then directly influences language learning outcomes such as proficiency and fluency. Figure 1 shows a portion of Gardner's (1985) model.

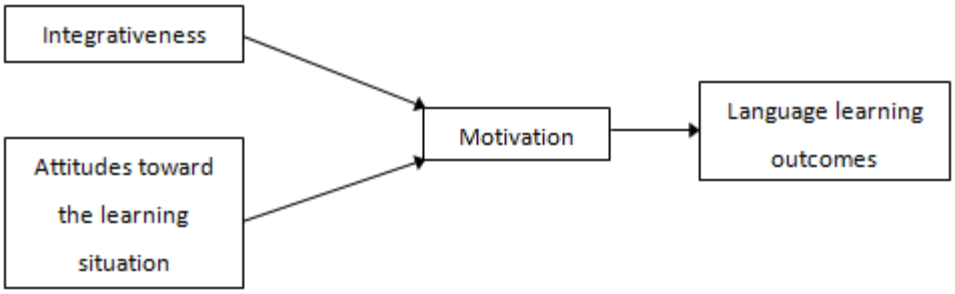


Figure 1. Portion of Gardner's (1985) model of L2 Communicative Competence. From "Personality, attitude, and affect as predictors of second language communication," by P. D. MacIntyre & C. Charos, (1996), *Journal of Language and Social Psychology, 15*(1), p. 5. Copyright 1996 by the *Journal of Language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

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]

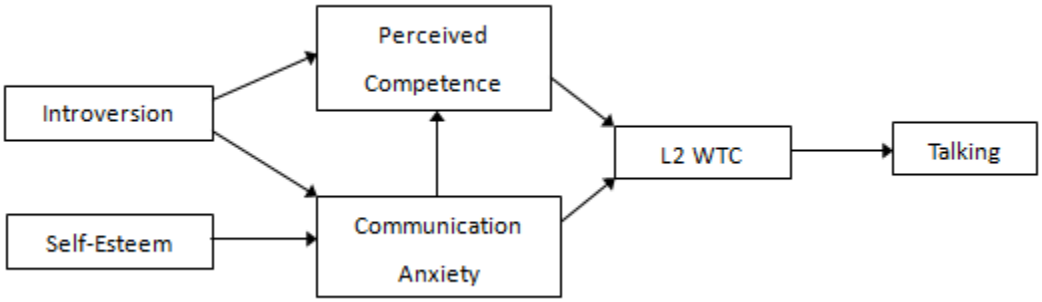


Figure 2. Portion of MacIntyre's (1994) willingness to communicate model. From P. D. MacIntyre & C. Charos, (1996), "Conceptualizing willingness to communicate in a L2: A situated model of confidence and affiliation." *Journal of Language and Social Psychology*, *15*(1), p. 8. Copyright 1996 by the *Journal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

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.

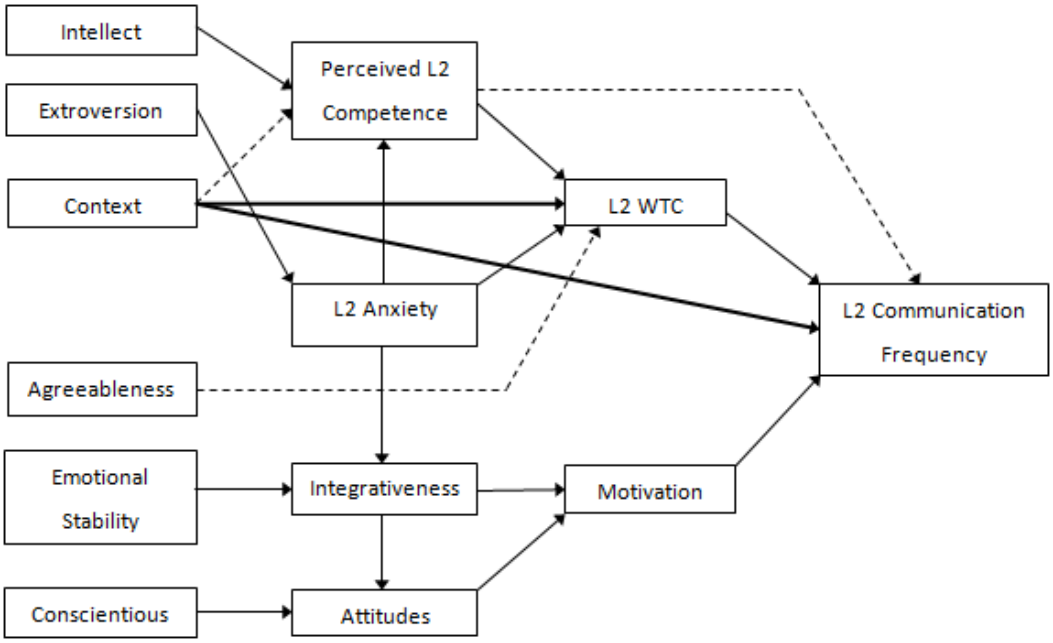


Figure 3. Final MacIntyre and Charos (1996) model of L2 willingness to communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 18. Copyright 1996 by *Journal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

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.

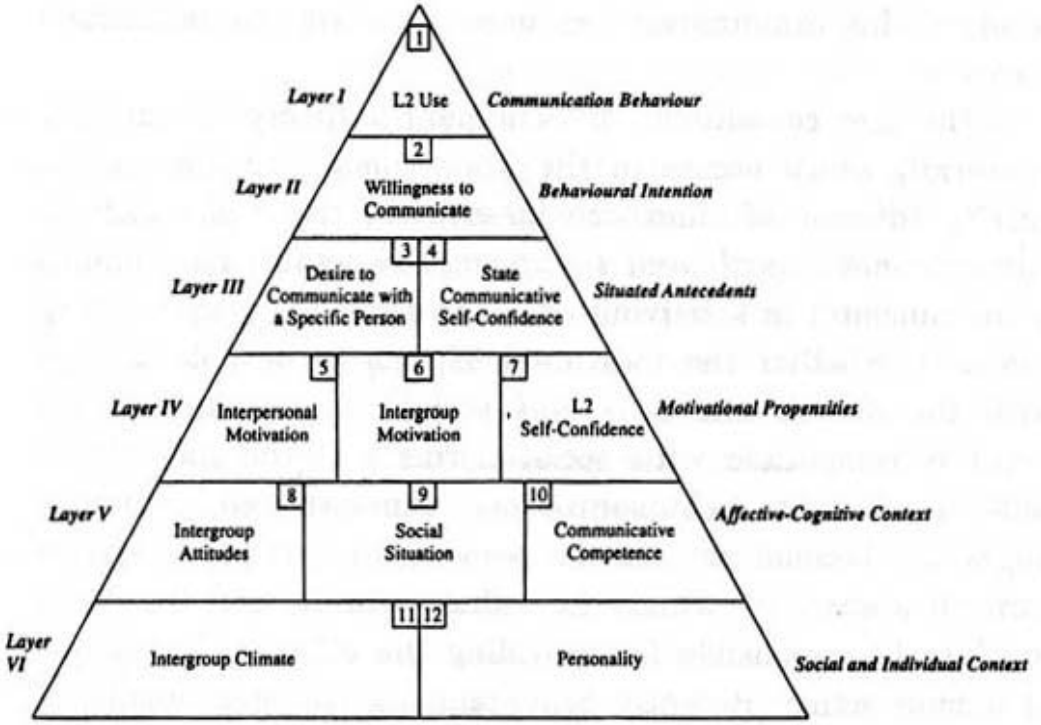


Figure 4. Schematic representation of the variables influencing L2 WTC. From "Conceptualizing Willingness to Communicate in a L2: A Situated Model of Confidence and Affiliation," by P. D. MacIntyre, R. Clément, Z. Dörnyei, and K. Noels, 1998, *Modern Language Journal, 82*, p. 547. Copyright 1998 by *The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

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.

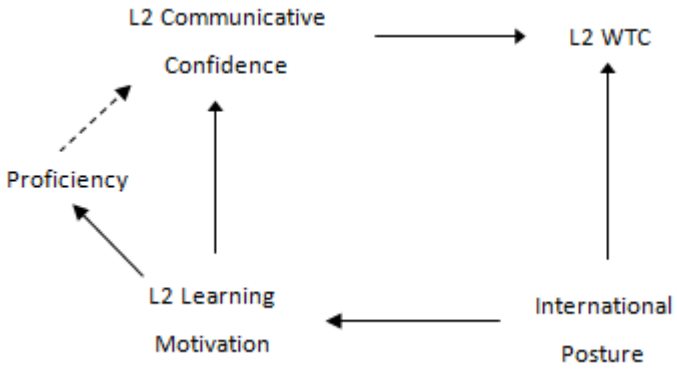


Figure 5. L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal*, *86*(1), 61. Copyright 2002 by *The Modern Language Journal.* Reprinted with permission. Note that the dashed path was hypothesized but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

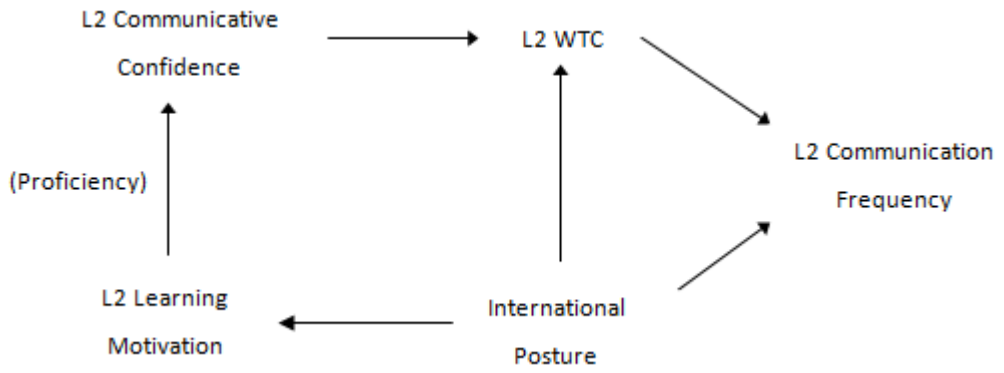


Figure 6. L2 communication model (minus proficiency). From "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Acquisition," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning, 54*(1), p. 127. Copyright 2004 by *Language Learning.* Reprinted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

ego permeability mediates the extent to which that person can assume new roles and thereby function adequately.

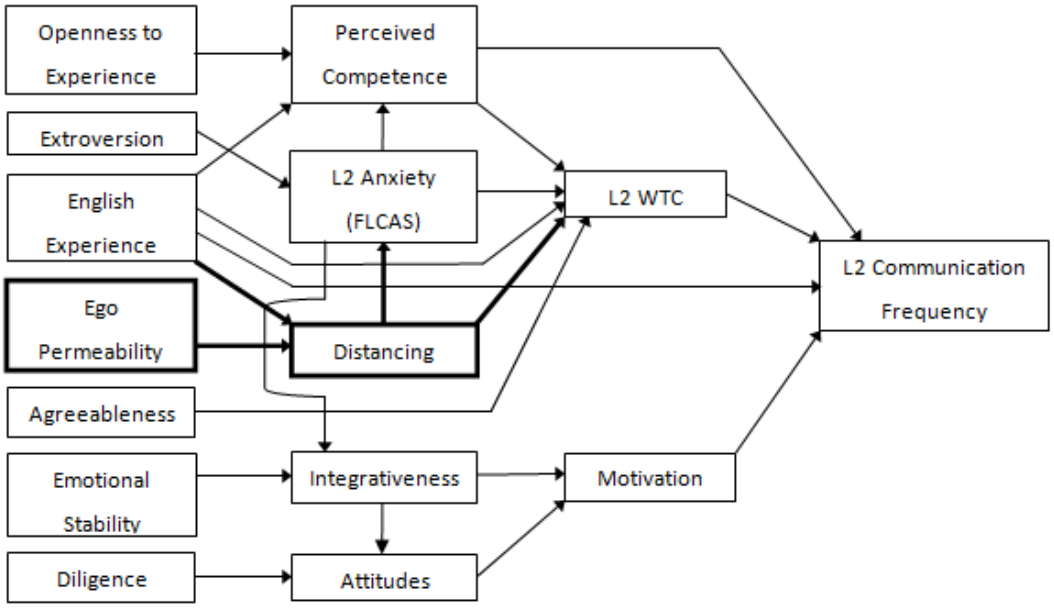


Figure 7. Proposed Model of L2 Willingness to Communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 12. Copyright 1996 by *Journal of Language and Social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

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).

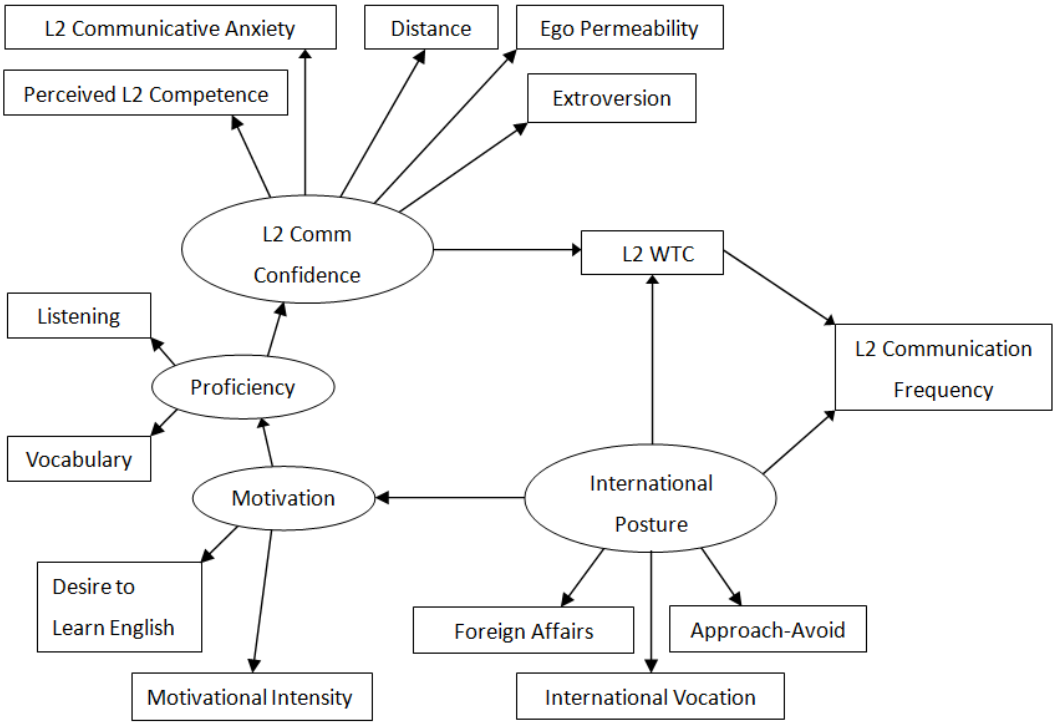
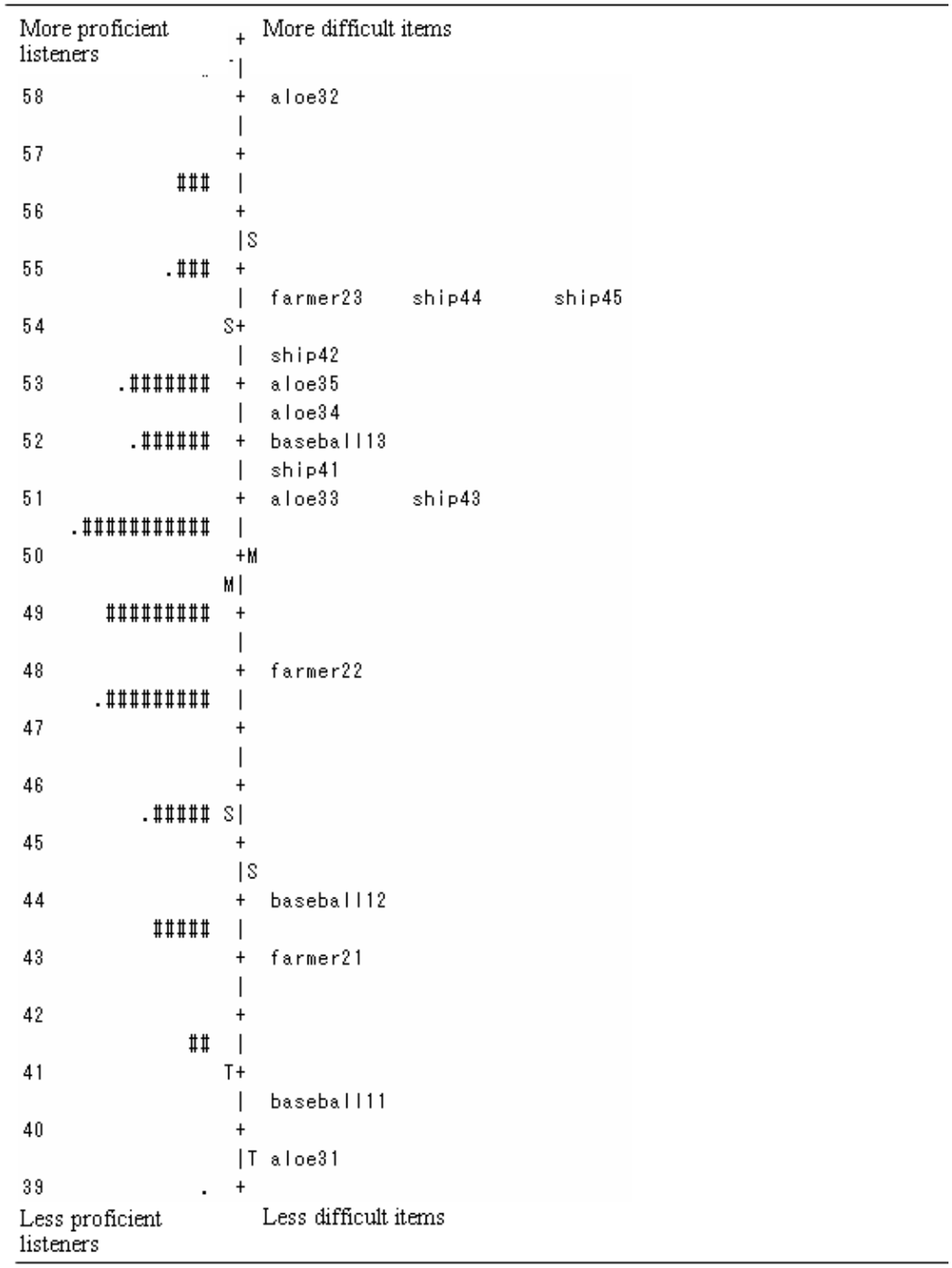


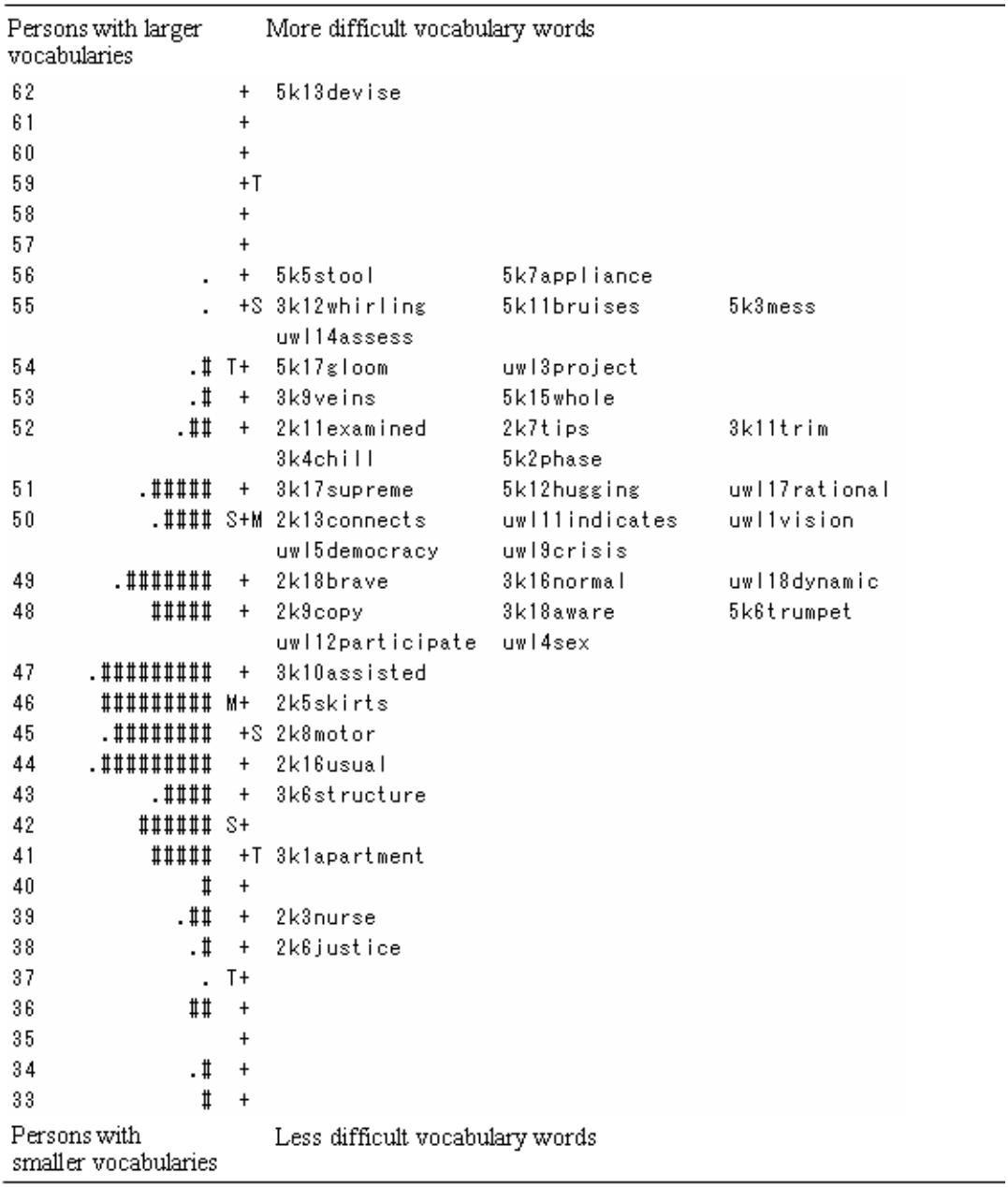
Figure 8. Proposed L2 Communication Model based on Yashima et al. (2004).

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



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

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

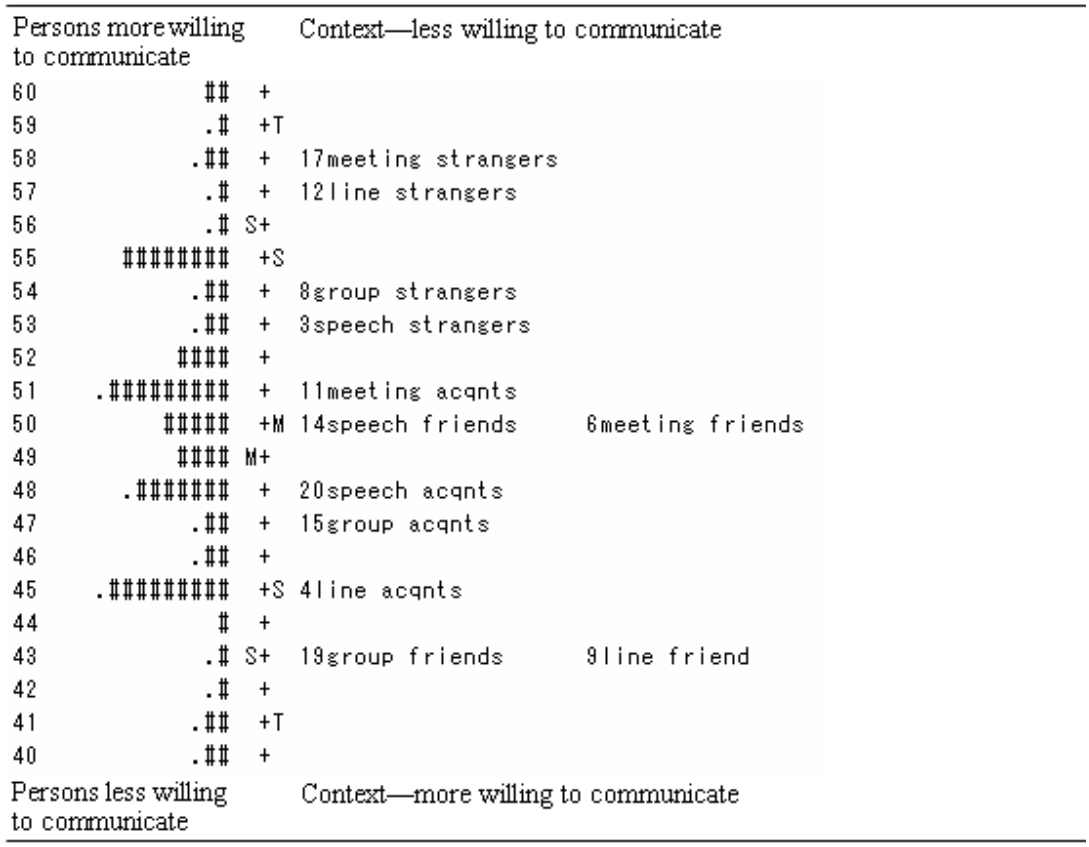


Note. M = mean, S = one standard deviation, T = two standard deviations

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 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.



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

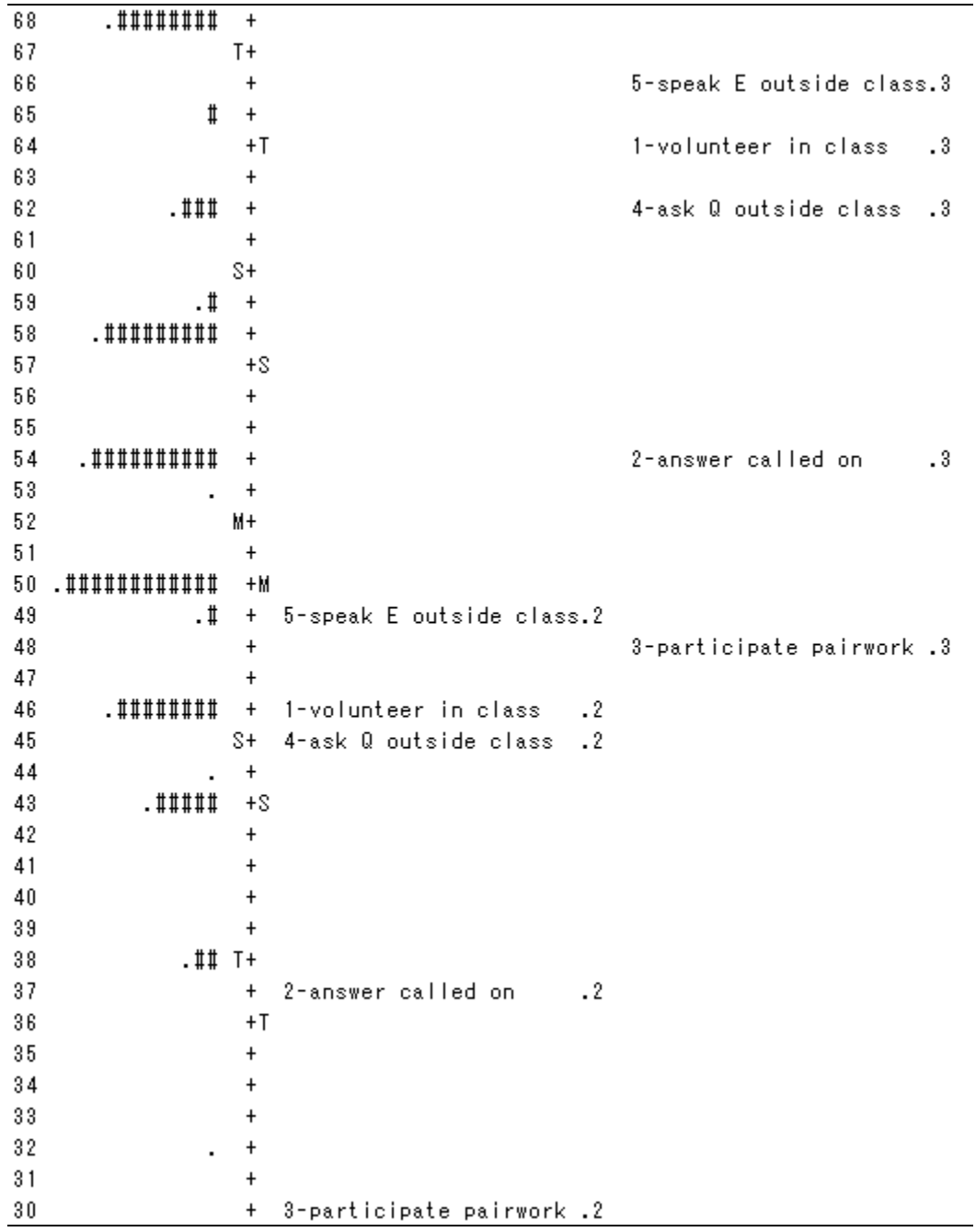


Figure 12. Item-person map with Rasch-Thurstone thresholds of the Frequency of L2 Communication instrument.

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

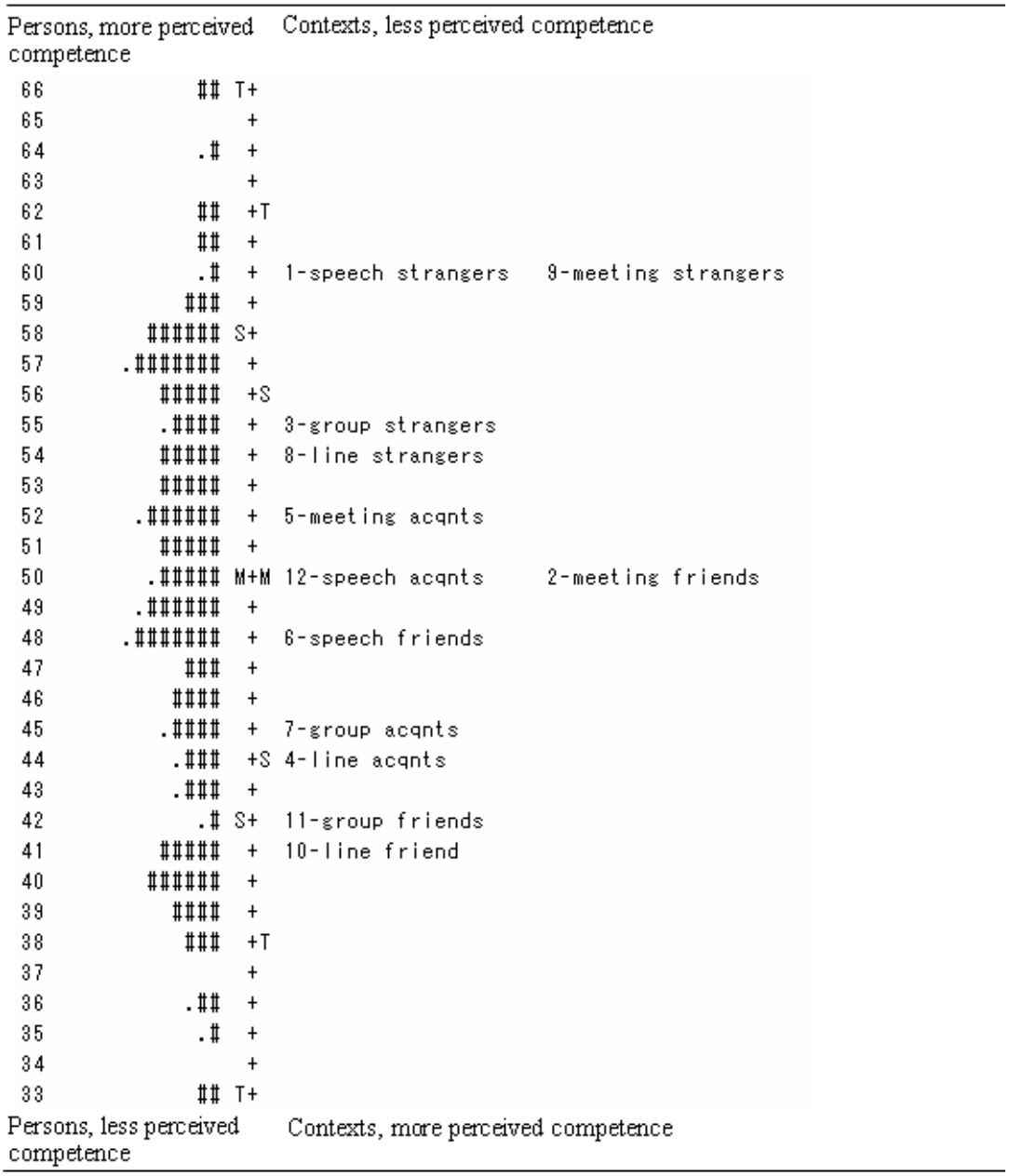
Table 18
Perceived Competence in English: Rasch Item Fit Statistics

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.



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 α = .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	
Positive loadings	Infit	Outfit
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	

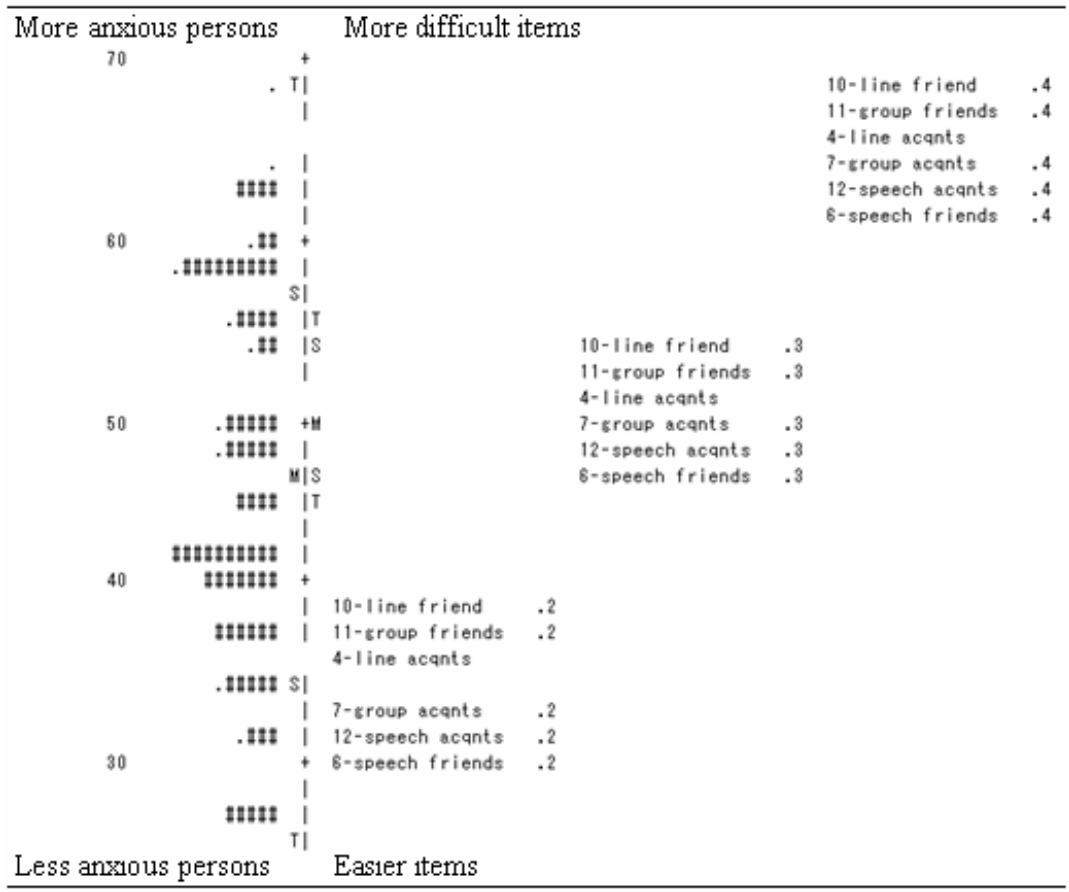
Note. $N = 252$, $k = 12$; Pt-M Corr = point-measure correlation; Misfitting values are indicated with an asterisk.

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.

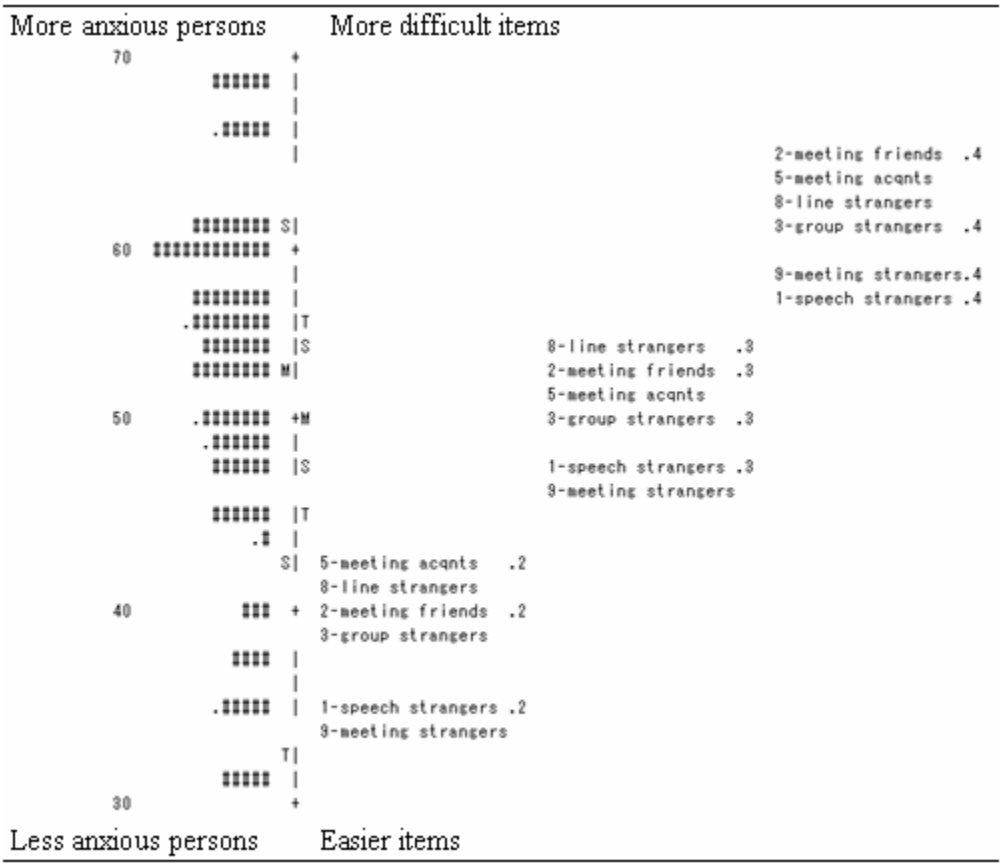
The item-person maps (Figures 14 and 15) indicated that although the range of the item means was somewhat limited when compared with the range of person ability estimates, the Rasch-Thurstone thresholds indicated adequate coverage. The difference between item difficulty and person ability means was only .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)

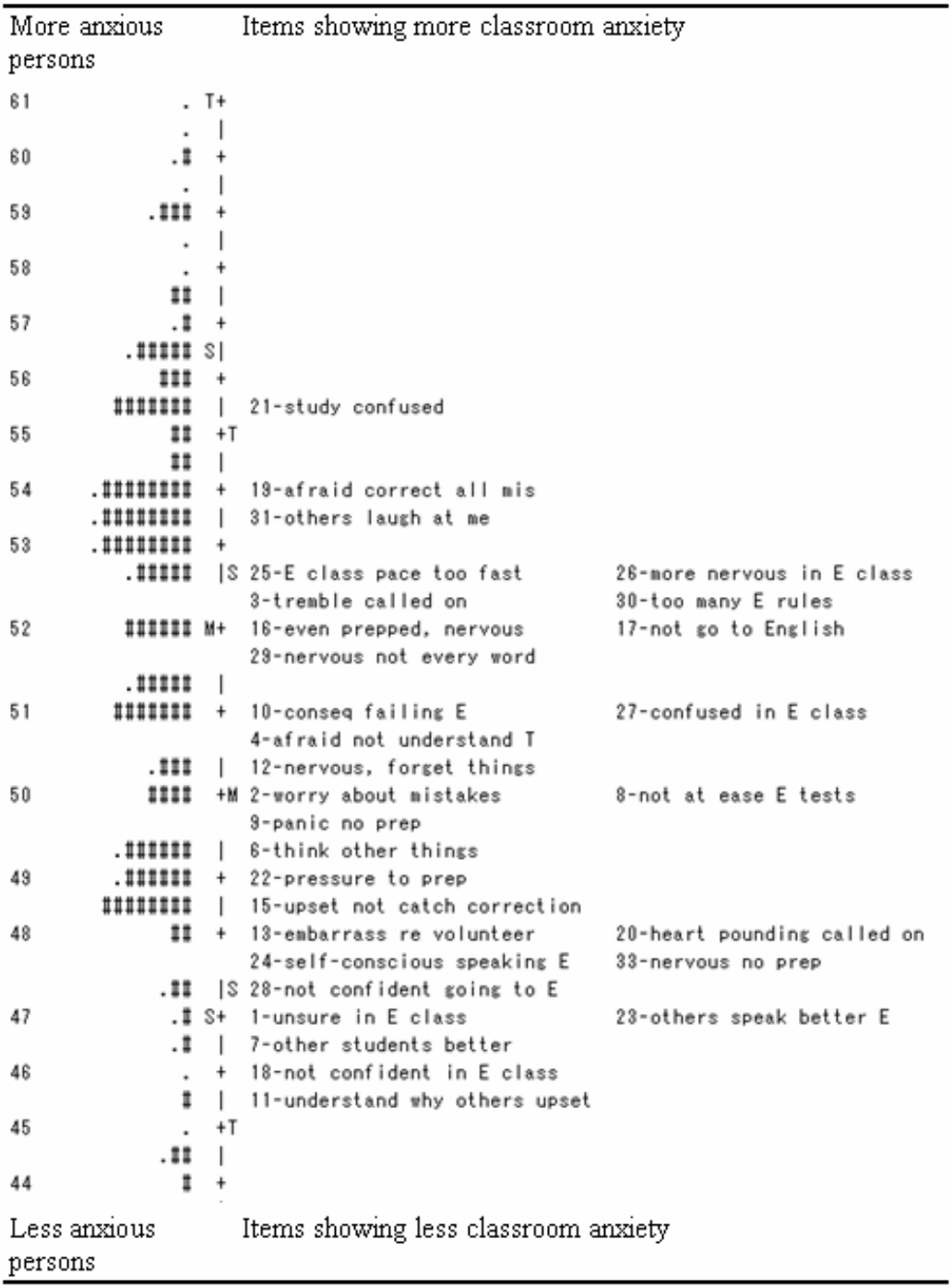
Table 24 (continued)
FLCAS Measure: Rasch Item Fit Statistics and Inter-Item Correlation

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.



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.

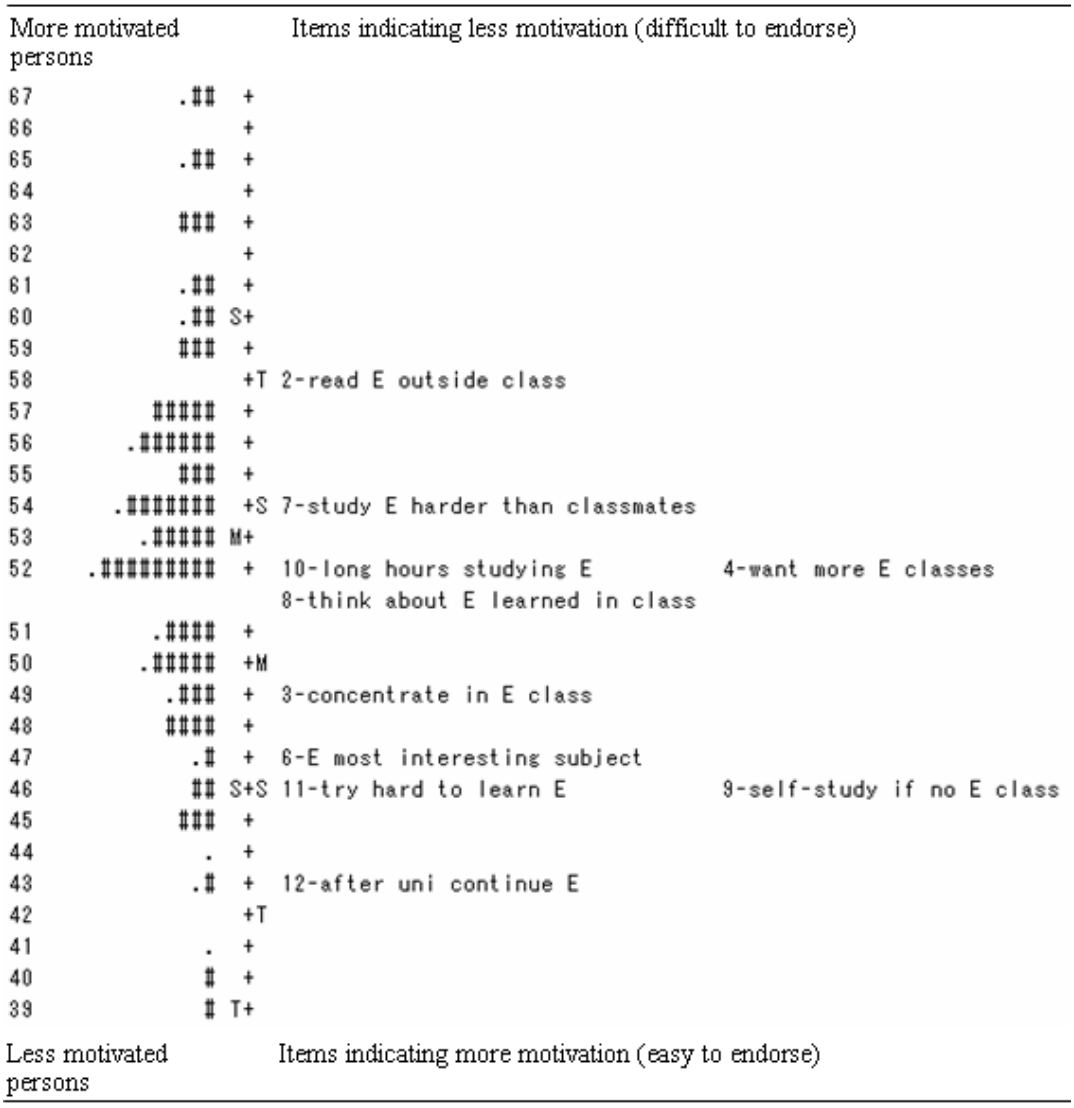
However, in looking at the content of the items, I'm not convinced that two distinct subscales are present. For example, Item 12 (the easiest item to endorse), which was originally in the Motivational Intensity subscale, deals with continuing to learn English after finishing college; however, it could 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



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 <u>not</u> 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.

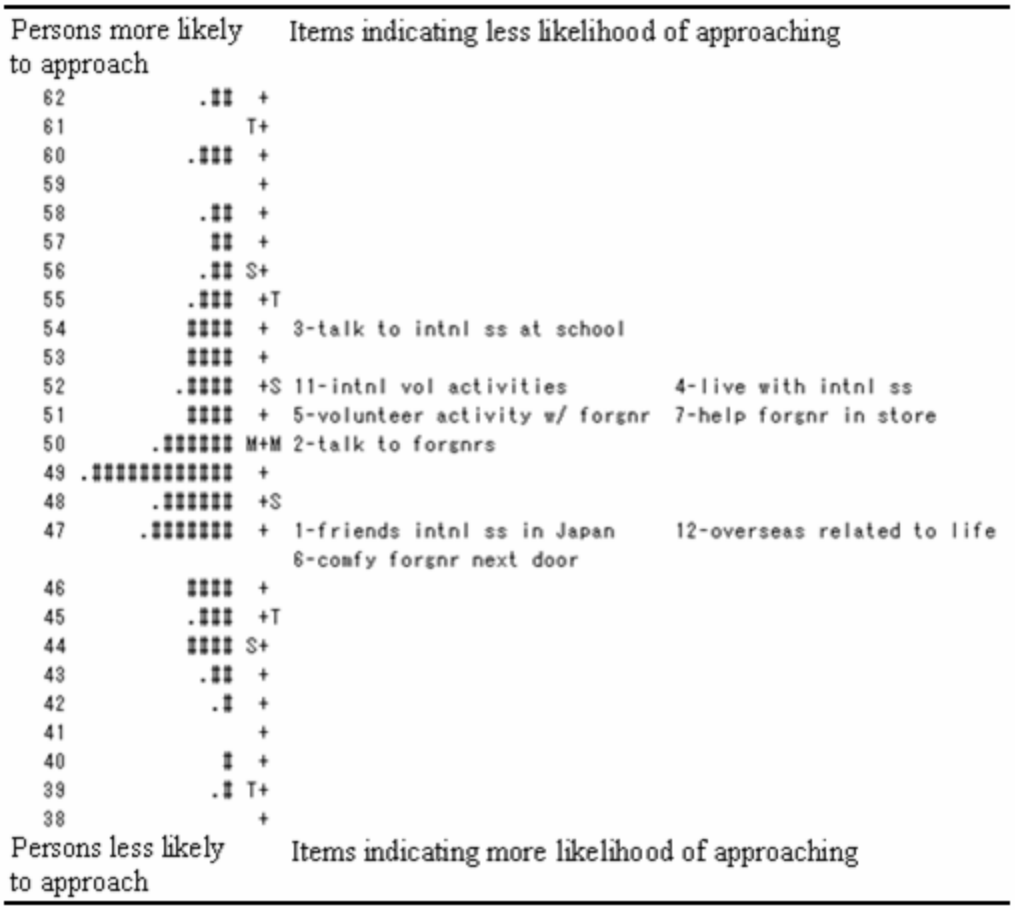
The dimensionality of the Intergroup Approach-Avoidance Tendency subscale was then investigated. The average inter-item correlation for the 9-item instrument was adequate (r = .40), and the internal reliability estimate was high (Cronbach's α = .87). An exploratory factor analysis yielded a two-component 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



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.

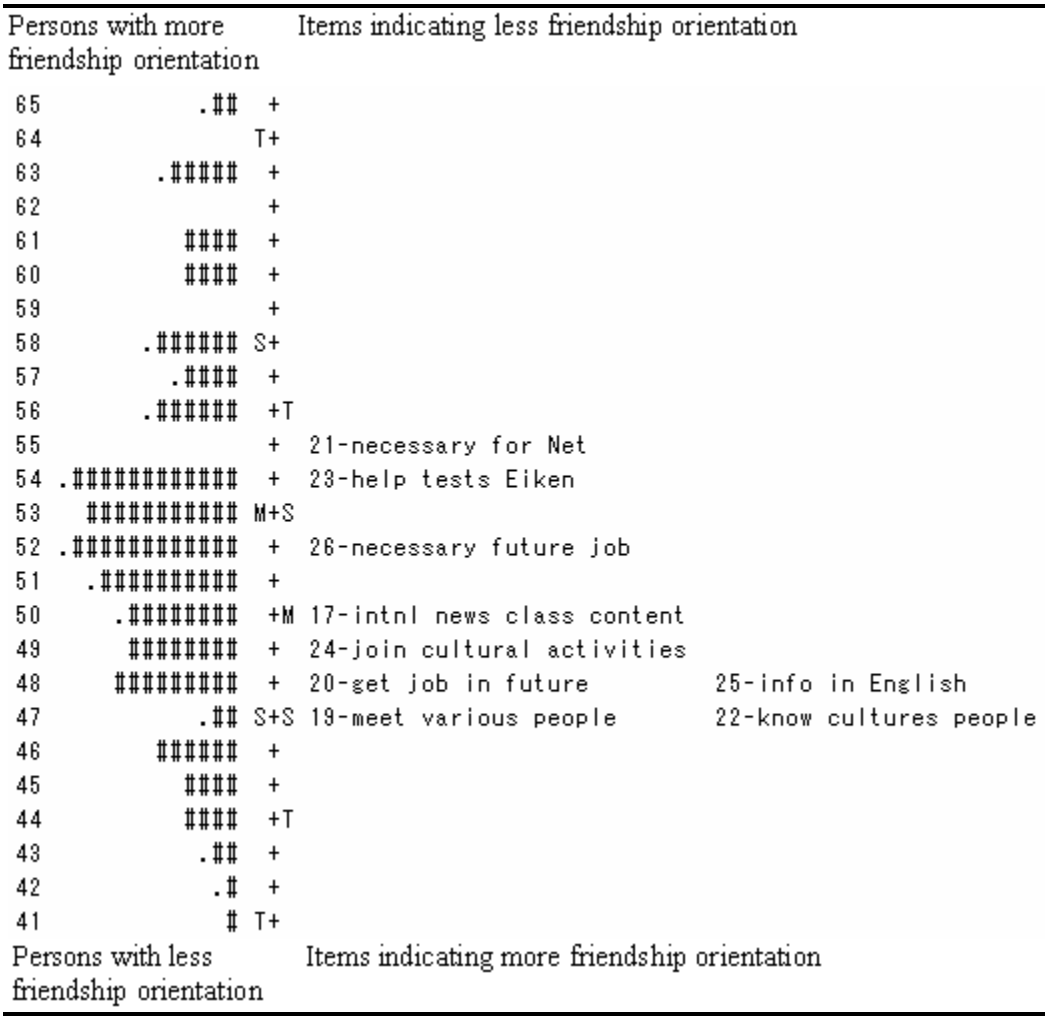
The Intercultural Friendship Orientation instrument was then examined with WINSTEPS. The scale had an item reliability estimate of .98, item separation of 6.53, a person reliability estimate of .77, person separation of 1.82, and thus a person strata statistic of 2.64. All nine items exhibited adequate 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



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 $\alpha = .73$). 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

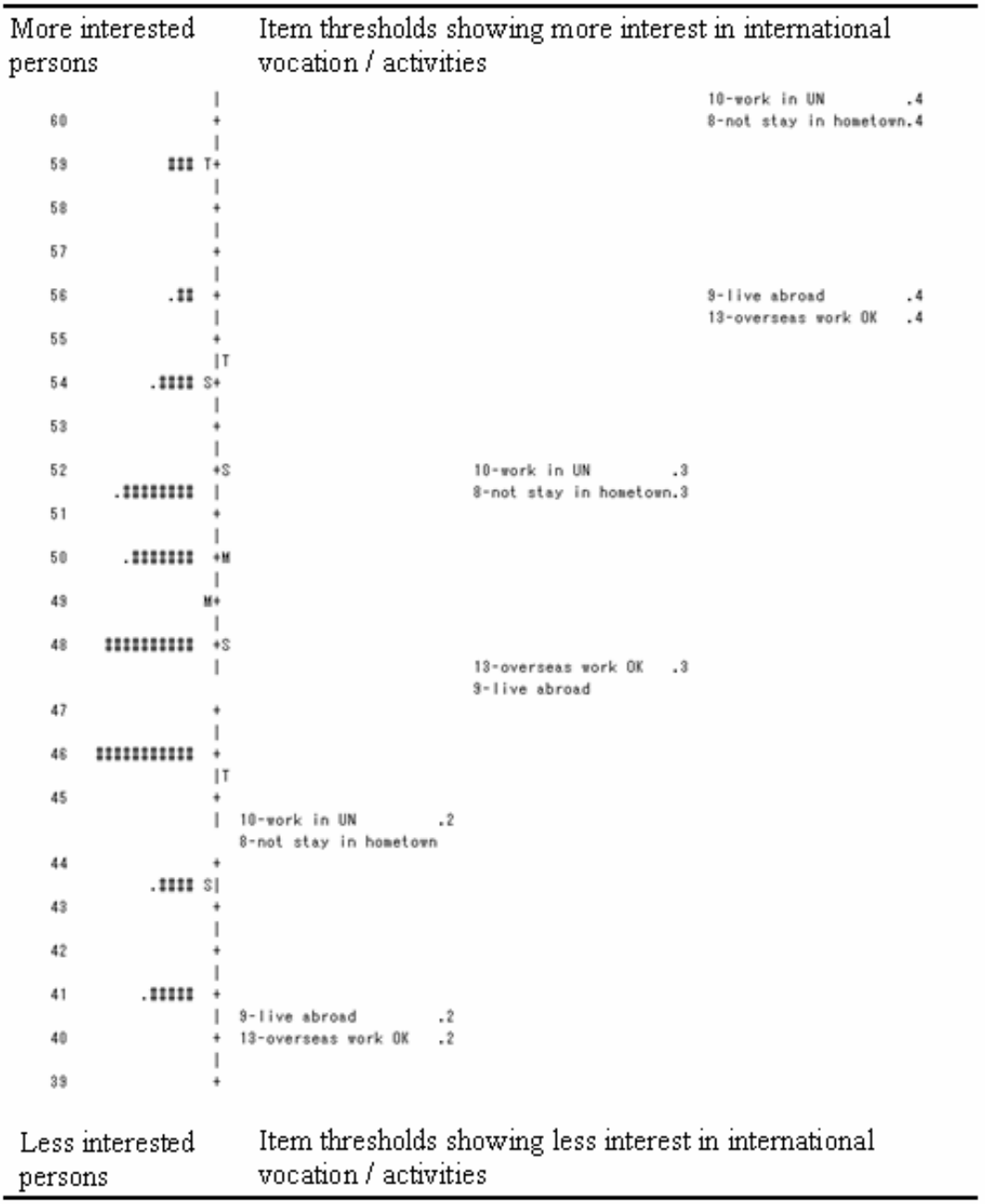
adequate, from .45 to .63. A PCA of item residuals showed that the disattenuated correlation of person measures derived from items with positive and negative residual loadings was .52, suggesting that this instrument was *not* unidimensional. However, splitting the subscale further would have resulted 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 $\alpha = .64$) 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.

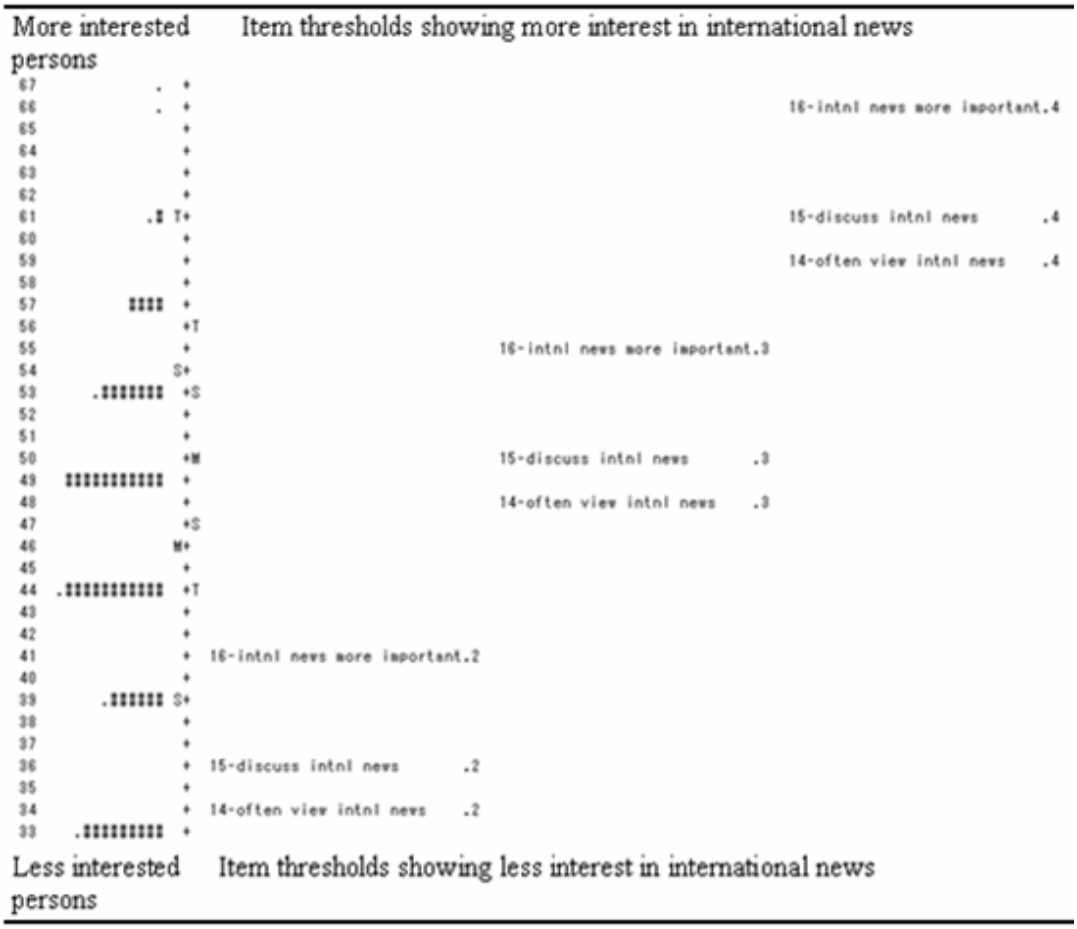
Table 38
Interest in Foreign Affairs Measure: Rasch Item Fit Statistics

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

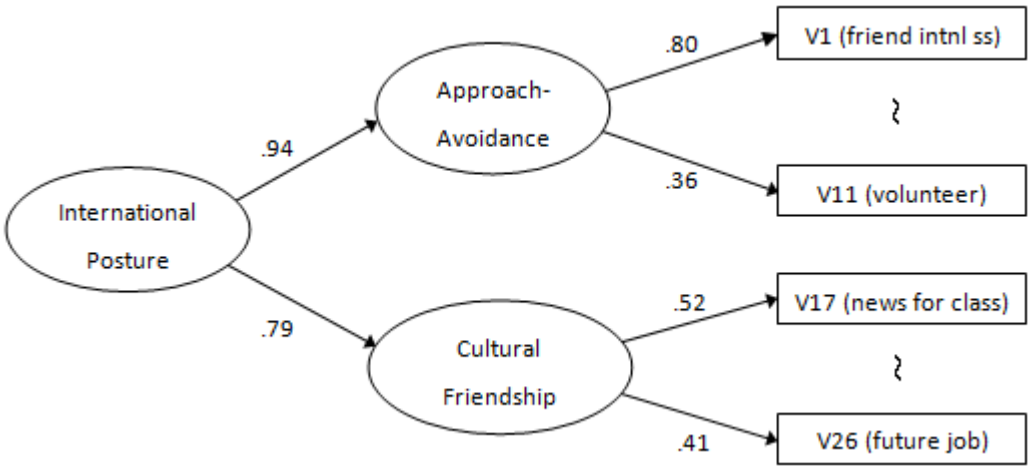


Figure 22. Standardized solution of the 2-factor International Posture instrument.

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*	.26					.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*		.27				.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*					.34	.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

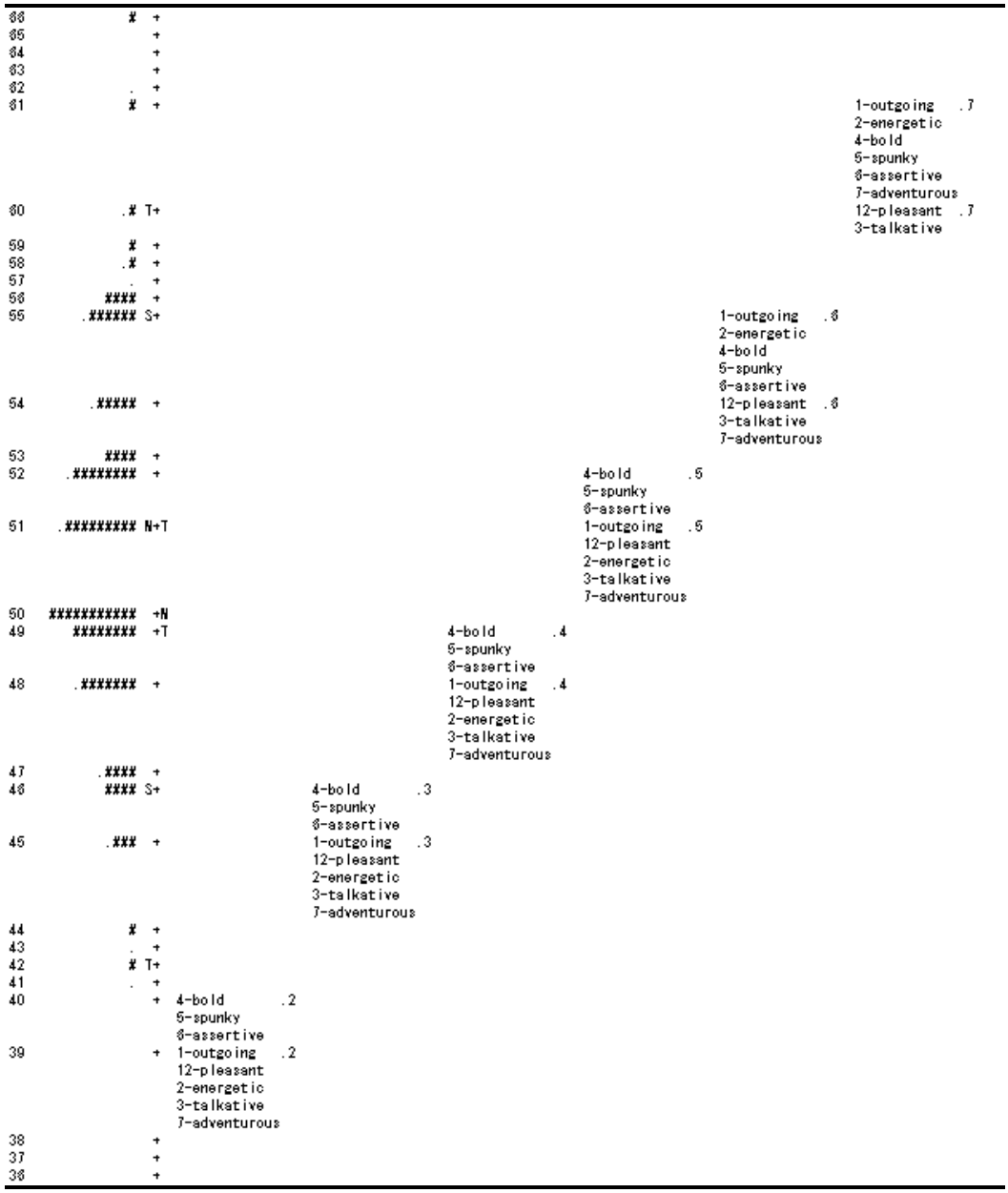
Item	Measure	*SE*	Infit MNSQ	Infit *t*	Outfit MNSQ	Outfit *t*	Pt-M Corr
4-bold	50.45	.25	1.04	.1	1.06	.8	.67
6-assertive	50.43	.25	.90	-1.2	.93	-.8	.70
5-spunky (active)	50.40	.25	.77	-2.7	.77	-2.9	.76
1-outgoing	50.10	.25	.86	-1.8	.85	-1.8	.74
2-energetic	49.92	.25	.97	-.1	.99	-.1	.70
12-pleasant (agreeable)	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 $\alpha = .87$). 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%).

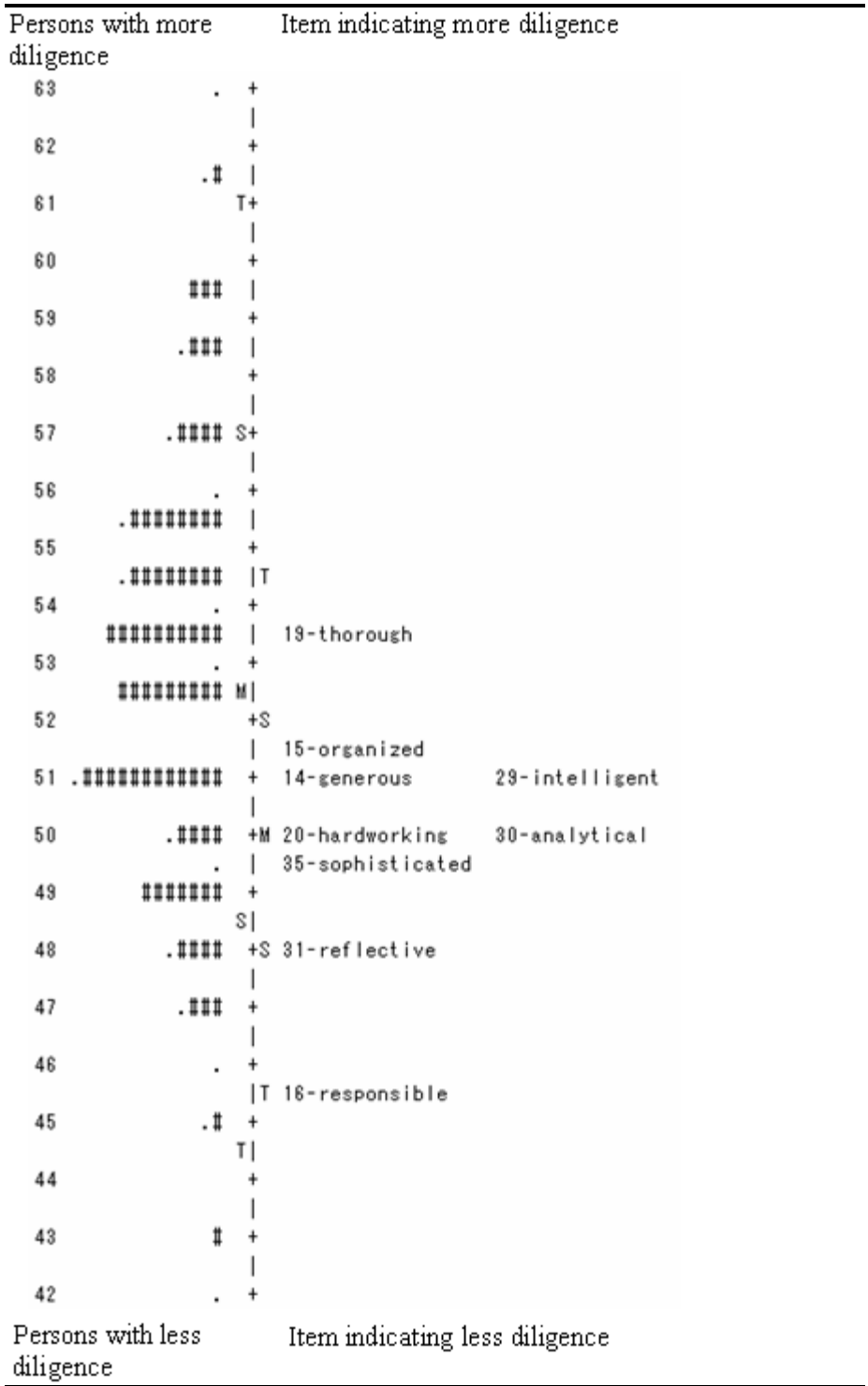
Rasch statistics yielded a Rasch item reliability estimate of .96, item separation of 4.83, a Rasch person reliability estimate of .61, person separation of 1.24, and a person strata statistic of 1.99. As shown in Table 45, eight of the nine items exhibited adequate fit and reasonable point-measure correlations. 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.



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 $\alpha = .87$). The initial EFA yielded a 1-component solution that accounted for 52.9% of the

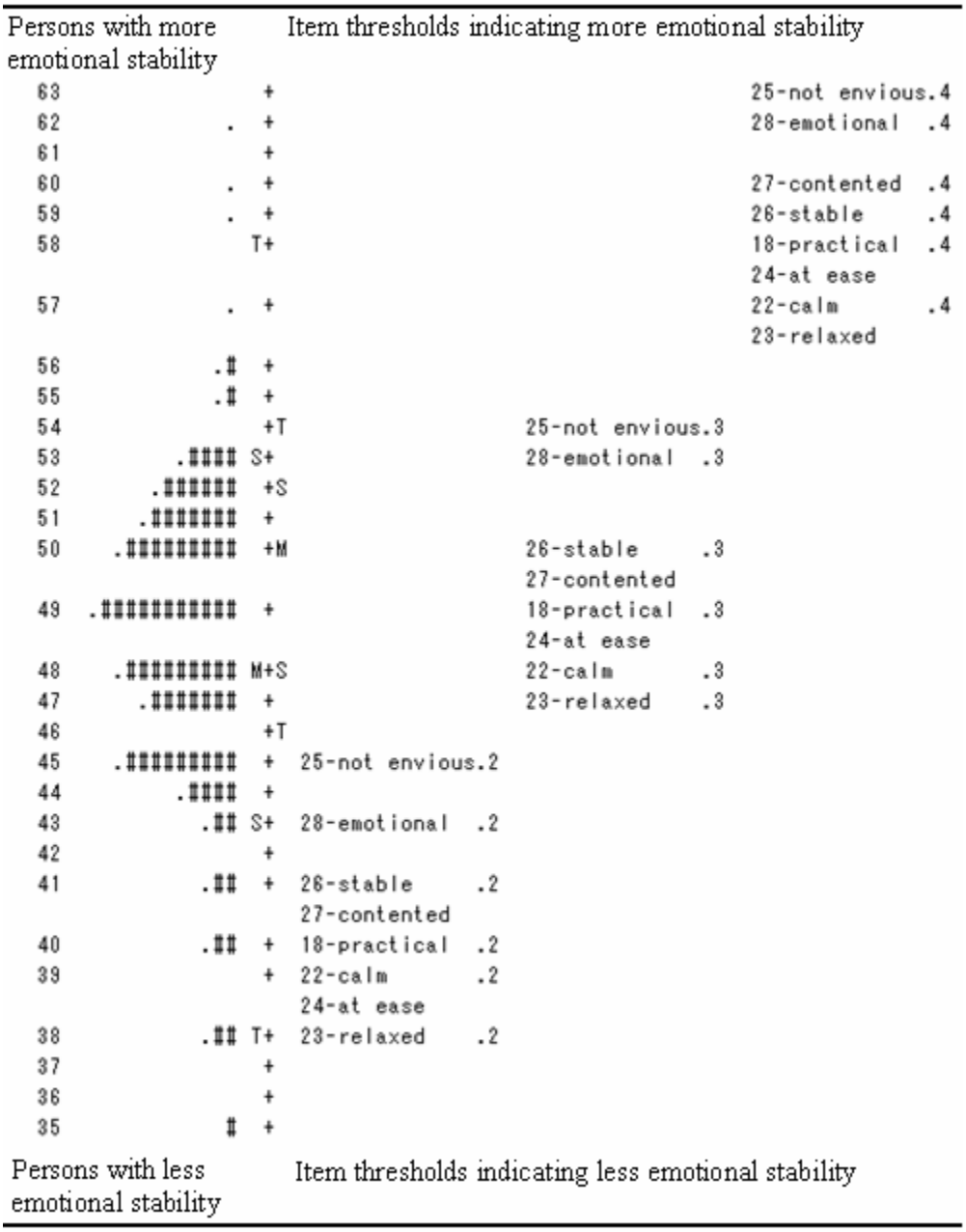
Table 47
Emotional Stability Subscale Measure: Rasch Item Fit Statistics

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.



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 $\alpha = .87$). 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

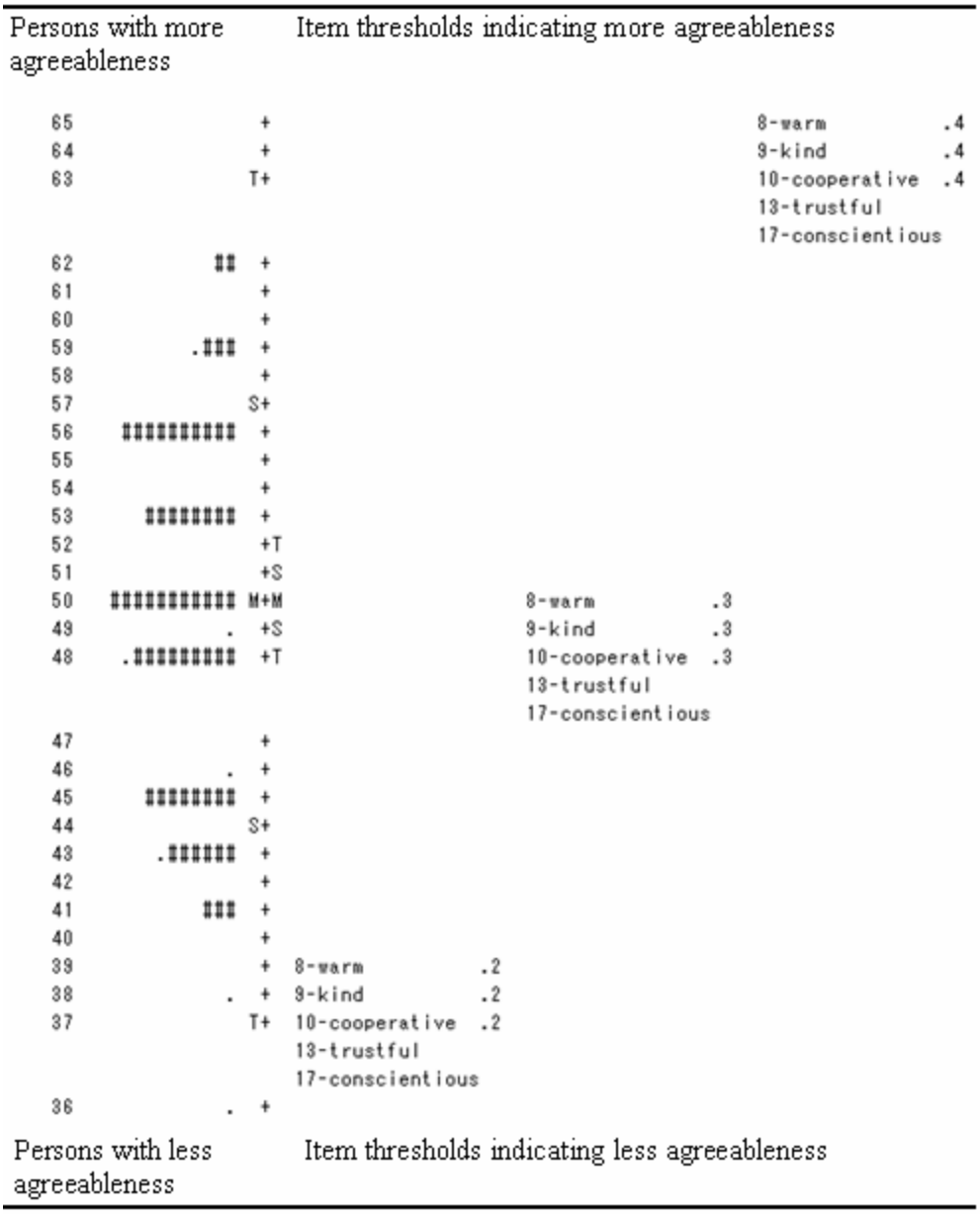
positive and negative residual loadings was .86, suggesting that this instrument was fundamentally unidimensional. In addition, the PCA of item residuals indicated that the Rasch model accounted for 47.3% of the variance. The unexplained variance accounted for by the first residual contrast was 1.5 units (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.



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).

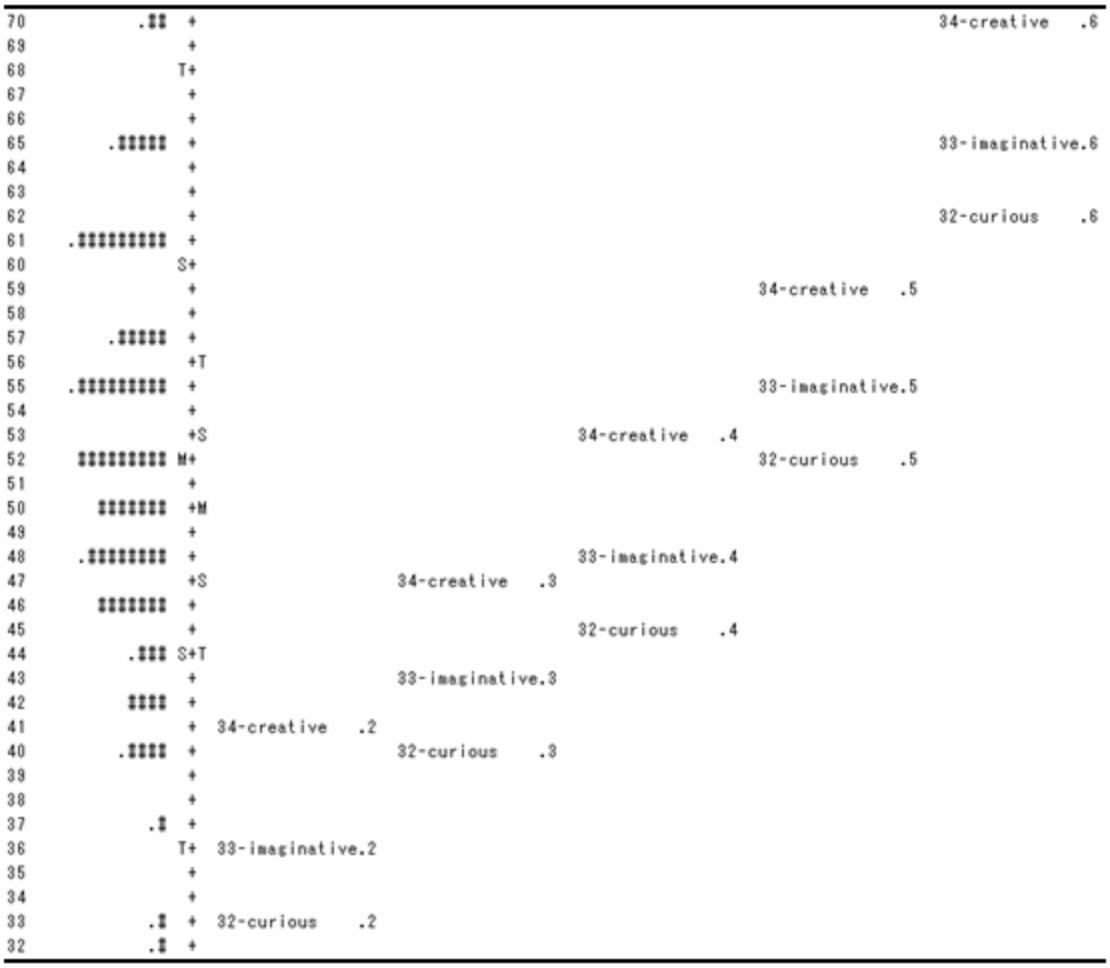
Table 51
Openness to Experience Subscale Measure: Rasch Item Fit Statistics

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 $\alpha = .82$). 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.

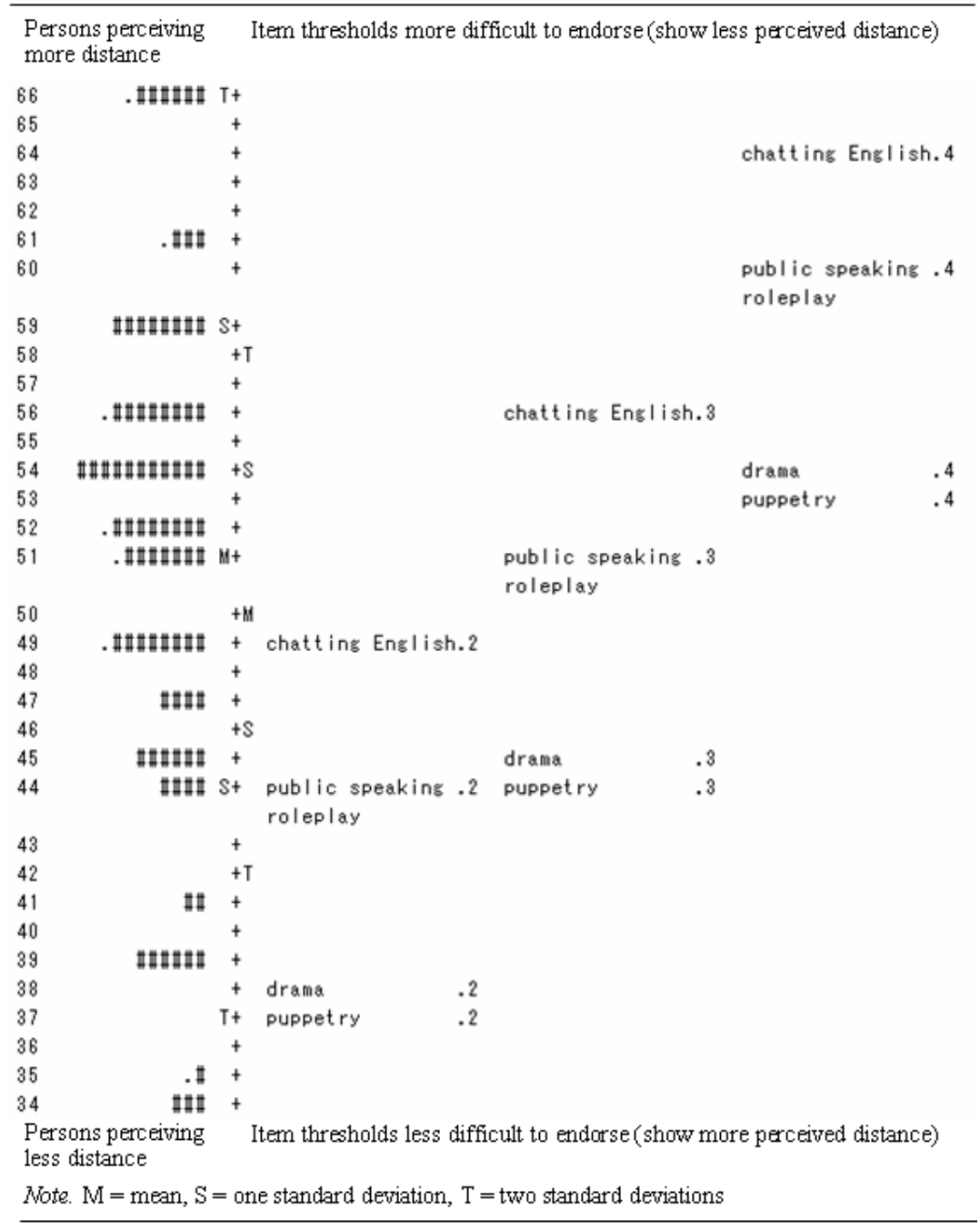


Figure 28. Item-person Rasch-Thurstone threshold map of the Perceived Distance instrument.

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 $\alpha = .87$). 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.

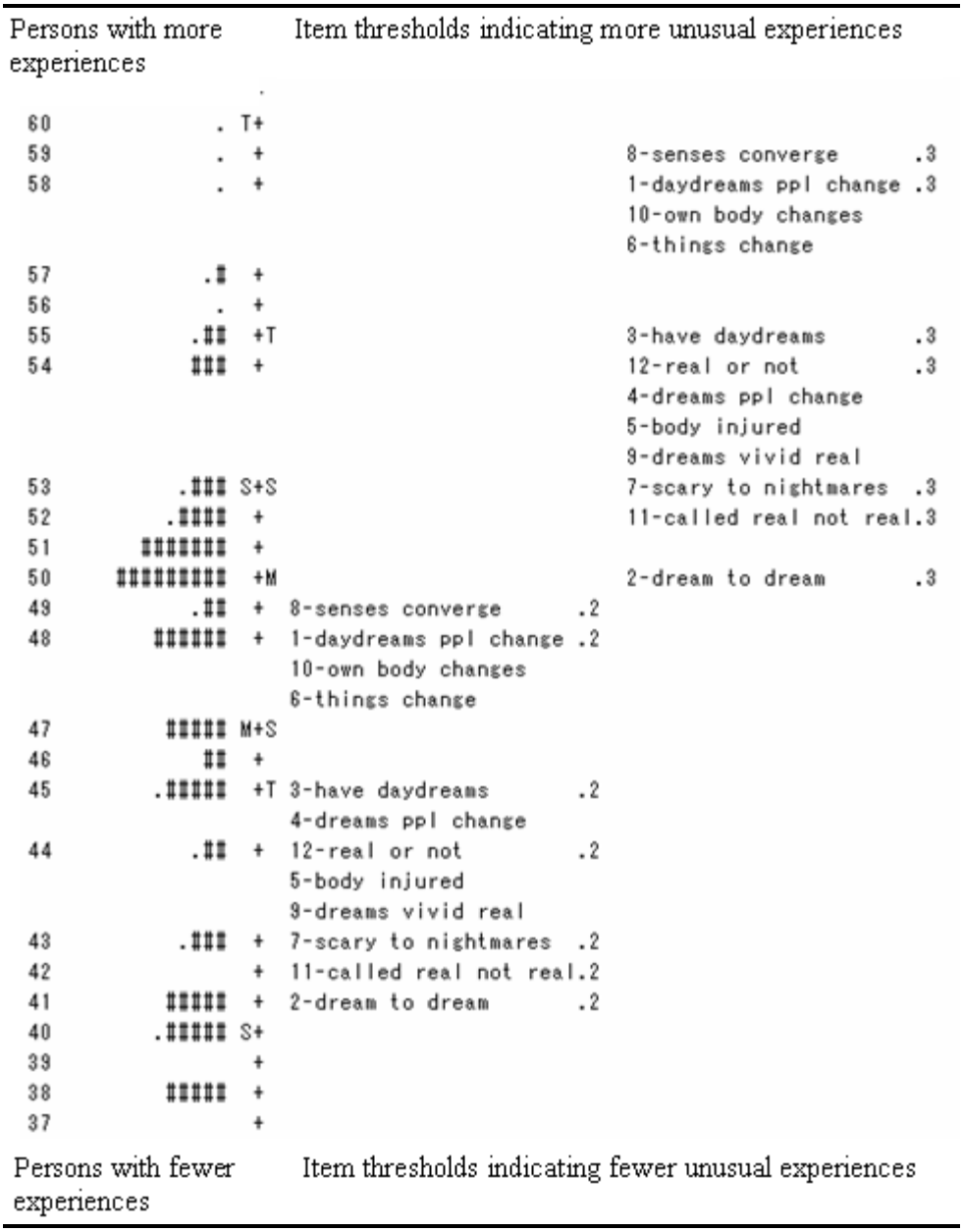
The items in the *Unusual Experiences* subscale were positioned as expected. Items 1, 6, and 10 dealt with people or things changing form, whereas Item 8 queried sensory convergence in which, for example, a person perceives a color to have sound. As expected, these items were difficult to endorse. Items 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.



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).

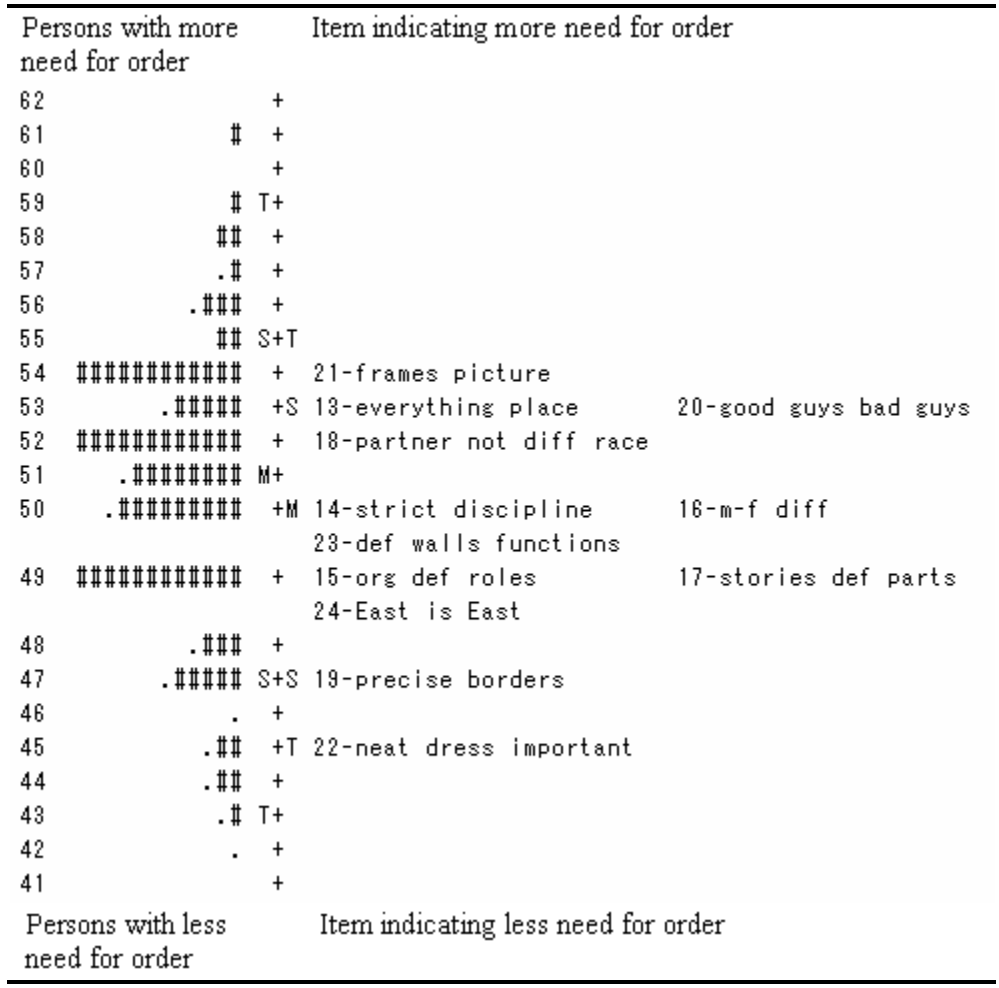
Table 59
Ego Permeability, Need for Order Subscale Measure: Rasch Item Fit Statistics

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



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 $\alpha = .70$). 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

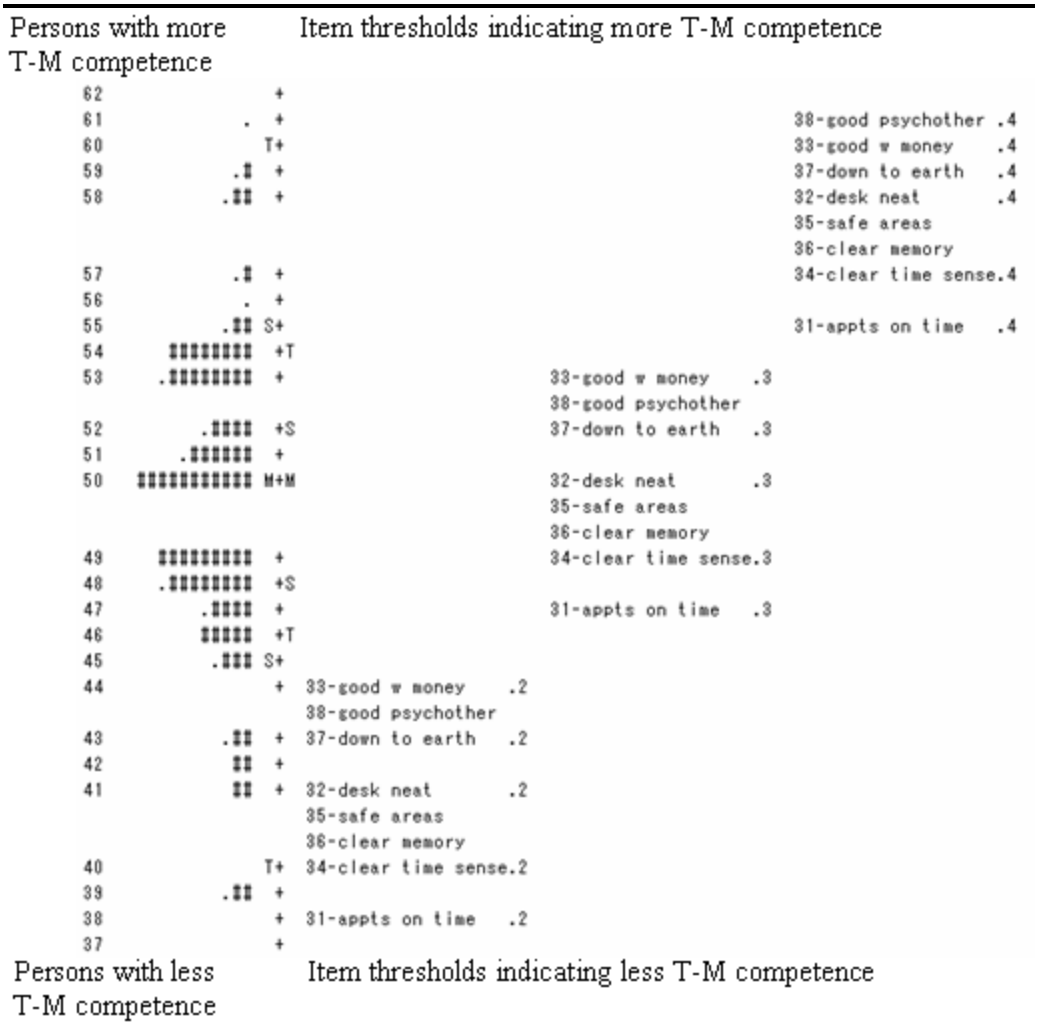
31, *I get to appointments right on time*) were the easiest to endorse, as was expected; Japanese are generally meticulous about time.

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.



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 $\alpha = .74$). 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

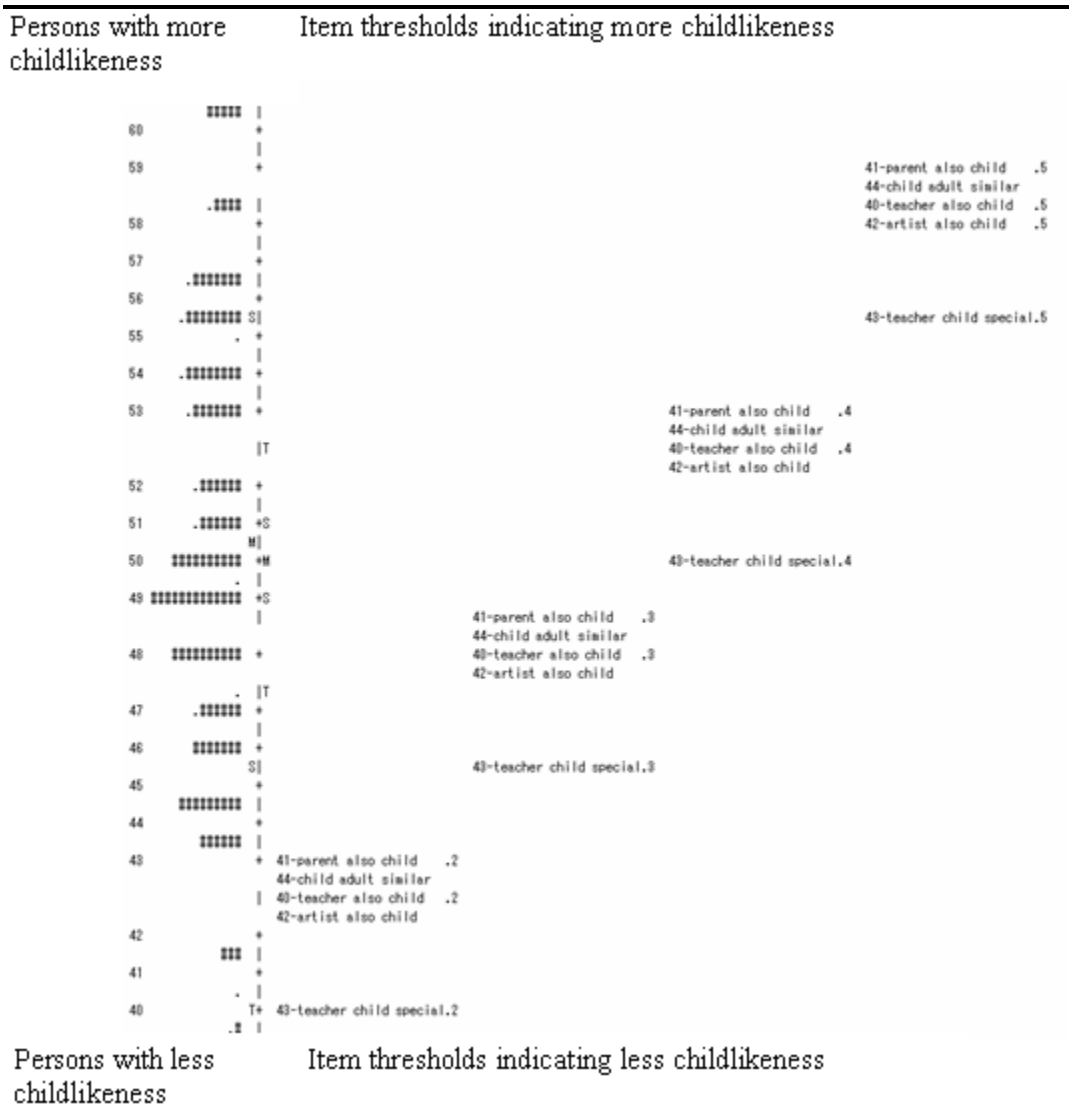
when deleted only improved the fit statistics to 2.07 and 2.53, respectively. Item 39 was thus deleted, and the Childlikeness subscale as originally postulated (Items 40-44) yielded satisfactory category function with a Rasch item reliability estimate of .91, item separation of 3.26, a Rasch person reliability 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.



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.

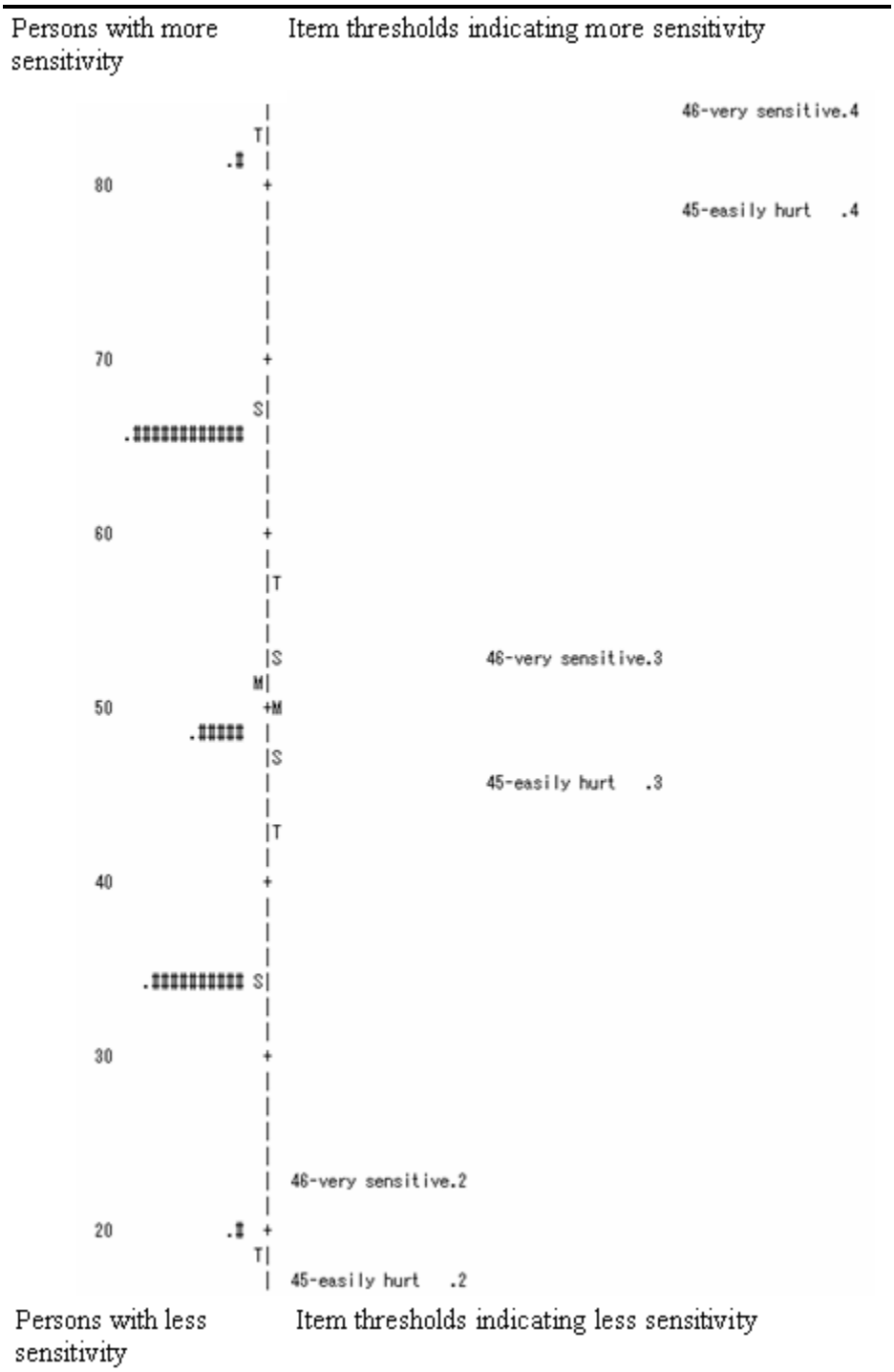


Figure 33. Item-person map with Rasch-Thurstone thresholds for the Sensitiveness subscale.

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: Unususal 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

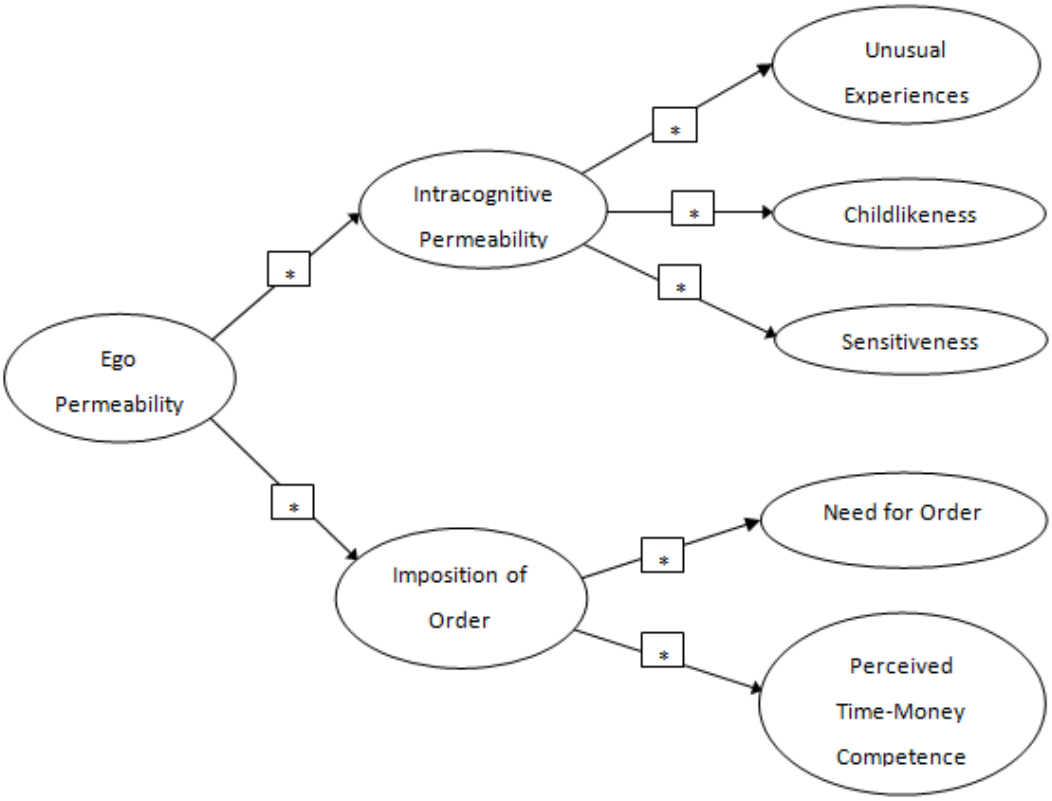


Figure 34. Hypothesized 2-factor model of Ego Permeability with Imposition of Order and Intracognitive Permeability.

[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.

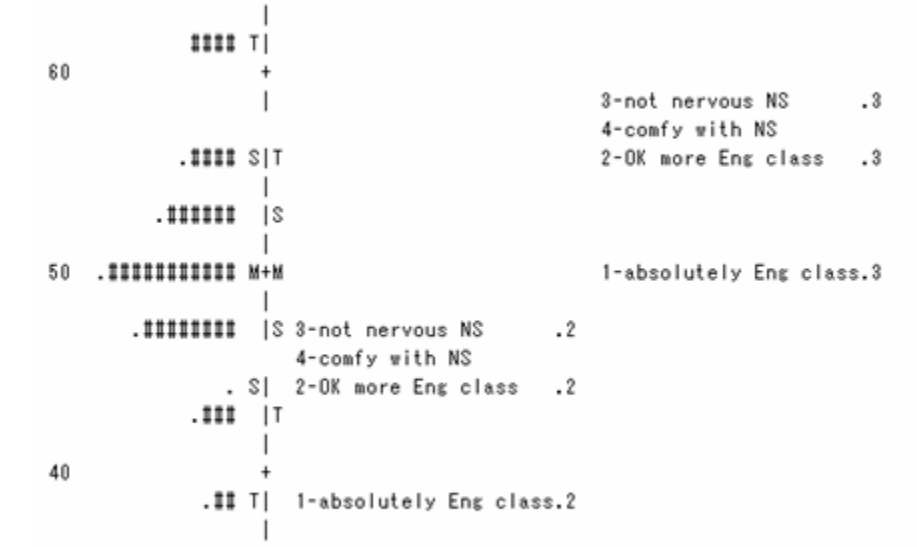


Figure 35. Item-person map with Rasch-Thurstone thresholds for the Attitudes about the Learning Situation instrument.

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).

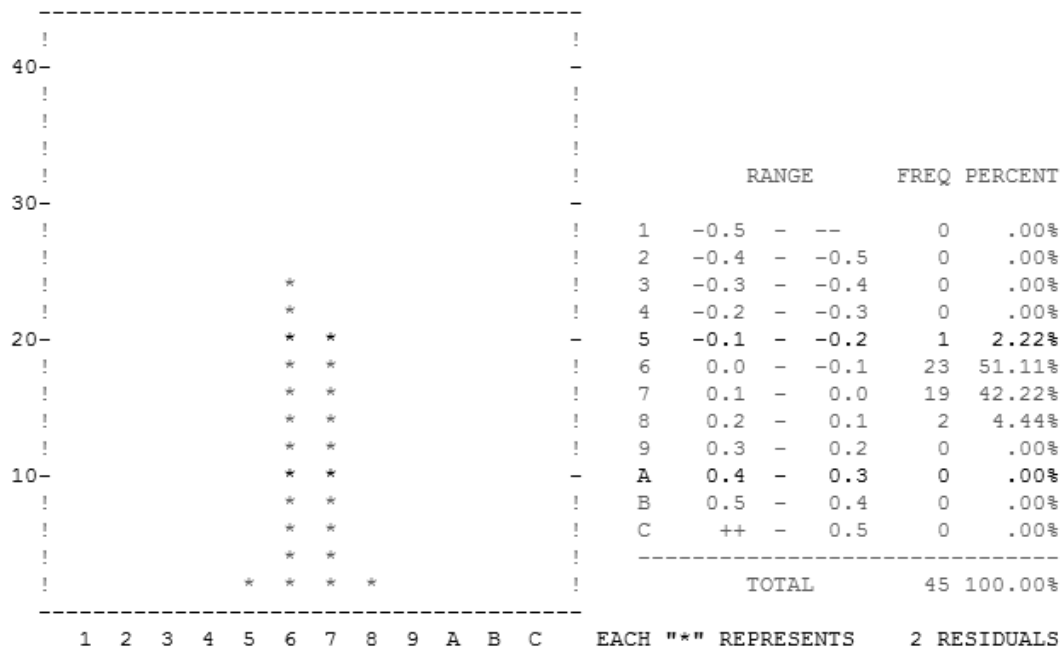


Figure 36. Distribution of standardized residuals for the Intercultural Friendship Orientation variable.

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.

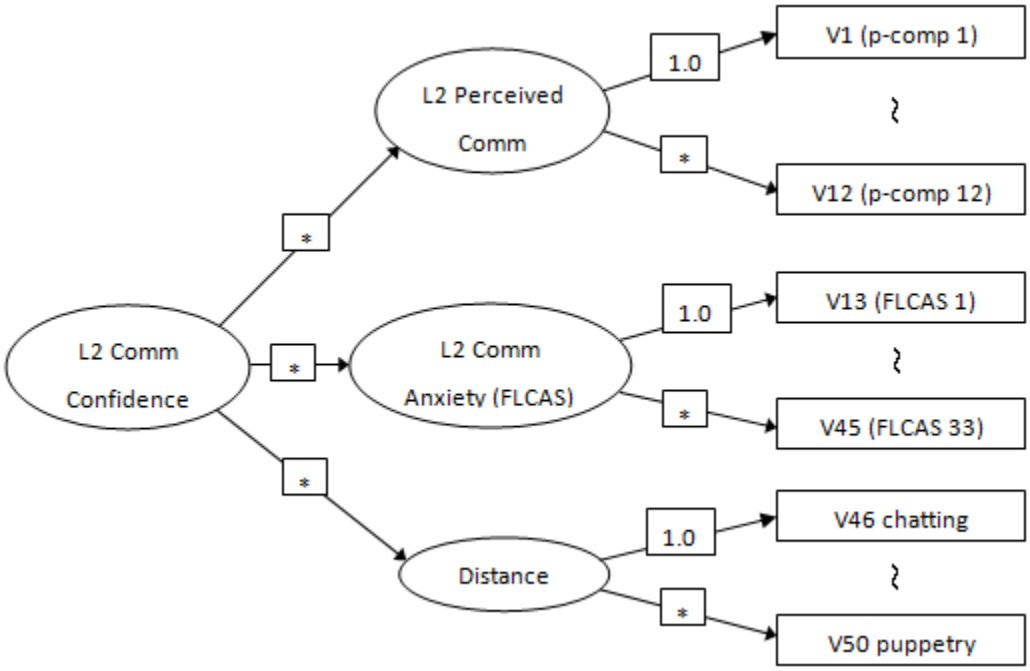


Figure 37. L2 Communicative Communication configuration with the addition of Perceived Distance. The three factors consisted of 12, 30, and 5 items, respectively, but only the first and last items are shown. Disturbances and error terms are not shown for the sake of clarity.

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).

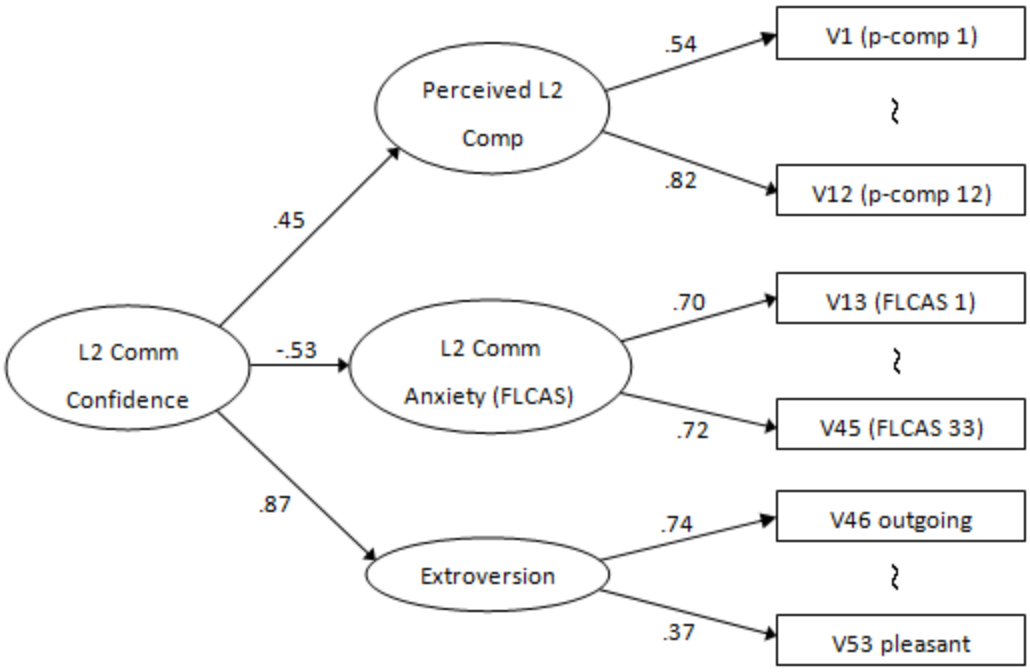


Figure 38. Standardized solution of the L2 Communicative Communication configuration with the addition of Extroversion. The three factors consisted of 12, 30, and 8 items, respectively, but only the first and last items are shown. Disturbances and error terms are also not shown for the sake of clarity. 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.

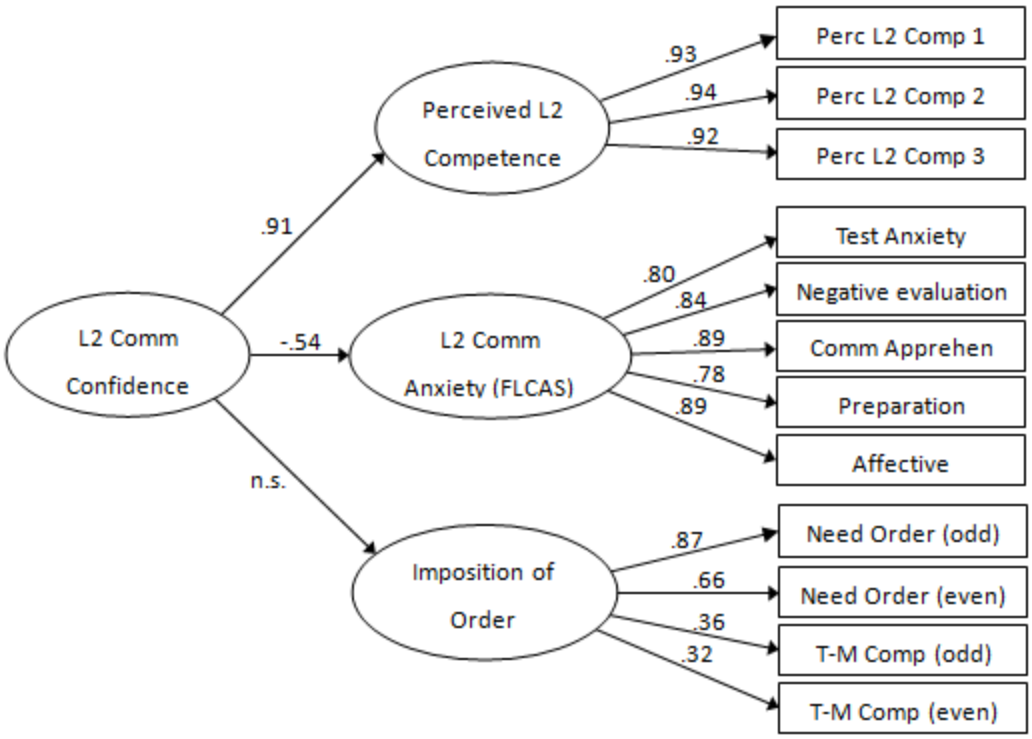


Figure 39. Standardized solution for the L2 Communicative Confidence configuration with the addition of Imposition of Order (Ego Permeability). Disturbances and error terms are not shown for the sake of clarity. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2 = 112.980$, $p < .01$, 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

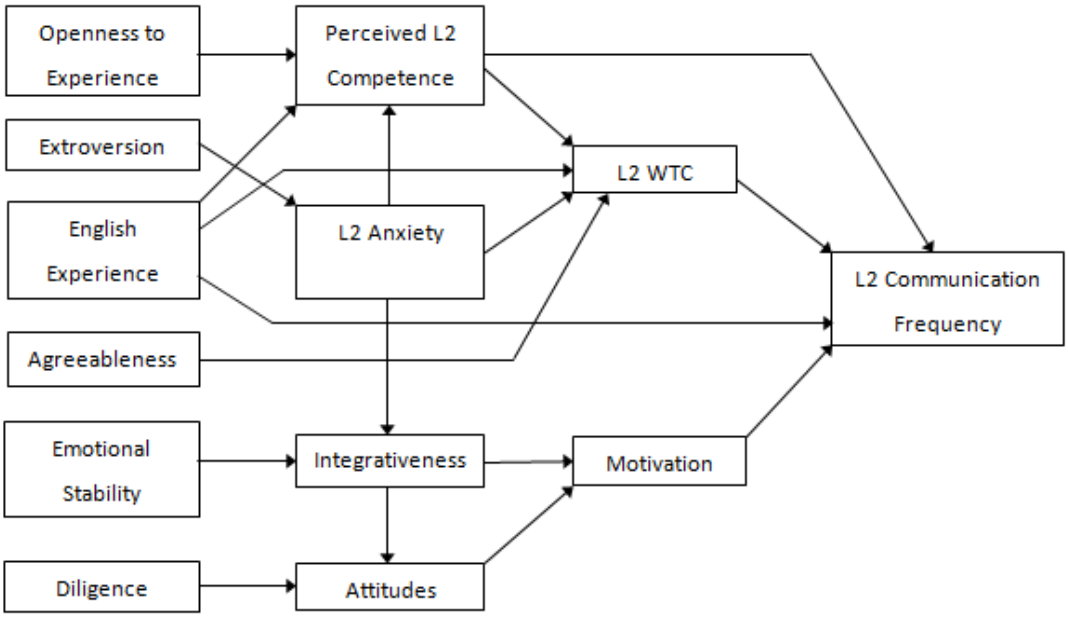


Figure 40. Revised MacIntyre and Charos (1996) model of L2 Willingness to Communicate. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre, and C. Charos, 1996, *Journal of Language and Social Psychology, 15*(1), p. 12. Copyright 1996 by *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

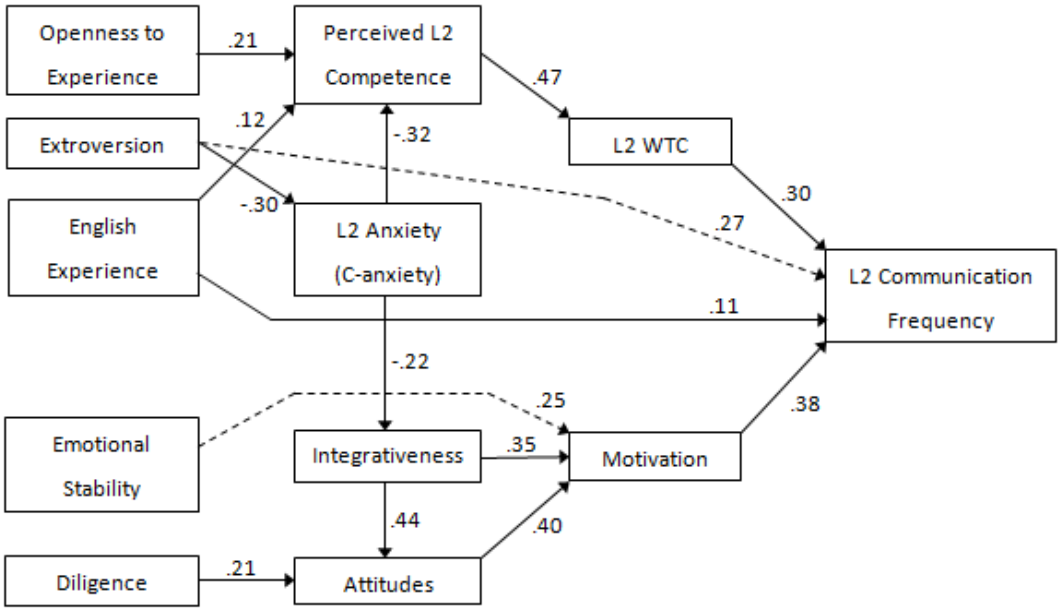


Figure 41. Standardized solution of the revised path-analytic model using communicative anxiety: Personality, attitudes, and affect as predictors of foreign language communication. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre and 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.

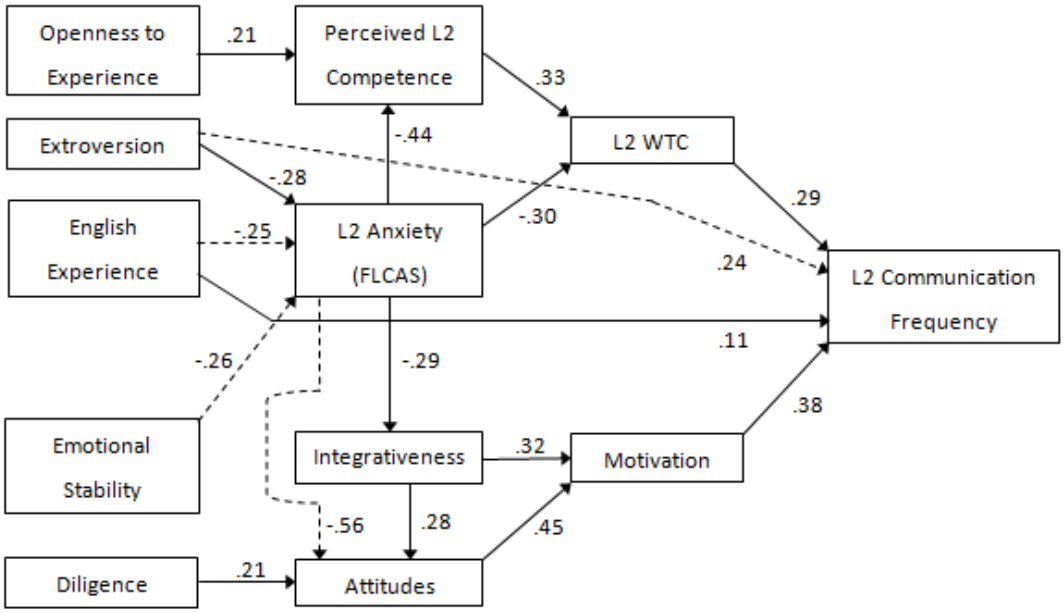


Figure 42. Revised path-analytic model using FLCAS: Personality, attitudes, and affect as predictors of foreign language communication. Data-driven additions to the model are shown as dashed lines. Adapted and reprinted with permission. Numerical values indicate that path coefficients were significant 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.

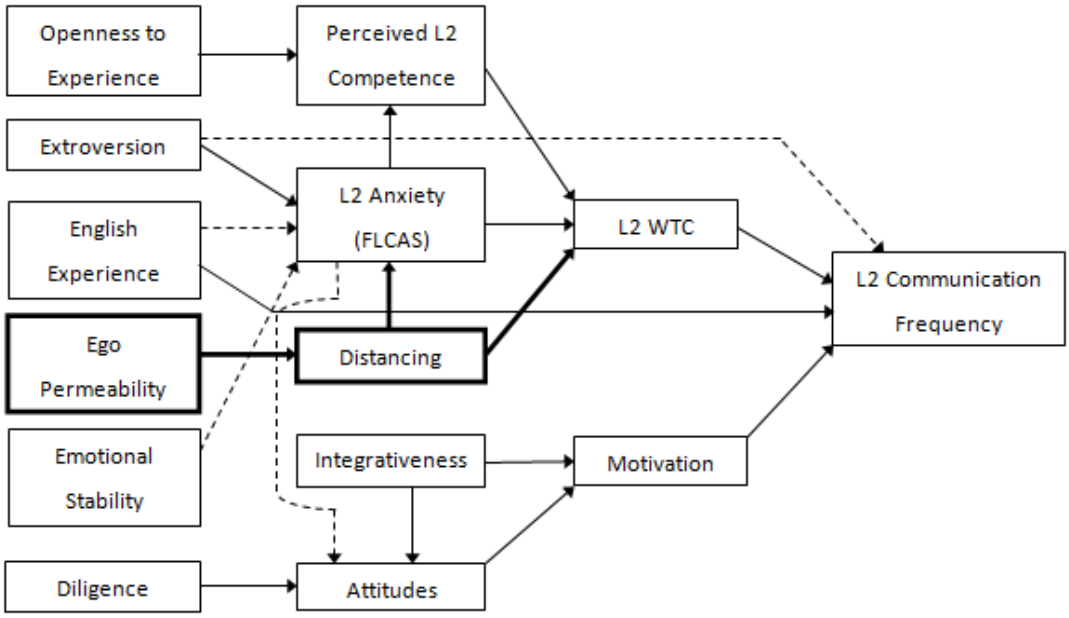


Figure 43. Hypothesized model of L2 communication with ego permeability and distancing added. Dashed lines represent data-driven additions. Bold lines show the hypothesized additions. Adapted from "Personality, Attitudes, and Affect as Predictors of Second Language Communication," by P. D. MacIntyre, 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

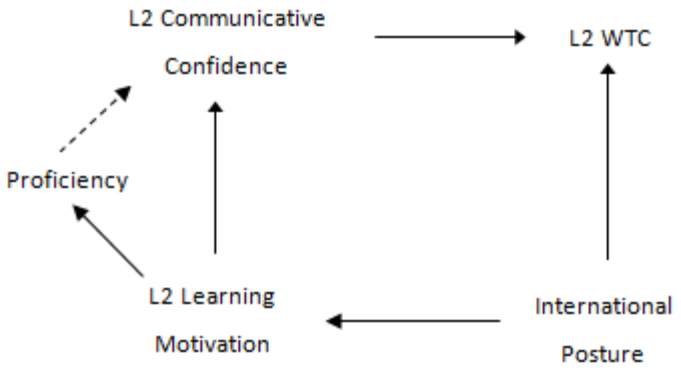


Figure 44. Core of the Yashima (2002) L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal, 86*(1), 61. Copyright 2002 by *The Modern Language Journal.* Reprinted by permission. Note that the dashed 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

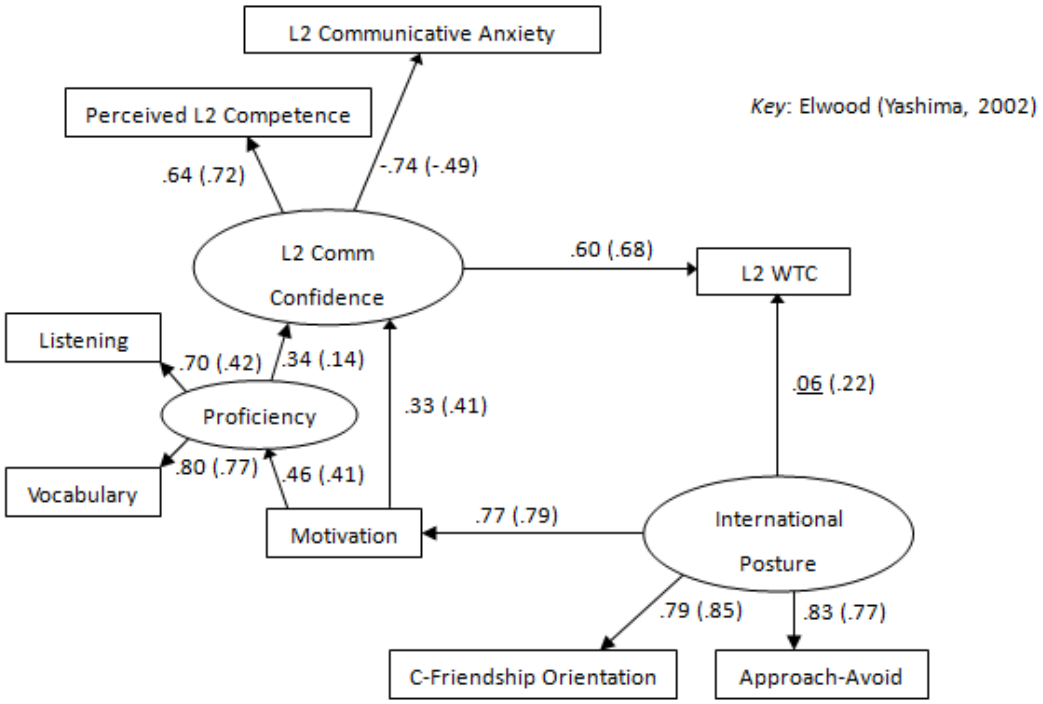


Figure 45. Standardized solution of the original Yashima (2002) model of L2 communication with standardized estimates. Numerical values list the value from the current first and the value from Yashima (2002) parenthetically. Path coefficients were significant at $p < .01$ with the exception of the underlined 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.

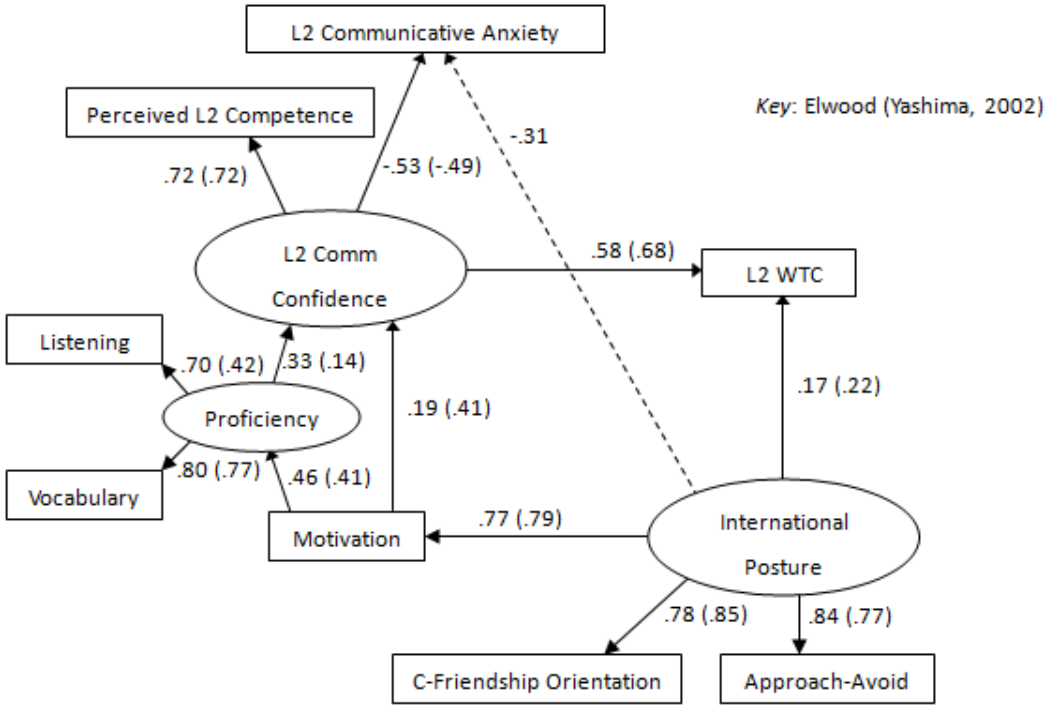


Figure 46. Respecified original model of L2 communication with standardized estimates. Numerical values list the value from the current first and the value from Yashima (2002) parenthetically. Path coefficients were significant at $p < .01$. χ^2 (15) = 27.759, p = .023, CFI = .980, RMSEA = .058, 90% 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.

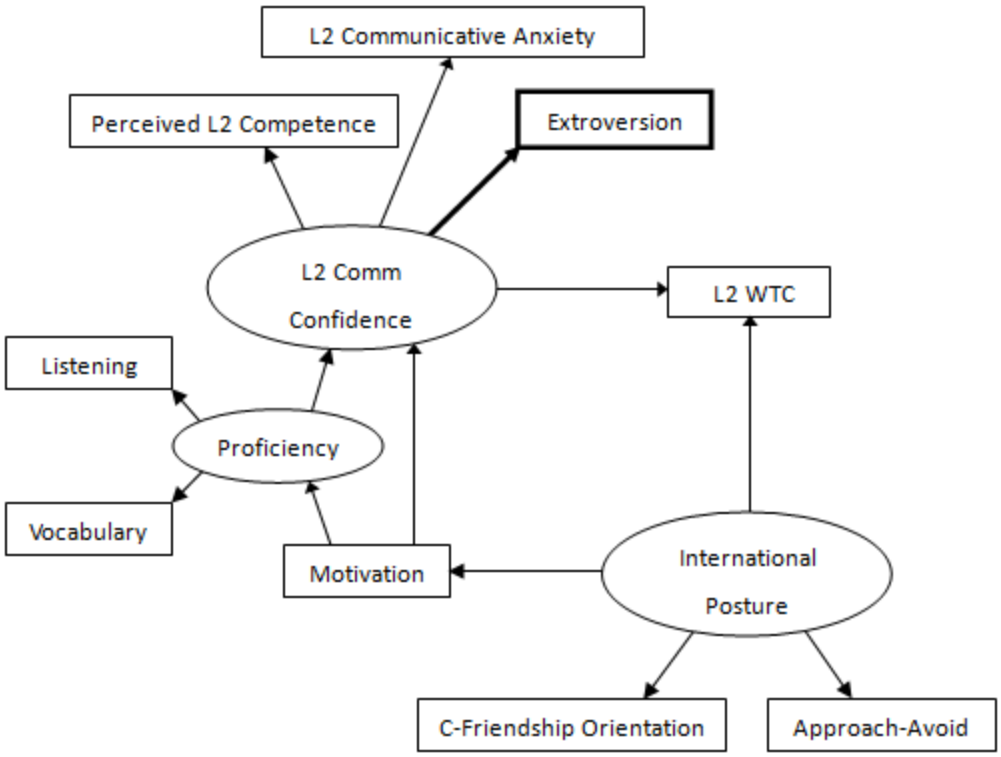


Figure 47. Revised Yashima (2002) L2 communication model. From "Willingness to Communicate in a Second Language: The Japanese EFL Context," by T. Yashima, 2002, *The Modern Language Journal*, *86*(1), 61. Copyright 2002 by *The Modern Language Journal.* Reprinted by permission.

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.

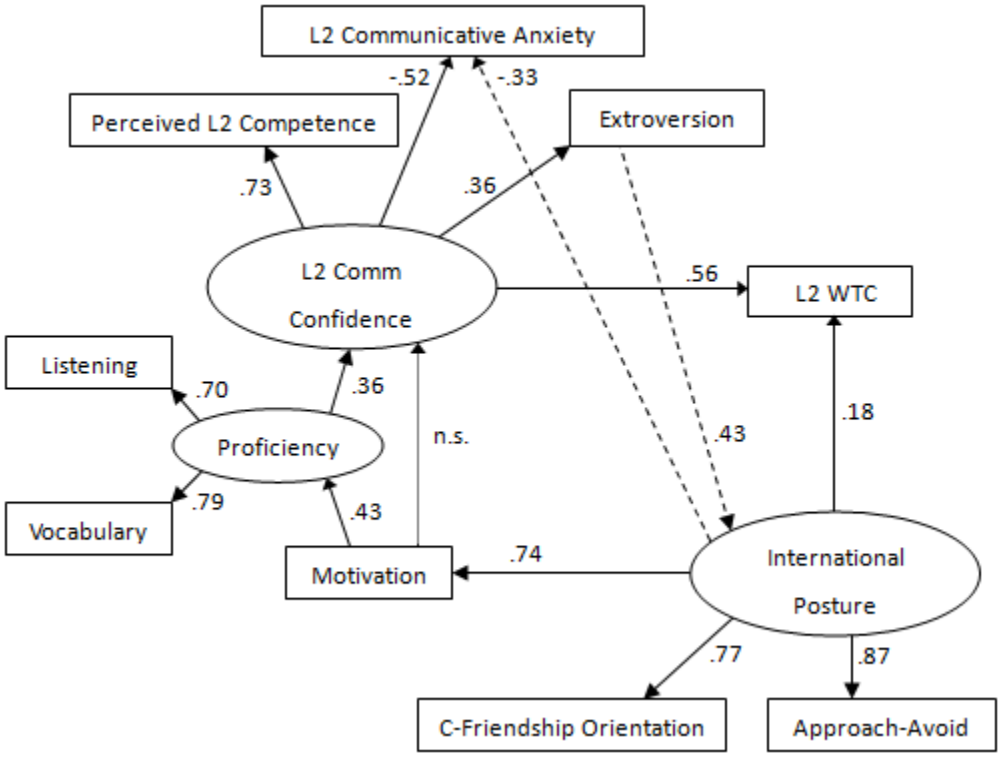


Figure 48. Standardized solution of the revised Yashima (2002) model. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2(21) = 44.31$ ($p < .01$), CFI = .968, RMSEA = .067, 90% C.I. = .039-.094.

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.

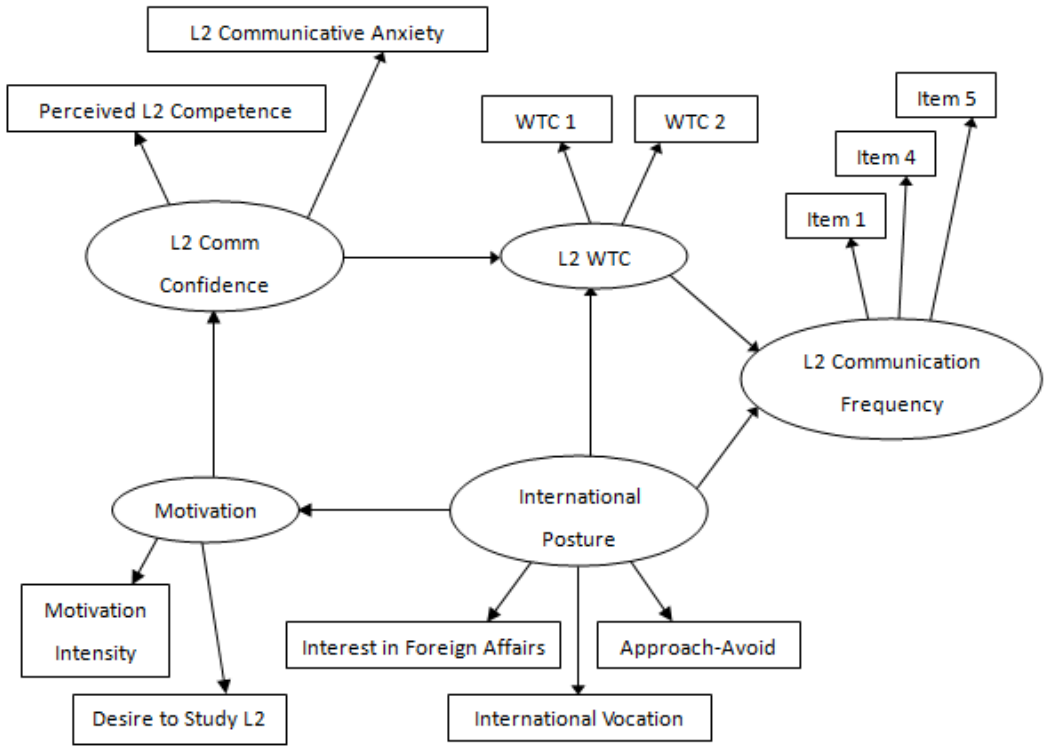


Figure 49. Model of L2 communication. Reprinted from "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Communication," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning, 54*(1), p. 134. Copyright 2004 by *Language Learning.* Reprinted with 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.

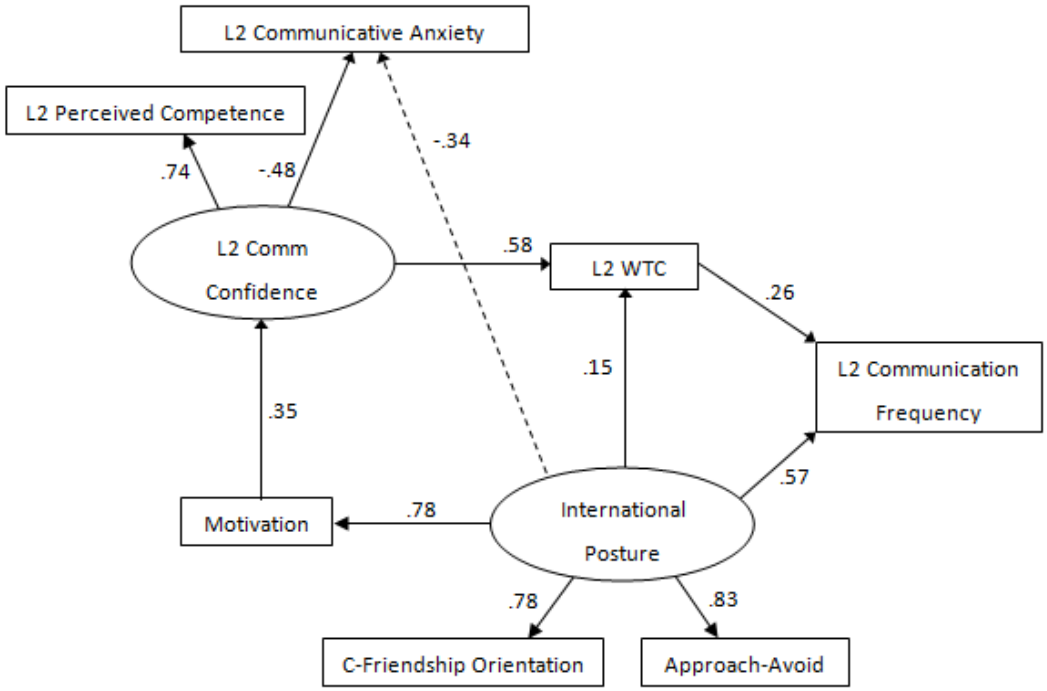


Figure 50. Results of SEM: Respecified revised model of L2 communication with standardized estimates. Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2 = 29.754$, $p < .01$, CFI = .970, RMSEA = .089, 90% C.I. = .053-.126.

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

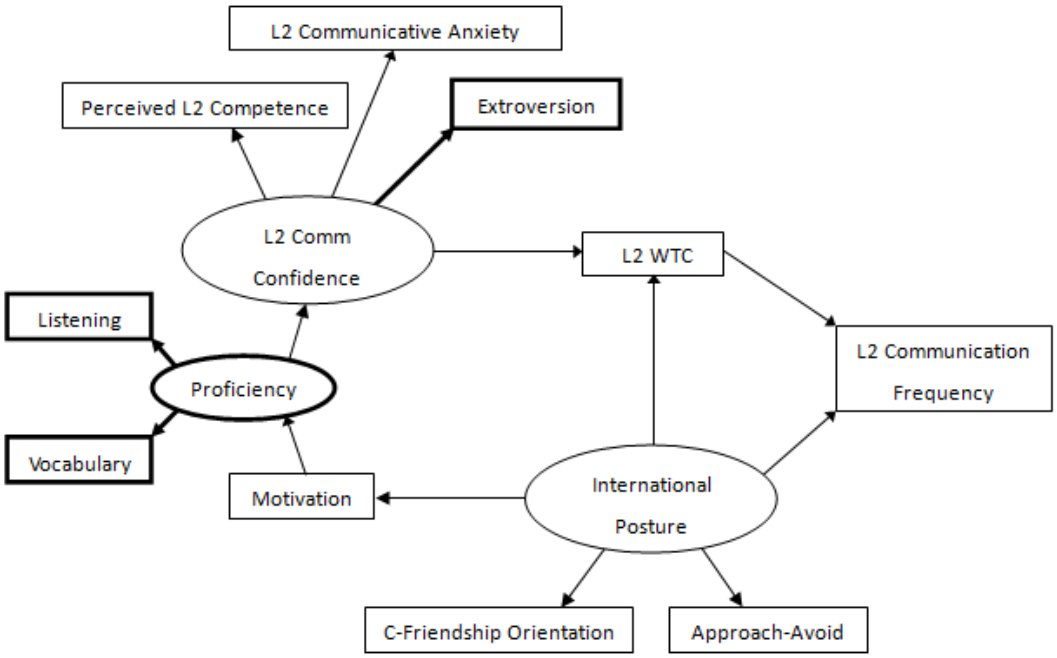


Figure 51. Revised model of L2 communication based on Yashima et al. (2004). Adapted from "The Influence of Attitudes and Affect on Willingness to Communicate and Second Language Communication," by T. Yashima, L. Zenuk-Nishide, and K. Shimizu, 2004, *Language Learning*, *54*(1), p. 134. Copyright 2004 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

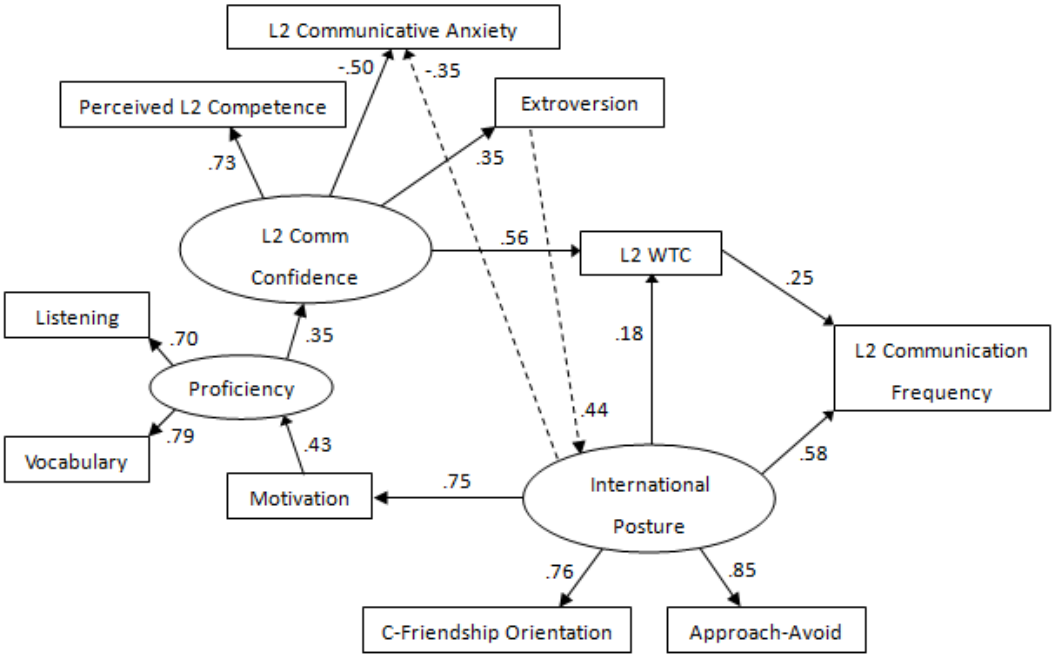


Figure 52. Standardized solution of the revised model of Yashima et al. (2004). Numerical values indicate that path coefficients were significant at $p < .01$. $\chi^2(29) = 59.656$ ($p < .01$), CFI = .965, RMSEA = .065, 90% C.I. = .041-.088.

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