	65.58	-.21	.98
L2 WTC	49.07	8.70	23.57	76.74	-.20	2.00
Perceived L2 Comp	47.89	10.56	20.37	80.12	-.74	1.09
Frequency L2 Comm	53.66	8.72	32.05	74.27	.62	.13
Cultural Friendship	54.97	9.04	25.78	75.31	.42	.42
Approach-Avoid	50.55	6.57	27.15	73.30	.73	2.24
Motivation	54.06	8.59	23.21	77.16	.23	1.64
Need for Order	50.90	3.77	31.98	58.89	-.72	2.11
T-M Competence	49.90	4.63	32.72	63.17	-.50	.89
Unusual Experiences	46.08	7.29	28.34	65.75	-.56	.07
Childlikeness	51.33	5.58	30.54	69.71	.82	1.38
Sensitiveness	52.57	21.08	11.07	89.95	.01	-.72
Openness to Exper	52.37	3.88	37.34	62.12	-.31	.72
Extroversion	50.90	4.07	34.59	61.92	-.03	.80
Diligence	47.86	4.68	31.80	60.23	-.54	1.00
Emotional Stability	52.68	3.85	40.52	65.55	-.05	.73
Agreeableness	50.35	6.50	31.59	69.95	.20	.56
Attitude	51.16	7.10	31.92	67.74	.26	.51
SE					.15	.31

Note. $N = 252$; skew = skewness; kurt = kurtosis; T-M = time-money; Exper = experience. All statistics are based on Rasch CHIPS measures.

Homoscedasticity. Homoscedasticity was also examined with scatterplots. In a bivariate distribution, scedasticity refers to the extent that the variance in one variable is the same at all values of the second variable. Homoscedasticity refers to variance that is the same, while heteroscedasticity denotes variance that is not the same. Violations of homoscedasticity are investigated by examining scatterplots; an oval shape is indicative of homoscedasticity, whereas a shape like a rounded triangle is indicative of skewness in one of the variables and thus of heteroscedasticity. The scatterplots examined exhibited no particular indication of heteroscedasticity.

Multicollinearity and singularity. Multicollinearity refers to an excessively high correlation between two variables, a situation which makes matrix inversion unstable due to excessively small determinants. Multicollinearity was investigated by examining the correlation matrix of the 22 variables. Correlations ranged from -.584 to .653, which was less than the .90 criterion indicative of multicollinearity (Tabachnick & Fidell, 2001, p. 83). Although examining bivariate scatterplots is also prudent, with 22 variables and 231 possible permutations that task becomes impractical. However, several potentially problematic combinations were examined (Tabachnick & Fidell, p. 79), and in particular, the distancing and extroversion permutation was carefully scrutinized. In the cases examined, scatterplots were not indicative of any particular problems.

Singularity refers to a situation in which variables are redundant, which prohibits matrix inversion. Although an assumption of SEM, the lack of singularity is confirmed post ipso facto. In short, if the model converges when the SEM analysis is conducted, then no singularity was present. In the present study, one measurement model (the L2 Communicative Confidence plus Ego Permeability model) did not converge when analyzed with the Rasch-corrected raw data, yet a careful examination of the variables indicated no excessively high correlations. This problem was addressed by parceling the data into 12 parcels (Hau & Marsh, 2004), which yielded a model that converged satisfactorily.

Residuals. Residuals should be small and symmetrically distributed around the mean. This is addressed by examining the distribution of residuals of covariances, an example of which is shown in Figure 36. The residuals are symmetric around the zero midpoint with 93.33% falling in the ±.1 range, which indicates that the model in this case was reasonably well specified (Byrne, 2006, p. 174).

```
   -----------------------------------------------------
   !                                                   !
 40-                                                   -
   !                                                   !
   !                                                   !
   !                                                   !
   !                                                   !           RANGE       FREQ PERCENT
 30-                                                   -
   !                                                   !      1   -0.5  -  --       0     .00%
   !                                                   !      2   -0.4  -  -0.5     0     .00%
   !                       *                           !      3   -0.3  -  -0.4     0     .00%
   !                       *                           !      4   -0.2  -  -0.3     0     .00%
 20-                       *   *                       -      5   -0.1  -  -0.2     1    2.22%
   !                       *   *                       !      6    0.0  -  -0.1    23   51.11%
   !                       *   *                       !      7    0.1  -   0.0    19   42.22%
   !                       *   *                       !      8    0.2  -   0.1     2    4.44%
   !                       *   *                       !      9    0.3  -   0.2     0     .00%
 10-                       *   *                       -      A    0.4  -   0.3     0     .00%
   !                       *   *                       !      B    0.5  -   0.4     0     .00%
   !                       *   *                       !      C     ++  -   0.5     0     .00%
   !                       *   *                       !      -------------------------------
   !                   *   *   *   *                   !           TOTAL           45 100.00%
   -----------------------------------------------------
       1   2   3   4   5   6   7   8   9   A   B   C     EACH "*" REPRESENTS    2 RESIDUALS
```
Figure 36. Distribution of standardized residuals for the Intercultural Friendship Orientation variable.

Measurement Models

As Byrne (2006) noted, an important first step in the analysis of full latent variable models is to test the validity of the measurement model(s). Three measurement models were treated in the previous chapter in the discussions of dimensionality of the respective scales; those models included Motivation, International Posture, and Ego Permeability. Posited to consist of two, four, and five subscales, respectively, they were instead found to be best represented as one, two, and two subscales. Furthermore, the two subscales in the Ego Permeability scale constituted a construct more akin to and thus labeled Imposition of Order. In the following section I treat the L2 Communicative Confidence instrument that appeared in the Yashima (2002) and the Yashima et al. (204) model.

L2 Communicative Confidence Baseline Model

Based on the work of Clement and Kruidenier (1997), the original configuration of L2 Communicative Confidence consisted of two subscales, Perceived Competence in English and L2 Communicative Anxiety. This configuration was initially evaluated twice, first using the L2 Communicative Anxiety data and then with the FLCAS data. Next, pursuant to Yashima's supposition that non-linguistic factors such as gender, personality, and L1 communication tendency (2002, p. 62) might also influence L2 communicative confidence, three personality variables were posited to enhance the construct: Perceived Distance, Ego Permeability, and Extroversion. The three posited additions were then added one by one and the respective 3-factor L2 Communicative Confidence measurement models were evaluated with confirmatory factor analyses using EQS. The three models were L2 Communicative Confidence plus Distancing, L2 Communicative Confidence plus Extroversion, and L2 Communicative Confidence plus Ego Permeability (Imposition of Order).

Inasmuch as the data set included both the FLCAS data and the L2 Communicative Anxiety data, the original 2-factor measurement model was evaluated twice. As shown in Table 73, the FLCAS-data model fit the data better than the L2 Communicative Anxiety-data model. For the FLCAS model, the χ^2/df ratio was just 1.842, while for the L2 Communicative Anxiety model the χ^2/df ratio was much higher at 4.258. Although CFI and IFI were indicative of poor fit for both models, the RMSEA for the FLCAS model was adequate at .058 while for L2 Communicative Anxiety it was poor at .115.

Table 73
Summary of Fit Indices for L2 Communicative Confidence 2-Factor Model (FLCAS and L2 Communicative Anxiety Data)

	FLCAS	L2 Comm
Reliability Coefficient (rho)	.910	.881
Multivariate Kurtosis		
Mardia's coefficient	165.224	94.794
Normalized estimate	20.044	21.171
Residuals		
Average absolute standardized residuals	.057	.059
Average off-diagonal absolute standardized residuals	.060	.066
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 990, 276)	4162.737	5540.794
Satorra-Bentler scaled χ^2 (df = 942, 247)	1734.789	1051.758
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	1.842	4.258
Fit Indices		
Comparative fit index (CFI)	.846	.842
Incremental fit index (IFI)	.848	.846
Root mean-square error of approximation (RMSEA)	.058	.115
RMSEA 90% confidence interval	.054-.062	.124-.138

Note. For the L2 Communicative Anxiety model, four error covariances were added, but only two error covariances were added for the FLCAS model.

Here we find an interesting anomaly as CFI and IFI values indicate poor fit of the proposed model although the RMSEA value indicates adequate fit. These apparently contradictory results deserve explication. Comparing CFI and RMSEA, Rigdon (1996) showed that CFI and other incremental fit indices are less stable across different estimation methods because a null model is involved in the calculation of the indices. On the other hand, RMSEA is robust against changes in sample size, especially when the sample size is large. Rigdon thus suggested that "CFI [is] better suit[ed] to more exploratory, small sample cases, and RMSEA [is] better suited to more confirmatory, large sample situations" (p. 376). Because the focus in the present study is confirmatory and the sample size is not small (N = 252), RMSEA is considered more appropriate than CFI for evaluating the model fit in this case in which CFI and IFI values differ markedly from the RMSEA value.

Based on the finding that the model fit the data much better with the FLCAS data than with the L2 Communicative Anxiety data, the model of L2 Communicative Confidence (FLCAS data) was treated as the baseline L2 Communicative Confidence model (hereafter Baseline Model). To the Baseline Model, the three personality variables (Perceived Distance, Ego Permeability, and Extroversion) were added individually and the respective 3-factor models were evaluated via confirmatory factor analysis using EQS. The three latent variables in the first model tested included Perceived L2 Competence, L2 Communicative Anxiety (FLCAS), and Perceived Distance (labeled simply 'Distance); the configuration is shown in Figure 36. In the event that more than one of these variables had improved the model, the 4- or 5-factor model of L2 communicative Confidence would have been analyzed next.

L2 Communicative Confidence with Perceived Distance

In the first model investigated, Perceived Distance (labeled Distance in the figures) was added to the Baseline Model so that L2 Communicative Confidence consisted of L2 Perceived Competence, L2 Communicative Anxiety, and Distance (Figure 37). Results for the 3-factor model yielded fit statistics very similar to the Baseline Model, but the path coefficient for the Perceived Distance–L2 Communicative Confidence path was not significant. Perceived Distance was thus deleted from further analyses.

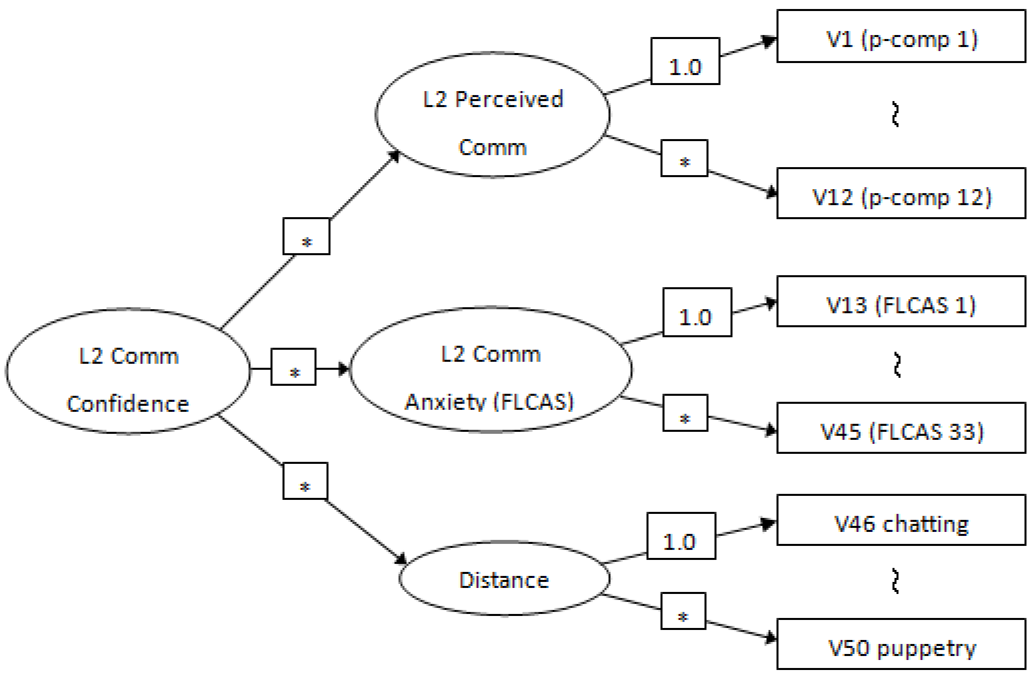


Figure 37. L2 Communicative Communication configuration with the addition of Perceived Distance. The three factors consisted of 12, 30, and 5 items, respectively, but only the first and last items are shown. Disturbances and error terms are not shown for the sake of clarity.

L2 Communicative Confidence with Extroversion

In the second model investigated, Extroversion was added to the Baseline Model so that L2 Communicative Confidence consisted of Perceived Competence, L2 Communicative Anxiety, and Extroversion (Figure 38).

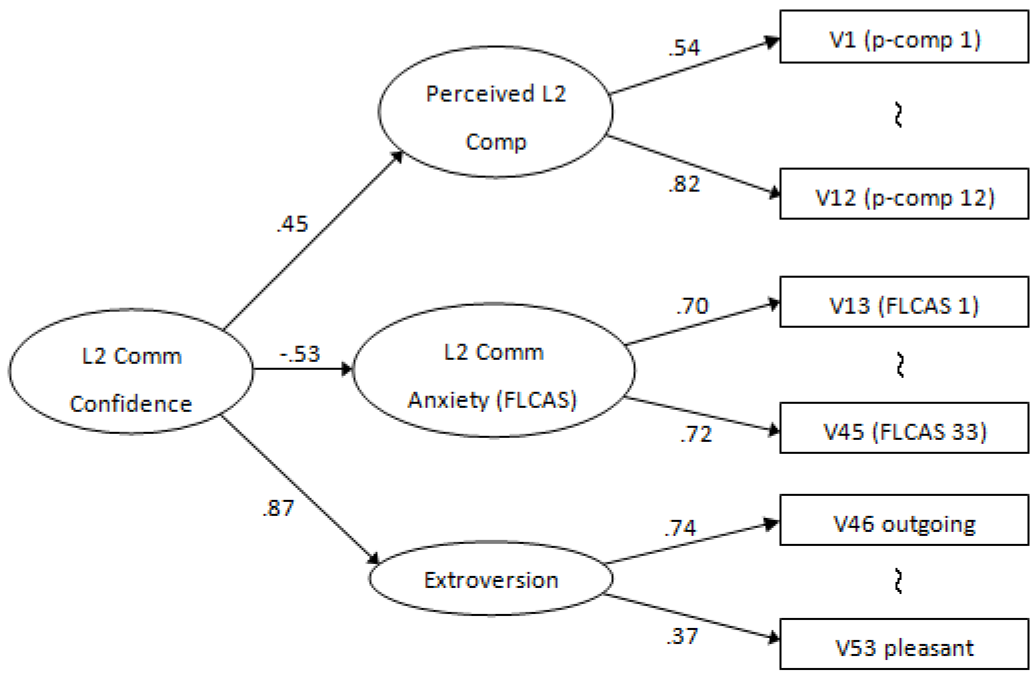


Figure 38. Standardized solution of the L2 Communicative Communication configuration with the addition of Extroversion. The three factors consisted of 12, 30, and 8 items, respectively, but only the first and last items are shown. Disturbances and error terms are also not shown for the sake of clarity. Numerical values indicate that path coefficients were significant at $p < .01$. Satorra-Bentler scaled χ^2 (1171) = 1920.356 ($p < .01$), CFI = .867, RMSEA = .051, C.I. = .046-.055.

The results for the 3-factor model indicated better fit than for the Baseline Model, which indicates that Extroversion is a significant addition. Again, although CFI and IFI were suggestive of poor fit, RMSEA values were indicative of adequate fit. Of particular note is the strength of the path regression coefficient (.87) from L2 Communicative Confidence to Extroversion, which is considerably larger than for either Perceived L2 Competence (.45) or L2 Communicative Anxiety (-.53). This offers further support that Extroversion is a prudent addition to the construct. Detailed statistics for the Baseline Model and the Baseline Plus

Extroversion Model are shown in Table 74, and the complete solution appears in Appendix AE.

Table 74
Summary of Fit Indices for the Baseline L2 Communicative Confidence Model and the Baseline + Extroversion Model

	Baseline Model	Plus Extroversion Model
Reliability Coefficient (rho)	.894	.868
Multivariate Kurtosis		
Mardia's coefficient	198.791	192.523
Normalized estimate	21.031	21.149
Residuals		
Average absolute standardized	.065	.063
Average off-diagonal absolute standardized	.068	.065
Model χ^2		
Model estimation method	ML, Robust	ML, Robust
Independence model χ^2 (*df* = 944, 1225)	7764.889	6849.705
χ^2 (*df* = 941, 1171)	2272.941	1920.356
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	2.415	1.640
Fit Indices		
Comparative fit index (CFI)	.851	.867
Incremental fit index (IFI)	.852	.868
Root mean-square error of approx (RMSEA)	.053	.051
RMSEA 90% confidence interval	.049-.056	.046-.055

Note. Because of the large degree of kurtosis (standardized Mardia's coefficient = 21.149), robust ML estimation was stipulated.

L2 Communicative Confidence with Ego Permeability (Imposition of Order)

In the third model investigated, Imposition of Order (Ego Permeability) was added to the Baseline Model so that L2 Communicative Confidence consisted of three factors: Perceived Competence, L2 Communicative Anxiety (FLCAS), and Imposition of Order. In order to render the model as a second-order model, the Need for Order and Perceived Time-Money Competence subscales were treated as measured variables while Perceived L2 Competence and the FLCAS remained as latent variables with 12 and 30 items, respectively. However, the model did not converge, thus indicating a problem with singularity or multicollinearity. To address this, all three subscales were divided into parcels (Hau & Marsh, 2004).

The Perceived L2 Competence subscale was split into three 4-items parcels, and the FLCAS was divided into five parcels reflecting the original theoretical composition (Horwitz et al., 1986, pp. 127-128) and the researcher’s intuition: test anxiety, fear of negative evaluation, comprehension apprehension, (lack of) preparation, and affective reactions. The Need for Order and Perceived Time–Money Competence subscales were resolved into two parcels respectively with odd-even splits (Figure 39). The results for the 3-factor model indicated adequate fit of the model to the data, but the L2 Communicative Confidence–Imposition of Order (Ego Permeability) path was not statistically significant. Thus, the Imposition of Order (Ego Permeability) instrument was deleted.

Based on the results of the measurement models, L2 Communicative Confidence was best represented by a 3-factor model consisting of Perceived L2 Competence, L2 Communicative Anxiety, and Extroversion. This configuration was used in subsequent analyses.

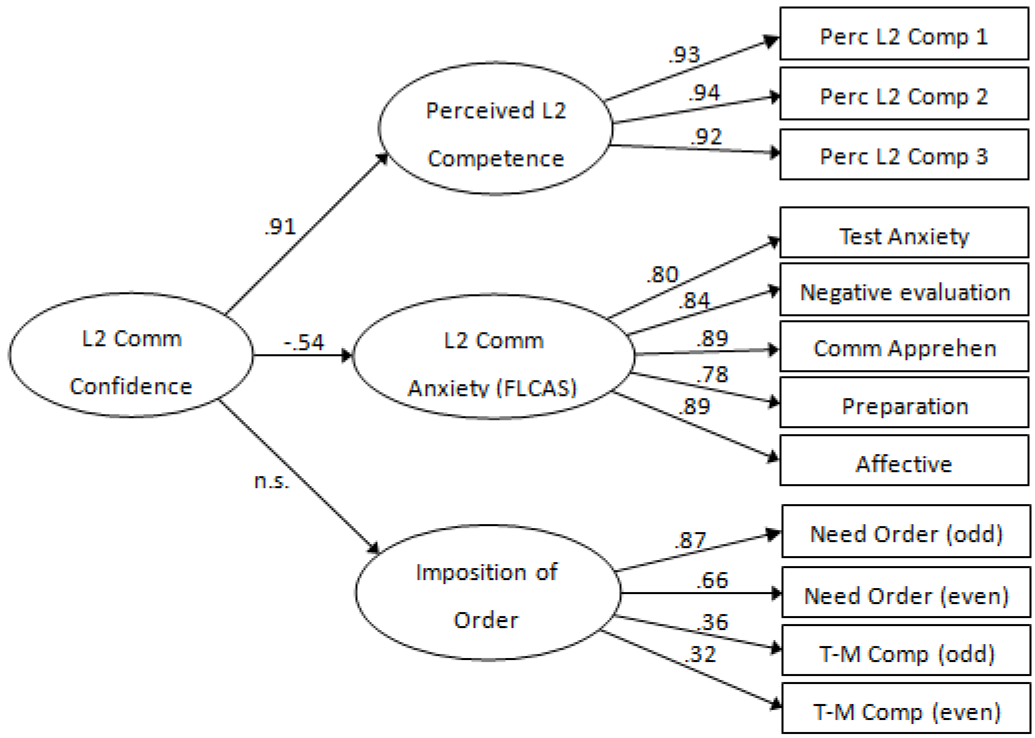


Figure 39. Standardized solution for the L2 Communicative Confidence configuration with the addition of Imposition of Order (Ego Permeability). Disturbances and error terms are not shown for the sake of clarity. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2 = 112.980$, $p < .01$, CFI = .954, RMSEA = .069, 90% C.I. = .051-.086.

Path Analysis of Models Based on MacIntyre and Charos (1996)

The first two research questions dealt with the assessment of (a) the replication of the two original L2 communication models, and of (b) the hypothesized modifications of the two models. The first research question dealt with replication of two earlier WTC models: "Will the WTC models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) be replicated in this context?" The second research question concerned modifications of those same models: "Do the above L2 communication models benefit from the addition of personality variables such as distancing, ego perm, and introversion-extroversion?"

The following section presents the results of the path analysis assessments of models based on the MacIntyre and Charos (1996) model.

Original MacIntyre and Charos (1996) Model

Based on theoretical considerations and analyses of the scales, the original MacIntyre and Charos (1996) model was slightly modified (Figure 40). The far left level includes the five subscales from the Big 5 Personality instrument: Openness to Experience, Extroversion, Agreeableness, Emotional Stability, and Diligence. The original model included Context, which has been replaced with English Experience in the present study. This was done under the assumption that what role context plays is essentially the same for all participants given the homogeneity of English education and the relative dearth of opportunities to speak English in Japan. The new English Experience variable includes a series of events that could supplement the amount of exposure to English. The list includes such activities as having lived abroad, studied abroad, traveled abroad, and attended an English conversation in Japan.

The second level includes Perceived Competence, L2 Anxiety, Integrativeness (the Cultural Friendship Orientation subscale), and Attitudes, which in turn underpin L2 WTC and Motivation. The model culminates in L2 Communication Frequency, which is posited to be defined by paths from English Experience, Perceived L2 Competence, L2 WTC, and Motivation.

The initial path analysis yielded χ^2 (49, N = 252) = 175.911, p < .000, with a total of six non-significant paths; this result was significantly better than the result

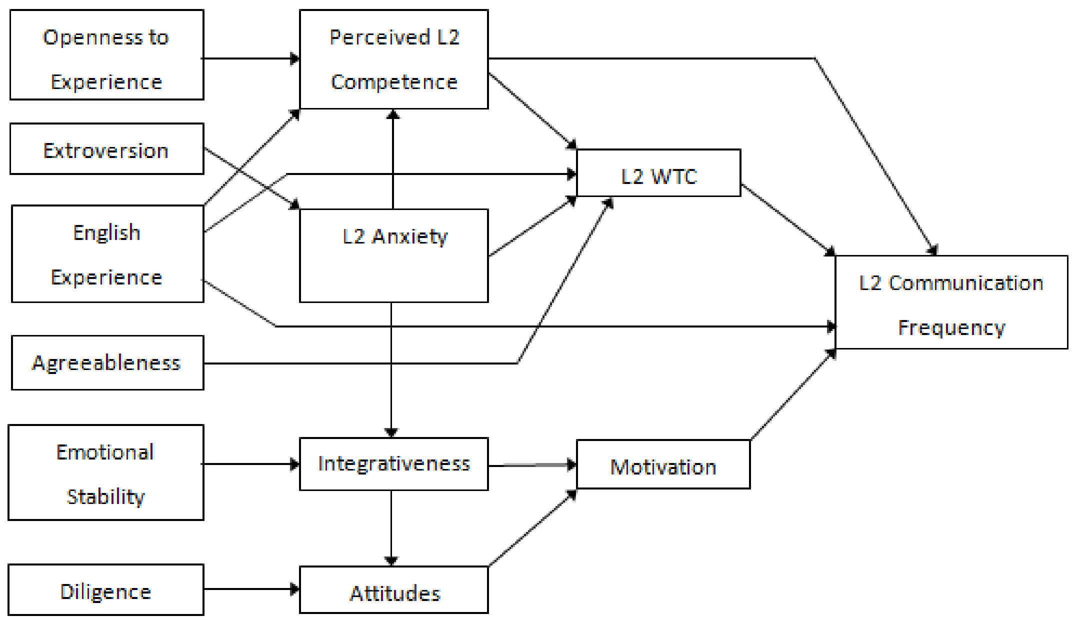


Figure 40. Revised MacIntyre and Charos (1996) model of L2 Willingness to Communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre, and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 12. Copyright 1996 by *Journal of Language and Social Psychology.* Reprinted with permission.

for the independence model, χ^2 (78, N = 252) = 843.700, p < .000. Skewness was adequately small for all the variables; some degree of kurtosis was present (Mardia's standardized coefficient = 17.967), and two cases with large contributions to kurtosis were deleted, but with the large degree of kurtosis, robust maximum likelihood estimation was requested. The standardized residuals reflected a substantial degree of non-normality with just 56.04% in the ±.1 interval. As noted, the χ^2 value was significant and fit indices showed poor fit of the model to the data with CFI = .814, IFI = .823, RMSEA = .089, and 90% C.I. = .073-.105.

Because of the poor fit statistics, the model was modified based on the Lagrange multiplier test results, with logical paths added one at a time. First, a path was added from Emotional Stability to Motivation; lack of emotional stability could imperil motivation if, for example, strongly emotional reactions undermine attention to and enthusiasm for the learning task at hand. Second, a direct path was added from Extroversion to L2 Communicative Frequency. This is a prudent addition, for a good attitude logically leads to more participation and thus greater frequency of communication. For both steps the change in χ^2 was statistically significant. Finally, the five non-significant paths were deleted: Emotional Stability to Integrativeness, Agreeableness to L2 WTC, L2 anxiety to L2 WTC, English experience to WTC, and Perceived L2 Competence to Frequency (Table 75). The reader should note that with the deletion of the Agreeableness to L2 WTC path, the Agreeableness subscale no longer plays any role in the model and is therefore absent in Figure 41.

The standardized solution for the final model is shown in Figure 41; the two data-driven additions are indicated as dashed lines (the standardized structural equations, standard errors, and squared multiple correlations [R^2] are shown in Appendix AF). Although the RMSEA of .070 indicates adequate fit, both CFI and IFI are somewhat low (.893 and .900, respectively). The χ^2/df ratio is also greater than 2; values of 2 are suggestive of good fit (Tabachnick & Fidell, 2004, p. 698).

Based on these results, the original MacIntyre and Charos model with the noted modifications was considered to fit the data somewhat poorly. The statistics for the

Table 75

Step-by-Step Procedure for Revising the Original MacIntyre-Charos Model (L2 Communicative Anxiety) with Data-Driven Paths

Model	S-B χ^2	*df*	CFI	IFI	RMSEA
Original model	163.944	45	.790	.801	.103
Add Emotional Stability – Motivation path	139.227	44	.832	.841	.093
Add Extroversion – Frequency path	118.880	43	.866	.874	.084
Delete five non-significant paths	124.661	48	.864	.871	.080

Note. S-B χ^2 = Satorra-Bentler scaled χ^2; CFI = Comparative fit index; RMSEA = root mean square error of approximation.;

Table 76

Summary of Fit Indices for Revised and Respecified MacIntyre-Charos Models (L2 Communicative Anxiety)

	Original Model	Final Model
Reliability Coefficient (Cronbach alpha)	.723	.723
Multivariate Kurtosis		
Mardia's coefficient	43.759	43.759
Normalized estimate	17.967	17.967
Residuals		
Average absolute standardized residuals	.091	.073
Average off-diagonal absolute standardized residuals	.105	.083
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (*df* = 78)	625.527	625.527
Satorra-Bentler scaled χ^2 (*df* = 45, 47)	163.014	105.658
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	3.623	2.248
Fit Indices		
Comparative fit index (CFI)	.784	.891
Incremental fit index (IFI)	.797	.897
Root mean-square error of approximation (RMSEA)	.100	.070
RMSEA 90% confidence interval	.084-.116	.053-.088

revised model and the final, respecified model are shown in Table 76. An unexpected result was that that the path from L2 anxiety to WTC was not significant (although there was an indirect influence with a path weight of -.15 via Perceived L2 Competence). The absence of a direct path is counterintuitive, and

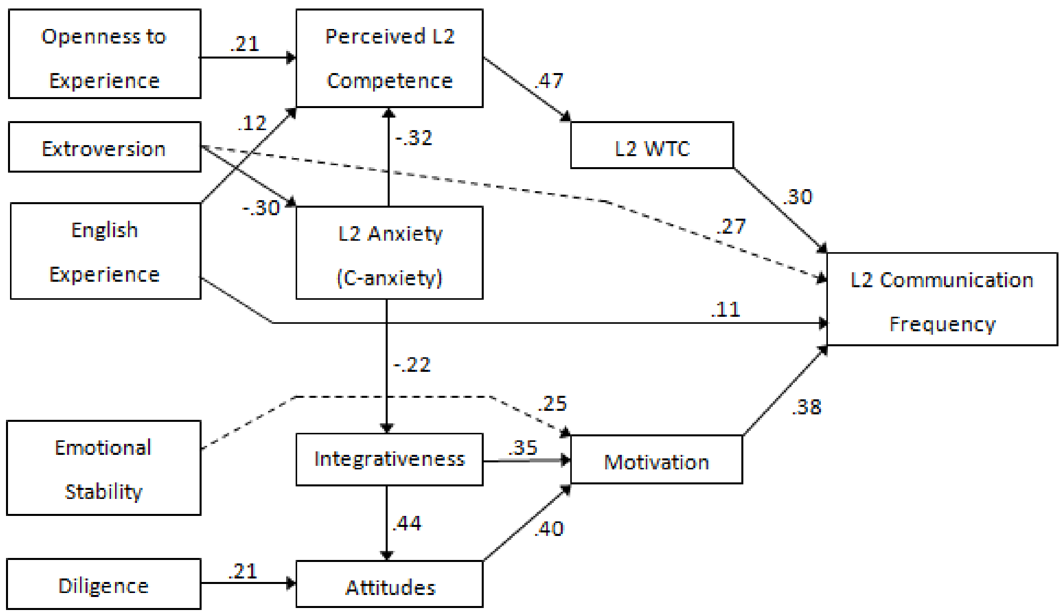


Figure 41. Standardized solution of the revised path-analytic model using communicative anxiety: Personality, attitudes, and affect as predictors of foreign language communication. Adapted from “Personality, Attitudes, and Affect as Predictors of Second Language Communication,” by P. D. MacIntyre and C. Charos, 1996, *Journal of Language and Social Psychology*, *15*(1), p. 12. Copyright 1996 by *Journal of Language and Social Psychology*. Adapted and reprinted with permission. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2 = 76.396$, $p < .01$, CFI = .926, RMSEA = .075, 90% C.I. = .053-.095.

given the satisfactory Rasch analysis results for the L2 Communicative Anxiety instrument, the use of the L2 Communicative Anxiety instrument in this context appears to be questionable. The lack of statistical significance might be due to a mismatch, inasmuch as the participants’ L2 communication is primarily in classroom contexts and the L2 Communicative Anxiety instrument deals mostly with contexts beyond the classroom.

Original MacIntyre and Charos (1996) Model with FLCAS Data

The original model MacIntyre and Charos (1996) model was next reanalyzed with FLCAS data instead of the L2 Communicative Anxiety data. As noted above, the use of the FLCAS to measure L2 communicative anxiety might be more appropriate in this EFL context because opportunities to interact in English are limited outside of school; indeed, the largest number of opportunities is probably in the compulsory English classrooms in junior high and senior high school. Initial results indicated that skewness was again not problematic while kurtosis was excessive. Moreover, the model fit the data very poorly, $\chi^2 = 197.540$, $p < .01$, CFI = .807, RMSEA = .106, 90% C.I. = .091-.121, suggesting that the model was poorly specified for this sample and context (Table 77).

Given the poor fit statistics, the model was modified based on the Lagrange multiplier and Wald test results, with logical paths added one at a time and non-significant paths then deleted en masse. First, a path was added from Attitudes to L2 Anxiety, which markedly improved the fit statistics (Table 78). This was a negative coefficient, as it makes sense that a positive attitude would correspond to less L2 anxiety. Second, a direct path was added from Extroversion to Frequency, which also makes sense as a more extroverted person should communicate more

Table 77
Summary of Fit Indices for Original and Revised MacIntyre-Charos Models (FLCAS)

	Original Model	Revised Model
Reliability Coefficient (Cronbach alpha)	.745	.743
Multivariate Kurtosis		
Mardia's coefficient	24.993	27.268
Normalized estimate	10.005	10.960
Residuals		
Average absolute standardized residuals	.104	.081
Average off-diagonal absolute standardized residuals	.120	.092
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 78)	754.547	751.837
Satorra-Bentler scaled χ^2 (df = 45, 47)	245.445	109.457
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	5.454	2.329
Fit Indices		
Comparative fit index (CFI)	.704	.907
Incremental fit index (IFI)	.718	.911
Root mean-square error of approximation (RMSEA)	.134	.072
RMSEA 90% confidence interval	.117-.150	.054-.089

frequently. Next, a path was added from Emotional Stability to L2 Anxiety. MacIntyre and Charos (1996, p. 19) noted that emotional stability is not strongly related to general trait anxiety, but the addition of a path is both logical (greater stability likely corresponds with less anxiety) and suggestive that L2 anxiety in this context might better be viewed as a trait rather than a state. Finally, a path was added from English Experience to L2 Anxiety. This is yet another logical alteration because increased exposure to English and therefore greater familiarity with the language should lead to lower levels of L2 anxiety.

Finally, the six non-significant paths were deleted: English Experience to Perceived Competence, L2 Anxiety to Integrativeness, Emotional Stability to

Integrativeness, Agreeableness to L2 WTC, English Experience to L2 WTC, and Perceived Competence to L2 Communication Frequency. As with the previous model, deleting the Agreeableness to L2 WTC path removed Agreeableness from the model. The sequence of steps undertaken in revising the model is shown in Table 78.

This model is the more logical of the two because of the direct effect of anxiety on WTC (Figure 42). The standardized solution for the final model is shown in Figure 41, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) are shown in Appendix AG. Agreeableness was

Table 78
Step-by-Step Procedure for Revising the Original MacIntyre-Charos Model (FLCAS) with Data-Driven Paths

Model	S-B χ^2	*df*	CFI	IFI	RMSEA
Original model	245.445	45	.704	.718	.134
Add *Attitudes – L2 Anxiety* path	150.099	44	.843	.854	.098
Add *Extroversion – Frequency* path	132.257	43	.868	.875	.091
Add *Emotional Stability – L2 Anxiety* path	117.548	42	.888	.894	.085
Add *English Experience – L2 Anxiety* path	99.899	41	.913	.917	.076
Delete six non-significant paths	105.006	47	.914	.918	.070

Note. S-B χ^2 = Satorra-Bentler scaled χ^2; CFI = Comparative fit index; RMSEA = root mean square error of approximation.

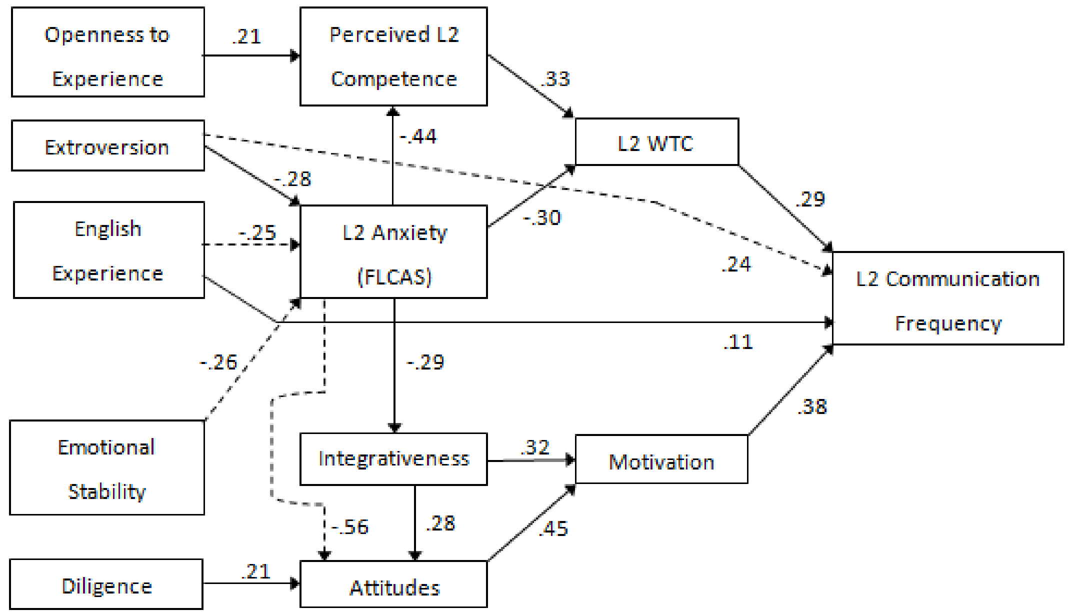


Figure 42. Revised path-analytic model using FLCAS: Personality, attitudes, and affect as predictors of foreign language communication. Data-driven additions to the model are shown as dashed lines. Adapted and reprinted with permission. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2(47) = 105.006$ ($p < .01$), CFI = .914, RMSEA = .070, 90% C.I. = .054-.089.

again deleted. The RMSEA of .072 indicates adequate fit, and both CFI and IFI (.907 and .911, respectively) are closer to reasonable fit than in the above model that used L2 Communication Anxiety rather than FLCAS data. The χ^2/df ratio is also just slightly greater than 2, which is suggestive of good fit. Moreover, this model includes the logical path from L2 anxiety to WTC.

The results for the two models are shown in Table 79, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) are shown in Appendix Y. As indicated, the modified MacIntyre and Charos model had better fit when anxiety was operationalized using the FLCAS instead of the L2

Table 79

Summary of Fit Indices for Revised MacIntyre-Charos Model (Communicative Anxiety) and Revised MacIntyre-Charos Model (FLCAS)

	L2 Comm Anxiety Model	FLCAS Model
Reliability Coefficient (Cronbach alpha)	.723	.743
Multivariate Kurtosis		
Mardia's coefficient	43.759	34.479
Normalized estimate	17.967	14.071
Residuals		
Average absolute standardized residuals	.091	.081
Average off-diagonal absolute standardized residuals	.105	.092
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 78)	625.527	751.837
Satorra-Bentler scaled χ^2 (df = 45, 47)	105.658	109.457
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	2.348	2.329
Fit Indices		
Comparative fit index (CFI)	.891	.907
Incremental fit index (IFI)	.897	.911
Root mean-square error of approximation (RMSEA)	.070	.072
RMSEA 90% confidence interval	.053-.088	.054-.089

Communicative Anxiety instrument. This result is not entirely unexpected, for students in Japan have far more interaction in English in foreign language classrooms than in the situations included in the L2 Communicative Anxiety instrument.

Extended MacIntyre and Charos (1996) Model

The extended MacIntyre and Charos (1996) model using the FLCAS data was used as the base model for the next step, in which Ego Permeability was added as a first-order variable and Distance was added as a second-order variable. Ego Permeability was hypothesized to positively affect Distance, as greater ego flexibility likely predicts greater ability to perceive distance. Distancing was posited to (a) negatively affect L2 Communicative Anxiety, as greater distance might act as a safe haven, and (b) positively affect L2 WTC because the ability to perceive distance from one's core self should provide greater freedom to communicate. In Figure 43, the hypothesized variables and paths are shown in bold.

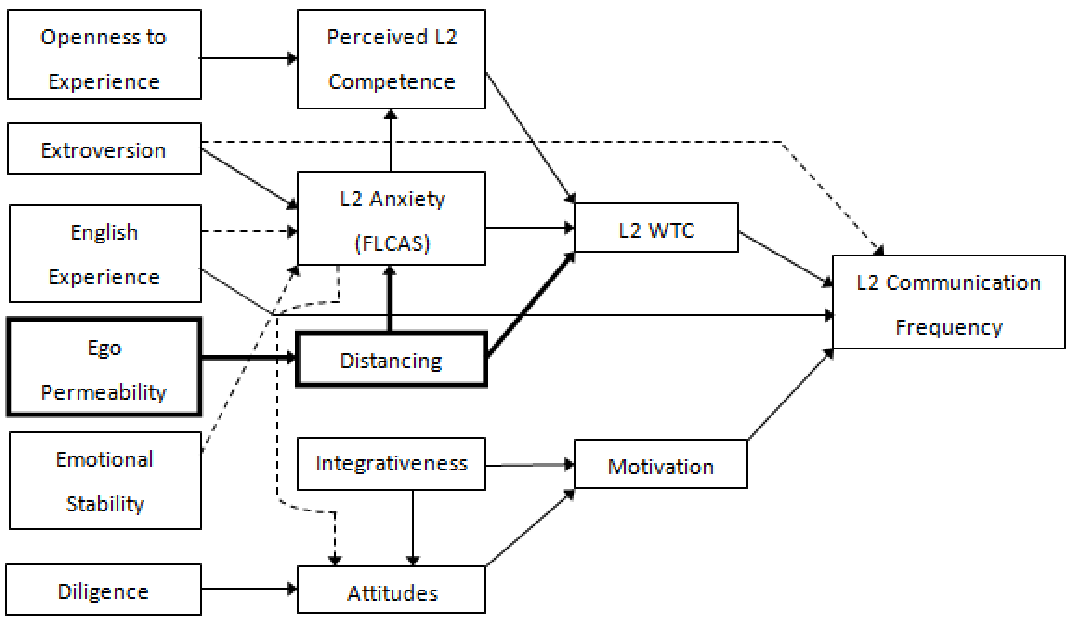


Figure 43. Hypothesized model of L2 communication with ego permeability and distancing added. Dashed lines represent data-driven additions. Bold lines show the hypothesized additions. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre, and C. Charos, 1996, *Journal of Language and Social Psychology*, *15*(1), p. 12. Copyright 1996 by *Journal of Language and Social Psychology*. Reprinted with permission.

The overall fit of the hypothesized model to the data was marginally acceptable with $\chi^2(71) = 228.307$ ($p < .01$), CFI = .846, RMSEA = .094, and 90% C.I. = .080-.107. However, the path analysis result for this model showed that all hypothesized paths associated with Ego Permeability and Distance were not significant, thus indicating that the baseline model did not benefit from the posited additions of Distance and Ego Permeability.

SEM Assessment of Models Based on Yashima (2002)

The following section presents structural equation modeling assessments of several models based on the model of Yashima (2002). These models include the original with minor modifications of the underlying variables and a revised model that includes Extroversion.

Original Yashima (2002) Model

The core model of L2 communication shown in Figures 43 (transposed 180 degrees about the Y-axis from the original) illustrates the importance of International Posture. In this conceptualization, International Posture directly influences Willingness to Communicate in the L2 and L2 Learning Motivation. L2 Learning Motivation in turn influences L2 Communicative Confidence with Proficiency playing an indeterminate, mediating role (the role of proficiency in the model was not specified in the original study). L2 Communicative Confidence directly influences L2 WTC (Figure 44).

Based on analyses in this study, three substantial modifications of the original model were undertaken. First, the International Posture factor was modified: composed of the original four subscales of the International Posture instrument in Yashima (2002), the Interest in Foreign Affairs subscale and the

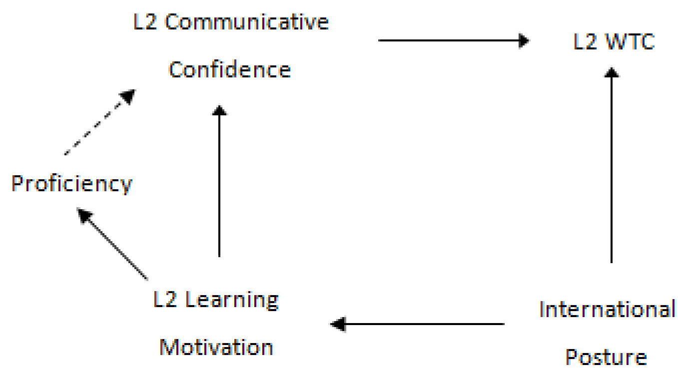


Figure 44. Core of the Yashima (2002) L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal*, *86*(1), 61. Copyright 2002 by *The Modern Language Journal*. Reprinted by permission. Note that the dashed path was hypothesized but found to be non-significant.

Interest in International Vocation/Activities were deleted and the Intercultural Friendship Orientation subscale was added based on the results of a confirmatory factor analysis. Thus, in the modified model International Posture consisted of Approach-Avoidance Tendency and Intercultural Friendship Orientation. Second, the 2-factor Motivation instrument was demonstrated to consist of a single dimension, so it entered the model as a measured variable instead of a latent variable. Third, L2 WTC was rendered as a measured variable rather than latent variables; in the original study L2 WTC was divided into two parcels that were used as indicators.

As shown in Figure 45, many of the path coefficients in the original configuration were similar (the lefthand value is from the current study, and the righthand parenthetical value is from Yashima, 2002). Two, however, differed in terms of statistical significance. In the original study, the path from Proficiency to

L2 Communicative Confidence was not significant at .14, but in the current study it was significant with a beta-weight of .34. This is a logical change, for increased proficiency generally corresponds with higher confidence levels. On the other hand, in the original study the path from International Posture to L2 WTC was significant albeit weak at .22, yet in the current study it was not significant at .06. This is an odd finding, for in the presence of a higher degree of international posture, in

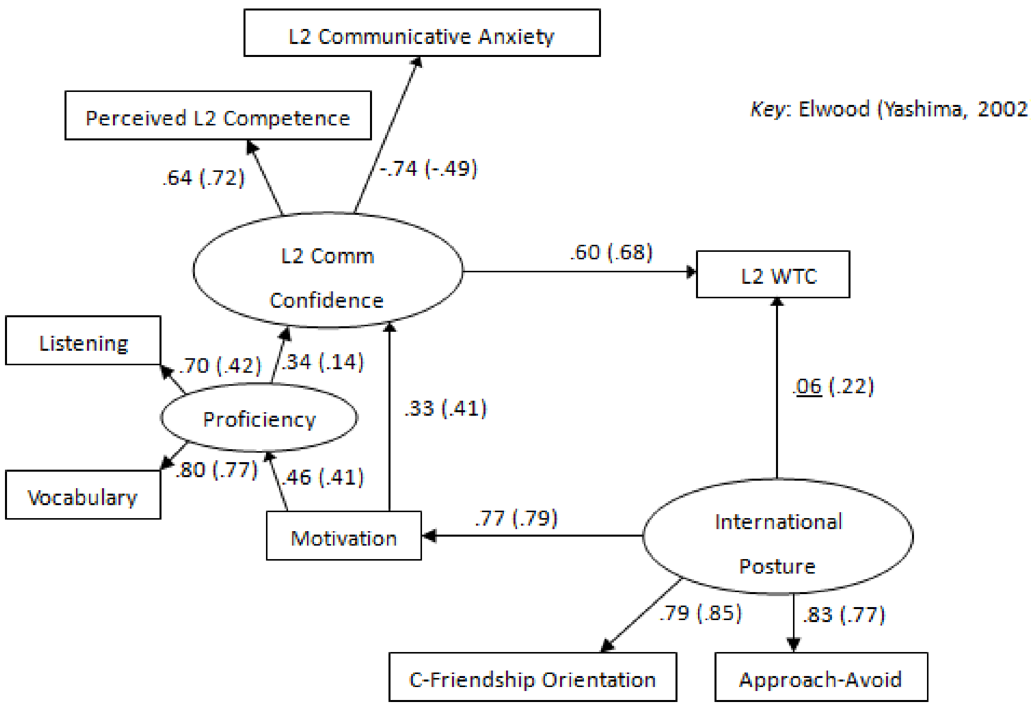


Figure 45. Standardized solution of the original Yashima (2002) model of L2 communication with standardized estimates. Numerical values list the value from the current first and the value from Yashima (2002) parenthetically. Path coefficients were significant at $p < .01$ with the exception of the underlined value (.06) for the path from International Posture to L2 WTC. χ^2 (16) = 43.941, $p < .01$, CFI = .957, RMSEA = .084, 90% C.I. = .055-.114.

which “learners are more interested in or have more favorable attitudes toward what English symbolizes” (Yashima, 2002, p. 57), such learners should have a greater willingness to engage in communicative acts, but with this particular data set and this model, that was not the case. A further consideration is that the two variables might represent somewhat of a mismatch: L2 WTC deals with very discrete situations, whereas the latent International Posture variable could represent more of an abstract ideal.

Although the original model had good fit, the Lagrange multiplier test suggested adding a path from International Posture to L2 Anxiety. The addition of this path resulted in a statistically significant decrease in χ^2 of 16.182, and the path had a value of -.31 (Table 80). This is a logical addition because a favorable disposition toward things international should correspond with lower anxiety about them. With the addition of this path, the analysis yielded good fit statistics with χ^2 (15) = 27.759 (p = .023), CFI = .980, RMSEA = .058, and 90% C.I. = .021-.092. Shown in Figure 46, these values are similar to those reported in Yashima (2002).

Table 80
Step-by-Step Procedure for Revising the Original Yashima et al. Model with Data-Driven Paths

Model	χ^2	*df*	CFI	IFI	RMSEA
Original model	43.941	16	.957	.958	.084
Add *International Posture – L2 Anxiety* path	27.759	15	.980	.981	.058

Note. CFI = Comparative fit index; RMSEA = root mean square error of approximation.

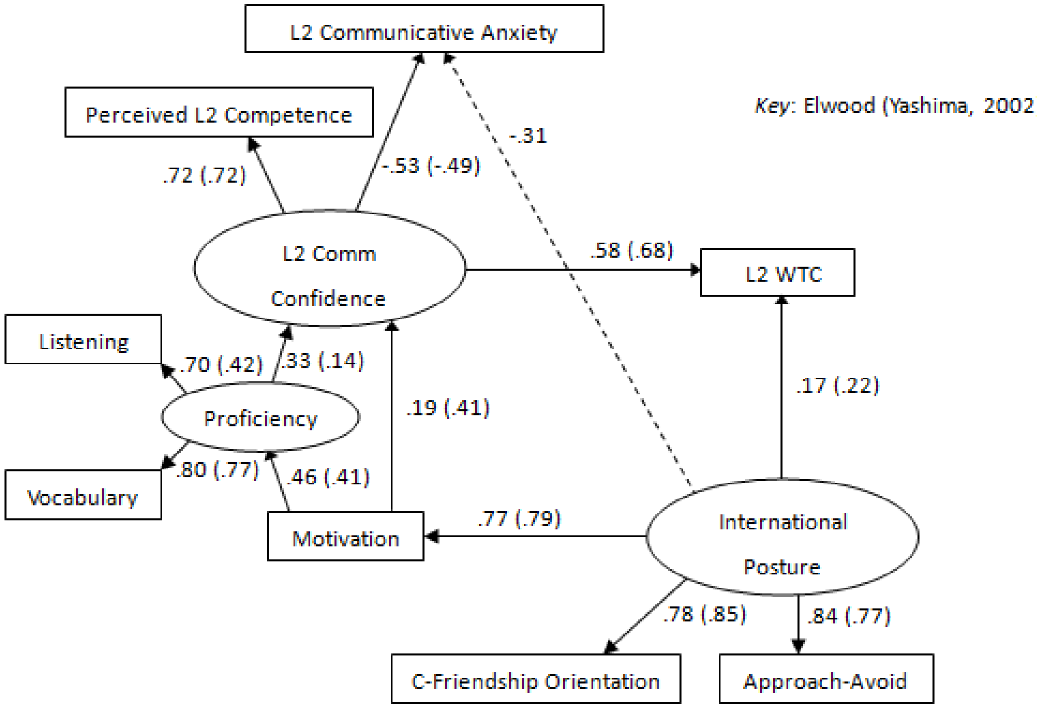


Figure 46. Respecified original model of L2 communication with standardized estimates. Numerical values list the value from the current first and the value from Yashima (2002) parenthetically. Path coefficients were significant at $p < .01$. χ^2 (15) = 27.759, p = .023, CFI = .980, RMSEA = .058, 90% C.I. = .021-.092.

In addition, most of the path coefficients in the current study were similar to those of the original study, differing with two exceptions by .10 at most. The first exception was the Proficiency–L2 Communicative Confidence path, with a value of .33 in the current study compared to .14 in the original study. The stronger coefficient in the current study is appealing because a higher level of proficiency logically correlates with a higher level of confidence. The second difference in path coefficients was that the path from Motivation to L2 Communicative Confidence was just .19 after the addition of the International Posture–L2 Anxiety path, whereas it was a much stronger .41 in the Yashima (2002) study.

The path from International Posture to L2 WTC (.22 in Yashima, 2002) was smaller in the current study (.17) but still significant. Recall that in the initial iteration (Figure 44) this path was *not* significant, yet it became significant when the International Posture–L2 Anxiety path was added. Two explanations are plausible, one of which is that this path is indeed very ‘fragile’. The second possibility is that the weakness of this path could be an anomaly in this data set.

Detailed statistics for both the original model and the modified model are shown in Table 81, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) appear in Appendix AH. In both cases, the model fit the data well, offering support for the robustness of the Yashima (2002) model.

Table 81
Summary of Fit Indices for Original 2002 Yashima Model

	Original Model	Respecified Model
Reliability Coefficient (rho)	.793	.799
Multivariate Kurtosis		
Mardia’s coefficient	11.798	13.445
Normalized estimate	7.374	7.554
Residuals		
Average absolute standardized residuals	.039	.036
Average off-diagonal absolute standardized residuals	.050	.043
Model χ^2		
Model estimation method	ML	ML
Independence model χ^2 (df = 36)	676.764	756.661
χ^2 (df = 16, 22)	43.941	27.759
Probability value for the χ^2 statistic	.000	.023
χ^2/df ratio	2.746	1.851
Fit Indices		
Comparative fit index (CFI)	.957	.980
Incremental fit index (IFI)	.958	.981
Standardized root mean square residual (SRMR)	.064	.050
Root mean-square error of approximation (RMSEA)	.084	.058
RMSEA 90% confidence interval	.055-.114	.021 -.092

Revised Yashima (2002) Model

Next, the original Yashima model was modified based on theoretical considerations and on the results of the Rasch analyses. Extroversion was added as a variable underpinning L2 Communicative Confidence, which then consisted of Perceived L2 Competence, L2 Communicative Anxiety, and Extroversion. Recall that according to the evaluation of the L2 Communicative Confidence measurement model, this 3-factor configuration was found to have the best fit of the various configurations that were evaluated. The hypothesized model is shown in Figure 47, with bold lines and the bold arrow indicating the additions.

This revised model was evaluated with regular ML estimation because of the moderate level of kurtosis (standardized Mardia's coefficient = 7.186). Initial results indicated adequate fit with $\chi^2(23) = 75.907$ ($p < .01$), CFI = .927, RMSEA = .096, 90% C.I. = .072-.120. Paths were similar to the original model results, but the path from International Posture to WTC was again not significant.

In lieu of the moderately good fit statistics, the model was modified based on the Lagrange multiplier and Wald test results, with logical paths added one at a time and non-significant paths then deleted en masse. First, a path was added from Extroversion to International Posture because a more extroverted personality should predict a favorable propensity toward international things. Second, a path was added from International Posture to L2 Anxiety (as was done above in the original model). Detailed in Table 82, this sequence resulted in substantially better fit: $\chi^2(29) = 68.175$ ($p < .01$), CFI = .955, RMSEA = .074, 90% C.I. = .051-.096.

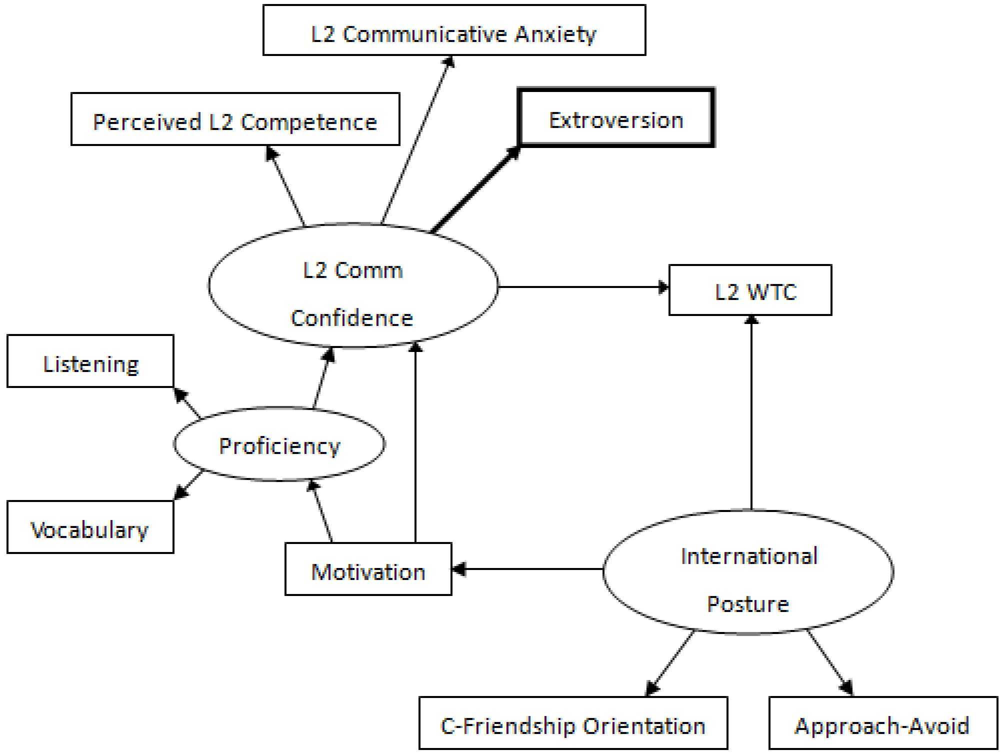


Figure 47. Revised Yashima (2002) L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal*, *86*(1), 61. Copyright 2002 by *The Modern Language Journal*. Reprinted by permission.

The standardized solution is shown in Figure 48. The hypothesized path from L2 Communicative Confidence to Extroversion was statistically significant (β = .36). The two data-driven additions from International Posture to Anxiety and Extroversion were fairly strong at -.33 and .43, respectively. With three exceptions, the original path coefficients are similar to the original Yashima (2002) model (Figure 47), differing by a maximum of .06. In this model, the fragile International Posture–L2 WTC path was again slightly weaker than in the original Yashima (2002) results.

Table 82
Step-by-Step Procedure for Respecifying the Revised Yashima et al. Model with Data-Driven Paths

Model	χ^2	*df*	CFI	IFI	RMSEA
Original model	75.907	23	.927	.928	.096
Add *Extroversion – International Posture*	62.440	22	.944	.945	.086
Add *International Posture–Anxiety* path	44.309	21	.968	.968	.067

Detailed statistics of the initial and final solutions are shown in Table 83, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) are shown in Appendix AJ.

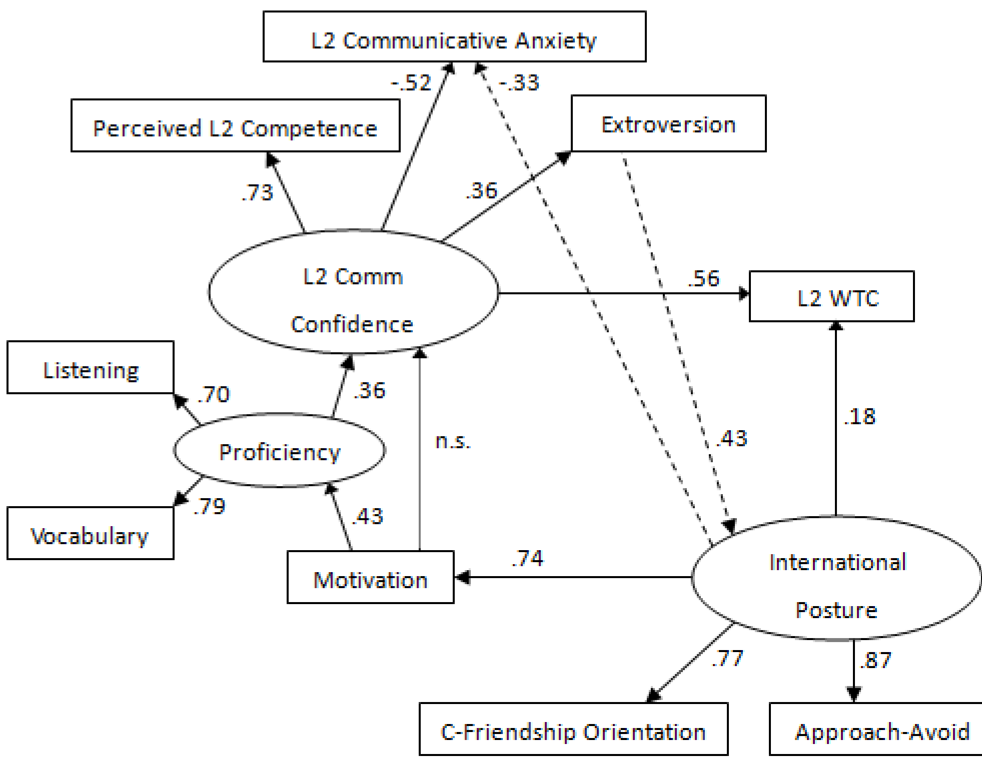


Figure 48. Standardized solution of the revised Yashima (2002) model. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2(21) = 44.31$ ($p < .01$), CFI = .968, RMSEA = .067, 90% C.I. = .039-.094.

Comparison of Original and Revised Yashima (2002) Models

We now arrive at a numerical comparison of the two final models (Table 84). For both models, reliability was adequate, and because of the kurtosis, robust statistics were requested. Residuals were normally distributed around the midpoint. The χ^2 value was lower for the original model, yet with more degrees of freedom,

Table 83
Summary of Fit Indices for the Revised Yashima (2002) Models

	Original Model	Final Model
Reliability Coefficient (Cronbach alpha)	.818	.808
Multivariate Kurtosis		
Mardia's coefficient	13.445	13.445
Normalized estimate	7.554	7.554
Residuals		
Average absolute standardized residuals	.046	.036
Average off-diagonal absolute standardized residuals	.058	.044
Model χ^2		
Model estimation method	ML	ML
Independence model χ^2 (df = 36)	756.661	756.661
χ^2 (df = 23, 21)	75.907	44.309
Probability value for the χ^2 statistic	.000	.002
χ^2/df ratio	3.908	2.110
Fit Indices		
Comparative fit index (CFI)	.927	.968
Incremental fit index (IFI)	.928	.968
Standardized root mean square residual (SRMR)	.078	.053
Root mean-square error of approximation (RMSEA)	.096	.067
RMSEA 90% confidence interval	.072-.120	.039-.096

the χ^2/df ratio was better for the revised model. The fit statistics were better for the revised model. In conclusion, while both models had adequate fit statistics, the revised model had better fit, which indicates that the addition of the latent proficiency variable and the extroversion variable improved the fit of the model to the data. In addition, these results yielded support for the robustness of the basic configuration of the Yashima (2002) model.

Table 84

Summary of Fit Indices for the Original and Revised Yashima 2002 Models

	Original model	Revised model
Reliability Coefficient (rho)	.780	.837
Multivariate Kurtosis		
Mardia's coefficient	16.647	28.470
Normalized estimate	12.048	14.930
Residuals		
Average absolute standardized residuals	.053	.060
Average off-diagonal absolute standardized residuals	.069	.072
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 21, 45)	416.677	655.129
Satorra-Bentler scaled χ^2 (df = 12, 31)	47.857	76.396
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	3.988	2.464
Fit Indices		
Comparative fit index (CFI)	.909	.926
Incremental fit index (IFI)	.911	.927
Standardized root mean square residual (SRMR)	(.084)	(.093)
Root mean-square error of approximation (RMSEA)	.107	.075
RMSEA 90% confidence interval	.076-.139	.053-.095

SEM Assessment of Models Based on Yashima et al. (2004)

The following section presents structural equation modeling assessments of several models based on the Yashima et al. (2004) model. These models include the original with minor modifications of the underlying variables and a revised model that included L2 proficiency and Extroversion.

Original Yashima et al. (2004) Model

The configuration of the original Yashima et al. (2004) model of L2 communication is shown in Figure 48. The original configuration of L2 WTC was defined by two observed variables (i.e., parcels) created from the odd- and even-numbered items, respectively. International Posture was defined by three of the original four subscales: Approach-Avoid Tendency, Interest in International Vocation/Activities, and Interest in Foreign Affairs. Motivation was treated as a latent variable consisting of Motivational Intensity and Desire to Learn English. Based on Clement and Kruidenier (1985), L2 communication confidence was posited to consist of (a lack of) L2 Communicative Anxiety and Perceived L2 Competence. The model culminates with L2 Communication Frequency underpinned by L2 WTC and International Posture (L2 Communication Frequency was absent in the 2002 model). The model was evaluated using EQS, Build 6.0 (Bentler, 2007a).

Based on analyses in this study, three substantial modifications of the original model were undertaken. First, the International Posture factor was modified: Composed of three of the original four subscales of the International Posture instrument in Yashima et al. (2004), the Interest in Foreign Affairs subscale and the Interest in International Vocation/Activities were deleted and the Intercultural Friendship Orientation subscale was returned to the model based on confirmatory factor analysis results. Thus, in the modified model International Posture consisted of Intergroup Approach-Avoidance Tendency and Intercultural

Friendship Orientation. Second, the original 2-factor Motivation instrument was demonstrated to consist of a single dimension, so it entered the model as a measured variable instead of a latent variable. Third, L2 WTC and L2 Communication Frequency were rendered as measured variables rather than latent variables; in the original study L2 WTC was divided into two parcels that were used as indicators, and Frequency of L2 Communication was defined by three items.

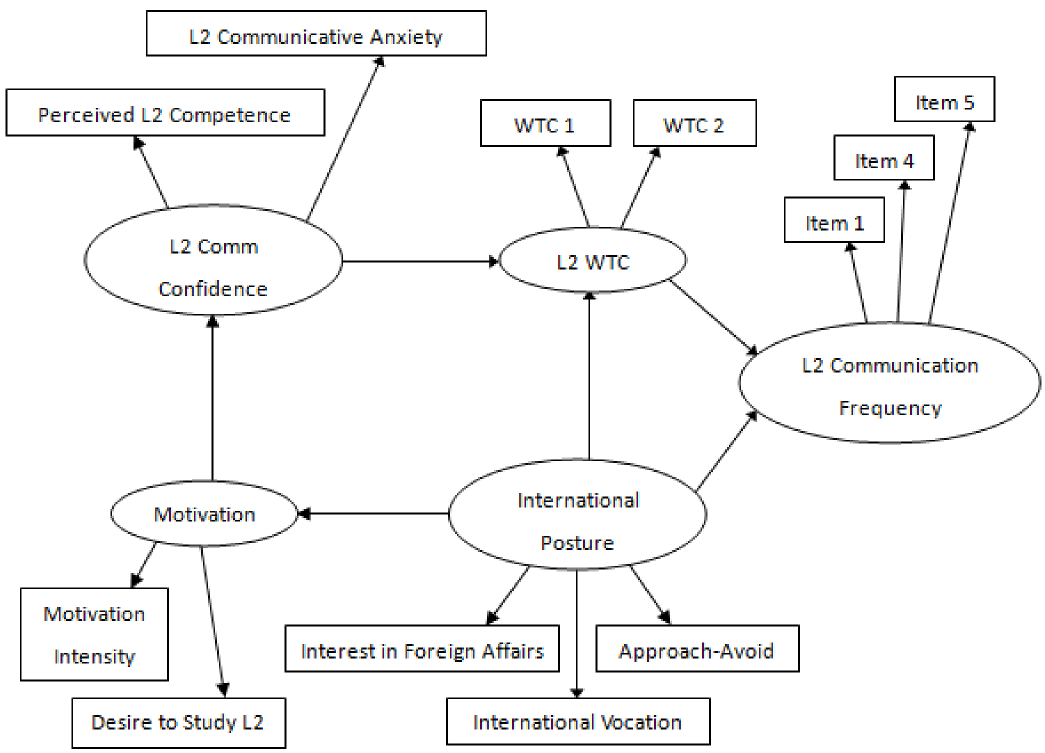


Figure 49. Model of L2 communication. Reprinted from "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Communication," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning*, *54*(1), p. 134. Copyright 2004 by *Language Learning*. Reprinted with permission.

SEM analysis of the original Yashima model indicated that the model fit the data reasonably well, but based on the Lagrange multiplier test, a path was added from International Posture to L2 Communicative Anxiety (Figure 50). This is a logical addition because a favorable disposition toward things international should correspond with lower anxiety about them. With the addition of this path, the analysis yielded good fit statistics with $\chi^2(10) = 29.754$ ($p < .01$), CFI = .970, RMSEA = .089, and 90% C.I. = .080-.125 (Table 85); these values are very similar to those reported in Yashima et al. (2004). In addition, the coefficients of the original paths were similar to those of Yashima et al., differing by .12 at most. The path from International Posture to L2 WTC (.27 in Yashima et al.) was weaker in the current study (.15) but still significant. However, in the initial iteration this path was not significant, yet it became significant when the International Posture–Anxiety path was added. Two explanations are plausible, one of which is that this path is indeed very fragile. The second possibility is that the weakness of this path could be an anomaly in this data set.

Table 85
Step-by-Step Procedure for Revising the Original Yashima et al. 2004 Model with Data-Driven Paths

Model	χ^2	*df*	CFI	IFI	RMSEA
Original model	49.856	11	.941	.942	.119
Add *International Posture – L2 Anxiety* path	29.754	10	.970	.964	.089

Note. CFI = Comparative fit index; RMSEA = root mean square error of approximation.

Detailed statistics for both the original model and the modified model are shown in Table 86, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) appear in Appendix AK. In both cases, the model fit the data well, offering strong support for the robustness of the Yashima et al. (2004) model.

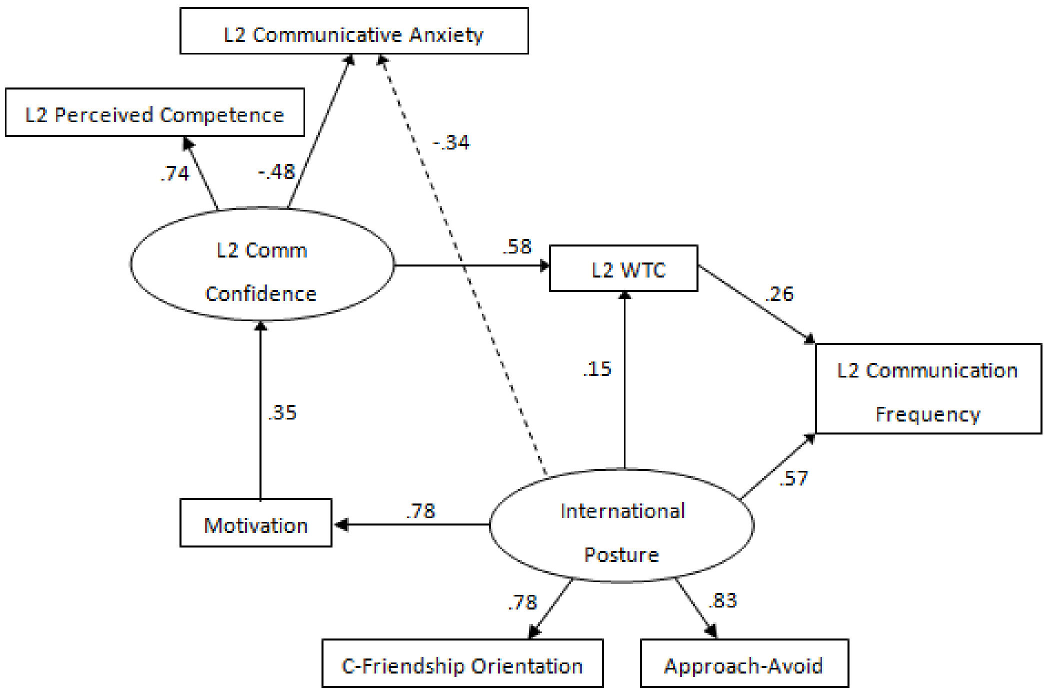


Figure 50. Results of SEM: Respecified revised model of L2 communication with standardized estimates. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2 = 29.754$, $p < .01$, CFI = .970, RMSEA = .089, 90% C.I. = .053-.126.

Table 86
Summary of Fit Indices for Original Yashima et al. 2004 Model

	Original model	Final model
Reliability Coefficient (rho)	.806	.806
Multivariate Kurtosis		
Mardia's coefficient	11.636	11.636
Normalized estimate	8.211	8.211
Residuals		
Average absolute standardized residuals	.051	.040
Average off-diagonal absolute standardized residuals	.067	.050
Model χ^2		
Model estimation method	ML	ML
Independence model χ^2 (df = 21)	675.748	675.748
χ^2 (df = 11, 11)	49.856	29.754
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	4.532	2.705

Table 86 (continued)
Summary of Fit Indices for Original Yashima et al. 2004 Model

	Original Model	*Final Model*
Fit Indices		
Comparative fit index (CFI)	.941	.970
Incremental fit index (IFI)	.942	.970
Standardized root mean square residual (SRMR)	.083	.059
Root mean-square error of approximation (RMSEA)	.119	.089
RMSEA 90% confidence interval	.086-.153	.053 -.126

Revised Yashima et al. (2004) Model

Next, the original Yashima et al. (2004) model was modified based on theoretical considerations and on analyses in this study, resulting in two substantial modifications of the original model. First, Extroversion was added as a variable underpinning L2 Communicative Confidence, which then consisted of Perceived L2 Competence, L2 Communicative Anxiety, and Extroversion. Recall that in the evaluation of the L2 Communicative Confidence measurement model, this 3-factor configuration was found to have the best fit of the various configurations that were evaluated.

Second, Proficiency was added as a latent variable defined by Listening Comprehension and Breadth of Vocabulary Knowledge. Although the role of proficiency lying between motivation and L2 communicative confidence was implied in Yashima et al.'s (2004) study, proficiency was not incorporated into the model (p. 147, Note 7). In the current study, however, proficiency was added as a latent variable. As noted above, the configuration of Proficiency with listening and vocabulary components but with no speaking component was utilized for two reasons. First, English education in Japanese secondary schools focuses heavily on grammar and receptive skills (i.e., listening and reading), which are crucial for the all-important university entrance exams. As such, first-year university students such as the majority of the sample in the current study typically have quite limited speaking proficiency. Second, the task of evaluating speaking proficiency of 252 respondents would have been a daunting job even if the time had been available. The hypothesized model is shown in Figure 51, with bold lines and arrows indicating the additions.

This revised model was evaluated, and initial results indicated barely adequate fit with $\chi^2(31) = 121.136$ ($p < .01$), CFI = .897, RMSEA = .108, and 90% C.I. = .082-.123. In lieu of the marginal fit statistics, the model was modified based on the Lagrange multiplier and Wald test results, with logical paths added one at a

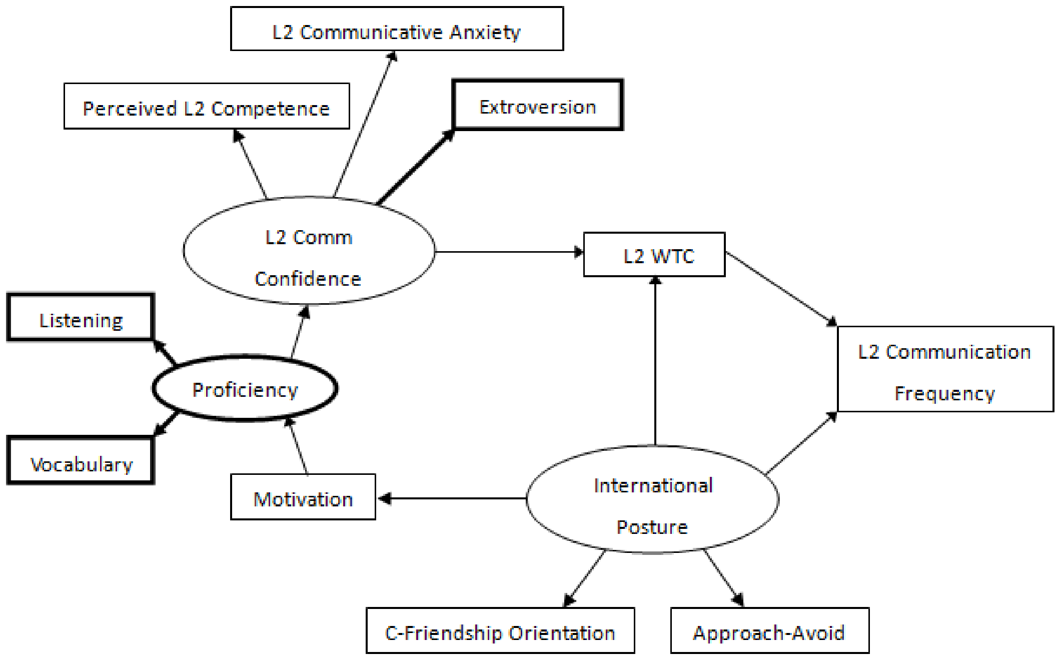


Figure 51. Revised model of L2 communication based on Yashima et al. (2004). Adapted from "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Communication," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning*, *54*(1), p. 134. Copyright 2004 by *Language Learning.* Reprinted with permission.

time and non-significant paths then deleted en masse. First, a path was added from International Posture to L2 Anxiety (as was done above in the Yashima [2002] model). Second, a path was added from Extroversion to International Posture inasmuch as a more extroverted person likely has a stronger propensity toward things international (Figure 50). Detailed in Table 87, this model resulted in substantially better fit: $\chi^2(29) = 68.175$ ($p < .01$), CFI = .955, RMSEA = .074, 90% C.I. = .051-.096.

Table 87
Step-by-Step Procedure for Respecifying the Revised Yashima et al. 2004 Model with Data-Driven Paths

Model	χ^2	*df*	CFI	IFI	RMSEA
Original model	121.136	31	.897	.899	.108
Add *International Posture – L2 Anxiety* path	101.173	30	.919	.920	.098
Add *Extroversion – International Posture* path	68.175	29	.955	.956	.074

The standardized solution is shown in Figure 52. The path coeffieients are similar to the original Yashima et al. model, with the co-occurring paths differing by a maximum of .06. In this model, the fragile International Posture–L2 WTC

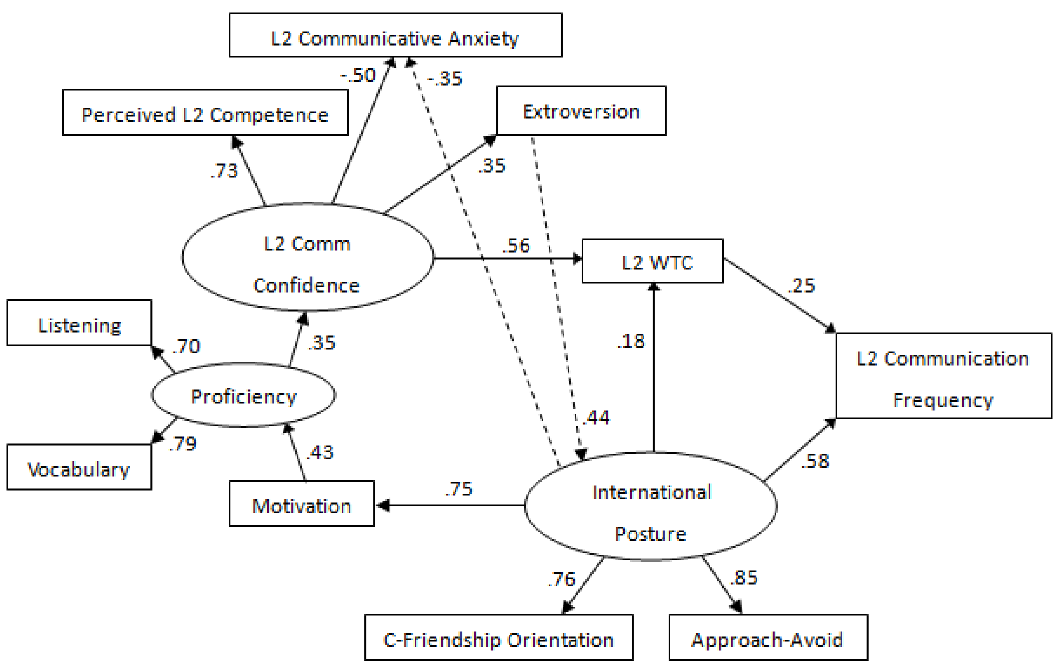


Figure 52. Standardized solution of the revised model of Yashima et al. (2004). Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2(29) = 59.656$ ($p < .01$), CFI = .965, RMSEA = .065, 90% C.I. = .041-.088.

path (.18) was somewhat weaker than the value of .25 reported in Yashima et al. (2004). The new paths from International Posture to L2 Communicative Anxiety (-.35) and from Extroversion to International Posture (.45) were fairly strong.

Detailed statistics of the initial and final solutions are shown in Table 88, and the standardized structural equations, standard errors, and squared multiple correlations (R^2) are shown in Appendix AL. In addition to the improved fit statistics, the average standardized residuals are considerably smaller, offering further support for the modified model having better fit than the original (Byrne, 2006, p. 93).

Table 88
Summary of Fit Indices for the Revised Yashima et al. (2004) Models

	Revised model	Final model
Reliability Coefficient (Cronbach alpha)	.818	.828
Multivariate Kurtosis		
Mardia's coefficient	15.971	15.971
Normalized estimate	8.150	8.150
Residuals		
Average absolute standardized residuals	.077	.050
Average off-diagonal absolute standardized residuals	.093	.058
Model χ^2		
Model estimation method	ML	ML
Independence model χ^2 (df = 45)	922.877	922.877
χ^2 (df = 31, 29)	121.136	68.175
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	3.908	2.351
Fit Indices		
Comparative fit index (CFI)	.897	.955
Incremental fit index (IFI)	.899	.956
Standardized root mean square residual (SRMR)	.119	.071
Root mean-square error of approximation (RMSEA)	.108	.074
RMSEA 90% confidence interval	.088-.128	.051-.096

Note. Based on the moderate level of kurtosis, regular ML estimation was used.

Comparison of Original and Revised Models of Yashima et al. (2004)

Finally, we arrive at a numerical comparison of the two final models (Table 89). For both models, reliability was adequate, and because of the kurtosis, robust statistics were requested for both models. Residuals were normally distributed around the midpoint. The χ^2 value was lower for the original model, yet with more degrees of freedom the χ^2/df ratio was better for the revised model. The fit statistics were better for the revised model. In conclusion, while both models had adequate fit statistics, the revised model had better fit, which indicates that the addition of

Table 89
Summary of Fit Indices for the Original and Revised Yashima et al. (2004) Models

	Original model	Revised model
Reliability Coefficient (rho)	.780	.837
Multivariate Kurtosis		
Mardia's coefficient	16.647	28.470
Normalized estimate	12.048	14.930
Residuals		
Average absolute standardized residuals	.053	.060
Average off-diagonal absolute standardized residuals	.069	.072
Model χ^2		
Model estimation method	ML (Robust)	ML (Robust)
Independence model χ^2 (df = 21, 45)	416.677	655.129
Satorra-Bentler scaled χ^2 (df = 12, 31)	47.857	76.396
Probability value for the χ^2 statistic	.000	.000
χ^2/df ratio	3.988	2.464
Fit Indices		
Comparative fit index (CFI)	.909	.926
Incremental fit index (IFI)	.911	.927
Standardized root mean square residual (SRMR)	(.084)	(.093)
Root mean-square error of approximation (RMSEA)	.107	.075
RMSEA 90% confidence interval	.076-.139	.053-.095

the Proficiency latent variable and the Extroversion variable improved the fit of the model to the data. However, these results yielded support for the robustness of the basic configuration of the Yashima et al. (2004) model.

Summary

In this chapter, the primary results of this study were described. The SEM results indicated that the MacIntyre and Charos (1996) model underwent considerable revision, as was the case in the 1996 study. The Yashima (2002) model and the Yashima et al. (2004) model, however, proved to be robust although both underwent minor revisions and benefitted from the addition of Extroversion to the L2 Communicative Confidence measurement model. Those results are discussed in Chapter 9.

CHAPTER 9

DISCUSSION

In this chapter the findings of the current study are discussed. Many of the details have been covered in the Preliminary Results chapters and the Results chapter, but in this chapter I attempt to construct a coherent narrative. To review for a moment, the general objectives of the current study were (a) to explore the psychometric properties of the various instruments used, (b) to replicate and extend the models of L2 communication, and (c) to explore the addition of personality variables to those models.

Research Question 1: Configuration of the L2 Communicative Confidence Construct

The first research question dealt with the configuration of an important higher-level factor, L2 Communicative Confidence, in the Yashima models. Specifically, this research question asked, "To what extent will the 2-factor structure of the L2 Communicative Confidence factor be replicated in this university EFL context? Will additional personality variables enhance this factor?"

As noted in the previous chapter, the 2-factor configuration displayed good fit. Based on Yashima's (2002) suggestion that L2 communicative confidence could be influenced by or composed of such additional factors as gender and personality, the three personality variables (Extroversion, Ego Permeability, and

Distancing) were added to the baseline 2-factor configuration one by one and the resulting 3-factor measurement models were evaluated with confirmatory factor analysis. Distancing resulted in a model with roughly equivalent fit statistics, but the path from Distance to L2 Communicative Confidence was not significant. Although speculative, Distancing could be subsumed by one of the other variables, among which ego permeability is a prime candidate. Based on the results of the current study, however, no definitive answer is possible, but this could be addressed in future research.

The second addition was Extroversion, which resulted in a 3-factor configuration with good fit statistics and strong path coefficients. The addition of extroversion is logical because an outgoing, extroverted personality should correspond with higher levels of confidence. For some time, Dewaele and colleagues (e.g., Dewaele, 2005; Dewaele & Furnham, 1999) have contended that extroversion is a crucial element in L2 acquisition, and the findings in this study support the importance of extroversion in FL contexts.

Ego Permeability was the third addition, but as noted above, the results indicated that it was configured differently than originally hypothesized. Based on the confirmatory factor analysis of the original five subscales, only two remained: Perceived Time-Money Competence and Need For Order. The Ego Permeability – L2 Communicative Confidence path was not significant.

Thus, of the three posited additions to the L2 Communicative Confidence factor, Extroversion was the sole statistically significant addition. The trifurcate L2

Communicative Confidence factor thus consisted of L2 Anxiety (as measured by the FLCAS), Perceived L2 Competence, and Extroversion.

Research Questions 2 and 3:
Replication and Extension of Three Models of L2 Communication

The second and third research questions addressed the replication and extension of the three models of L2 communication. In particular, the second research question asked, "To what degree will the L2 communication models of MacIntyre and Charos (1996), Yashima (2002), and Yashima et al. (2004) be replicated in this university EFL context? To what extent do data-driven additions improve the models?" The third research question asked, "How much will the three L2 communication models be improved by the addition of perceived distance, extroversion, and ego permeability?" The three models are discussed below in chronological order.

MacIntyre and Charos (1996) Model

The path analysis results indicated that the model had adequate fit to the data. The posited changes in the variables (e.g., the change from Context to English Experiences) functioned well. Four of the five personality subscales were statistically significant; only the path from Agreeableness to L2 WTC was not significant. In the original (1996) study by MacIntyre and Charos, this was a data-driven addition to the model, and based on the non-significant result in the current study, it is possible that the path was a product of a chance characteristic in the data. A second possibility is that the path is subsumed by another variable, which in this case could be extroversion: An agreeable nature should correspond with an extroverted personality, and these two subscales correlated at .33.

The scale modifications appeared to function well, and the model functioned much the same as the original. Of the additions to the model, two observations are in order. First, English Experience (the sequel to Context) was directly although not strongly linked with several variables. Second, L2 Anxiety was a hub in the center of the model with direct paths to six different variables. The sheer number of significant paths underscores the importance of English Experience and L2 Anxiety in the model.

On the other hand, the number of additions, while logical and statistically justified, indicates that this model was not originally a well-specified model. In the original MacIntyre and Charos (1996) study, five data-driven paths were added, and in the current study four more were added. This indicates that the model, in spite of the adequate fit indices, was not optimally specified in either study or that the instruments were suspect.

From the outset of this study, I hypothesized that the additions of Ego Permeability and Distancing would exert a positive effect in the MacIntyre and Charos model, but the addition of those two variables did not improve the model. As noted, the fit statistics in this study were worse than those for the original model, and no path coefficients associated with the additions were significant. This might be due to the two variables being subsumed by other variables or combinations of variables. For example, perceived distance could underpin L2 communicative anxiety, similar to its hypothesized position as a first-level variable in the revised MacIntyre and Charos (1996) model (L2 anxiety was a second-level variable in that model). This could be an avenue for further research.

Yashima (2002) Model

The SEM results indicated that the Yashima 2002 model was robust, with both the original model and the revised model displaying good fit to the data. The path coefficients were similar to those reported in the original (2002) study, but two path coefficients bear mentioning. The L2 Proficiency to L2 Communicative Confidence path was substantially stronger in the current study, while the direct Motivation **–** L2 Communicative Confidence path that bypasses L2 Proficiency was substantially weaker but still significant.

The path coefficient from International Posture to WTC was not significant in the first iteration (before respecification of the model) and barely significant after respecification. The change in significance could be related to a masking effect in which the path coefficient was suppressed by the misspecification of the L2 Communicative Confidence factor (Cheung & Lau, 2009). The general malaise in this path is puzzling, for International Posture should be strongly predictive of WTC. Two possible explanations come to mind: first, the relative dearth of opportunities for Japanese university students to communicate in English could mean that WTC is seen to be unimportant. Second, affective responses in particular situations—akin to ‘performance anxiety’—could overwhelm the underlying propensity toward things international.

The most notable departure from the original model was the data-driven respecification in which a path was added from International Posture to L2 Anxiety. Intuitively this is justified, for L2 Anxiety is underpinned by both L2 Communicative Confidence and International Posture (i.e., L2 anxiety would be lower for students with greater confidence and inclination toward things international). The path coefficients are negative, indicating that higher levels of confidence and international posture correspond with lower L2 anxiety. In addition, the significance of this path underscores the crucial role of International Posture in this model of L2 communication: Five paths originate from International Posture.

The a priori changes posited for this model were, on the whole, more successful than those hypothesized for the MacIntyre and Charos model. Both of the proficiency variables had strong paths, as did the two International Posture subscales. Motivation was recast as a measured variable, and its performance was satisfactory.

Finally, the addition of Extroversion to the L2 Communicative Confidence variable was shown to be a positive step. Data-driven paths from International Posture to Extroversion and L2 Communicative Anxiety were added.

Yashima et al. (2004) Model

The SEM results indicated that the Yashima et al. (2004) model was quite robust, with both the original model and the revised model displaying good fit. The path coefficients were similar to those reported by Yashima et al. (2004), and the SEM results indicated that the path from International Posture to WTC was again barely significant, as was true for the replication of the Yashima (2002) model.

In this model as well as in the replication of the Yashima (2002) model, a data-driven respecification resulted in a path being added from International Posture to L2 Anxiety. In the revised model, a further path was added from Extroversion to International Posture. Again, these paths are indicative of the crucial role of International Posture in this model of L2 communication: Four paths originate from International Posture.

The additions posited for this model were, on the whole, more successful than those for the MacIntyre and Charos model. The addition of the proficiency variables and extroversion improved the fit of the respective models to the data, indicating that both should be included in L2 communication models in the future.

Finally, one more path change deserves note: In the respecified and revised Yashima et al. (2004) model, the direct path from Motivation to L2 Communicative Confidence was not significant, nor did the Lagrange multiplier test indicate that adding it would be prudent. However, in the final revised (2002) model, that path was weakly significant (.18).

Theoretical Implications

One important result of the current study was that extroversion was an important addition to the models of L2 communication. As Dewaele and Furnham (1999) noted, while extroversion is a highly regarded and well-researched variable in psychology, its place in SLA research had at that time received much less attention, but the results in this study indicate that it plays an important role in models of L2 communication.

As noted above, the FLCAS was found to be more appropriate than the L2 Communicative Anxiety scale. This was not an unexpected result inasmuch as opportunities to speak English are limited except for mandatory classes in secondary schools, and even those opportunities fall victim to an increasing grammar-oriented test preparation focus in high school English courses.

Finally, the Ego Permeability construct underwent a transformation. While the instrument emerged virtually unscathed from the Rasch analysis, with only one item misfitting and the configuration of the five subscales remaining otherwise intact, when the overall configuration (i.e., the measurement model) of the instrument was evaluated via a confirmatory factor analysis, the results suggested a 2-factor structure rather than the original 5-factor configuration. The two factors Need for Order and Perceived Time-Money Competence, appear to represent a propensity toward imposing order on one’s personal life and—to the extent possible—on the world at large.

The measurement model was further investigated with a 2-factor, second-order model based on the initial finding. In the hypothesized model, the top-level Ego Permeability factor consisted of Imposition of Order (composed of Need for Order and Perceived Time-Money Competence) and Cognitive Flexibility (made up of Unusual Experiences, Childlikeness, and Sensitivity). However, the model had poor fit, leading to the conclusion that for this particular sample, Ego Permeability was best represented by the new 2-factor Imposition of Order construct.

This newly-dubbed Imposition of Order factor would thus be the diametric opposite of ego permeability, and it seems close to the notion of tolerance of ambiguity, which Furnham and Ribchester (1995) defined as "the way an individual (or group) perceives and processes information about ambiguous situations or stimuli when confronted by an array of unfamiliar, complex, or incongruent clues" (p. 179). Building on the early work of Frenkel-Brunswik (1948, 1949), Budner (1962) asserted that tolerance of ambiguity was indeed a personality variable, and in psychology it represents an individual difference of interest (Anderson & Schwartz, 1992; Nutt, 1993; Tsui, 1993). In the second language acquisition literature, tolerance of ambiguity has received some attention with, for example, Chapelle and Roberts (1986) finding that tolerance of ambiguity and field independence were significant predictors of ESL proficiency. It has also been found to be of significance in the use of L2 learning strategies (Ely, 1989; Zhang, 2004), listening comprehension (Zhou, 2000), and vocabulary retention (Grace, 1998). Indeed, the closeness of the two is highlighted in Ehrman's (1999) comment on the relationship of ego boundaries and tolerance of ambiguity: "[T]hose who tolerate ambiguity are likely to have much less difficulty with experiencing themselves in a variety of ways and seeing themselves through the eyes of other people" (p. 76). Thus, ego boundaries and tolerance of ambiguity are intrinsically related, but the results of this study suggest that tolerance of ambiguity is more appropriate than ego permeability in this context.

Pedagogical Implications

While the primary implications of this study concern theoretical issues, one pedagogical implication should be noted. With extroversion having been found to play an important in the models of L2 WTC that this study addressed, it would behoove language instructors to systematically use distance-inducing activities in EFL classes. I grant that puppetry might not suit some instructors, but roleplay, drama, and public speaking can play useful roles in the EFL classroom in this regard.

Methodological Innovations

While the focus of many researchers is on the theoretical or empirical findings, methodological innovations are also a legitimate result of research. Having said that, the current study includes some innovations that could be useful for future researchers. The use of Rasch analysis and SEM is more illustrative than innovative, but in L2 research the use of those two techniques is not as common as could be. I hope that this study serves as an example of how those two powerful techniques can be incorporated into L2 research.

A useful analysis was the extrapolation exercise involving the category separation criteria for Rasch categories. Assuming that 5-, 6-, and 7-category scales exist (they do) and can be examined with Rasch analysis (they can), a more complete set of separation criteria is thus necessary.

The number of categories in scales is another finding of the current study. Although a greater number of categories allows finer discrimination of responses while shorter scales have greater reliability (Preston & Colman, 2000), the results in this study indicate that employing fewer categories is preferred to a greater number of categories because Likert scales of five or more categories can result in underutilized categories. This finding corroborates results from Cowan's (2000) study, in which mental storage capacity was found to average four chunks of information.

Summary

In this chapter, the results obtained in the current study were discussed. The results of this study indicated that the L2 Communicative Confidence construct was best configured as trifurcate with L2 Communicative Anxiety (FLCAS), Perceived L2 Competence, and Extroversion.

Although the replication and extension of the MacIntyre and Charos (1996) model both yielded satisfactory fit, the models required substantial respecification, which indicates that the initial specification is suspect. However, the Yashima (2002) and Yashima et al. (2004) replications and extensions yielded excellent fit, findings which point to the robustness of the underlying Yashima model.

Of the three personality variables hypothesized to strengthen the respective models, extroversion was the sole survivor that did so. This offers support for the body of work of Dewaele and suggests that extroversion should assume a more prominent role in future research.

Under theoretical implications, the reconfiguration of the Ego Permeability instrument was indicative that its conceptualization could be revisited. Furthermore, the FLCAS was found to be the more appropriate of the two anxiety scales used in this FL context.

Finally, the results concerning the number of categories yielded two findings. First, the separation scale was extended to more fully cover the range of possible numbers of categories. On the other hand, the second finding of importance showed that fewer than five categories are generally necessary. Nonetheless, in those uncommon instances in which a larger number of categories has adequate separation, the minimum separation scale is now available.

In Chapter 10, the limitations of this study, suggestions for future research, and concluding remarks are presented.

CHAPTER 10

CONCLUSION

This final chapter consists of three sections. First, the limitations of the study are discussed. Second, recommendations for future research are outlined. Finally, I offer a brief epilogue.

Limitations of the Study

In the course of conducting this study, several shortcomings that could restrict the interpretability of the results emerged, and it would behoove the reader to remain cognizant of them. First, the use of two of the instruments in this study was suspect. As noted above, the results indicated that the L2 Communicative Anxiety instrument was bidimensional although it was originally posited to be unidimensional, and the replication of the MacIntyre and Charos (1996) model using the L2 Communicative Anxiety instrument yielded an odd model in which anxiety did not directly predict L2 WTC. The second suspect instrument was the Ego Permeability instrument. The configuration of each subscale proved to be robust, but the overall variable consisting of five subscales was not supported by the analyses.

Second, the reliability of several instruments was low (e.g., the Interest in Foreign Affairs subscale of the International Posture instrument). Low reliability of instruments affects the SEM results, generally causing underestimation of causal effects (Kline, 2005).

Directions for Future Research

With the limitations listed above in mind, in this section I offer suggestions for future research.

Replication

The first category involves replication. Nesselroade (1991) offered a succinct summary of general areas that can be the focus of replication studies: time, location, and individuals. A larger sample would permit cross-validation of the results, which would lessen the possibility that the results are due to chance. In the current study, the sample size of 252 was too small to allow for cross-validation; sample sizes of 600 or more permit cross-validation as well as greater power in the analyses. Browne and Cudeck (1989) asserted that their use of a cross-validation coefficient represented an estimate of a function of population parameters, which could then be estimated from the single sample. However, I find the notion of cross-validation with a subsample from the same population questionable: If a primary result is based on some chance characteristic in the population (i.e., a function of population parameters), then any sample drawn from that same population runs a higher risk of having that anomalous characteristic than would a sample from a different population. Replication using unrelated samples is preferable to evaluating a second group from the original sample.

The current study used a cross-sectional design, but the questions addressed in this study might be better addressed using a longitudinal design as in the second section of Yashima et al.'s (2004) study. A useful analytical technique in such a longitudinal study would be latent growth curve analysis.

Second, the models could be tested with different groups in Japan: In the current study, the sample was primarily made up of first-year university students, but upperclassmen might have different orientations toward English (or another L2). As suggested in Yashima (2002), another natural dyad would be to replicate the studies by gender. Mirroring the Yashima et al. (2004) study, investigating these models with internationally oriented students (e.g., those majoring in international studies, English, or tourism) would shed further light on the robustness of the model. In his work on extroversion, Dewaele (2005) noted that many researchers target university students and called for consideration of other populations that represent different "different ethnic or linguistic background, age, ability, and so on" (2005, p. 4), which could include working members of society, teachers, and younger students (e.g., junior high school students). Similar consideration for L2 communication models would be prudent.

Replicating this study with samples from other countries would also be an excellent step. Among the variables used in the current study, for example, L2 WTC has been evaluated in Korea (Kim, 2004) and China (Cao & Philp, 2006;

Wen & Clément, 2003), and similar studies in other Asian contexts would broaden knowledge on the process of L2 communication. Given sufficient sample sizes, the invariance of the model(s) could be tested across different national contexts using multigroup SEM (Lu, Cheung, & Wang 2006).

To Nesselroade's triad I would add 'tools'. These are addressed in the next two sections.

Research with Reconfigured Variables

The second general area for further research concerns parts of the measurement models investigated in the current study. The investigation of two in particular would strengthen this line of research. The first is a detailed analysis of the nature of foreign language anxiety and whether anxiety is best viewed as a state, trait, or combination of the two. I believe that it is a combination and should be manifest on a continuum. Related to this is the question of what type of anxiety instrument is most appropriate in Japanese EFL contexts; in the current study, the FLCAS appeared to be the more appropriate anxiety instrument.

L2 anxiety. As noted above, the FLCAS was considered to be more appropriate than the L2 Communicative Anxiety scale. This was not an unexpected result inasmuch as opportunities to speak English are limited except for mandatory classes in secondary schools, and even that opportunity falls victim to the grammar-oriented test preparation focus in high school English courses. An interesting aside beyond the scope of this study is whether the FLCAS and the L2 Communicative Anxiety instrument could be combined into a hybrid FL anxiety scale. As noted, the two scales address fundamentally different FL contexts, and the FLCAS could be treated as a collection of several minor dimensions. Recall that the original conceptualization of the FLCAS (Horwitz et al., 1986) emerged from consideration of comprehension apprehension, fear of negative evaluation, and test anxiety. In addition, as noted in the Results chapter, several items addressed the notion of anxiety based on insufficient preparation. The hybrid scale would thus include the four minor dimensions and the L2 Communicative Anxiety scale; items would run the gamut from explicitly classroom-oriented items dealing with tests and preparation to items dealing with situations outside the classroom, for example, talking with a stranger while waiting in line (Item 8 of the L2 Communicative Anxiety scale). Such an instrument would cover more of the possible L2 anxiety-inducing contexts than either the classroom-focused FLCAS or the L2 Communicative Anxiety instrument, which addresses some contexts that could occur in a classroom (giving a speech) and some that could not (speaking in line).

In addition, further research into the structure of the FLCAS would be prudent. The original configuration consisted of three factors (omprehension apprehension, fear of negative evaluation, and test anxiety), which Liu and Jackson (2008) also found in their study of Chinese EFL learners. As noted in Chapter 8, the FLCAS could also be partitioned into five subscales with the addition of (lack of) preparation and affective reactions.

Ego permeability. The second construct that could benefit from further research is ego permeability. In the current study, the ego permeability construct was found to be best configured as a 2-factor Imposition of Order construct rather than the 5-factor configuration of the original shortened form. In the work of Madeline Ehrman and colleagues (e.g., Ehrman & Oxford, 1996), ego permeability was associated with a number of interesting results, but in the current study the ego permeability instrument concerned tangible objects, not more nebulous cognitive aspects. Because the instrument was less than robust in the current study than in Ehrman's work, a replication study would help ascertain if ego permeability is fundamentally different in this Japanese EFL context than in the contexts in which it was originally validated.

Another fruitful path would be to compare the original 5-factor ego permeability configuration with tolerance of ambiguity instruments. One instrument could be MacDonald's (1970) ambiguity tolerance instrument that was an extension of an earlier instrument developed by Rydell and Rosen (1966).

Regarding the extroversion instrument, in this study a series of adjective pairs was used, but an instrument similar to that used by Dewaele and Furnham (1999) in which participants indicated the degree of agreement with sentences might work more effectively because a single lexeme is might allow for more interpretation by the respondent than a sentence would.

Openness to Experience. Although again beyond the scope of the present study, further investigating the composition of the respective personality subscales in a Japanese context would be prudent. Of the five subscales, two emerged nearly intact (extroversion and emotional stability), but the other three subscales underwent considerable revision. In particular, the diligence subscale lost three items and gained five for a revised total of nine items, which might indicate that the Japanese notion of diligence differs from that in North American or other contexts. The same might be true of the Openness to Experiences subscale, which lost four items. To evaluate such queries, one might pursue multi-level structural equation modeling as suggested by Lu, Cheung, and Wang (2006) for evaluating invariance across cultures.

Sensitiveness. The shortest personality subscale, Sensitiveness, included just two items, so measurement derived from this is imprecise, given that the person ability estimates have high standard errors. As with the Frequency of L2 Communication instrument, in future studies it is necessary to pilot and incorporate additional items. The two extant items specifically target perception of one's *own* sensitivity, but adding items addressing sensitiviy about external things would broaden the scope from a holistic, speaker-focused “I am sensitive” focus to items dealing with discrete objects or situations in the form of “I react in a sensitive way to [an external object or event].” For example, items could be created to ask about sensitivity toward beauty, death, injury, and accomplishment on the lines of “I feel sad when I see someone crying," or "I feeling very moved when I witness a remarkable [sports/musical/artistic] performance."

Frequency of L2 Communication. Naturally, in future studies a longer scale would be prudent. A further series of questions paralleling the WTC items would broaden the scope of the scale to include asking about frequency of speaking English outside of academic (i.e., school-related) contexts, for example, while shopping or using public transportation such as trains and buses. Using some of the venues from the original eight WTC filler items (e.g., frequency of communication with a salesperson or office personnel) would also be a possibility.

International posture. Third, investigating international posture in more detail would be an excellent step. Two of the subscales (Interest In International News and Interest In International Vocation/Activities) emerged as rather short instruments, and it appears that the Intergroup Approach-Avoid Tendency and Intercultural Friendship Orientation could subsume the two smaller subscales. As noted above, International Posture was vitally important in the Yashima models, serving as a hub.

English Experience. The English Experience variable in this study incorporated the experiences that an EFL learner in Japan might undergo, either as part of his or her compulsory education (English classes in secondary education) or based upon conscious choice by the learner or the learner's parents (e.g., attending a cram school or traveling overseas). However, exposure to English can also be incidental, as when English is encountered in the media or on a sign in public, or it can be the result of an impulsive decision (e.g., a spur-of-the-moment decision to watch a movie on television or to go to a theater). As noted above, the participants in this study were from urban areas in Japan and thus likely had similar exposure to incidental English in everyday life, but in the future, researchers can attempt to quantify this type of English Experience to investigate whether it also plays a role in L2 WTC.

Research with Reconfigured Models

As correctly noted in MacCallum and Austin (2000), there is always the possibility in SEM that alternative models fit the data equally well. The current study was primarily devoted to replicating several models of L2 communication, but other models might fit these or similar data as well if not better.

On the other hand, the number of additions, while logical and statistically justified, indicates that the MacIntyre and Charos (1996) model was not originally a well-specified model. In that study, five data-driven paths were added, and in the current study four more were added. This indicates that the model, in spite of the adequate fit indices, was not optimally specified in either study or that somehow the instruments were suspect. Pursuant to this chimera-like quality, future researchers should certainly replicate the MacIntyre and Charos (1996) study and the current study.

Jekyll and Hyde

The Jekyll and Hyde situation that provided some of the impetus for this study deserves further research. One step would be to examine the behavior of groups with different levels of extroversion in terms of L2 WTC and L2 Communicative Confidence. Such research could include self-perceptions as in the current study in addition to observation of actual L2 behavior (e.g., Cao & Philp, 2006). Moreover, qualitative assessment via interviews would further deepen the data.

Final Remarks

At this point, I must thank the reader for an extraordinary amount of patience and stamina in reading this manuscript. In the course of this study I have learned a great deal, and I hope that the reader has found something of interest and usefulness in these pages. In those long preliminary results chapters, the various instruments used in this study were validated, and it is hoped that they will be used and investigated more fully. In addition, the L2 communication models of Yashima and colleagues (Yashima, 2002; Yashima et al., 2004) were found to be very robust. My hope that the addition of personality variables would improve these models of

L2 communication was partially borne out, and the role of extroversion in such models for Japanese EFL contexts is clearer now.

REFERENCES

Ahmed, S. M. (1990). Psychometric properties of the boredom proneness scale. *Perceptual and Motor Skills, 71*, 963-966.

Andersen, C. (2004). Learning in "as-if" worlds: Cognition in drama in education. *Theory into Practice, 43*(4), 281-286.

Anderson, S., & Schwartz, A. (1992). Intolerance of ambiguity and depression. *Social Cognition, 10*, 271-298.

Baddeley, A. (1994). The magical number seven: Still magic after all these years? *Psychological Review*, *101*(2), 353-356.

Baker, S., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning, 50,* 311-341.

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

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497-529.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P. M. (2006a). *EQS 6* [Computer software]. Encino, CA: Multivariate Software.

Bentler, P. M. (2006b). *EQS 6 structural equations program manual*. Encino, CA: Multivariate Software.

Berkowitz, P. A. (1982). Foreign language: The university's sleeping giant. *Canadian Modern Language Review, 38,* 311-314.

Bettelheim, B. (1975). *The uses of enchantment.* New York: Vintage Books.

Biddle, A., & Thomas, B. (1966). *Role theory: Concepts and research.* New York: Wiley & Sons.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body-motion communication.* Philadelphia: University of Pennsylvania Press.

Bowman, S. L. (2007). The psychological power of the roleplaying experience. *Journal of Interactive Drama, 2*(1), 7-21.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106-148.

Brown, H. D. (1973). Affective variables in second language acquisition. *Language Learning, 23*(2), 231-244.

Brown, H. D. (1980). The optimal distance model of second language acquisition. *TESOL Quarterly, 14*(2), 157-164.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-159). Thousand Oaks, CA: Sage.

Budner, S. (1962). Intolerance of ambiguity as a personality variable. *Journal of Personality*, *30*, 29-50.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming (2nd ed.)*. Mahwah, NJ: Erlbaum.

Cao, Y., & Philp, J. (2006). Interactional context and willingness to communicate: A comparison of behavior in whole class, group and dyadic interaction. *System, 34*, 480-493. DOI: 10.1016/j.system.2006.05.002

Carroll, T. (2005). Beyond keigo: Smooth communication and the expression of respect in Japanese as a Foreign Language. *International Journal of the Sociology of Language. 2005*(175-176), 233–247. DOI: 10.1515/ijsl.2005.2005.175-176.233,

Chang, L. (1994). A psychometric evaluation of four-point and six-point Likert-type scales in relation to reliability and validity. *Applied Psychological Measurement, 18*, 205-215.

Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field independence as predictors of proficiency in English as a Second Language. *Language Learning, 36* (1), 27-45. DOI: 10.1111/j.1467-1770.1986.tb00367.x

Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning*, *25*, 153-161.

Chen, C., Lee, S.-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science, 6*(3), 170-175.

Cheung, G. W., & Lau, R. S. (2009). Testing mediation and suppression effect of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods, 11*(2), 296-325. DOI: 10.1177/1094428107300343

Choi, I., & Choi, Y. (2002). Culture and self-concept flexibility. *Personal and Social Psychology Bulletin, 28*(11), 1508-1517.

Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The effect of number of rating scale categories on levels of inter-rater reliability: A Monte-Carlo investigation. *Applied Psychological Measurement, 9*(1), 31-36.

Clément, R. (1980). Ethnicity, contact and communicative competence in a second language. In H. Giles, W. P. Robinson, & P. Smith (Eds.), *Language: Social psychological perspectives* (pp. 147-154). Oxford: Pergamon Press.

Clément, R. (1986). Second language proficiency and acculturation: an investigation of the effects of language status and individual characteristics. *Journal of Language and Social Psychology, 5*(4), 271-290. DOI: 10.1177/0261927X8600500403

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*, 190-209.

Clément, R., & Kruidenier, B. G. (1985). Aptitude, attitude and motivation in second language proficiency: A test of Clément's model. *Journal of Language and Social Psychology, 4*(1) 21-37.

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning, 44*(3), 417–448. DOI: 10.1111/j.1467-1770.1994.tb01113.x

Cohen J. A. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*, 213-220.

Cohen, R. (2004). Role distance: On state and on the merry-go-round. *Journal of Dramatic Theory and Criticism, Fall 2004.*

Coleman, J. A. (1997). Residence abroad within language study. *Language Teaching, 30*(1), 1-20.

Combs, C. (1988). Theatre and drama in education: A laboratory for actual, virtual or vicarious experience. *Youth Theatre Journal, 2*(3), 9-10.

Costa, P. T., Jr., & McCrae, R. R. (1992). Four ways five factors are best. *Personality and Individual Differences*, *13*, 656-665.

Coste, J., Guillemin, F., Pouchot, J., & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50*(3), 247-252.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform.* Cambridge, UK: Cambridge University Press.

Csikszentmihalyi, M. (2000). *Between boredom and anxiety.* San Francisco: Jossey-Bass.

Curtis, A. M. (1999). Communicating with bereaved children: A drama therapy approach. *Illness, Crisis, & Loss*, *7*(2), 183-190.

DeKeyser, R. (2007). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208-226). New York: Cambridge University Press.

Deppe, R. K., & Harackiewicz, J. M. (1996). Self-handicapping and intrinsic motivation: Buffering intrinsic motivation from the threat of failure. *Journal of Personality and Social Psychology, 70*, 868-876.

Dewaele, J.-M., & Furnham, A. (1999). Extraversion: The unloved variable in applied linguistic research. *Language Learning*, *49*(3), 509-544.

Dewaele, J.-M. (2005). Investigating the psychological and emotional dimensions of instructed language learning: Obstacles and possibilities. *The Modern Language Journal, 89*(3), 367-380.

Digman, J. M. (1989). Five robust trait dimensions: Development, stability, and utility. *Journal of Personality, 57*, 195-214.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. In M. R. Robinson & L. W. Porter (Eds.), *Annual Review of Psychology, 41* (pp. 417-440). Palo Alto, CA: Annual Reviews.

Dörnyei, Z. (1990). Conceptualizing motivation in foreign-language learning. *Language Learning*, *40*(1), 45-78.

Dörnyei, Z. (1994).Understanding L2 motivation: On with the challenge! *Modern Language Journal*, *78*(4), 515-523.

Dörnyei, Z. (2003). Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications. In Z. Dörnyei (Ed.), *Attitudes, orientations, and motivations in language learning: Advances in theory, research, and applications* (pp. 3-32). Malden, MA: Blackwell.

Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics, 23*(4), 421-462.

Dovey, J., & Kennedy, H. W. (2006). *Game culture: Computer games as new media*. Maidenhead, Berkshire, UK: Open University Press.

Dunkel, H. B. (1947). The effect of personality on language achievement. *Journal of Educational Psychology, 38*(3), 177-182.

Ehrman, M. E. (1993). Ego boundaries revisited: Toward a model of personality and learning. In J. E. Alatis (Ed.), *Strategic interaction and language acquisition: Theory, practice, and research* (pp. 331-362). Washington, DC: Georgetown University.

Ehrman, M. E. (1996). *Understanding second language learning difficulties.* Thousand Oaks, CA: Sage.

Ehrman, M. E. (1999). Ego boundaries and tolerance of ambiguity in second language learning. In J. Arnold (Ed.), *Affect in language learning* (pp. 68-86). Cambridge: Cambridge University Press.

Ehrman, M. E., & Oxford, R. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal*, *74*, 311-327.

Ehrman, M. E., & Oxford, R. (1996). Cognition plus: Correlates of adult language proficiency. *Modern Language Journal, 79*(3), 67-89.

Ellis, R. (1985). *Understanding second language acquisition.* Oxford: Oxford University Press.

Elwood, J. (2005). *Moose and the virtual Rubicon: Foreign language anxiety in one Japanese EFL context.* Unpublished manuscript.

Ely, C. M. (1989). Tolerance of ambiguity and use of second language strategies. *Foreign Language Annals 22*(5), 437-445.

Emunah, R. (1994). *Acting for real: Drama therapy process, technique, and performance.* New York: Brunner/Mazel.

Eysenck, H. J. (1947). *The measurement of personality*. London: Kegan Paul, Trench, Trubner.

Eysenck, H. J. (1991). Dimensions of personality: 16, 5, or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences, 12*, 773-790.

Eysenck, S., & Zuckerman, M. (1978). The relationship between sensation-seeking and Eysenck's dimensions of personality. *British Journal of Psychology, 69*, 483-487.

Foss K., & Reitzel, A. (1988). A relational model for managing second language classroom anxiety. *TESOL Quarterly*, *22*, 437-454.

Frenkel-Brunswik, E. (1948). Intolerance of ambiguity as an emotional perceptual personality variable. *Journal of Personality*, *18*, 108-143.

Frenkel-Brunswik, E. (1949). Tolerance towards ambiguity as a personality variable. *American Psychologist, 3*, 268.

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2002). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences*, *14*(1), 47-64.

Furnham, A., & Ribchester, T. (1995). Tolerance of ambiguity: A review of the concept, its measurement and applications. *Current Psychology, 14*(3), 179-199. DOI:10.1007/BF02686907

Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology, 13*, 266-272.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning.* Rowley, MA, US: Newbury House.

Gardner, R. C. (1979). Social psychological aspects of second language acquisition. In H. Giles & R. St. Clair (Eds.), *Language and Social Psychology* (pp. 193-220). Oxford: Blackwell.

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

Gardner, R. C. (2001). Integrative motivation and second language acquisition. In Z. Dörnyei & R. Schmidt (Eds.), *Motivation and second language learning* (pp. 1-20). Honolulu, HI: University of Hawai'i Press.

Gardner, R. C. & MacIntyre, P. D. (1991). An instrumental motivation in language study: Who says it isn't effective. *Studies in Second Language Acquisition, 13*, 57-72.

Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning, 43*, 157-194.

Garson, G. D. (2007). Structural equation modeling. Retrieved June 28, 2007, from http://www2.chass.ncsu.edu/garson/pa765/structur.htm

Giles, H., & Byrne, J. (1982). An intergroup approach to second language acquisition. *Journal of Multilingual and Multicultural Development, 3*, 17-40.

Goffman, E. (1959). *The presentation of self in everyday life.* New York: Anchor.

Goffman, E. (1961). *Encounters: Two studies in the sociology of interaction.* New York: Macmillan.

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26-42.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34.

Gollob, H. S., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80-92.

Gorsuch, G. J. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly, 34*(4) 675-710.

Grace, G. (1998). Personality type, tolerance of ambiguity, and vocabulary retention in CALL. *CALICO Journal, 15*, 19-45.

Guiora, A. Z. (1972). Construct validity and transpositional research: Toward an analytical study of psychoanalytical concepts. *Comprehensive Psychiatry, 13*, 131-150.

Guiora, A. Z., & Acton, W. R. (1979). Personality and language: A restatement. *Language Learning, 29*(1), 193-204.

Guiora, A. Z., Beit-Hallahmi, B., Brannon, R. C., Dull, C. Y., & Scovel, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study. *Comprehensive Psychiatry, 13*(5), 421-428.

Guiora, A. Z., Brannon, R. C., & Dull, C. Y. (1972). Empathy and second language learning. *Language Learning, 22*, 111-130.

Guiora, A. Z., Paluszney, M., Beit-Hallahmi, B., Catford, J. C., Cooley, R. E., & Dull, C. Y. (1975). Language and person studies in language behavior. *Language Learning, 25*(1), 43-61.

Hall, L, Woods, S., Aylett, R., Newall, L., & Paiva, A. (2005). Achieving empathic engagement through affective interaction with synthetic characters. In J. Tao, T. Tan, & R.W. Picard (Eds.), *Affective computing and intelligent interaction: Lecture notes in computer science, 2005/3784*, 731-738. DOI: 10.1007/11573548_94

Hancock, G. R., & Klockars, A. J. . (1991). The effect of scale manipulations on validity: Targeting frequency rating scales for anticipated performance levels. *Applied Ergonomics, 22*, 147-154.

Hartmann, E. (1991). *Boundaries of the mind: A new psychology of personality.* New York: Basic Books.

Hatano, G., & Inagaki, I. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262-272). San Francisco: W. H. Freeman & Company.

Hau, K.-T., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology, 57*, 327-351.

Heine, S. J., Takata, T., & Lehman, D. R. (2000). Beyond self-presentation: Evidence for self-criticism among Japanese. *Personality and Social Psychology Bulletin, 26*(1), 71-78.

Heath, S. B. (1993). Inner city life through drama: Imagining the language classroom. *TESOL Quarterly, 27*(2), 177-192.

Hilleson, M. (1996). “I want to talk with them, but I don’t want them to hear”: An introspective study of second language anxiety in an English-medium classroom. In K. Bailey & D. Nunan (Eds.), *Voices from the language classroom* (pp. 248-275). Cambridge: Cambridge University Press.

Holliday, A. (1999). Small cultures. *Applied Linguistics, 20*, 237-264.

Horwitz, E. (1988). The beliefs about language learning of beginning university foreign language students. *Modern Language Journal, 72*, 284-293.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*(2), 125-132.

Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, (6)*1, 1-55.

Jones, K. W. (2001). “I’ve called ‘em tom-ah-toes all my life and I’m not going to change!”: Maintaining linguistic control over English identity in the US. *Social Forces, 79*(3), 1061-1094.

Jones, P. (1996). *Drama as therapy: Theatre as living.* New York: Routledge.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: User’s guide* (3rd ed.). Mooresville, IN: Scientific Software.

Jung, C. (1969). *Man and his symbols.* Garden City, NY: Doubleday.

Kanno, Y. (2003). *Negotiating bilingual and bicultural identities: Japanese returnees betwixt two worlds.* Mahwah, NJ: Erlbaum.

Kawczynski, A. S. (1951). The two psychological types of language students. *Modern Language Journal, 35*, 113-118.

Kazin, A. (1951). *A walker in the city.* New York: Harcourt, Brace, & World.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

Kim, S. J. (2004). *Exploring willingness to communicate (WTC) among Korean EFL (English as a Foreign Language) students in Korea: WTC as a predictor of success in second language acquisition*. Unpublished doctoral dissertation. Akron, OH: The Ohio State University.

Kimball, A., & Holyoak, A. (2000). Transfer and expertise. In E. Tulving, & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109-122). Oxford: Oxford University Press.

Kipper, D. A. (1996). The emergence of role playing as a form of psychotherapy. *Journal of Group Psychotherapy, Psychodrama, & Sociometry, 49*(3), 99-119.

Kline, R. (2005). *Principles and practices of structural equation modeling,* (2nd ed.). New York: The Guilford Press.

Koole, S. L. (2004). Volitional shielding of the self: Effects of action orientation and external demands on implicit self-evaluation. *Social Cognition, 22*(1), 100-125. (Special issue: The Art and Science of Self-Defense.)

Kormos, J., & Csizér, K. (2007). An interview study of inter-ethnic contact and its role in language learning in a foreign language environment. *System, 35*(2), 241-258.

Labrie, N., & Clément, R. (1986). Ethnolinguistic vitality, self-confidence and second language proficiency: An investigation. *Journal of Multilingual and Multicultural Development, 7*, 269-282.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*(2), 202-220.

Lahad, M. (1999). The use of drama therapy with crisis intervention groups, following mass evacuation. *The Arts in Psychotherapy*, *26*(1), 27-33. DOI: 10.1016/S0197-4556(98)00045-8

Lalonde, R. N., & Gardner, R. C. (1984). Investigating a causal model of second language acquisition: Where does personality fit? *Canadian Journal of Behavioural Science, 16*, 224-237.

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, *24*(1), 35-63. DOI: 10.1177/0267658307082981

Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, *16*(1), 36-55. DOI: 10.1177/026553229901600103

Leary, M. R. (2004). The self we know and the self we show: Self-esteem, self-presentation, and the maintenance of interpersonal relationships. In M. Brewer, & M. Hewstone (Eds.), *Emotion and motivation* (pp. 204-224). Malden, MA: Blackwell.

Lederman, L. M., & Teresi, D. (1993). *The god particle: If the universe is the answer, what is the question?* New York: Dell.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*, 103-122.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85-106.

Linacre, J. M. (2006a). WINSTEPS Rasch measurement computer program (Version 3.363.0) [Computer software]. Chicago: Winsteps.com

Linacre, J. M. (2006b). *A user's guide to WINSTEPS.* Chicago, IL: Winsteps.com

Linacre, J. M. (n.d.). Dimensionality: Contrasts & variances. Retrieved October 28, 2007, from http://www.winsteps.com/winman/principalcomponents.htm

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal, 92*(1), 71-86.

Lounsbury, J. W., Sundstrom, E., Loveland, J. M., & Gibson, L. W. (2003). Intelligence, "Big Five" personality traits, and work drive as predictors of course grade. *Personality and Individual Differences, 35*(6), 1231-1239.

MacCallum, R. C., Browne, M. W., & Sugahara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

MacDonald, A. (1970). Revised scale for ambiguity tolerance: Reliability and validity. *Psychological Reports, 26*, 791-798.

MacIntyre, P. (1994). Variables underlying willingness to communicate: A causal analysis. *Communication Research Reports, 11*, 135-142.

MacIntyre, P. (2007). Willingness to communicate in the second language: Understanding the decision to speak as a volitional process. *The Modern Language Journal, 91*, 564-576.

MacIntyre, P. D., & Charos, C. (1996). Conceptualizing willingness to communicate in a L2: A situated model of confidence and affiliation. *Journal of Language and Social Psychology, 15*(1), 3-26.

MacIntyre, P., Clément, R., Baker, S. C., & Conrod, S. (2001). Willingness to communicate, social support and language learning orientations of immersion students. *Studies in Second Language Acquisition, 23*, 369-388.

MacIntyre, P., Clément, R., Dörnyei, Z., & Noels, K. (1998). Conceptualizing willingness to communicate in a L2: A situated model of confidence and affiliation. *Modern Language Journal, 82*, 545-562.

MacIntyre, P. D. & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning, 41*, 513-534.

MacIntyre, P. D. & Gardner, R. C. (1994a). The effects of induced anxiety on cognitive processing in computerised vocabulary learning. *Studies in Second Language Acquisition, 16*, 1-17.

MacIntyre, P. D., & Gardner, R. C. (1994b). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44* (2), 283-305. DOI: 10.1111/j.1467-1770.1994.tb01103.x

Makita-Discekici, Y. (1999). A touch of … class! Creative skit activity in the Japanese language classroom. *Canadian Modern Language Review, 32*(3), 224-243.

Mantero, M. (Ed.). (2006). Identity and second language learning: Culture, inquiry, and dialogic activity in educational contexts. Charlotte, NC, US: Information Age.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41*, 954-969.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38*, 299-337.

Marsella, S. C., Johnson, W. L., & LaBore, C. M. (2003). Interactive pedagogical drama for health interventions. In H. U. Hoppe et al. (Eds.), *Artificial intelligence in education* (pp. 341-348).

Marsh, H. W., Balla, J. R., & Hau, K-T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 315-353). Mahwah, NJ: Erlbaum.

McCrae, R. R., & Costa, P. T., Jr. (1989). The structure of interpersonal traits: Wiggins' circumplex and the five-factor model. Journal of Personality and Social Psychology, 56(4), 586-595.

McCroskey, J. C. (1992). Reliability and validity of the willingness to communicate scale. *Communication Quarterly, 40*, 16-25.

McCroskey, J. C., & Richmond, V. P. (1991). Willingness to communicate: A cognitive perspective. In M. Booth-Butterfield (Ed.), *Communication, cognition, and anxiety* (pp. 114-127). Rowley, MA: Newbury House.

MEXT [Ministry of Education, Sports, Science, Culture, and Technology]. (2003). *The course of study for foreign languages.* Tokyo: Author. Retrieved February 13, 2005, from http://www.mext.go.jp/english/shotou/030301.htm

Miller, G. A. (1956). The magic number seven, plus or minus two: Some constraints on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Mulaik, S. A., James, L. A., Van Alstine, J., Bennett, N. Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430-443.

Muller, R., & Buttner, P. (1994). A critical discussion of intraclass correlation coefficients. *Statistics in Medicine, 13*(23-24), 2465-2476.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Newsom, J. (2007). Some clarifications and recommendations on fit indices. Retrieved September 6, 2008, from http://www.upa.pdx.edu/IOA/newsom/semclass/ ho_fit.doc

Norem, J. K., & Cantor, N. (1990). Cognitive strategies, coping and perceptions of competence. In R. J. Sternberg and J. Kolligian, Jr. (Eds.), *Competence considered* (pp. 190-201). New Haven, CT: Yale University Press.

Norton, B., & Toohey, K. (2002). Identity and language learning. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 115-123). Oxford: Oxford University Press.

Nutt, P. (1993). Flexible decision styles and the choices of top executives. *Journal of Management Studies, 30*, 695-721.

Oaster, A. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills, 68*, 549-550.

Onwuegbuzie, T., Bailey, P., & Daley, C. E. (1999). Factors associated with foreign language anxiety. *Applied Psycholinguistics, 20*, 217-239.

Pavlenko, A. (2002). Narrative study: Whose story is it, anyway? *TESOL Quarterly, 36*, 213-218.

Pavlenko, A., & Blackledge, P. (2004). *Negotiation of identities in multilingual contexts.* Clevedon, UK: Multilingual Matters.

Peabody, D. G., & Lawrence, R. (1989). Some determinants of factor structures for personality-trait descriptors. *Journal of Personality and Social Psychology, 57*, 552-567.

Pennycook, A. (1998). *English and the discourses of colonialism.* New York: Routledge.

Perls, F. (1969). *Gestalt therapy verbatim.* New York: Bantam Books.

Piedmont, R. L., & Hyland, M. E. (1993). Inter-item correlation frequency distribution analysis: A method for evaluating scale dimensionality. *Educational and Psychological Measurement, 53*, 369-378.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667-686.

Pratt, M. L. (1991). Arts of the contact zone. *Profession, 91*, 33-40.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*, 1-15.

Psychological Assessment Resources, Inc. (2005). NEO Personality Inventory-Revised. Available from http://www3.parinc.com/products/product.aspx?Productid=NEO-PI-R

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Ratey, J. J. (2002). *A user's guide to the brain: Perception, attention, and the four theaters of the brain.* New York: Vintage Books.

Rawlings, D. (2001). An exploratory factor analysis of Hartmann's Boundary Questionnaire and an empirically-derived short version. *Imagination, Cognition, and Personality, 21*(2), 131-144.

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire items and scale properties. In P. Fayers and R. Hays (Eds.), *Assessing quality of life in clinical trials: Methods and practice* (2nd ed.) (pp. 55-73). Oxford: Oxford University Press.

Ritchie, T. D., Skowronski, J. J., Wood, S. E., Walker, W. R., Vogl, R. J., & Gibbons, J. (2006). Event self-importance, event rehearsal, and the fading affect bias in autobiographical memory. *Self and Identity, 5*, 172-195.

Ross, M., & Wilson, A. E. (2002). It feels like yesterday: Self-esteem, valence of personal past experiences, and judgments of subjective distance. *Journal of Personality and Social Psychology, 82*(5), 792–803.

Rothstein, M. G., Paunonem, S. V., Rush, J. C., & King, G. A. (1994). Personality and cognitive ability predictors of performance in graduate business school. *Journal of Educational Psychology, 86*(4), 516-530.

Rouse, J. (1989). *Brecht and the West German theater: The practice and politics of interpretation*. Ann Arbor, MI: UMI Press.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in *Personality and Social Psychology Bulletin*. *Personality and Social Psychology Bulletin, 28*(12), 1629-1646.

Rydell, S., & Rosen, E. (1966). Measurement and some correlates of need cognition. *Psychological Reports, 19*, 139-165.

Samimy, K. K., & Tabuse, M. (1992). Affective variables and a less commonly taught language: A study in beginning Japanese classes. *Language Learning, 42*(3), 377-398. DOI: 10.1111/j.1467-1770.1992.tb01341.x

Schlenker, B. R. & Trudeau, J. V. (1990). Impact of self-presentation on private self-beliefs: Effects of prior self-beliefs and misattribution. *Journal of Personality and Social Psychology, 58*(1), 22-32.

Schumacker, R. E., & Muchinsky, P. M. (1996). Disattenuating correlation coefficients. *Rasch Measurement Transactions, 10*(1), 479.

Schumann, J. (1975). Affective factors and the problem of age in second language acquisition. *Language Learning, 25*(2), 209-235.

Schumann, J. (1976). Social distance as a factor in second language acquisition. *Language Learning, 26*(1), 135-143.

Sick, J. (2006). *The learner's contribution: Individual differences in language learning in a Japanese high school.* Unpublished doctoral dissertation, Temple University.

Siroka, R. W., Siroka, E., & Schloss, G. (Eds.). (1971). *Sensitivity training and group encounter: An introduction.* New York: Grossett and Dunlop.

Smith, Jr., E. V. M. (2001). Evidence for the reliability of measures and the validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2*, 281-311.

Smith, R. E., Sarason, I. G., & Sarason, B. R. (1982). *Psychology: The frontiers of behavior.* New York: Harper & Row.

Sogunro, O. A. (2004). Efficacy of role-playing pedagogy in training leaders: Some reflections. *Journal of Management Development 23*(4), 355-371.

Spolin, V. (1986). *Theater games for the classroom.* Evanston, IL, US: Northwestern University Press.

Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041-1053.

Statistical Package for the Social Sciences [SPSS]. (2004). [Computer software]. Chicago: Author.

Steiger, J. H. (1990). Structural equation modeling and evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*, 173-180.

Stephen, J. (2004). *Ego boundaries re-interpreted: The contemporary relevance of a psychoanalytic construct for understanding individual differences in psychopathology and attention processing.* Unpublished doctoral dissertation, York University.

Stern, S. L. (1980). Drama in second language learning from a psycholinguistic perspective. *Language Learning, 30*(1), 77-100.

Stern, S. L. (1983). Why drama works: A psycholinguistic perspective. In J.W. Oller & P.A. Richard-Amato (Eds.), *Methods that work: A smorgasbord of ideas for language teachers* (pp. 207–225). Rowley, MA: Newbury House.

Stern, S. L. (1993). Why drama works: A psycholinguistic perspective. In J. Oller (Ed.), *Methods that work: Ideas for literacy and language teachers* (pp. 70-83). Boston: Heinle & Heinle.

Sutton, L.A. (2001). The principle of vicarious interaction in computer-mediated communications. *International Journal of Educational Telecommunications, 7*(3), 223-242.

Tabachnick, B., & Fidell, L. (2001). *Multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Tajfel, H. (1982). *Social identity and intergroup relations.* Cambridge: Cambridge University Press.

Taylor, S. E. Pham, L. B., Rivkin, I. D., & Armor, D. A (1998). Harnessing the imagination: Mental stimulation, self-regulation, and coping. *American Psychologist, 53*(4), 429-439.

Tedeschi, J. T. (1981). *Impression management theory and social psychological research*. New York: Academic Press.

Thompson, T. (1994). Self-worth protection: Review and implications for the classroom. *Educational Review, 46*(3), 259-274.

Tice, D. M., & Baumeister, R. F. (1990). Self-esteem, self-handicapping, and self-presentation: The strategy of inadequate practice. *Journal of Personality, 58*, 443-464.

Tsui, J. (1993). Tolerance for ambiguity, uncertainty audit qualifications and bankers' perceptions. *Psychological Reports, 72*, 915.

Vandrick, S. (1997). The role of hidden identities in the postsecondary ESL classroom. *TESOL Quarterly, 31*(1), 153-157.

van Ments, M. (1999). *The effective use of role play: Practical techniques for improving learning* (2nd ed.). New York: Kogan Page Publishers.

Waternaux, C. M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika, 63*(3), 639-645. Retrieved March 27, 2010, from http://www.jstor.org/stable/2335746

Wen, W. P., & Clément, R. (2003). A Chinese conceptualization of willingness to communicate in ESL. *Language, Culture, and Curriculum, 16*(1), 18-38.

Wilson, R. G. (1978). Introversion/extraversion. In H. London & Je. E. Exner (Eds.), *Dimensions of personality* (pp. 217-261). New York: John Wiley.

Wilson, R. G., & Lynn, R. (1990). Personality, intelligence components and foreign language attainment. *Educational Psychology, 10*, 57-71.

Wright, B. D., & Masters, G. N. (2002). Number of person or item strata. *Rasch Measurement Transactions, 16*(3), 888. Retrieved April 20, 2009, from http://www.rasch.org/rmt/rmt163f.htm

Wright B. D., & Stone, G. N. (1979). *Basic test design: Rasch measurement*. Chicago: MESA Press.

Wolfe, E. N., & Smith, E. V., Jr. (2007). 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. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling*, *3*(1), 3-24.

Yashima, T. (2000). Orientations and motivation in foreign language learning: A study of Japanese college students. *JACET Bulletin, 31*, 121–133.

Yashima, T. (2002). Willingness to Communicate in a second language: The Japanese EFL context. *The Modern Language Journal, 86*(1), 54-66.

Yashima, T., & Zenuk-Nishide, L. (2008). The impact of learning contexts on proficiency, attitudes, and L2 communication: Creating an imagined international community. *System, 36*(4), 566-585.

Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language acquisition. *Language Learning*, *54*(1), 119-152.

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does the language anxiety research suggest? *Modern Language Journal, 75,* 426-439.

Zhang, Q. (2004). Effects of tolerance of ambiguity on the selection of language learning strategies. *Foreign Language Teaching and Research, 6.*

Zhou, Y. (2000). The influence of ambiguity tolerance on listening comprehension. *Journal of Pla University of Foreign Languages, 4*. DOI: cnki:ISSN:81-5019.0.2000-04-018

APPENDICES

APPENDIX A
DEMOGRAPHIC INFORMATION (JAPANESE)

この調査は研究目的であり、皆さんがこの調査で提供する個人情報は厳密に管理されます。また、皆さんの回答は成績や単位に全く影響はありません。ご協力に感謝します。

(a)学科_____ (b) 学年 _____ (c) 性別：男・女

(d) これまでに受験した英語に関する試験とそのスコア（STEP, TOEFL, TOEIC,その他）__________________________

(e)留学、語学研究など海外で経験したことがありますか？（はい、いいえを〇で囲む）

(f)（はいの場合、どの国にどのくらいの期間勉強しましたか？）

____

(g)ホームステイをしたことはありますか？（はい、いいえを〇で囲む）

(h)（はいの場合、どの国にどのくらいの期間ホームステイをしましたか？）

____

(i)海外に住んだことはありますか？（はい、いいえを〇で囲む）

(j)（はいの場合、どの国にどのくらいの期間住みましたか？）

____

(k)海外旅行をしたことはありますか？（はい、いいえを〇で囲む）

(l)（はいの場合、どの国にどのくらいの期間旅行しましたか？）

____

(m)今までに英会話学校へ通ったことはありますか？（はい、いいえを〇で囲む）

(n)（はいの場合、どのくらいの期間通いましたか？）

____

(o) 予備校や塾へ通ったことはありますか？（はい、いいえを〇で囲む）

(p) （はいの場合、どのくらいの期間通いましたか？）

____

(q) 何歳のときに英語の勉強を始めましたか？ _____

APPENDIX B
DEMOGRAPHIC INFORMATION (ENGLISH)

This survey is for research. The information you provide will be held in the strictest confidence and will in no way affect your grade. Thank you for your time!

(a) Major _____ (b) Year in school _____ (c) Gender: Male / Female

(d) English test scores (STEP, TOEFL, TOEIC, etc.) ________________________

(e) Have you ever studied abroad? Yes / No

(f) If "yes," where and how long? _______________________________________

(g) Have you ever done a homestay? Yes / No

(h) If "yes," where and how long? _______________________________________

(i) Have you ever lived abroad? Yes / No

(j) If "yes," where and how long?____

(k) Have you ever traveled abroad? Yes / No

(l) If "yes," where and how long? _______________________________________

(m) Have you ever attended an eikaiwa? Yes / No

(n) If "yes," where and how long? _______________________________________

(o) Have you ever attended or are you now attending a yobiko or a juku? Yes / No

(p) If "yes," where and how long? _______________________________________

(q) At what age did you begin studying English? _____

APPENDIX C
BREADTH OF VOCABULARY KNOWLEDGE

Complete the missing vocabulary word.
(example) The girl is sk______ on the ice. ➔ The girl is sk<u>ating</u> on the ice.

2K-level.

3. The nu____ was helping the doctor in the operating room.

5. This year long sk_____ are fashionable again.

6. Laws are based on the principle of jus_____.

7. He is walking on the ti_____ of his toes.

8. The mechanic had to replace the mo_____ of the car.

9. There is a co_____ of the original report in the file.

11. The doctor ex_____ the patient thoroughly.

13. The railway con_____ London with its suburbs.

16. This work is not up to your usu_____ standard.

18. You must have been very br_____ to participate in such a dangerous operation.

3K-level

1. I live in a small apa_______ on the second floor.

4. It was a cold day. There was a ch______ in the air.

6. Anthropologists study the struc_______ of ancient societies.

9. Some aristocrats believed that blue blood flowed through their ve______.

10. The secretary assi_______ the boss in organizing the course.

11. His beard was too long. He decided to tr______ it.

12. People were whir_______ around on the dance floor.

16. Crying is a nor_______ response to pain.

17. The Emperor of China was the supr_______ ruler of his country.

18. You must be awa_______ that very few jobs are available.

5K-level

2. After finishing his degree, he entered a new ph______ in his career.

3. The workmen cleaned up the me______ before they left.

5. I saw them sitting on st______ at the bar drinking beer.

6. His favorite musical instrument was a tru______.

7. The building is heated by a modern heating appl______.

11. After falling off his bicycle, the boy was covered with bru_______.

12. The child was holding a doll in her arms and hu______ it.

13. We'll have to be inventive and de______ a scheme for earning more money.

15. Nuts and vegetables are considered who______ food.

17. Many people feel depressed and gl_______ about the future of mankind.

University Word Level

1. I've had my eyes tested and the optician says my vi______ is good.

3. In their geography class, the children are doing a special pro______ on North America.

4. In a free country, people are not discriminated against on the basis of colour, age, or s______ .

5. A true dem______ should ensure equal rights and opportunities for all citizens.

9. Governments often cut budgets in times of financial cri______ .

11. Research ind______ that men find it easier to give up smoking than women.

12. In a lecture, a lecturer does most of the talking. In a seminar, students are expected to part______ in the discussion.

14. It's difficult to ass______ a person's true knowledge by one or two tests.

17. His decision to leave home was not well thought out. It was not based on rat_______ considerations.

18. The challenging job required a strong, successful, and dyn______ candidate.

APPENDIX D
LISTENING COMPREHENSION

Part 1. Dialogues

Dialogue 1. (Train directions for a foreigner)

A: Excuse me, but you seem to be wondering about something.
B: Well, now that you ask, yes, I am. This is the first time I've ridden the subways here, and I really don't know how to get where I want to go.
A: Which is…?
B: This station called Korakuen—I want to see a Giants' game!
A: OK. First take the Yamanote Line (it's a JR train line) to Ikebukuro and then take the Marunouchi Line (a subway line marked in red on the signs) to Korakuen.
B: Thank you so much.
(88 words)

Dialogue 2. (Wheat field farmer chat)

Frank: Howdy, Ted.
Ted: 'Lo, Frank.
Frank: Good looking crop this year.
Ted: Yep.
Frank: S'pose prices'll stay up this year?
Ted: Hard telling.
Frank: Well, rumor says we'll prob'ly ship lots overseas. Glad lots of folks like bread.
Ted: Yep.
Frank: Well, good seeing ya.
Ted: Yep. Later.
(39 words)

Dialogue 3. (Office visit)

A: Welcome! Come on in!
B: Thanks! Not too busy, are you?
A: Not at all. Always a pleasure to see you.
B: Every time I visit your office, I am so amazed!
A: Why is that?
B: You always have something new—if it's not your puppets, then it's something else. Take this plant, for example.
A: Yep, I repotted that one about three weeks ago, and it's doing great.
B: What is it? It looks familiar…
A: Oh, that's an aloe vera. It's great for burns, and I've heard people even eat it—but I really don't want to try it.
B: Good for burns, you say?
A: Sure. Shall I make a little one for you?

B: Many thanks! Please write down how to care for it, too, or I might kill the poor little guy! *(127 words)*

Part 2. Listening passage

Shipbuilding in Charleston

(Elwood; adapted from *Smithsonian*, April, 2003, pp. 30-34)

Under a soaring white tent, two volunteers coated in sawdust pull a 5-meter-long oak plank though a 2.5-meter-tall band saw. Gas-powered planers and circular saws howl and screech. The deafening noise suits the half-dozen men armed with power tools in a makeshift shipyard near the waterfront in Charleston, South Carolina. When you build a tall ship, you want people to notice.

Today, a 30-meter ship is rising, board by board. Detailed plans were obtained from the Smithsonian Museum, and local people are faithfully following the so-called "line drawings" of the original ship. Built along the lines of the *Frances Elizabeth,* an 1879 ship, the new *Spirit of South Carolina* is designed to carry 20 young people on extended educational sailing voyages sponsored by the South Carolina Maritime Heritage Foundation.

The construction of the *Spirit of South Carolina* has sparked renewed interest in the city's maritime history. Local volunteers found a picture of the original ship racing in 1889, and the great-grandson of the shipyard owner where the *Frances Elizabeth* was built has played an important role in development of the project. Moreover, the place where the original ship sank was discovered in 1993, and an exploratory dive is planned.
(198 words)

Dialogue 1.	
1. Why is Person A looking confused? a. He has been drinking. b. He is using the train system for the first time. c. He doesn't know which team to cheer for. d. He wants to walk to the stadium.	3. How many transfers will Person A make? a. None. b. One. c. Two. d. Three.
2. What is the easiest route to go to Korakuen Station? a. To use JR trains. b. To take a bus. c. To take the subway. d. To use both JR trains and the subway.	
Dialogue 2.	
1. What are these two men talking about? a. Good-looking women. b. Baseball. c. Telling stories. d. Their crops. 2. What kind of crop are they talking about? a. Wheat. b. Apples. c. Onions. d. Cattle.	3. Where will they send much of their production? a. To their girlfriends. b. To foreign countries. c. To the local stores. d. To domestic companies.
Dialogue 3.	
1. Where are the two people speaking? a. On a street corner. b. On a bus. c. In an office. d. At a restaurant.	4. What does the host offer to do for his visitor? a. Give him a plant. b. Make coffee. c. Heal his burns. d. Help him decide a name for the plant.
2. What does the visitor mention before the latest new thing? a. A new coffee mug. b. Nothing. c. His host's new necktie. d. The human-like dolls.	5. The visitor is a. Good at taking care of plants. b. Afraid of plants. c. Very, very busy. d. Not good at taking care of plants.
3. The plant is especially good for what? a. Eating. b. Healing burns. c. Being green. d. Talking to.	

Listening Passage	
1. Why are local people building a wooden sailing ship? a. To export local products. b. To provide an educational sailing experience for young people. c. To improve their carpentry skills. d. Because they have lots of free time.	4. What have local people <u>not</u> done in conjunction with the renewed interest in the city's maritime history? a. Changed the name of the city. b. Planned to explore more. c. Involved relatives of long-ago shipbuilding people. d. Obtained historical plans to build the new ship.
2. The name of the new ship is what? a. *Spirit of South Carilina* b. *Francis Elizabeth* c. *Smithsonian* d. It doesn't have a name yet.	5. The new ship is being built of a. Steel. b. Wood. c. Cement. d. Aluminum.
2. How long ago was the original ship built? a. About 25 years ago. b. About 2.5 years ago. c. About 300 years ago. d. About 130 years ago.	

APPENDIX E
FREQUENCY OF L2 COMMUNICATION (YASHIMA, 2002)

あなたは、次のような状況でどの程度英語でコミュニケーションをとりましたか。大学に入学してからの約３ヶ月を振り返って、自分の行動に最も近いところに○をしてください。

全くなかった		たまにあった		時々あった		よくあった		非常に頻繁にあった
1		2		3	4	5	6	7

fr1	英語の授業中自分から**すすんで**発表したり質問したりした。	1－2－3－4－5－6－7
fr2	英語の授業中先生に指名されて発言した。	1－2－3－4－5－6－7
fr3	英語の授業中、ペアワークなどの活動に参加した。	1－2－3－4－5－6－7
fr4	授業外で先生に英語で質問したり話をした。	1－2－3－4－5－6－7
fr5	学校外の友人や知り合いと英語で話をした。	1－2－3－4－5－6－7

APPENDIX F
FREQUENCY OF L2 COMMUNICATION (ENGLISH) (YASHIMA, 2002)

あなたは、次のような状況でどの程度英語でコミュニケーションをとりましたか。
大学に入学してからの約３ヶ月を振り返って、自分の行動に最も近いところに〇
をしてください。

(fr1). I volunteered to answer or ask questions in class.

(fr2). I answered when I was called upon by the teacher.

(fr3). I participated in classroom activities such as pair work.

(fr4). I asked teachers questions or talked to them outside the class period.

(fr5). I talked with friends or acquaintances outside school in English.

APPENDIX G
L2 WILLINGNESS TO COMMUNICATE (L2 WTC; MCCROSKEY, 1992)

英語でコミュニケーションをするかしないかを選択するような状況を、下に 20 示します。
コミュニケーションをするかしないかは、全くあなたの自由だと仮定してください。それぞれの状況において、どれぐらいのパーセンテージでコミュニケーションすると思うか考えてください。（日本の国内もしくは外国でおこりうる状況です）
下線の上に、そのパーセンテージを記入してください。その状況で絶対にしないという場合は 0、必ずするという場合は 100 となります。すべて英語で話す状況です。

0% == 100%
絶対コミュニケーションしない　　　　必ずコミュニケーションする

英語で話す状況 (こういう状況を経験したことがなくても<u>想像で</u>回答して下さい。)

(wtc1) _____ % * ガソリンスタンドの店員と話す

(wtc2) _____ % *医者と話す

(wtc3) _____ %知らない人たちにスピーチ（プレゼンテーション）をする

(wtc4) _____ %列になって待っているとき知り合いと会話をする

(wtc5) _____ % *店で店員と話をする

(wtc6) _____ %友人の大きな集まり（会議）で発言する

(wtc7) _____ % *警察官・婦人警官と話をする

(wtc8) _____ %知らない人の小グループで会話をする

(wtc9) _____ %列になって待っているとき友人と会話をする

(wtc10) _____ % *レストランでウェイター・ウェイトレスと話をする

(wtc11) _____ %知り合いの大きな集まり（会議）で発言する

(wtc12) _____ % 列になって待っているとき知らない人と会話をする

(wtc13) _____ % *秘書と話をする

(wtc14) _____ %友人の一団にスピーチ（プレゼンテーション）をする

(wtc15) _____ %知り合いの小グループで会話をする

(wtc16) _____ % *ごみを集めにくる係員と話す

(wtc17) _____ %知らない人の大きな集まり（会議）で発言する

(wtc18) _____ % *夫・妻（ボーイフレンド・ガールフレンド）と話す

(wtc19) _____ %友人の小グループで会話をすとき

(wtc20) _____ %知り合いにスピーチ（プレゼンテーション）をする

APPENDIX H
L2 WILLINGNESS TO COMMUNICATE (ENGLISH)
(WTC; MCCROSKEY, 1992)

(Note: filler items are marked with an asterisk)

In the 20 situations below, please indicate if you would or would not communicate in English. Please write a percentage in the blank. An answer of 0% indicates you would absolutely <u>not</u> communicate in English, and an answer of 100% indicates that you would <u>certainly</u> communicate in English. (These situations could have occurred either in Japan or abroad).

0% == 100%
Absolutely would not speak   Certainly would speak

(Even if you have never experienced such a situation, answer based on what you <u>imagine you would do</u> in that situation.)

(wtc1) _____ % *Talk with a service station attendant.

(wtc2) _____ % *Talk with a physician.

(wtc3) _____ % Present a talk to a group of strangers.

(wtc4) _____ % Talk with an acquaintance while standing in line.

(wtc5) _____ % *Talk with a salesperson in a store.

(wtc6) _____ % Talk in a large meeting of friends.

(wtc7) _____ % *Talk with a police officer.

(wtc8) _____ % Talk in a small group of strangers.

(wtc9) _____ % Talk with a friend while standing in line.

(wtc10) _____ % *Talk with a waiter/waitress in a restaurant.

(wtc11) _____ % Talk in a large meeting of acquaintances.

(wtc12) _____ % Talk with a stranger while standing in line.

(wtc13) _____ % *Talk with a secretary.

(wtc14) _____ % Present a talk to a group of friends.

(wtc15) _____ % Talk in a small group of acquaintances.

(wtc16) _____ % *Talk with a garbage collector.

(wtc17) _____ % Talk in a large meeting of strangers.

(wtc18) _____ % *Talk with a spouse (or boyfriend/girlfriend).

(wtc19) _____ % Talk in a small group of friends.

(wtc20) _____ % Present a talk to a group of acquaintances.

APPENDIX I
PERCEIVED L2 COMPETENCE (MACINTYRE & CHAROS, 1996)

英語で人とコミュニケーションをとるときに、あなたがどう感じるかについて、12 の状況（シチュエーション）について質問します。それぞれの状況について、コミュニケーションに対しての**自信がどの程度あると感じるか**を 0%から 100%のパーセンテージで下線に書いてください。回答には、正しい答えも間違った答えもありません。すばやく第一印象を記していくように心がけてください。

（例）

- 知らない人と英語で話す時、全く自信がなければ、0 や 10 などの数字を書いてください。→ (__10___ 1. 知らない人と話すとき)
- ほどほどの自信があれば、40、50、60 を記してください。→ (__50___1. 知らない人と話すとき)
- ほとんどいつも自信があれば、90 や 100 を記入してください。
  → (__100___1. 知らない人と話すとき)

(pcomp1) _____% 知らない人たちの前での一団にスピーチ（プレゼンテーション）をするとき

(pcomp2) _____% 親しい友人の大きな集まり（会議）で発言するとき

(pcomp3) _____%知らない人たちの小グループで会話をするとき

(pcomp4) _____% 並んで待っているときに知り合いと会話するとき

(pcomp5) _____% 知り合いの大きな集まり（会議）で発言するとき

(pcomp6) _____% 親しい友人の一団にスピーチ（プレゼンテーション）をするとき

(pcomp7) _____% 知り合いの小グループで会話をするとき

(pcomp8) _____% 並んで待っているときに知らない人と会話するとき

(pcomp9) _____% 知らない人の大きな集まり（会議）で発言するとき

(pcomp10) _____% 並んで待っているときに親しい友人と会話をするとき

(pcomp11) _____% 親しい友人の小グループで会話するとき

(pcomp12) _____% 知り合いの一団にスピーチ（プレゼンテーション）をするとき

APPENDIX J
PERCEIVED L2 COMPETENCE (ENGLISH)
(MACINTYRE & CHAROS, 1996)

When speaking English in the following 12 situations, please indicate the extent to which you feel competent with your English. Your answer should be as a percentage from 0% to 100%. There are no correct or incorrect answers; please answer with your first impression.

(Example)

- You have very little competence when speaking with a stranger in English. → 0% or 10%
- You feel somewhat competent → 40、50、60
- You feel very competent → 90 or 100%

(pcomp1) _____% I would feel competent presenting a speech to a group of strangers.

(pcomp2) _____% I would feel competent talking in a large meeting of friends.

(pcomp3) _____% I would feel competent talking in a small group of strangers.

(pcomp4) _____% I would feel competent talking with an acquaintance while standing in line.

(pcomp5) _____% I would feel competent talking in a large meeting of acquaintances.

(pcomp6) _____% I would feel competent presenting a speech to a group of friends.

(pcomp7) _____% I would feel competent talking in a small group of acquaintances.

(pcomp8) _____% I would feel competent talking with a stranger while standing in line.

(pcomp9) _____% I would feel competent talking in a large meeting of strangers.

(pcomp10) _____% I would feel competent talking with a friend while standing in line.

(pcomp11) _____% I would feel competent talking in a small group of friends.

(pcomp12) _____% I would feel competent presenting a speech to a group of acquaintances.

APPENDIX K
COMMUNICATION ANXIETY IN ENGLISH
(MACINTYRE & CHAROS, 1996)

英語で人とコミュニケーションをとるときに、あなたがどう感じるかについて、12 の状況について答えていただきます。 それぞれの状況において、どれくらいのパーセンテージで**コミュニケーションについての不安があるか**、下線の上に書いてください。正しい答えも間違った答えもありません。すばやく第一印象を記していくのが一番よいやり方です。

（例）全く不安がないのであれば、0 や10 などの数字を書いてください。
ほどほどの不安があるのであれば、40、50、60 などを記してください。
ほとんどいつも不安があるのであれば、90 や100 など記してください。

0% -- 100%

英語を喋るのに 決して不安を感じない　　　英語を喋るのに いつも不安を感じる

英語で話す状況 （日本の国内や外国でおこりうる状況です。こういう状況を経験したことがなくても想像で回答して下さい。）

(canx1) _____ % 知らない人の一団にスピーチ（プレゼンテーション）をするとき

(canx2) _____ % 友人の大きな集まり（会議）で発言するとき

(canx3) _____ % 知らない人の小グループで会話をするとき

(canx4) _____% 何かを待つ列に並んでいるとき知り合いと会話するとき

(canx5) _____% 知り合いの大きな集まり（会議）で発言するとき

(canx6) _____ % 友人の一団にスピーチ（プレゼンテーション）をするとき

(canx7) _____ % 知り合いの小グループで会話をするとき

(canx8) _____ % 何かを待つ列に並んでいるとき知らない人と会話するとき

(canx9) _____ % 知らない人の大きな集まり（会議）で発言するとき

(canx10) _____ % 何かを待つ列に並んでいるとき友人と会話をするとき

(canx11) _____ % 友人の小グループで会話をするとき

(canx12) _____ % 知り合いの一団にスピーチ（プレゼンテーション）をするとき

APPENDIX L
L2 COMMUNICATION ANXIETY (ENGLISH)
(MACINTYRE & CHAROS, 1996)

Please indicate how much anxiety you feel when communicating in English in the following 12 situations. In the blank provided, please indicate your level of anxiety as a percentage. There are no correct or incorrect answers; please just indicate your first impression.

For example, if you feel little or no anxiety, an answer of 0% or 10% would be appropriate. Some anxiety would correspond to 40%, 50% , or 60%, while feeling very anxious would be 90% or 100%.

0% -- 100%
Not anxious   Very anxious

The following situations concern speaking English. These could be in Japan or abroad; if you have never had such experiences, please imagine what you would do.

(canx1) _____ % I would feel anxious presenting a speech to a group of strangers.

(canx2) _____ % I would feel anxious talking in a large meeting of friends.

(canx3) _____ % I would feel anxious talking in a small group of strangers.

(canx4) _____% I would feel anxious talking with an acquaintance while standing in line.

(canx5) _____% I would feel anxious talking in a large meeting of acquaintances.

(canx6) _____ % I would feel anxious presenting a speech to a group of friends.

(canx7) _____ % I would feel anxious talking in a small group of acquaintances.

(canx8) _____ % I would feel anxious talking with a stranger while standing in line.

(canx9) _____ % I would feel anxious talking in a large meeting of strangers.

(canx10) _____ % I would feel anxious talking with a friend while standing in line.

(canx11) _____ % I would feel anxious talking in a small group of friends.

(canx12) _____ % I would feel anxious presenting a speech to a group of acquaintances.

APPENDIX M
FOREIGN LANGUAGE CLASSROOM ANXIETY SURVEY
(HORWITZ ET AL., 1986)

あってはまらない			どちもない	あってはまる		
非常に	普通	少し		少し	普通	非常に
1	2	3	4	5	6	7

flcas1	英語の授業で話している時、いつも自信がない。	1−2−3−4−5−6−7
flcas2	英語の授業で、間違いをしても気にしない。	1−2−3−4−5−6−7
flcas3	英語の授業で教師に質問されると分かると緊張して震えてしまう。	1−2−3−4−5−6−7
flcas4	英語の授業で教師が何を言っているのか分からないと怖い。	1−2−3−4−5−6−7
flcas5	今よりもっと英語科目の授業をとっても大丈夫だ。	1−2−3−4−5−6−7
flcas6	英語の授業中、授業以外のことをよく考える。	1−2−3−4−5−6−7
flcas7	他の学生が自分より英語が上手だといつも思っている。	1−2−3−4−5−6−7
flcas8	英語の授業の試験中緊張しない。	1−2−3−4−5−6−7
flcas9	英語の授業で準備なしに、話させられたら、パニック状態になる。	1−2−3−4−5−6−7
flcas10	英語の授業を落としてしまうことに不安がある。	1−2−3−4−5−6−7
flcas11	他の人が英語の授業に対して、どうして不安になってしまうのが分からない。	1−2−3−4−5−6−7
flcas12	英語の授業で緊張して知っていることを忘れてしまう。	1−2−3−4−5−6−7
flcas13	英語の授業で率先して答えるのは恥ずかしくてできない。	1−2−3−4−5−6−7
flcas14	ネイティヴ・スピーカーと英語で話すことについては緊張しない。	1−2−3−4−5−6−7
flcas15	英語の授業では、教師が何の間違いを指摘しているのか分からないと不安に感じる。	1−2−3−4−5−6−7
flcas16	英語の授業のために十分に準備しても、不安になる。	1−2−3−4−5−6−7

flcas17	英語の授業に、出席したくないときがしばしばある。	1－2－3－4－5－6－7
flcas18	英語の授業で話す時には、自信を持っている。	1－2－3－4－5－6－7
flcas19	英語の教師が間違いをすべて直そうとするのではないかと不安だ。	1－2－3－4－5－6－7
flcas20	英語の授業で指されそうになったらドキドキする。	1－2－3－4－5－6－7
flcas21	英語の試験のために勉強をすればするほど混乱する。	1－2－3－4－5－6－7
flcas22	英語の授業の準備をそれほどしなくても、心配にならない。	1－2－3－4－5－6－7
flcas23	英語の授業で他の学生が私よりうまくしゃべるといつも思う。	1－2－3－4－5－6－7
flcas24	他の学生の前で、英語を話すと、とても緊張する。	1－2－3－4－5－6－7
flcas25	英語の授業のペースが速すぎて、取り残されることが心配だ。	1－2－3－4－5－6－7
flcas26	他の授業より、英語の授業の法方が緊張するし、神経質になる。	1－2－3－4－5－6－7
flcas27	英語の授業で話す時に、緊張して混乱する。	1－2－3－4－5－6－7
flcas28	英語の授業に行くとき、自信がありリラックスしていると感じる。	1－2－3－4－5－6－7
flcas29	教師の言っていることば一つ一つが分からないと心配になる。	1－2－3－4－5－6－7
flcas30	英語を話すのに学ばなければならない文法の数にまいっている。	1－2－3－4－5－6－7
flcas31	英語の授業中、英語で話したら、皆に笑われるのではないか心配だ。	1－2－3－4－5－6－7
flcas32	英語のネイティヴ・スピーカーと一緒にいるとしても、たぶん緊張しない	1－2－3－4－5－6－7
flcas33	答えを準備していない質問を教師にされると緊張する	1－2－3－4－5－6－7

APPENDIX N
FOREIGN LANGUAGE CLASSROOM ANXIETY SURVEY (ENGLISH)
(HORWITZ ET AL., 1986)

(flcas1). I never feel quite sure of myself when I am speaking in my English class.
(flcas2). I don't worry about making mistakes in my English class.
(flcas3). I tremble when I know that I'm going to be called on in my English class.
(flcas4). It frightens me when I don't understand what the teacher is saying in my English class.
(flcas5). It wouldn't bother me at all to take more English language classes.
(flcas6). During English class, I find myself thinking about things that have nothing to do with the course.
(flcas7). I keep thinking that the other students are better at English than I am.
(flcas8). I am usually at ease during tests in my English class.
(flcas9). I start to panic when I have to speak without preparation in English class.
(flcas10). I worry about the consequences of failing my English class.
(flcas11). I don't understand why some people get so upset over English class.
(flcas12). In English class, I can get so nervous I forget things I know.
(flcas13). It embarrasses me to volunteer answers in my English class.
(flcas14). I would not be nervous speaking English with native speakers.
(flcas15). I get upset when I don't understand what the teacher is correcting.
(flcas16). Even if I am well prepared for English class, I feel anxious about it.
(flcas17). I often feel like not going to English class.
(flcas18). I feel confident when I speak in my English class.
(flcas19). I am afraid that my English teacher is ready to correct every mistake I make.
(flcas20). I can feel my heart pounding when I'm going to be called on in English class.
(flcas21). The more I study for an English test, the more confused I get.
(flcas22). I don't feel pressure to prepare very well for my English class.
(flcas23). I always feel that the other students speak English better than I do.
(flcas24). I feel very self-conscious about speaking English in front of other students.
(flcas25). English class moves so quickly I worry about getting left behind.
(flcas26). I feel more tense and nervous in my English class than in my other classes.
(flcas27). I get nervous and confused when I am speaking in my English class.
(flcas28). When I'm on my way to English class, I feel very sure and relaxed.
(flcas29). I get nervous when I don't understand every word my English teacher says.
(flcas30). I feel overwhelmed by the number of rules you have to learn to speak English.
(flcas31). I am afraid that the other students will laugh at me when I speak English.
(flcas32). I would probably feel comfortable around native speakers of English.
(flcas33). I get nervous when my English teacher asks questions which I haven't prepared in advance.

APPENDIX O
MOTIVATION (YASHIMA, 2002)

あってはまらない			どちもない	あってはまる		
非常に	普通	少し		少し	普通	非常に
1	2	3	4	5	6	7

Desire to Learn English (Yashima, 2002)

mot1	英語の宿題はできるだけ早く取り組むほうだ	1－2－3－4－5－6－7
mot2	授業に関係がなくても英語で新聞や雑誌を読む	1－2－3－4－5－6－7
mot3	英語の勉強中は内容に興味をもって集中できる	1－2－3－4－5－6－7
mot4	できることなら学校の英語の時間を増やしてほしい	1－2－3－4－5－6－7
mot5	英語は学校で必ず教えられるべきである	1－2－3－4－5－6－7
mot6	他の科目に比べて英語は興味がもてる	1－2－3－4－5－6－7

Motivational Intensity (Yashima, 2002)

mot7	自分は同級生と比べてよく英語を勉強する	1－2－3－4－5－6－7
mot8	英語の授業で習ったことや英語についてよく考える	1－2－3－4－5－6－7
mot9	学校で教科として英語がなくても自分で学習したい	1－2－3－4－5－6－7
mot10	平均すると英語の勉強に時間をかける方である	1－2－3－4－5－6－7
mot11	私は英語を学習する気が十分にある	1－2－3－4－5－6－7
mot12	大学卒業後も英語を勉強したり、なんらかの形で英語力の向上に勉めたい	1－2－3－4－5－6－7

APPENDIX P
MOTIVATION (ENGLISH) (YASHIMA, 2002)

Desire to Learn English (Yashima, 2002)
(mot1) When I have assignments to do in English, I try to do them immediately.
(mot2) I would read English newspapers or magazines outside my English course work.
(mot3) During English classes I'm absorbed in what is taught and concentrate on my studies.
(mot4) I would like the number of English classes school increased.
(mot5) I believe absolutely English should be taught at school.
(mot6) I find studying English more interesting than other subjects.

Motivational Intensity (Yashima, 2002)
(mot7) Compared to my classmates, I think I study English relatively hard.
(mot8) I often think about the words and ideas that I learn about in my English classes.
(mot9) If English were not taught at school, I would study it on my own.
(mot10) I think I spend fairly long hours studying English.
(mot11) I really try to learn English.
(mot12) After I graduate from college, I will continue to study English and try to improve.

APPENDIX Q
INTERNATIONAL POSTURE (YASHIMA, 2002)

INTERNATIONAL POSTURE: INTERGROUP APPROACH-AVOIDANCE TENDENCY SUBSCALE

あってはまらない			どちもない	あってはまる		
非常に	普通	少し		少し	普通	非常に
1	2	3	4	5	6	7

Intergroup Approach-Avoidance Tendency 以下の質問に、あなたにどの程度当てはまりますか。これまでと同じように自分にあてはまるところに〇をしてください。

ip1	日本に来ている留学生など外国人と（もっと）友達になりたい。	1−2−3−4−5−6−7
ip2	外国人と話すのを避けられれば避ける方だ。	1−2−3−4−5−6−7
ip3	日本の学校で留学生がいれば気軽に声かけようと思う。	1−2−3−4−5−6−7
ip4	留学生や外国人の学生と寮やアパートなどでルームメートになってもよいと思う。	1−2−3−4−5−6−7
ip5	地域の外国人を世話するようなボランティア活動に参加してみたい。	1−2−3−4−5−6−7
ip6	もし、隣に外国人が越してきたら困ると思う。	1−2−3−4−5−6−7
ip7	レストランや駅で言葉が通じず困っている外国人がいれば進んで助けると思う。	1−2−3−4−5−6−7

INTERNATIONAL POSTURE INTEREST IN INTERNATIONAL VOCATION / ACTIVITIES SUBSCALE

ip8	故郷にずっとすみ続けたい。	1－2－3－4－5－6－7
ip9	日本以外の国に住んでみたい。	1－2－3－4－5－6－7
ip10	国連など国際機関で働いてみたい。	1－2－3－4－5－6－7
ip11	青年海外協力隊など参加するなど、途上国でのボランティア活動に興味がある。	1－2－3－4－5－6－7
ip12	海外の出来事は私の日常生活にあまり関係ないと思う。	1－2－3－4－5－6－7
ip13	海外出張の多い仕事は避けたい。	1－2－3－4－5－6－7

INTERNATIONAL POSTURE: INTEREST IN FOREIGN AFFAIRS SUBSCALE

ip14	外国に関するニュースをよく見たり、読んだりする。	1－2－3－4－5－6－7
ip15	外国の情勢や出来事について家族や友人とよく話し合うほうだ。	1－2－3－4－5－6－7
ip16	(Elwood) 外国に関するニュースのほうが地域のニュースより大切だ。	1－2－3－4－5－6－7
ip17	(Elwood) 外国に関するニュースを授業中での使用するのは、面白くて、効果的な教材になる。	1－2－3－4－5－6－7
ip18	(Elwood) 外国に関するニュースがは理解が難しい。	1－2－3－4－5－6－7

INTERNATIONAL POSTURE:
INTERCULTURAL FRIENDSHIP ORIENTATION SUBSCALE

Intercultural Friendship Orientation あなたにとって英語の学習にはどのような意味がありますか。次に示す理由がどの程度自分に当てはまるか、最もあなたの気持ちをよく表すところに○をしてください。

英語を勉強する理由として、英語が。。。

ip19	より多くの多様な人々と会って話ができるから。	1－2－3－4－5－6－7
ip20	将来就職に役立つと思うので。	1－2－3－4－5－6－7
ip21	インターネットを使うために必要だから。	1－2－3－4－5－6－7
ip22	英語が話せるといろんな文化を知り、文化背景の異なる人々と知り合えるから。	1－2－3－4－5－6－7
ip23	英検などの試験に挑戦するから。	1－2－3－4－5－6－7
ip24	英語が話せると異文化の人々の活動に自由に参加できるから。	1－2－3－4－5－6－7
ip25	英語で情報や知識を得ることができるから。	1－2－3－4－5－6－7
ip26	英語は自分が将来つきたい仕事に必要だから。	1－2－3－4－5－6－7

APPENDIX R
INTERNATIONAL POSTURE (ENGLISH) (YASHIMA, 2002)

INTERNATIONAL POSTURE: INTERGROUP APPROACH-AVOIDANCE TENDENCY SUBSCALE

(ip1). I want to make friends with international students studying in Japan.

(ip2). I try to avoid talking with foreigners if I can.

(ip3). I would talk to an international student if there were one at school.

(ip4). I wouldn't mind sharing an apartment or room with an international student.

(ip5). I want to participate in a volunteer activity to help foreigners living in the surrounding community.

(ip6). I would feel somewhat uncomfortable if a foreigner moved in next door.

(ip7). I would help a foreigner having trouble communicating in a restaurant or at a station.

INTERNATIONAL POSTURE: INTEREST IN INTERNATIONAL VOCATION / ACTIVITIES SUBSCALE

(ip8). I would rather stay in my hometown.

(ip9). I want to live in a foreign country.

(ip10). I want to work in an international organization such as the United Nations.

(ip11). I'm interested in volunteer activities in developing countries such as participating in Youth International Development Assistance.

(ip12). I don't think what's happening overseas has much to do with my daily life.

(ip13). I'd rather avoid the kind of work that sends me overseas frequently.

INTERNATIONAL POSTURE:
INTEREST IN FOREIGN AFFAIRS SUBSCALE

(ip14). I often read and watch news about foreign countries.

(ip15). I often talk about situations and events in foreign countries with my family and/or friends.

(ip16). *(Elwood)* International news is more important than local news.

(ip17). *(Elwood)* International news makes interesting, useful content for school classes.

(ip18). *(Elwood)* International news is too difficult to understand.

INTERNATIONAL POSTURE:
INTERCULTURAL FRIENDSHIP ORIENTATION SUBSCALE

As a reason to study English:

(ip19). It will allow me to meet and converse with more and varied people.

(ip20). It will help me get a job in the future.

(ip21). It's necessary for using the Internet.

(ip22). It will allow me to get to know various cultures and peoples.

(ip23). It will help on such tests as the Eiken.

(ip24). I will be able to participate more freely in the activities of other cultural groups.

(ip25). I want to be able to get information and knowledge in English.

(ip26). It's necessary for me to find a job I want in the future.

APPENDIX S
PERCEIVED DISTANCE QUESTIONNAIRE

以下の５項目に示した英語を使う状況で、実際に英語を話す時に日本語で話す時の自分自身の性格とは違う性格に変わることありますか？それぞれの状況において、どれぐらいのパーセンテージで自分自身が変わるか記入してください。

0% == 100%
性格が変わらない 性格が完全に変わる
（普通の自分のまま） （他人になる）

例: __70%__ 1. 英語で英語の先生と話す（70%ぐらい性格が変わる）

例: _______ % クラスメートと日本語でしゃべる

(dis1) _____ % 1.クラスメートと英語でしゃべる

(dis2) _____ % 2.ロールプレーを英語でする

(dis3) _____ % 3.スピーチを英語でする

(dis4) _____ % 4.劇を英語でする

(dis5) _____ % 5.手人形を使いながら、英語で話す

APPENDIX T
PERCEIVED DISTANCE QUESTIONNAIRE (ENGLISH)

Indicate the "distance" you feel from your normal self while doing [conversation, public speaking, roleplay, drama, puppetry] in English."

When speaking English in the five situations listed below, do you feel your character changes from your character when speaking Japanese? For each activity, please indicate as a percentage the extent to which your character changes (to which you feel distance)。

0% == 100%
I don't change. I change completely.
(I'm the same as always.) (I become a different person)

Example: __70％__ 1. When speaking English with my English teacher (my character changes about 70%)

Example: _______ % When speaking Japanese with a classmate.

(dis1) _____ % 1. When speaking English with a classmate.

(dis2) _____ % 2. When doing a roleplay.

(dis3) _____ % 3. When making a speech in English.

(dis4) _____ % 4. When performing in a drama.

(dis5) _____ % 5. When using puppets.

APPENDIX U
PERCEIVED DISTANCE EXPLANATION AND SCRIPTS

1. Japanese (L1) conversation: ふたりで母国語でおしゃべりして下さい。課題は自由ですが、必ず日本語で話して下さい。

2. English (L2) conversation: ふたりで英語でおしゃべりして下さい。課題は自由ですが、必ず英語で話して下さい。

3. English (L2) public speaking: Next, you will do an excerpt from Martin Luther King, Jr.'s speech, "I Have a Dream."

    And so even though we face the difficulties of today and tomorrow, I still have a dream. It is a dream deeply rooted in the American dream.

    I have a dream that one day this nation will rise up and live out the true meaning of its creed: "We hold these truths to be self-evident, that all men are created equal."

    I have a dream that one day on the red hills of Georgia, the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood.

    I have a dream that one day even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into an oasis of freedom and justice.

    I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character.

4. L2 roleplay: 二人で、ロールプレーして下さい。一人は寝ようとしているけれども、隣にすんでいる方の犬が長く吠えてて、あなたが寝られません。もう一人は犬の持ち主です。こういう問題を解決してみて下さい。

5. L2 drama: You and your partner will perform the following excerpt from Shakespeare's "Romeo and Juliet."

    (JULIET appears above at a window)

    JULIET
    Oh, my Romeo, why is our situation like this?
    Please, my Romeo, change your name and deny your family.
    Or if you cannot, then I will change mine…

ROMEO
Just look at how she leans on her hand
If I could be a glove on that hand
So I could touch her cheek!

ROMEO
(quietly) Should I say something, or should I listen more? Say you love me,

JULIET
A name is only a name, but the Montague name is my enemy,
Indeed, what is in a name?
A rose by any other name would still smell sweet
And my Romeo would likewise be sweet…

ROMEO
I believe you, Juliet:
Say you love me, and I'll change my name—
I will never be Romeo again

JULIET
What? Who is listening?
Who is there, hiding and listening?

ROMEO
I do not want to say my name, dear Juliet,
because my name is the name of your enemy.
I would change my name for you…

JULIET
I have only heard a few words, but I know you—
Aren't you my dear Romeo?

ROMEO
Yes, I am, but if you don't like my name,
I will change it, dear Juliet.

JULIET
Why are you here?
The walls are very high and hard to climb,
And my family would kill you if they found you here.

ROMEO
Because I love you, I could fly over these walls!
Love can overcome anything, so your family doesn't scare me.

6. L2 puppetry: You and three partners will perform the following excerpt from the "Bremen Town Musicians," a German folktale. Each person will be one of the animals (a donkey, a dog, a cat, and a rooster).

A certain old donkey was walking down a road, talking to himself…
DONKEY: There in Bremen, I can surely be a town-musician. *(he sees a dog, lying on the road)* What's wrong, sir?

DOG: Woof. I am old, and every day I become weaker. I can no longer hunt, so my master wanted to kill me. I ran away; but now I have nothing to eat!

DONKEY: You know, I am going to Bremen to be a town-musician there; go with me and be a musician, too. I will play the lute, and you play the drum.

DOG: What a great idea! I have large ears and big paws, so perhaps I'll be a good musician! Let's go!

(soon they see a cat, sitting on the path, with a very sad face)

DONKEY: Now then, Miss Cat, what's wrong?

CAT: Who can be happy when his life is in danger? Because I am now getting old, my teeth are not sharp and I like to sit by the fire rather than chase mice. My lady wants a new cat, so I ran away. But now I have no friends. Where can I go?

DOG: Go with us to Bremen. You understand music, so you can be a musician, too.

CAT: I'd love to, but I have no talent as a musician. I'll do my best, however.

(soon they come to a farmyard, where the rooster is sitting on the gate, crowing with all his might…)

CAT: Your crowing is a lonely sound. What's the matter?

ROOSTER: I have been foretelling fine weather, because today Our Lady washes shirts, and she wants to dry them, but guests are coming Sunday, so I will become chicken soup. Off with my head, so while I still have it, I am crowing for all I'm worth.

DOG: Ah, you had better come away with us. We are going to Bremen. You have a good voice, and if we make music together it will be wonderful!

ROOSTER: Good idea—certainly better than dying.

(much later, after walking for a long, long time…)

DONKEY: Hey, I'm tired. Let's sleep here, under this tree!

DOG: Good idea—woof! I'll join you here.

CAT: No, thanks. I'll be up in the tree on a branch—much cozier up there.

ROOSTER: Bye, all. I'll be up on top of the tree since I can fly up there. G'night.

ROOSTER: Yo, I see a light, perhaps a house!"

DONKEY: If so, we should go on, for this isn't a four-star hotel.

DOG: Woof. Good idea! A few bones with some meat on them would be delicious!

The four arrive at the little house and DONKEY looks in the window…

ROOSTER: What do you see?

DONKEY: What do I see? I see a table covered with good things to eat and drink, and robbers sitting at it enjoying themselves.

CAT: Ah, that sounds delicious—for us!

A short discussion follows…and they decide…

DONKEY: I'll stand by the window.

DOG: And I'll climb on the donkey's back.

CAT: And I'll climb up on the dog.

ROOSTER: And I'll fly up and sit on the cat's head.

EVERYONE: Then we'll all make music together—really loud music!

The four musicians begin making their very loud music and burst through the window into the room, so that the glass shattered! At this horrible noise, the robbers jumped up, thinking that a ghost had come in, and ran away into the forest.

EVERYONE: munch, munch, chew, chew, eat, eat, etc.

EVERYONE: Gosh, I'm really full. Yawn.

EVERYONE: Zzzzzzzz… (*snoring*)

APPENDIX V
PERCEIVED DISTANCE EXPLANATION (ENGLISH)

1. Japanese (L1) conversation: You and your partner are chatting in your first language, Japanese. You may speak about any topic, but please do so only in Japanese.
2. English (L2) conversation: You and your partner are chatting in your second language, English. You may talk about any subject, but please speak only in English.
3. English (L2) public speaking: Next, you will do an excerpt from Martin Luther King, Jr.'s speech, "I Have a Dream."
4. L2 roleplay: You and your partner should roleplay the following situation in English. One of you is trying to sleep, but your partner's dog has been barking and is keeping you from sleeping.
5. L2 drama: You and your partner will perform the following excerpt from Shakespeare's "Romeo and Juliet."
6. L2 puppetry: You and three partners will perform the following excerpt from the "Bremen Town Musicians," a German folktale. Each person will be one of the animals (a donkey, a dog, a cat, and a rooster).

APPENDIX W
PERSONALITY BIPOLAR SCALE OF GLOBAL PERSONALITY TRAITS (GOLDBERG, 1992)

今現在の自分自身の性格（将来なりたい自分ではなく）を示して下さい。同年代の友人知人と比較し、普段の自分を正確に記して下さい。それぞれの項目に、1から7までの尺度から自分の性格に当てはまる数字を選び、丸をつけて下さい。

あなたはどんな人ですか？

例：内向的 〜 外交的

大変内向的			どちらでもない			大変外交的
１	２	３	４	５	６	７

あなたはどんな人ですか？

(per1)	内向的	1−2−3−4−5−6−7	外向的
(per2)	精力的な	1−2−3−4−5−6−7	消極的
(per3)	静か	1−2−3−4−5−6−7	おしゃべり
(per4)	大胆	1−2−3−4−5−6−7	臆病
(per5)	保守的な	1−2−3−4−5−6−7	活発
(per6)	控えめ	1−2−3−4−5−6−7	自己主張が強い
(per7)	非冒険的・保守的な	1−2−3−4−5−6−7	冒険的
(per8)	心が冷たい	1−2−3−4−5−6−7	心が温かい・優しい
(per9)	親切	1−2−3−4−5−6−7	不親切
(per10)	非協力的	1−2−3−4−5−6−7	協力的
(per11)	利己的でない？	1−2−3−4−5−6−7	利己的
(per12)	気むずかしい	1−2−3−4−5−6−7	感じの良い
(per13)	信じやすい	1−2−3−4−5−6−7	人を信じない
(per14)	けち	1−2−3−4−5−6−7	気前のよい
(per15)	手際が悪い	1−2−3−4−5−6−7	手際が良い
(per16)	責任感がない	1−2−3−4−5−6−7	責任感がある
(per17)	誠実	1−2−3−4−5−6−7	なげやり
(per18)	実用的な	1−2−3−4−5−6−7	融通が利がない
(per19)	そそっかしい	1−2−3−4−5−6−7	注意深い
(per20)	勤勉	1−2−3−4−5−6−7	怠け者
(per21)	贅沢な	1−2−3−4−5−6−7	質素な・堅実な
(per22)	穏やかな	1−2−3−4−5−6−7	怒りっぽい
(per23)	堅苦しい	1−2−3−4−5−6−7	のんきな
(per24)	のんびりしている	1−2−3−4−5−6−7	神経質
(per25)	人を羨む	1−2−3−4−5−6−7	人を羨まない
(per26)	気分が安定してい	1−2−3−4−5−6−7	気分が不安定

	る		
(per27)	不満がある	1－2－3－4－5－6－7	満足している
(per28)	感情的でない	1－2－3－4－5－6－7	感情的である
(per29)	知的でない	1－2－3－4－5－6－7	知的である
(per30)	論理的である	1－2－3－4－5－6－7	論理的でない
(per31)	思慮深い	1－2－3－4－5－6－7	思慮が浅い
(per32)	好奇心が弱い	1－2－3－4－5－6－7	好奇心がある
(per33)	想像力が乏しい	1－2－3－4－5－6－7	想像力に富む
(per34)	創造性がない	1－2－3－4－5－6－7	創造的である
(per35)	洗練されている	1－2－3－4－5－6－7	野暮ったい

APPENDIX X
PERSONALITY BIPOLAR SCALE OF GLOBAL PERSONALITY TRAITS (ENGLISH) (GOLDBERG, 1992)

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly the same age. After each trait, please write a number indicating how accurately that trait describes you, using the following rating scale.

What kind of person are you?

Example: introverted – extroverted

Very introverted			Neither			Very extroverted
1	2	3	4	5	6	7

(per1)	introverted	1 − 2 − 3 − 4 − 5 − 6 − 7	extroverted
(per2)	energetic	1 − 2 − 3 − 4 − 5 − 6 − 7	low-key
(per3)	quiet	1 − 2 − 3 − 4 − 5 − 6 − 7	talkative
(per4)	daring, bold	1 − 2 − 3 − 4 − 5 − 6 − 7	timid
(per5)	conservative	1 − 2 − 3 − 4 − 5 − 6 − 7	spunky
(per6)	passive	1 − 2 − 3 − 4 − 5 − 6 − 7	assertive
(per7)	not adventurous	1 − 2 − 3 − 4 − 5 − 6 − 7	adventurous
(per8)	aloof (cold-hearted)	1 − 2 − 3 − 4 − 5 − 6 − 7	warm-hearted
(per9)	kind	1 − 2 − 3 − 4 − 5 − 6 − 7	not kind
(per10)	not cooperative	1 − 2 − 3 − 4 − 5 − 6 − 7	cooperative
(per11)	not selfish	1 − 2 − 3 − 4 − 5 − 6 − 7	selfish
(per12)	difficult, unpleasant	1 − 2 − 3 − 4 − 5 − 6 − 7	pleasant, agreeable
(per13)	trustful	1 − 2 − 3 − 4 − 5 − 6 − 7	not trustful
(per14)	stingy	1 − 2 − 3 − 4 − 5 − 6 − 7	generous
(per15)	coarse, lack finesse	1 − 2 − 3 − 4 − 5 − 6 − 7	refined
(per16)	lack responsible	1 − 2 − 3 − 4 − 5 − 6 − 7	responsible
(per17)	conscientious	1 − 2 − 3 − 4 − 5 − 6 − 7	deceitful
(per18)	practical	1 − 2 − 3 − 4 − 5 − 6 − 7	impractical
(per19)	careless	1 − 2 − 3 − 4 − 5 − 6 − 7	thorough
(per20)	diligent	1 − 2 − 3 − 4 − 5 − 6 − 7	lazy
(per21)	extravagant	1 − 2 − 3 − 4 − 5 − 6 − 7	simple, solid
(per22)	calm	1 − 2 − 3 − 4 − 5 − 6 − 7	easily angered
(per23)	formal, prim	1 − 2 − 3 − 4 − 5 − 6 − 7	easygoing
(per24)	at ease	1 − 2 − 3 − 4 − 5 − 6 − 7	on edge, not at ease
(per25)	envious	1 − 2 − 3 − 4 − 5 − 6 − 7	not envious
(per26)	stable	1 − 2 − 3 − 4 − 5 − 6 − 7	not stable
(per27)	not contented	1 − 2 − 3 − 4 − 5 − 6 − 7	contented

(per28)	not emotional	1 − 2 − 3 − 4 − 5 − 6 − 7	emotional
(per29)	dull, not intelligent	1 − 2 − 3 − 4 − 5 − 6 − 7	intelligent
(per30)	analytical	1 − 2 − 3 − 4 − 5 − 6 − 7	non-analytical
(per31)	reflective	1 − 2 − 3 − 4 − 5 − 6 − 7	shallow
(per32)	not curious	1 − 2 − 3 − 4 − 5 − 6 − 7	curious
(per33)	not imaginative	1 − 2 − 3 − 4 − 5 − 6 − 7	imaginative
(per34)	not creative	1 − 2 − 3 − 4 − 5 − 6 − 7	creative
(per35)	sophisticated	1 − 2 − 3 − 4 − 5 − 6 − 7	coarse, unsophisticated

APPENDIX Y
EGO PERMEABILITY HARTMANN BOUNDARY QUESTIONNAIRE, SHORTENED FORM (BQ-SH) RAWLINGS, 2001-2002

Unusual Experiences (UE) 不思議な経験

ego1 白昼夢の中で、複数の人間が別の一人の人間に変身したり、一人の人が別の人に変身したりする。
ego2 夢から起きたら、直ちに次の夢を見始める。夢から覚めてはまた別の夢を見る。
ego3 白昼夢を見る。
ego4 夢の中で、人々は互いになりかわったり、別の人間になったりする。
ego5 私の身体や別の誰かの体が刺されたり、傷つけられたり、引き裂かれたりする夢、白昼夢、悪夢を見る。
ego6 私の周りでは物の大きさや形が変わるような気がする。
ego7 何か怖いことが起こるたびに、その怖かったことに関係したことが、悪夢や幻想や記憶の再現(フラッシュバック）の中に現れる。
ego8 たとえば、色がにおったり、音が見えたり、匂いが聞こえたりするような、異なった感覚が一緒に起こるような経験をすることがよくある。
ego9 私の夢はとても鮮明なので、目が覚めているときに起こった事実とあとになっても区別がつかない。
ego10 自分の体の大きさや形が時々変わるような気がする。
ego11 誰かが自分の名前を呼んだり言ったりして、それが本当に起こったことなのか自分の空想なのか区別がつかないという経験がある。
ego12 自分が何かを想像しているだけなのか、それとも本当にそのことが起こっているのか分からないという経験をしたことがある。

Need for Order (NFO)秩序の必要性

ego13 すべての物事には定位置があり、すべてはその定位置にあるべきだ。
ego14 子供には厳しいしつけが必要だと思う。
ego15 組織では、全員が自分の定まった地位と役割を持つべきである。
ego16 男性は男性であり、女性は女性である。その区別を保つことはとても大切だ。
ego17 私は、きちんと起承転結のある話が好きだ。
ego18 私には、他民族の人と一緒に住んだり、結婚することは想像できない。
ego19 私は物事にはっきりけじめのあることを好む。
ego20 一番好きなテレビ番組と映画は、善人と悪人がいて、それが誰なのかいつもはっきりしているものだ。
ego21 写真や絵には、立派でしっかりした額がとても大切だ。
ego22 きちんとして清潔な身なりをすることはとても大切だ。
ego23 部屋が壁でしっかり区切られ、それぞれの部屋が定まった用途を持った家が好きだ。
ego24 東は東、西は西。その二つは決して交わることはない。

Trust (TR)信頼

ego25　私はとても率直な人である。
ego26　人をすぐ信頼する。
ego27　私はいつも少し気が張っている。
ego28　私は、人に会うとすぐその人を完全に信頼するので、初対面でも自分のことを何でも教えることができる。
ego29　私は他人にある程度の距離を保ってもらいたい。
ego30　私は、本当に気心が知れるまでは他人に言うことに気を使う。

Perceived Competence (PC)自分で見た自分の能力

ego31　いつも時間を守る。
ego32　自分の机や作業台をきちんと整頓しておく。
ego33　私は小遣い帳をつけたり(口座を管理したり)、自分のお金の記録をつけるのが得意だ。
ego34　私ははっきりとした時間の感覚がある。
ego35　街のどこが安全でどこが危険な箇所なのかをはっきりと知っている。
ego36　自分の過去についてはっきりとした記憶がある。何年に何が起こったかよく言える。
ego37　私は現実的で、馬鹿な考えを持たないタイプの人間である。
ego38　自分はよい心理療法士になれると思う。
ego39　人間には、正常な人、問題を抱えた人、精神に異常をきたした人などというはっきりした区別はない。

Childlikeness (Ch) 子供らしさ

ego40　よい教師には、ある程度、子供の部分が残っていないといけないと思う。
ego41　いい親には少しは子供の部分がなくてはいけない。
ego42　芸術家には、ある程度、子供の部分が残っていないといけないと思う。
ego43　よい教師は、子供がその子らしくいられるよう手助けすべきである。
ego44　子供と大人には共通点がたくさんある。大人の役割、子供の役割というはっきりとした役割を離れて一緒にいる機会を互いに作るべきである。

Sensitivity (Se) 感性

ego45　私は傷つきやすい。
ego46　私はとても繊細な人間である。

APPENDIX Z
EGO PERMEABILITY HARTMANN BOUNDARY QUESTIONNAIRE, SHORTENED FORM (BQ-SH) (ENGLISH) (RAWLINGS, 2001-2002)

Unusual Experiences (UE)

ego1 In my daydreams, people kind of merge into one another or one person turns into another.
ego2 I wake from one dream into another.
ego3 I have daydreams.
ego4 In my dreams, people sometimes merge into each other or become other people.
ego5 I have dreams, daydreams, nightmares in which my body or someone else's body is being stabbed, injured, or torn apart.
ego6 Things around me seem to change their size and shape.
ego7 Every time something frightening happens to me, I have nightmares or fantasies or flashbacks involving the frightening event.
ego8 I have often had the experience of different senses coming together. For example, I have felt that I could smell a color, or see a sound, or hear an odor.
ego9 My dreams are so vivid that even later I can't tell them from waking reality.
ego10 My body sometimes seems to change its size and shape.
ego11 I have had the experience of someone calling me or speaking my name and not being sure whether it was really happening or I was imagining it.
ego12 I have had the experience of not knowing whether I was imagining something or it was actually happening.

Need for Order (NFO)

ego13 There is a place for everything and everything should be in its place.
ego14 I think children need strict discipline.
ego15 In an organization, everyone should have a definite place and a specific role.
ego16 A man is a man and a woman is a woman; it is very important to maintain that distinction.
ego17 I like stories that have a definite beginning, middle, and end.
ego18 I cannot imagine living with or marrying a person of another race.
ego19 I like clear, precise borders.
ego20 The movies and TV shows I like the best are the ones where there are good guys and bad guys and you always know who they are.
ego21 Good solid frames are very important for a picture or a painting.
ego22 Being dressed neatly and cleanly is very important.
ego23 I like houses where rooms have definite walls and each room has a definite function.
ego24 East is East and West is West, and never the twain shall meet. (Kipling)

Trust (TR)

ego25 I am a very open person.
ego26 I trust people easily.
ego27 I am always at least a bit on my guard.
ego28 Sometimes I meet someone and trust him or her so completely that I can share just about everything about myself at the first meeting.
ego29 I expect other people to keep a certain distance.
ego30 I am careful about what I say to people until I get to know them really well.

Perceived Competence (PC)

ego31 I get to appointments right on time.
ego32 I keep my desk and worktable neat and well organized.
ego33 I am good at keeping accounts and keeping track of my money.
ego34 I have a clear and distinct sense of time.
ego35 I know exactly what parts of town are safe and what parts are unsafe.
ego36 I have a clear memory of my past. I could tell you pretty well what happened year by year.
ego37 I am a down-to-earth, no-nonsense kind of person.
ego38 I think I would be a good psychotherapist.
ego39 There are no sharp dividing lines between normal people, people with problems, and people who are considered psychotic or crazy.

Childlikeness (Ch)

ego40 I think a good teacher must remain in part a child.
ego41 A good parent has to be a bit of a child, too.
ego42 I think an artist must in part remain a child.
ego43 A good teacher needs to help a child remain special.
ego44 Children and adults have a lot in common. They should give themselves a chance to be together without any strict roles.

Sensitivity (Se)

ego45 I am easily hurt.
ego46 I am a very sensitive person.

Note. Items ego27, ego29, ego30, and ego39 are reverse-scored to produce subscale totals. For calculation of the BQ-Sh total score, subscale scores on NFO and PC are first reversed by subtracting sums from ego13 and 36, respectively. TR (Trust subscale) is not included in total score. Total score for BQ-Sh is then UE + NFO + PC + Ch + Se.

APPENDIX AA
RESEARCHER'S INTRODUCTION SCRIPT

Thank you for your participation in my research project. My name is Jim Elwood, and I am a professor at Tsukuba University. Your teacher, Mr. _____, has kindly consented to allowing me to visit and conduct this research, which is for my dissertation at Temple University Japan.

Please allow me to explain my research. This project is investigating the role of distancing in second language acquisition. Basically, I am asking this question: when you speak English, are you 100% the same person as when you speak Japanese? To more fully understand the relationships among distancing and affective variables, you are being asked to fill out numerous questionnaires and participate in a series of classroom activities involving oral communication: chatting, public speaking, roleplay, drama, and puppetry.

This research is strictly voluntary, and there is no relationship with your course grade (although you will receive bonus points for helping). You may, of course, choose to not participate. There is no physical risk involved; all responses will remain confidential and all participants will be anonymous.

If you have questions, please feel free to ask at any time.

APPENDIX AB
EXPLANATION AND CONSENT FORM

Thank you for your participation in my research project. My name is Jim Elwood, and this project is investigating the role of distancing in second language acquisition. Basically, I am asking this question: when you speak English, are you 100% the same person as when you speak Japanese? To more fully understand the relationships among distancing and affective variables, you are being asked to fill out numerous questionnaires and participate in a series of classroom activities involving oral communication: chatting, public speaking, roleplay, drama, and puppetry. Your next class with Mr./Ms. ____________ will be conducted by the researcher, and you will be asked to finish the questionnaire at home after the class and return it the following week to Mr./Ms. ____________.

This research is strictly voluntary, and there is no relationship with your course grade (although you will receive bonus points for helping). You may, of course, choose to not participate. There is no physical risk involved; all responses will remain confidential, and all participants will be anonymous.

I, ______________________, hereby agree to participate in Mr. Elwood's research.

Date ____________________________

Participant __________________________________

Researcher __________________________________

APPENDIX AC
CORRELATION AND COVARIANCE MATRICES FOR INSTRUMENTS FOR MACINTYRE AND CHAROS (1996) MODELS

Table

Correlation and Covariance Matrix for MacIntyre and Charos (1996) Models

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 openness		6.473	6.883	2.485	4.254	11.117	-2.875	6.177	7.173	6.552	8.852	9.165	.972
2 extroversion	.411		9.465	5.502	2.026	13.420	-6.923	6.033	10.396	8.375	7.685	14.651	1.683
3 agreeableness	.274	.389		6.977	4.857	6.150	-7.238	9.849	14.014	6.420	15.479	14.332	1.621
4 emot stability	.132	.278	.221		.360	12.071	-6.336	1.356	7.093	7.397	6.214	8.806	.139
5 diligence	.285	.130	.195	.019		5.669	-3.064	3.047	5.661	5.896	11.890	7.550	-.606
6 perc comp	.272	.313	.090	.235	.140		-20.861	12.930	22.887	42.774	26.018	32.204	4.303
7 FLCAS	-.167	-.382	-.250	-.292	-.179	-.443		-8.649	-20.381	-17.898	-17.130	-19.894	-2.912
8 integrative	.242	.226	.231	.042	.121	.187	-.295		20.983	14.258	30.778	25.355	3.123
9 attitudes	.261	.361	.305	.205	.208	.306	-.645	.450		21.570	36.074	32.697	4.225
10 WTC	.195	.237	.114	.175	.177	.467	-.463	.250	.350		21.144	36.584	2.859
11 motivation	.266	.221	.278	.149	.361	.288	-.449	.546	.594	.284		38.962	4.033
12 frequency	.272	.414	.254	.208	.226	.351	-.513	.443	.530	.458	.522		5.170
13 English exp	.115	.189	.114	.013	-.072	.187	-.299	.217	.272	.150	.215	.272	
Mean	52.375	50.896	50.355	47.772	52.684	47.892	51.921	52.722	51.163	49.069	54.055	53.662	2.810
SD	3.884	4.072	6.503	4.880	3.854	10.561	4.464	6.588	7.104	8.701	8.588	8.720	2.193

Note. Correlations, means, and standard deviations are shown below the diagonal, and covariances are shown above the diagonal.

APPENDIX AD
CORRELATION AND COVARIANCE MATRICES FOR INSTRUMENTS FOR THE YASHIMA (2002) MODELS AND THE YASHIMA ET AL. (2004) MODELS

Table

Correlation and Covariance Matrix for Yashima (2002) models and Yashima et al. (2004) models

	1	2	3	4	5	6	7	8	9	10
1 listening		9.734	10.785	-6.079	1.777	6.103	12.724	7.955	6.264	11.604
2 vocabulary	.542		9.184	-6.336	1.415	7.125	12.920	10.255	8.422	8.727
3 perceived comp	.233	.212		-20.962	13.474	42.945	26.122	21.563	22.926	32.332
4 anxiety	-.304	-.339	-.435		-6.830	-17.995	-16.264	-14.278	-14.221	-19.418
5 extroversion	.100	.085	.313	-.367		8.408	7.715	10.603	11.430	14.710
6 WTC	.160	.200	.467	-.453	.237		22.228	22.332	21.520	34.722
7 motivation	.338	-.368	.288	-.415	.221	.284		49.326	35.911	39.117
8 cultural friendship	.201	.277	.226	-.346	.288	.284	.635		39.015	37.927
9 approach-avoid	.217	.313	.330	-.474	.427	.376	.636	.657		31.756
10 frequency	.304	.245	.351	-.488	.414	.458	.522	.481	.554	
Mean	49.723	45.828	47.892	51.909	50.896	49.069	54.055	51.965	50.555	53.662
SD	4.384	4.094	10.541	4.565	4.072	8.701	8.588	9.042	6.572	8.720

Note. Correlations, means, and standard deviations are shown below the diagonal, and covariances are shown above the diagonal.

APPENDIX AE
STRUCTURAL EQUATIONS WITH STANDARD ERRORS FOR THE L2 COMMUNICATIVE CONFIDENCE MEASUREMENT MODEL

An asterisk indicates the variable's path coefficient was set to unity to define the factor's scale.

Equations	*SE*	R^2
Factor 1: Perceived Competence		
SPCHSTRN = V1 = .543 F1 + .840 E1	*	.295
MTNGFRND = V2 = .812*F1 + .584 E2	.174	.659
GRPSTRNG = V3 = .704*F1 + .710 E3	.125	.496
LINEACQ = V4 = .775*F1 + .632 E4	.194	.601
MTNGACQ = V5 = .845*F1 + .535 E5	.171	.714
SPCHFRND = V6 = .837*F1 + .548 E6	.201	.700
GRPACQ = V7 = .801*F1 + .599 E7	.187	.642
LINESTRN = V8 = .633*F1 + .774 E8	.141	.401
MTNGSTRN = V9 = .555*F1 + .832 E9	.081	.308
LINEFRND = V10 = .743*F1 + .670 E10	.202	.551
GRPFRNDS = V11 = .772*F1 + .635 E11	.205	.597
SPCHACQ = V12 = .819*F1 + .574 E12	.167	.670
Factor 2: FLCAS		
UNSUREEC = V13 = .702 F2 + .712 E13	*	.493
WORRYMST = V14 = .413*F2 + .911 E14	.091	.170
TRMBLCAL = V15 = .677*F2 + .736 E15	.095	.458
AFRDNOTU = V16 = .666*F2 + .746 E16	.094	.444
THINKOTH = V18 = .330*F2 + .944 E18	.083	.109
OTHERBET = V19 = .370*F2 + .929 E19	.091	.137
WORRYETE = V20 = .348*F2 + .937 E20	.103	.121
PANICNOP = V21 = .663*F2 + .748 E21	.085	.440
CONSEQFA = V22 = .522*F2 + .853 E22	.092	.273
WHYUPSET = V23 = .431*F2 + .902 E23	.089	.186
WORRYFOR =V24 = .571*F2 + .821 E24	.087	.326
EMBRSVLN = V25 = .653*F2 + .758 E25	.081	.426
UPSETNOT = V27 = .502*F2 + .865 E27	.091	.252
EVENPREP = V28 = .671*F2 + .742 E28	.089	.450
NOTGO2EC = V29 = .540*F2 + .841 E29	.085	.292
NOCONFEC = V30 = .561*F2 + .828 E30	.065	.315
CHECKALL = V31 = .571*F2 + .821 E31	.079	.326
HRTPNDCA = V32 = .710*F2 + .704 E32	.085	.504
STUDYCNF = V33 = .530*F2 + .848 E33	.075	.280
PRSSR2PR = V34 = .382*F2 + .924 E34	.087	.146
OTHERSSP = V35 = .411*F2 + .912 E35	.092	.169

SLFCNSCS = V36 = .698*F2 + .717 E36	.064	.487
PACE2FAS = V37 = .658*F2 + .753 E37	.080	.433
MORENVSE = V38 = .684*F2 + .729 E38	.098	.468
CNFSDECL = V39 = .752*F2 + .659 E39	.085	.566
NOTCNFGO = V40 = .567*F2 + .824 E40	.078	.321
WORRYNOT = V41 = .569*F2 + .822 E41	.093	.324
2MANYERU = V42 = .586*F2 + .810 E42	.089	.343
OTHERSLA = V43 = .703*F2 + .711 E43	.083	.494
WORRYNOP = V45 = .719*F2 + .695 E45	.079	.517
Factor 4: Extroversion		
OUTGOING = V46 = .736 F4 + .677 E46	*	.541
ENERGETI = V47 = .658*F4 + .753 E47	.087	.433
TALKATIV = V48 = .644*F4 + .765 E48	.066	.415
BOLD = V49 = .601*F4 + .799 E49	.083	.362
SPUNKY = V50 = .788*F4 + .615 E50	.082	.621
ASSERTIV = V51 = .683*F4 + .731 E51	.081	.466
ADVENTUR = V52 = .641*F4 + .768 E52	.092	.411
PLEASANT = V53 = .370*F4 + .929 E53	.074	.137
Factor Equations		
P_COMPET = F1 = .453*F6 + .891 D1	.043	.205
FLCAS = F2 = -.523*F6 + .852 D2	.057	.273
EXTROVER = F4 = .869*F6 + .495 D4	.092	.755

APPENDIX AF
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE ORIGINAL MACINTYRE AND CHAROS (1996) MODEL USING L2 COMMUNICATIVE ANXIETY DATA

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations					R^2
PERC-COM=V6 = .214*V1 - .319*V7 + .118*V13 + .901E6					
SE	.163	.102	.223		.188
L2ANXIET = V7 = -.297*V2 + .955E7					
SE	.138				.088
INTEGRAT = V8 = -.222*V7 + .975E8					
SE	.046				.049
ATTITUDE = V9 = .206*V5 + .443*V8 + .870E9					
SE	.084	.059			.242
WTC = V10 = .466*V6 + .838 E10					
SE	.058				.217
MOTIVATI = V11 = .245*V4 + .352*V8 + .401*V9 + .725E11					
SE	.091	.078	.065		.475
FREQUENC= V12 = .265*V2 + .298*V10 + .380*V11 + .111*V13 + .793E12					
SE	.105	.046	.051	.168	.371

APPENDIX AG
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE REVISED MACINTYRE AND CHAROS (1996) MODEL USING FLCAS DATA

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations					R^2
PERC-COM=V6 = .209*V1 - .437*V7 + .856 E6					
SE	.141	.155			.267
L2ANXIET = V7 = -.280*V2 - .263*V4 - .247*V13 + .849 E7					
SE	.062	.063	.111		.280
INTEGRAT = V8 = -.285*V7 + .959 E8					
SE	.088				.081
ATTITUDE = V9 = -.563*V7 + .279*V8 + .718 E9					
SE	.073	.050			.485
WTC = V10 = .331*V6 - .304*V7 + .838 E10					
SE	.056	.107			.298
MOTIVATI = V11 = .320*V8 + .449*V9 + .755 E11					
SE	.082	.071			.430
FREQUENC= V12 = .239*V2 + .290*V10 + .381*V11 + .105*V13 + .761 E12					
SE	.109	.048	.051	.169	.421

APPENDIX AH
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE ORIGINAL YASHIMA (2002) MODEL (INITIAL SOLUTION)

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations			R^2
Factor 1: Proficiency			
LISTENIN = V1 = .696 F1 + .719 E1			
SE	.*		.484
VOCAB = V2 = .799*F1 + .602 E2			
SE	.171		.638
Factor 2: L2 Communicative Confidence			
PER_COM = V3 = .639 F2 + .769 E3			
SE	.*		.409
ANXIETY = V4 = -.741*F2 + .672 E4			
SE	.068		.549
Factor 3: International Posture			
MOTIVAT = V7 = .772*F3 + .635 E7			
SE	.078		.597
C-FRIEND = V8 = .793*F3 + .610 E8			
SE	.*		.628
APPROACH = V9 = .826*F3 + .564 E9			
SE	.065		.682
WTC = V6 = .602*F2 + .055*F3 + .781 E6			
SE	.124	.088	.389
MOTIVATAT = V7 = .772*F3 + .635 E7			
SE	.078		.597
PROFICIENC = F1 = .458*V7 + .889 D1			
SE	.030		.210
COMM_CON = F2 = .328*V7 + .341*F1 + .822 D2			
SE	.070	.225	.325

Note. Because F3 (International Posture) is an endogenous variable (i.e., only paths pointing away), there is no F3 equation and therefore no D3 disturbance term.

APPENDIX AI
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE ORIGINAL YASHIMA (2002) MODEL (FINAL SOLUTION)

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations			R^2
Factor 1: Proficiency			
LISTENIN = V1 = .696 F1 + .718 E1			
SE	.*		.485
VOCAB = V2 = .716*F1 + .603 E2			
SE	.174		.636
Factor 2: L2 Communicative Confidence			
PER_COM = V3 = .716 F2 + .698 E3			
SE	.*		.512
ANXIETY = V4 = -.529*F2 - .313*F3 + .733 E4			
SE	.058	.045	.463
Factor 3: International Posture			
C-FRIEND = V8 = .781*F3 + .625 E8			
SE	.*		.609
APPROACH = V9 = .841*F3 + .540 E9			
SE	.066		.708
WTC = V6 = .584*F2 + .165*F3 + .763 E6			
SE	.122	.090	.418
MOTIVATA = V7 = .766*F3 + .642 E7			
SE	.079		.587
PROFICIEN = F1 = .459*V7 + .720 D1			
SE	.030		.210
COMM_CON = F2 = .189*V7 + .328*F1 + .894 D2			
SE	.085	.266	.220

APPENDIX AJ
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE REVISED YASHIMA (2002) MODEL (FINAL SOLUTION)

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations			R^2
Factor 1: Proficiency			
LISTENIN = V1 = .700 F1 + .714 E1			
SE	.*		.490
VOCAB = V2 = .790*F1 + .614 E2			
SE	.172		.623
Factor 2: L2 Communicative Confidence			
PER_COM = V3 = .727 F2 + .687 E3			
SE	.*		.529
ANXIETY = V4 = -.522*F2 – .331*F3 + .726 E4			
SE	.065		.576
EXTROVE = V5 = .355*F2 + .919 E5			
SE	.048		.235
Factor 3: International Posture			
C-FRIEND = V8 = .766*F3 + .643 E8			
SE	.*		.586
APPROAC = V9 = .867*F3 + .498 E9			
SE	.065		.752
WTC = V6 = .561*F2 + .184*F3 + .773 E6			
SE	.119	.083	.402
MOTIVAT = V7 = .736*F3 + .665 E7			
SE	.078		.558
PROFICIENC = F1 = .459*V7 + .885 D1			
SE	.030		.216
COMM_CON = F2 = .362*V7 + .916 D2			
SE	.030		.161
I-POSTUR = F3 = .426*V7 + .889 D2			
SE	.068	.214	.210

APPENDIX AK
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE ORIGINAL YASHIMA ET AL. (2004) MODEL (FINAL SOLUTION)

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations				R^2
Factor 1: Proficiency				
LISTENIN	= V1 = .698 F1 + .716 E1			
SE	.*			.487
VOCAB	= V2 = .796*F1 + .606 E2			
SE	.174			.633
Factor 2: L2 Communicative Confidence				
PER_COM	= V3 = .728 F2 + .685 E3			
SE	.*			.531
ANXIETY	= V4 = -.530*F2 - .325*F3 + .728 E4			
SE	.055	.045		.470
EXTROV	= V5 = .261*F2 + .369*F3 + 866 E6			
SE	.041	.041		.250
Factor 3: International Posture				
C-FRIEND	= V8 = .770*F3 + .638 E8			
SE	.*			.594
APPROAC	= V9 = .863*F3 + .506 E9			
SE	.068			.744
WTC	= V6 = .567*F2 + .178*F3 + .776 E6			
SE	.113	.089		.402
MOTIVAT	= V7 = .748*F3 + .664 E7			
SE	.080			.582
PROFICIEN	= F1 = .459*V6 + .889 D1			
SE	.030			.221
COMM_CON	= F2 = .180*V7 + .314*F1 + .904 D2			
SE	.085	.267		.183

Note. Because F3 (International Posture) is an endogenous variable (i.e., only paths pointing away), there is no F3 equation and therefore no D3 disturbance term.

APPENDIX AL
STANDARDIZED STRUCTURAL EQUATIONS, STANDARD ERRORS, AND R^2 FOR THE REVISED YASHIMA ET AL. (2004) MODEL

Standard errors and squared multiple correlations (R^2) are shown in the row beneath each equation.

Equations			R^2
Factor 1: Proficiency			
LISTENIN = V1 = .700 F1 + .714 E1			
SE	.*		.490
VOCAB = V2 = .790*F1 + .614 E2			
SE	.160		.623
Factor 2: L2 Communicative Confidence			
PER_COM = V3 = .727 F2 + .687 E3			
SE	.*		.521
ANXIETY = V4 = -.522*F2 - .331*F3 + .726 E4			
SE	.053	.041	.473
EXTROVE = V5 = .355*F2 + .919 E5			
SE	.040	.039	.156
WTC = V6 = .561*F2 + .184*F3 + .773 E6			
SE	.112	.081	.402
Factor 3: International Posture			
MOTIVAT = V7 = .736*F3 + .665 E7			
SE	.079		.558
C-FRIEND = V8 = .766*F3 + .643 E8			
SE	.*		.586
APPROAC = V9 = .847*F3 + .498 E9			
SE	.064		.752
FREQUENC = V10 = .251*V6 + .578*F3 + .720 E10			
SE	.051	.079	.481
PROFICIENC = F1 = .475*V7 + .880 D1			
SE	.032		.226
COMM_CON = F2 = .414*F1 + .910 D2			
SE	.245		.172

Note. Because F3 (International Posture) is an endogenous variable (i.e., only paths pointing away), there is no F3 equation and therefore no D3 disturbance term.