

# A Rasch Model Analysis of the “Four L2 Anxieties”

Matthew T. Apple

## Introduction

Foreign or second language (L2) anxiety is considered as a situation-specific variable that arises within foreign language contexts, related to but separate from general trait/state anxiety and test anxiety. However, existing instruments that measure L2 anxiety have not been examined for item fit or unidimensionality and frequently produce different results from study to study, even within similar contexts. In addition, claims that different language-skill related L2 anxieties are independent have been based on traditional forms of statistical investigation that rely on correlational analysis to determine construct validity. The lack of valid unidimensional measurement instruments poses a problem for L2 teachers who need to identify sources of anxiety for their students. The present study aims to investigate (1) the construct validity of instruments measuring the “four L2 anxieties” and test anxiety, and (2) to what degree the four anxieties and testing anxiety are distinct constructs.

## Background of the Issue

### *Definitions of Anxiety*

The concept of anxiety as both a personality trait as well as a temporary state was established almost half a century ago and has been measured traditionally by the State-Trait Anxiety Inventory (STAI, Spielberger 1983). Gardner (1985) was one of

---

M.T. Apple (✉)  
Ritsumeikan University, Kyoto, Japan  
e-mail: mapple@fc.ritsumei.ac.jp

the first second language acquisition (SLA) researchers to develop an instrument solely designed to measure foreign language anxiety—a short five-item scale called “French classroom anxiety”—but it was the study by Horwitz et al. (1986) that popularized the term “foreign language anxiety” for L2 teaching. Horwitz et al. argued that L2 anxiety in the classroom was composed of three primary components—communication apprehension, fear of negative evaluation by the teacher, and test anxiety—and created the Foreign Language Classroom Anxiety Scale (FLCA), with 33 positively and negatively worded items.

Based on a series of studies conducted in the late 1980s and early 1990s (e.g., MacIntyre and Gardner 1989, 1991) L2 anxiety has been viewed as a “situation-specific” anxiety, e.g., a type of state anxiety that occurs only when using one’s L2 in certain social situations. Based on factor analysis, MacIntyre and Gardner (1991) argued that test anxiety and “general anxiety” (which they defined based on a combination of items concerning L1 usage in various social situations) differed from L2 anxiety. Other studies since then have proposed L2 skills-related anxieties and created new L2 anxiety instruments, such L2 speaking anxiety (Apple 2013), L2 listening anxiety (Kim 2000), L2 reading anxiety (Saito et al. 1999), and L2 writing anxiety (Cheng et al. 1999). However, problems remain in SLA anxiety research. Dewaele (2013) pointed out that existing L2 anxiety instruments do not correlate well, and the assumption that L2 anxiety is independent of trait anxiety is somewhat suspect. In fact, evidence that L2 anxiety is strongly related to first language (L1) anxiety and trait anxiety has already been attested in communication studies (Jung and McCroskey 2004), though seemingly ignored by SLA researchers. In addition, the language used in L2 anxiety measurement instruments, compounded by a lack of item fit analysis and construct validity, is a problem for cross-sample validity of L2 anxiety instruments.

### *Conceptual Problems*

A major issue concerning L2 anxiety is the very language used to describe it. As previously mentioned, the “trait” versus “state” distinction has been suggested. However, previous research has cast doubt upon the validity of the STAI (e.g., Kaipper et al. 2010; Tenenbaum et al. 1985). The terms “facilitative” as well as “debilitative” have been used to describe types of L2 anxiety, based on the idea that anxiety can both help and hinder language performance (e.g., Price 1991; Young 1999). However, the items used to measure L2 anxiety contain words that may or may not pertain to the measurement of anxiety, for example: afraid, annoyed, concerned, frightened, frustrated, nervous, panic, tense, uncertain, uncomfortable, uneasy, upset, and worried. Compared to a dictionary-definition of anxiety (Fig. 1),

**Fig. 1** Anxiety definitions from Merriam-Webster (Source <http://m-w.com>)

- 1a** : painful or apprehensive uneasiness of mind usually over an impending or anticipated ill
- b** : fearful concern or interest
- c** : a cause of anxiety
- 2** : an abnormal and overwhelming sense of apprehension and fear often marked by physiological signs (as sweating, tension, and increased pulse), by doubt concerning the reality and nature of the threat, and by self-doubt about one’s capacity to cope with it

these words sometimes exemplify and sometimes are only tangentially related with traditional ideas of what constitutes “anxiety.” Key components missing from L2 anxiety measurement instruments are the anticipation of something painful and the idea that anxiety is abnormal. Physical descriptions (e.g., sweating, increased heart rate) are also typically missing. Thus, even before considering measurement problems, there are already conceptual, and thus content validity, concerns in L2 anxiety instruments.

### *Measurement Problems*

In addition to content validity, an overriding concern regarding the use of Likert-scale questionnaire instruments to measure L2 anxiety is construct validity, without which there is no way of knowing whether the results obtained tell us that the construct has actually been measured rather than the result of errors, disturbances, and other random statistical noise. An additional requirement of the L2 anxiety measurement instruments that has not yet been confirmed is unidimensionality; each L2 anxiety measurement instrument should measure only one construct, rather than several related constructs. L2 anxiety researchers who use multidimensional construct instruments cannot be certain which data come from which constructs, making generalization of their findings extremely difficult, if not impossible.

Traditionally, L2 anxiety measurement instruments have been “validated” through correlation to existing instruments and split-half reliability analysis using Cronbach’s alpha. However, neither correlational analysis nor Cronbach’s alpha measure construct validity, facts which have been known for nearly half a century (Green et al. 1977; Cortina 1993). Excepting for L2 speaking anxiety (Apple 2013), the separate L2 instruments have not been subject to item fit analysis and unidimensionality of construct has not been validated. Finally, the relationship among the four L2-skills anxieties and testing anxiety still need to be clarified.

## Methods

### *Participants*

This study comprised 315 first and second-year students in 12 intact EFL classes at two undergraduate universities in Kyoto, Japan. Fourteen questionnaires were discarded due to incomplete answers, for a total sample size of  $N = 298$ . There were 150 male and 148 female students; 126 were first year students and 172 were second-year students. Study participants had an intermediate English proficiency level as measured by the TOEIC ( $M = 569.34$ ,  $SD = 108.67$ ) and comprised three majors: social sciences (30.9 %), letters/humanities (39.3 %) and economics (29.9 %).

### *Instruments*

A Likert-type questionnaire instrument<sup>1</sup> with six points (1 = strongly disagree, 6 = strongly agree) was used, with 51 items chosen from five existing L2 anxiety instruments:

1. L2 speaking anxiety ( $k = 11$ ; Apple 2013).
2. L2 listening anxiety ( $k = 11$ ; Kimura 2008).
3. L2 reading anxiety ( $k = 11$ ; Matsuda and Gobel 2004).
4. L2 writing anxiety ( $k = 10$ ; Cheng et al. 1999).
5. Test anxiety ( $k = 8$ ; In'nami 2006).

Of the five instruments, only one (Apple 2013) had been created and analyzed using Rasch Analysis prior to the study. Aside from Apple (2013), the original instruments were constructed according to traditional statistic-based methods, i.e., using large numbers of items to drive up the Cronbach's alpha estimate and analyzed according to factor analysis or correlational analysis using Pearson's coefficient. L2 Listening Anxiety originally consisted of 33 items. L2 Reading Anxiety originally consisted of 20 items.<sup>2</sup> L2 Writing Anxiety originally comprised 25 items. Testing Anxiety was originally a combined questionnaire with items from two separate instruments totalling 57 items (see Discussion for more on this instrument). In the interests of reducing questionnaire fatigue, only the top 11 items ( $<0.40$ ) from each instrument were selected based on existing traditional factor analysis (EFA) factor loadings.

### *Procedures*

A bilingual English–Japanese questionnaire was distributed via SurveyMonkey, an online survey tool that allowed study participants to complete the survey using

smartphone-based browser software. Three L1 teachers of English assisted the researcher in collecting data; teachers were asked to read the directions out loud (in English) to students, encouraged students to select responses honestly and without discussing with their classmates, and informed students that the survey was voluntary and would not affect their course grades. No student chose to opt out of the survey. Data obtained from the questionnaire were analyzed in WinSteps 3.75 using the Rasch rating scale model (Andrich 1978), which is a polytomous data form of the Rasch model (Rasch 1960). To judge Likert category functioning, the four criterion of Linacre (2002) were used: (1) At least 10 observations should be present for each step of the scale, (2) average person measures for each step should be higher than the average person measures of the previous step, (3) outfit mean squares of each step should be less than 2.0, and (4) gaps in step difficulties should be no fewer than 0.59 and no greater than 5 logits.

The criterion to determine item fit were mean squares from 0.7 to 1.3 (Bond and Fox 2007), with 1.0 denoting a perfect fit to the Rasch model’s expectations. Item-person maps were additionally requested as visual confirmation of the constructs. Rasch principal component analysis (PCA) of construct residuals were also conducted to check unidimensionality of the individual constructs using the technique known as Rasch factor analysis (Wright 1996). The criteria used to determine construct validity were 50 % of variance accounted for by the primary dimension (the Rasch model), with secondary dimension contrasts accounting for 10 % or less variance and having less than 3.0 eigenvalues (Linacre 2007). The combination of item fit analysis and Rasch PCA can thus be used to support claims of cross-sample generalization (Wolfe and Smith 2007).

## Results

As a preliminary step, all items were simultaneously input into the Rasch model and an item-person map was requested to provide a visual image of the overall targeting of the items (Fig. 2). In general, the items were appropriately targeted for the sample population; however, the item endorsability range tended to clump around the mean of the population and there were several participants whose anxiety levels could not be adequately measured by the items. Writing anxiety items (wa) tended to be more difficult to endorse, and listening anxiety items (la) tended to be easier to endorse compared to the other three anxieties (speaking, reading, testing), whose items were more evenly distributed. Rasch item reliability was 0.98 and item separation was 6.64, and Rasch person reliability was 0.96 and person separation was 4.76 for all items.

Since the various anxiety instruments were originally meant to measure a single construct, items from each instrument were input into Rasch model. Likert category analysis was conducted to determine the appropriate usage of the six Likert categories prior to item fit analysis and Rasch PCA for individual constructs.

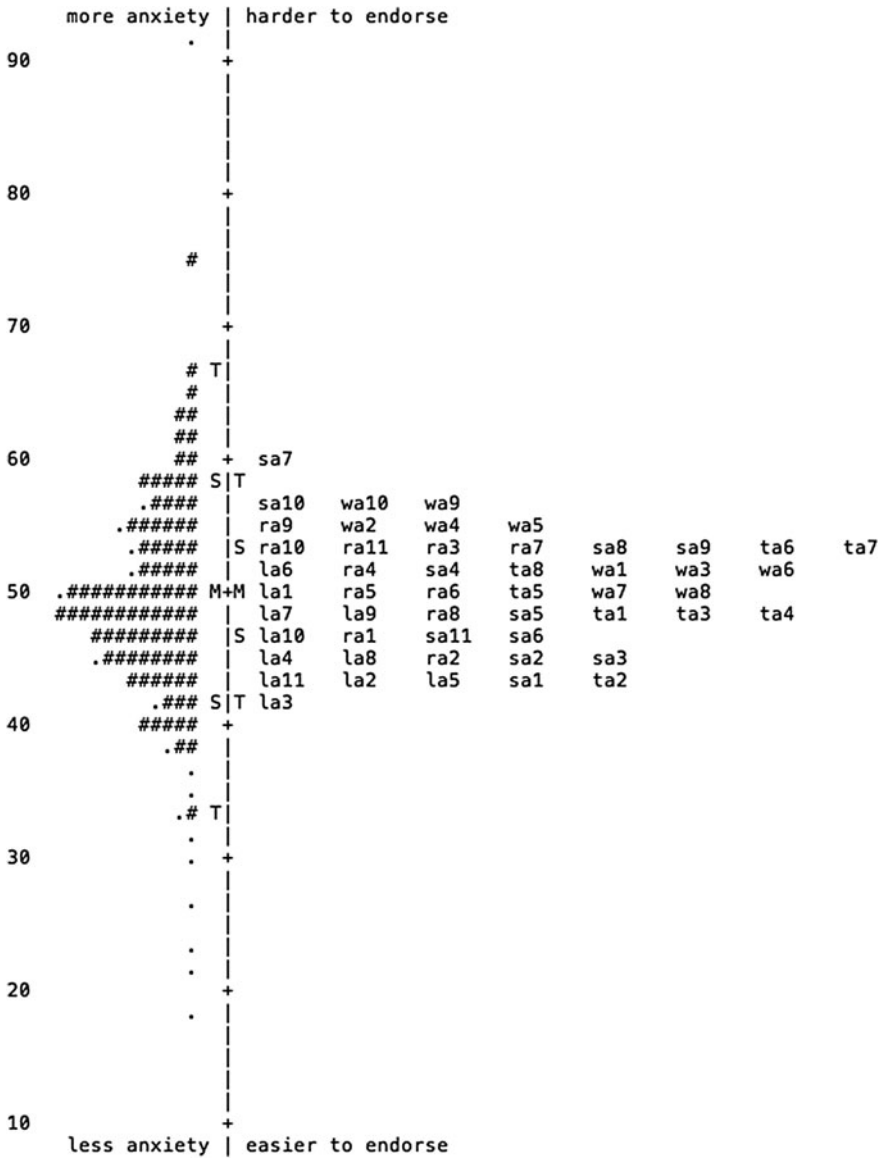


Fig. 2 All 51 items from the five anxiety-related instruments, prior to item fit analysis. *M* stands for mean. *S* stands for one standard deviation from the mean. *T* stands for two standard deviations from the mean; sa = speaking anxiety; la = listening anxiety; wa = writing anxiety; ra = reading anxiety; ta = testing anxiety; *N* = 298

### *Likert Category Utility and Item Misfit Analysis*

Likert category utility analysis showed misfit for the first category in L2 listening anxiety and for the sixth category in L2 writing anxiety, L2 reading anxiety, and Testing anxiety.<sup>3</sup> No Likert categories misfit in L2 speaking anxiety. Likert categories 1 and 2 were combined for L2 listening anxiety, and Likert categories 5 and 6 were combined for L2 writing anxiety, L2 reading anxiety, and Testing anxiety prior to item fit analysis.

Rasch model item analysis revealed that seven items misfit their intended constructs (Table 1). Item Sa1 (“I’m worried that other students in class speak English better than I do”) misfit the L2 speaking anxiety construct (infit MNSQ = 1.81, outfit MNSQ = 1.90). Item La1 (“I get stuck with one or two unfamiliar words”) misfit the L2 Listening Anxiety construct (infit MNSQ = 1.32, outfit MNSQ = 1.40). Two items misfit the L2 reading anxiety construct, item Ra8 (“It bothers me to encounter words I can’t pronounce while reading English,” infit MNSQ = 1.88, outfit MNSQ = 1.85) and item Ra10 (“By the time you get past the funny letters and symbols in English, it is hard to remember what you are reading about,” infit MNSQ = 0.64, outfit MNSQ = 0.64). Item Wa5 (“Discussing my English writing with others is unenjoyable”) misfit the L2 Writing Anxiety construct (infit MNSQ = 1.26, outfit MNSQ = 1.46). Two items misfit the Testing Anxiety construct, item Ta4 (“It seems to me that examination periods ought not to be made the tense situations which they are,” infit MNSQ = 1.96, outfit MNSQ = 2.00) and item Ta8 (“I feel less confident during tests,” infit MNSQ = 0.65, outfit MNSQ = 0.63).

For L2 Speaking Anxiety, the most difficult item and the easiest items to endorse were Sa7, “I’m afraid my partner will laugh when I speak English with a classmate in a pair” (Rasch item difficulty measure = 62.95) and Sa3, “I’m worried that my partner speaks better English than I do” (Rasch item difficulty measure = 42.21), respectively. These were also the most difficult and the easiest to endorse items for the entire questionnaire. Other “most difficult” and “easiest” items in terms of endorsability level were as follows: L2 Listening Anxiety most difficult, La6, “I get nervous and confused when I don’t understand every word in listening test situations” (Rasch item difficulty measure = 59.13), easiest, La 11, “The thought that I may be missing key words frightens me” (Rasch item difficulty measure = 45.40); L2 Reading Anxiety most difficult, Ra9, “I usually end up translating word by word when I’m reading in English” (Rasch item difficulty measure = 54.71), easiest Ra2, “When reading English, I often understand the words but still can’t understand what the author is saying” (Rasch item difficulty measure = 42.63); L2 Writing Anxiety most difficult, Wa10, “While writing in English, I’m nervous” (Rasch item difficulty measure = 56.38), easiest, Wa 8, “I worry that my English compositions are a lot worse than others” (Rasch item difficulty measure = 44.85); Testing Anxiety most difficult, Ta7, “I feel my heart beating very fast during tests” (Rasch item difficulty measure = 55.49), easiest, Ta2, “I wish examinations did not bother me so much” (Rasch item difficulty measure = 40.79).

**Table 1** Initial rasch item fit statistics for the five anxiety constructs

Item	Measure	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
<i>L2 speaking anxiety</i>						
Sa1	41.58	<b>1.81</b>	8.0	<b>1.90</b>	7.7	0.59
Sa2	42.99	1.10	1.3	1.30	3.0	0.75
Sa3	42.21	1.17	2.0	1.20	2.1	0.71
Sa7	62.95	0.91	-1.2	0.85	-1.6	0.77
Sa10	59.70	0.90	-1.2	0.87	-1.5	0.78
Sa5	48.04	0.88	-1.4	0.89	-1.3	0.79
Sa11	46.31	0.88	-1.5	0.84	-1.9	0.81
Sa9	54.41	0.84	-2.0	0.83	-2.0	0.81
Sa8	54.12	0.81	-2.4	0.81	-2.2	0.81
Sa4	52.39	0.81	-2.5	0.80	-2.4	0.82
Sa6	45.30	0.77	-3.0	0.76	-2.9	0.81
<i>L2 listening anxiety</i>						
La1	56.63	<b>1.32</b>	3.6	<b>1.40</b>	4.2	0.65
La2	46.94	1.23	2.5	1.15	1.7	0.73
La5	46.29	1.14	1.6	1.16	1.8	0.70
La3	42.80	1.09	1.0	1.00	0	0.74
La7	54.23	1.02	0.3	1.05	0.6	0.73
La4	47.23	0.99	-0.1	0.98	-0.2	0.76
La11	45.40	0.90	-1.2	0.88	-1.4	0.77
La8	48.91	0.85	-1.8	0.89	-1.3	0.78
La6	59.13	0.87	-1.7	0.88	-1.5	0.78
La9	52.49	0.83	-2.1	0.85	-1.5	0.79
La10	49.93	0.76	-3.0	0.75	-3.1	0.81
<i>L2 reading anxiety</i>						
Ra8	46.82	<b>1.88</b>	8.6	<b>1.85</b>	8.4	0.51
Ra9	54.71	1.19	2.2	1.18	2.1	0.68
Ra2	42.63	1.08	0.9	1.08	1.0	0.63
Ra1	44.55	1.06	0.7	1.06	0.8	0.71
Ra4	50.94	1.00	0	0.98	-0.3	0.74
Ra6	47.81	0.97	-0.4	0.96	-0.4	0.72
Ra11	53.64	0.88	-1.5	0.88	-1.5	0.70
Ra5	49.26	0.86	-1.8	0.85	-1.9	0.75
Ra3	53.96	0.71	-3.9	0.72	-3.8	0.77
Ra7	53.40	0.71	-4.0	0.71	-3.9	0.80
Ra10	52.29	<b>0.64</b>	-5.2	<b>0.64</b>	-5.0	0.75
<i>L2 writing anxiety</i>						
Wa5	51.47	1.26	3.0	<b>1.46</b>	4.8	0.70
Wa2	52.14	1.14	1.7	1.26	2.9	0.72
Wa10	56.38	1.19	2.2	1.21	2.4	0.72

(continued)



**Table 1** (continued)

Item	Measure	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	PMC
Wa4	51.59	1.03	0.4	1.15	1.7	0.75
Wa1	46.75	0.98	-0.2	0.98	-0.2	0.76
Wa8	44.85	0.91	-1.1	0.90	-1.1	0.78
Wa3	47.30	0.89	-1.4	0.89	-1.3	0.79
Wa9	55.89	0.89	-1.3	0.89	-1.3	0.76
Wa6	47.66	0.88	-1.5	0.88	-1.4	0.77
Wa7	45.96	0.72	-3.8	0.70	-3.8	0.81
<i>Testing anxiety</i>						
Ta4	48.68	<b>1.96</b>	9.1	<b>2.00</b>	9.2	0.45
Ta1	48.56	1.12	1.4	1.21	2.4	0.66
Ta2	40.79	1.15	1.7	1.15	1.5	0.64
Ta3	49.07	0.92	-1.0	0.96	-0.4	0.71
Ta5	49.97	0.75	-3.3	0.75	-3.3	0.77
Ta7	55.49	0.73	-3.7	0.73	-3.7	0.75
Ta6	55.02	0.71	-3.9	0.72	-3.8	0.76
Ta8	52.42	<b>0.65</b>	-4.9	<b>0.63</b>	-5.1	0.78

Notes Bolded numerals indicate misfit (-0.7 to 1.3 MNSQ criteria). sa = Speaking Anxiety; la = Listening Anxiety; wa = Writing Anxiety; ra = Reading Anxiety; ta = Testing Anxiety; N = 298

### *Rasch Item and Person Reliability and Separation*

Before removing the misfitting items, Rasch reliability and separation were obtained for the original 51 items by inputting items from each construct separately (Table 2). Rasch item reliability and separation were lowest for L2 writing anxiety (0.96, 4.72) and highest for L2 speaking anxiety (0.99, 9.39). Rasch person reliability and separation were lowest for Testing Anxiety (0.81, 2.07) and highest for L2 speaking anxiety (0.91, 3.27). In general, the five anxiety constructs showed reasonable person reliability. The person separation for four of the five constructs was below 3.00, indicating the ability of these instruments to separate participants

**Table 2** Rasch reliability and separation estimates among the anxiety constructs

Type of anxiety	k	Item		Person	
		Reliability	Separation	Reliability	Separation
L2 speaking	11	0.99	9.39	0.91	3.27
L2 listening	11	0.97	5.84	0.89	2.79
L2 writing	10	0.96	4.72	0.87	2.57
L2 reading	11	0.97	5.47	0.86	2.52
Testing	8	0.98	6.32	0.81	2.07

**Table 3** Rasch reliability and separation estimates excluding misfitting items

Type of anxiety	<i>k</i>	Item		Person	
		Reliability	Separation	Reliability	Separation
L2 speaking	10	0.99	9.71	0.92	3.36
L2 listening	10	0.97	5.55	0.88	2.71
L2 writing	9	0.96	5.09	0.86	2.46
L2 reading	9	0.97	6.23	0.86	2.48
Testing	6	0.98	6.70	0.77	1.83

into 3–4 groups. However, the person separation of Testing Anxiety was considerably lower than the L2 skills-related instruments and not ideal for an instrument related directly to testing. As a comparison, all misfitting items were removed from their respective constructs and the Rasch reliability and separation statistics were computed again (Table 3). Except for L2 Speaking Anxiety, all other constructs had lower person reliability and separation.

Rasch item-person maps were also obtained at this time to provide visual confirmation of the item hierarchy in each construct. The item-person maps used all items, prior to removing misfitting items (Figs. 3, 4 and 5). L2 speaking anxiety and L2 reading items were generally targeted appropriately for the sample population, while the mean endorsability level of L2 listening anxiety and L2 writing anxiety items were one standard deviation below the mean anxiety level of the participants. The endorsability level of testing anxiety items were almost two standard deviations below the anxiety level of participants. All five constructs showed some redundancy among items. L2 writing anxiety having five overlapping items and the least item spread, and L2 speaking anxiety having the greatest item spread.

### ***Rasch Principal Components Analysis of Residuals***

Misfitting items were kept for the initial Rasch Principal Components analysis of residuals (PCA or PCAR). Each individual construct was input separately to check construct dimensionality (Table 4). Rasch PCAR indicated support for unidimensionality across all five separate constructs, as the Rasch model explained well above 50 % for each construct and less than 10 % on principal contrasts. L2 Speaking Anxiety was the strongest construct (eigenvalue = 52.2, variance accounted for = 82.6 %), and L2 Reading Anxiety (eigenvalue = 17.3, variance accounted for = 61.1 %) and Testing Anxiety were the weakest (eigenvalue = 12.6, variance accounted for = 61.2 %). The eigenvalue was at or slightly above 2.0 for the principal contrast on three constructs (L2 Speaking Anxiety, L2 Listening Anxiety, and L2 Writing Anxiety); however, no contrasts explained more than 10 % variance for any construct. The principal contrast for Testing Anxiety explained 8.9 % of the variance, and the variance accounted for by the principal

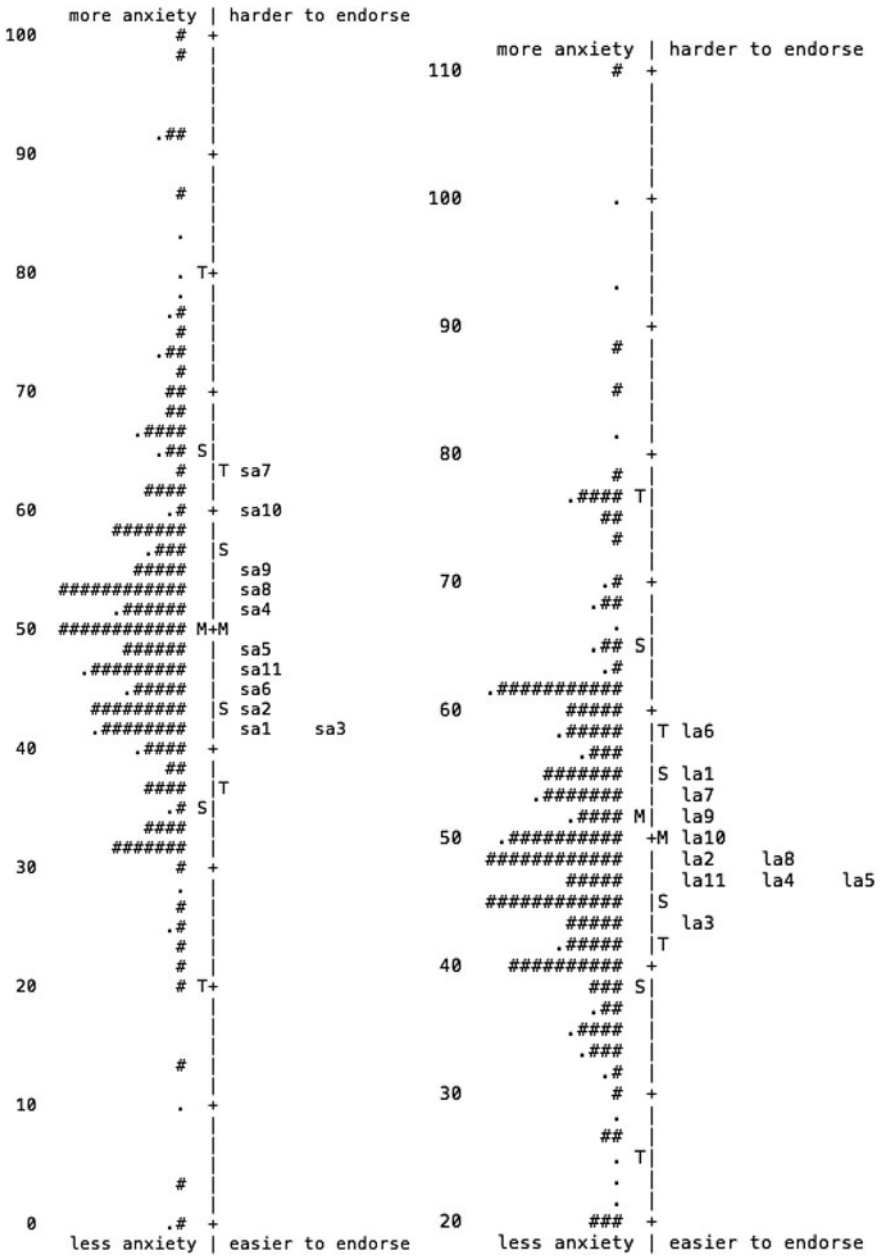


Fig. 3 The item-person maps of L2 speaking anxiety (left) and L2 listening anxiety (right). Each # is two people. N = 298

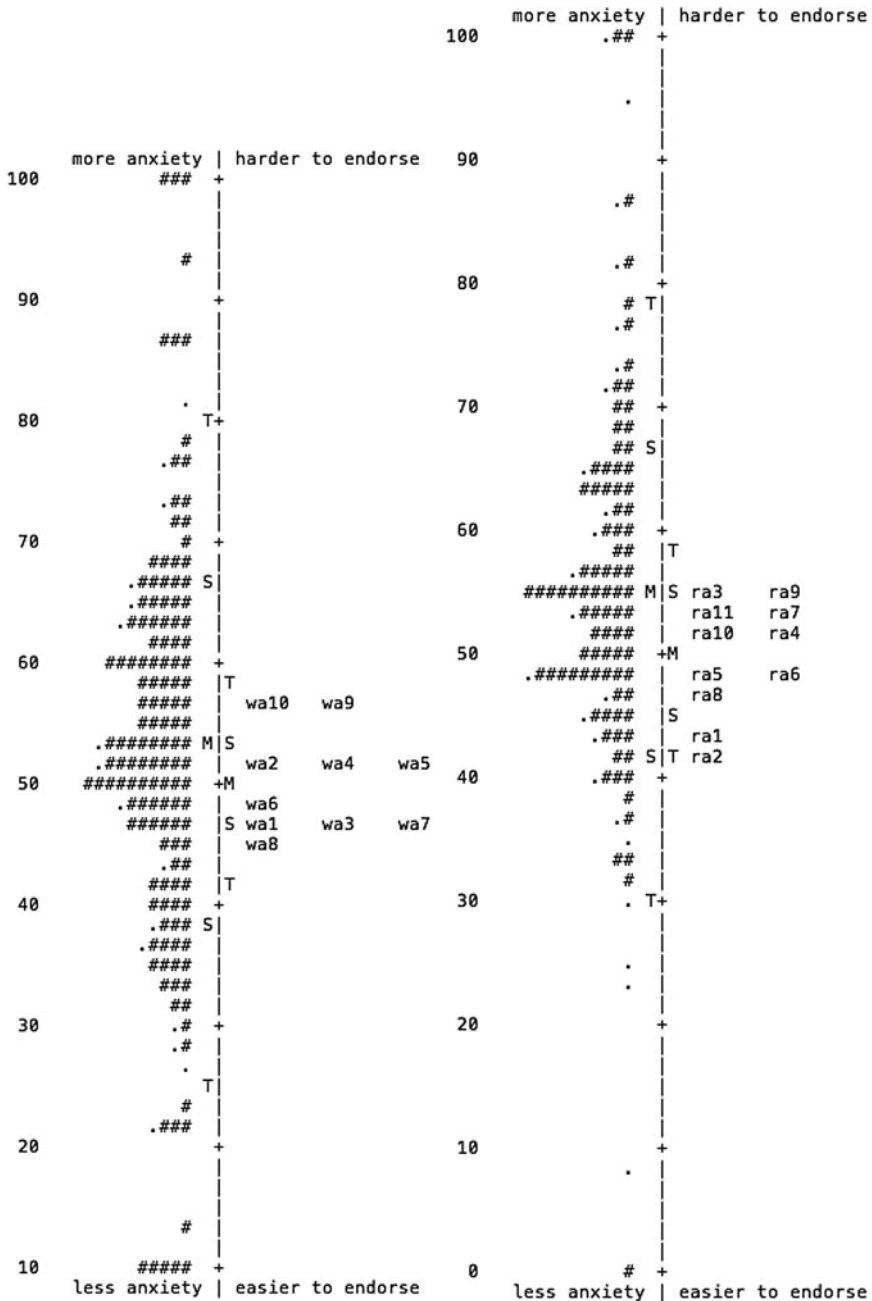
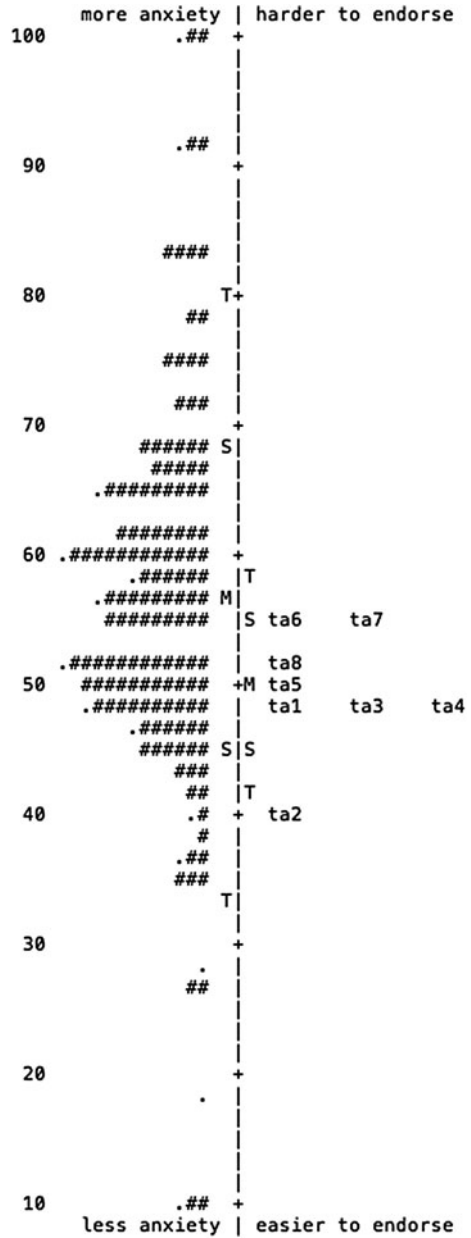


Fig. 4 The item-person maps of L2 Writing Anxiety (left) and L2 Reading Anxiety (right). Each # is two people for L2 Writing Anxiety and three people for L2 Reading Anxiety.  $N = 298$

**Fig. 5** The item-person map of Testing Anxiety. Each # is two people.  $N = 298$



contrasts to L2 Listening Anxiety, L2 Writing Anxiety, and L2 Reading Anxiety were above 5.0, a possible indication of measurement noise.

As with Rasch reliability and separation analysis, all seven misfitting items were removed and Rasch PCAR was conducted again on each construct for a comparison

**Table 4** Rasch PCA eigenvalues and variance accounted for ( $k = 51$ )

Type of anxiety	Rasch model		Principal contrast	
	Eigenvalue	% of var	Eigenvalue	% of var
L2 speaking	52.2	82.6	2.4	3.8
L2 listening	25.5	69.9	2.0	5.4
L2 writing	20.3	67.0	2.1	6.9
L2 reading	17.3	61.1	1.7	6.1
Testing	12.6	61.2	1.9	9.3

**Table 5** Rasch PCA eigenvalues and variance accounted for, excluding misfitting items ( $k = 44$ )

Type of anxiety	Rasch model		Principal contrast	
	Eigenvalue	% of var	Eigenvalue	% of var
L2 speaking	58.1	85.3	2.1	3.0
L2 listening	21.7	68.4	1.9	5.9
L2 writing	18.0	66.7	2.2	8.0
L2 reading	20.9	69.9	1.5	5.0
Testing	11.5	65.8	1.9	10.9

(Table 5). The eigenvalue and variance accounted for by the Rasch model was higher in L2 Speaking Anxiety and L2 Reading Anxiety but lower in L2 Listening Anxiety, L2 Writing Anxiety, and Testing Anxiety.

Although Rasch model analysis is designed for individual constructs, Rasch PCAR can also provide some indication of relatedness among constructs with a single dataset by examining whether the contrasting dimensions to the Rasch model cohere into separate dimensions or comprise random noise. Therefore, as an initial test of relatedness among the five anxiety instruments, data from all five constructs were input into the Rasch model simultaneously. L2 Speaking and L2 Writing Anxiety items all loaded onto the primary, “positive” dimension. L2 Listening Anxiety, L2 Reading Anxiety, and Testing anxiety items all loaded onto the secondary, “negative” dimension (Fig. 6).

As a final examination of relations among the five anxiety constructs, a correlational analysis was conducted using Rasch person measures. First, correlations were computed for the constructs before removing misfitting items (Table 6). All five constructs were significantly correlated ( $p < 0.05$ ). The strongest correlation was between L2 Listening Anxiety and L2 Reading Anxiety ( $r = 0.71$ ), and the weakest correlation was between L2 Speaking Anxiety and Testing Anxiety ( $r = 0.41$ ). After removing the seven misfitting items, the person measures were recalculated and the correlational analysis was conducted again (Table 7). Overall there were no appreciable differences in the Pearson’s coefficient  $r$  values.

CON-TRAST	LOADING	INFIT			OUTFIT		ENTRY		LOADING	MEASURE	MNSQ	MNSQ	NUMBER	ITEM
		MEASURE	MNSQ	MNSQ	NUMBER	ITEM								
1	.62	56.89	1.04	1.03	A	10	sa10	-.49	48.66	.87	.86	a	18	la7
1	.62	53.17	1.10	1.14	B	9	sa9	-.48	45.35	.81	.79	b	19	la8
1	.59	52.97	1.07	1.15	C	8	sa8	-.44	47.57	.80	.78	c	20	la9
1	.58	51.75	.98	.98	D	4	sa4	-.41	43.76	1.02	1.00	d	16	la5
1	.58	59.17	1.09	1.12	E	7	sa7	-.38	51.72	1.05	1.04	e	26	ra4
1	.56	47.43	1.08	1.06	F	11	sa11	-.38	53.51	.84	.84	f	29	ra7
1	.47	46.71	.91	.91	G	6	sa6	-.37	50.50	.79	.79	g	27	ra5
1	.43	45.03	1.21	1.19	H	2	sa2	-.36	45.98	.73	.71	h	21	la10
1	.43	48.66	.91	.90	I	5	sa5	-.35	49.45	.91	.90	i	28	ra6
1	.40	44.47	1.33	1.46	J	3	sa3	-.33	44.15	1.11	1.18	j	13	la2
1	.39	50.77	.79	.78	K	40	wa7	-.32	43.22	.87	.85	k	22	la11
1	.35	50.33	.82	.81	L	41	wa8	-.32	53.07	.85	.85	l	49	ta6
1	.31	56.65	.88	.90	M	42	wa9	-.32	51.82	.79	.80	m	17	la6
1	.30	51.38	.86	.86	N	34	wa1	-.31	53.51	.84	.84	n	50	ta7
1	.29	52.09	.84	.84	O	39	wa6	-.29	53.92	.82	.83	o	25	ra3
1	.28	44.01	1.68	1.81	P	1	sa1	-.29	49.86	1.04	1.03	p	48	ta5
1	.18	54.19	1.06	1.18	Q	38	wa5	-.28	51.45	.96	.96	q	51	ta8
1	.16	54.59	1.11	1.15	R	35	wa2	-.23	43.76	1.45	1.46	r	45	ta2
1	.08	51.72	1.06	1.05	S	36	wa3	-.23	41.66	.97	.98	s	14	la3
1	.05	57.06	.89	.87	T	43	wa10	-.22	48.73	1.56	1.58	t	30	ra8
1	.01	54.25	1.01	1.01	U	37	wa4	-.21	44.33	.88	.86	u	15	la4
								-.21	53.68	.78	.78	v	33	ra11
								-.20	47.09	.76	.74	w	23	ra1
								-.19	45.70	.81	.80	x	24	ra2
								-.14	50.20	.98	1.01	y	12	la1
								-.13	49.11	1.29	1.31	z	46	ta3
								-.10	48.90	1.09	1.11	Y	44	ta1
								-.09	52.70	.62	.62	X	32	ra10
								-.08	54.46	1.19	1.28	W	31	ra9
								-.06	48.94	1.69	1.87	V	47	ta4

Fig. 6 The Rasch PCA of residuals for all items ( $k = 51$ ). sa = Speaking Anxiety; la = Listening Anxiety; wa = Writing Anxiety; ra = Reading Anxiety; ta = Testing Anxiety

Table 6 Correlations among the five anxiety constructs ( $k = 51$ )

	SA	LA	WA	RA	TA
L2 speaking anxiety	–	0.55	0.56	0.47	0.41
L2 listening anxiety	.	–	0.51	0.71	0.51
L2 writing anxiety			–	0.69	0.58
L2 reading anxiety				–	0.57
Testing anxiety					–

Note Correlations were all significant ( $p < 0.01$ );  $N = 298$

Table 7 Correlations among the five anxiety constructs excluding misfitting items ( $k = 44$ )

	SA	LA	WA	RA	TA
L2 speaking anxiety	–	0.54	0.58	0.50	0.39
L2 listening anxiety	.	–	0.50	0.70	0.52
L2 writing anxiety			–	0.67	0.56
L2 reading anxiety				–	0.56
Testing anxiety					–

Note Correlations were all significant ( $p < 0.01$ );  $N = 298$

## Discussion

While the top performing items of existing L2 anxiety instruments overall function well, the Rasch item fit analysis demonstrates that there is still room for improvement. The L2 Speaking Anxiety item (Sa1) that misfit the construct seemed more related to perceptions of speaking competence rather than anxiety (“...other students speak better than I do.”). The L2 Listening Anxiety item (La1) that misfit the construct contained vague wording (“I get stuck with...”) that could apply to other types of anxiety, or be related to comprehension rather than anxiety. Of the two L2 Reading Anxiety item that misfit, one (Ra4) was more related to speaking than to reading (“...that I can’t pronounce), while the other (Ra10) was, to be blunt, extremely silly. The original item was written for Japanese learners just beginning to learn how to read English; the item was likely intended to tap into the orthographic challenge of reading alphabetic characters in contrast to Chinese ideograms (*kanji*) and Japanese syllabary characters (*kana*). However, the wording “funny letters and symbols” has little if anything to do with reading anxiety, and the Rasch item analysis bears this out.

The L2 Writing Anxiety had only one misfitting item (Wa5), which could be construed as a speaking item as well as writing (“Discussing my writing...”). Additionally, there is a question whether asking if discussing writing is “enjoyable” constitutes language anxiety or not. Finally, there were two items that misfit the Testing Anxiety construct. The first (Ta4) contained awkward wording, including a negation and a “seems to me” clause, neither of which gives the sense of anxiety. The second (Ta8) was about confidence rather than anxiety. Feeling “less confident” does not necessarily imply “anxiety”—as discussed in the background section, conceptually there are still wording issues with many anxiety instruments, as anxiety is seen in clinical psychology as abnormal and not simply being “nervous” or “concerned.” Since this construct had the lowest variance accounted for by the Rasch model as well as the lowest Rasch person reliability it is worth noting that the study from which these items were borrowed had an extremely low sample size ( $N = 79$ ) to go with a large number of items ( $k = 57$ ) which originally stemmed from two separate studies (Fujii 1993; Sarason 1975). The “test anxiety” instrument created used traditional factor analysis to reduce the 57 items to 12 items across three factors, one of which was related to mental and psychical exhaustion. The results in the present study indicate problems with item fit and construct validity, leading to the logical conclusion that this construct still needs seriously reworking before further usage.

While Rasch model analysis is designed for single constructs rather than a multi-construct questionnaire like the one used in this study, the Rasch PCAR was revealing concerning the relation among the five anxiety constructs. When all items were input into the Rasch model, the Rasch model correctly identified all L2 Speaking Anxiety items as a single construct; all SA items had “positive” loadings above 0.40. However, all the L2 Writing Anxiety items also loaded onto the primary dimension, an indication that study participants treated speaking and writing



items similarly; this may not be surprising, as speaking and writing are considered “active,” student-output related L2 skills. The Rasch PCAR also indicated the relationship between L2 Reading Anxiety and L2 Listening Anxiety; all the RA and LA items loaded “negatively” onto the secondary dimension. This result may be the result of study participants treating reading and listening as “passive,” input-related L2 skills. Testing Anxiety (TA) items also loaded negatively. This may not be too surprising given the context of the study: in Japan, university EFL programs often use standard exams such as the Test of English for International Communication (TOEIC), which is used by Japanese companies for hiring and internal promotion. While there is a version of the exam that tests speaking and writing, the TOEIC used in universities (TOEIC IP) typically only tests reading and listening abilities. Thus, it makes sense that Rasch identified L2 listening and reading anxieties as related strongly to test anxiety.

The correlational analysis of Rasch person measures among the five anxiety instruments contradicts previous assumptions in the L2 anxiety literature that test anxiety is not related to L2 anxiety. The results in this study indicate that the various L2 anxieties are, in fact, strongly related to each other as well as to testing anxiety. Indeed, this should come as no surprise: the wording of the L2 listening, writing, and reading items tend to focus on exam and assessment situations. In particular, nearly all L2 writing anxiety items are related to concerns about written essay assessment (and hence, testing). Given the relation of all four L2 anxieties to testing anxiety—which is supposed to be the same regardless of first or second language context—the results additionally suggest that the use of a second language may not represent a separate anxiety, agreeing with Dewaele (2013) and Jung and McCroskey (2004).

## Conclusion

The results of this study provide examples of a construct created through Rasch model principles (L2 Speaking Anxiety) compared to constructs created through correlation and factor analysis (L2 Listening Anxiety, L2 Reading Anxiety, L2 Writing Anxiety, Testing Anxiety). Whereas a measurement instrument created according to Rasch principles assumes that questionnaire takers with a high level of anxiety will have a higher probability of endorsing items that also indicate a higher degree of anxiety (Wilson 2005), a measurement instrument created through traditional statistics assumes that someone with a greater level of anxiety will endorse all items to a greater degree than someone with a lower level of anxiety. As Waugh and Chapman (2005) have pointed out, the fact that items correlate strongly in traditional factor analysis does not indicate construct validity. The use of multiple items to artificially inflate Cronbach’s alpha, rather than spread participants across a range of endorsability levels on the hypothesized construct, also does not lead to greater construct validity.

A limitation of the current study is that all study participants were in a single EFL context (Japan), as well as roughly the same age and level of education.

The sample size ( $N = 298$ ) was adequate, but given the number of total items ( $k = 51$ ) a larger sample size may be desirable. Finally, the number of Likert-type categories (six) had to be collapsed for the four of the five measurement instruments that were originally constructed using traditional statistical methods. This may have slightly suppressed the person reliabilities and separation for these constructs.

Whether the “four L2 anxieties” truly measure separate anxieties may still be open for debate. Certainly several items in each anxiety measurement, and particularly those of testing and L2 reading anxiety, are in need of revision before further investigation of the relationship among the varying hypothesized aspects of L2 anxiety. Deeper analysis of the relationship among the “active” and “passive” anxieties is also warranted, once individual “four anxieties” L2-skills based measurement instruments have been further examined and validated in other ESL or EFL contexts.

## Notes

1. The full 51-item questionnaire was omitted for space considerations but is available by request; please email the author.
2. The original L2 Reading Anxiety instrument was created by Saito et al. (1999), but as only descriptive statistics were used in that paper, Matsuda and Gobel's factor loadings were used to select items.
3. Likert category utility statistics were omitted for space considerations but are available by request.

## References

- Andrich, D. (1978). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement*, 38, 665–680.
- Apple, M. T. (2013). Using Rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety. *JALT Journal*, 35(1), 5–28.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum.
- Cheng, Y., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, 49, 417–446.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98–104.
- Dewaele, J.-M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi- and multilinguals. *Modern Language Journal*, 97(3), 670–684.
- Fujii, Y. (1993). Tesuto eikyuu inbentorii (TII) no sakusei [Construction of a Test Influence Inventory (TII)]. *Japanese Journal of Psychology*, 64(2), 135–139.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

- Green, S. B., Lissitz, R. W., & Mulait, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827–838.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70, 125–132.
- In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System*, 34, 317–340.
- Jung, H.-Y., & McCroskey, J. C. (2004). Communication apprehension in a first language as predictors of communication apprehension in a second language: A study of speakers of English as a Second Language. *Communication Quarterly*, 52(2), 170–181.
- Kaipper, M. B., Chachamovich, E., Hidalgo, M. P. L., Torres, I. L. S., & Caumo, W. (2010). Evaluation of the structure of Brazilian State-Trait Anxiety Inventory using a Rasch psychometric approach. *Journal of Psychosomatic Research*, 68, 223–233.
- Kim, J.-H. (2000). Foreign language listening anxiety: A study of Korean students learning English. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 30004305).
- Kimura, H. (2008). Foreign language listening anxiety: Its dimensionality and group differences. *JALT Journal*, 30(2), 173–195.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85–106.
- Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch-model computer program*. Chicago: MESA.
- MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning*, 39, 251–275.
- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second language. *Language Learning*, 41, 513–534.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32(1), 21–36.
- Price, M. L. (1991). The subjective experience of foreign language anxiety: Interviews with highly anxious students. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 101–108). Englewood Cliffs, NJ: Prentice-Hall.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago.
- Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign language reading anxiety. *Modern Language Journal*, 83, 202–218.
- Sarason, I. G. (1975). The Test Anxiety Scale: Concept and research. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 2, pp. 193–217). Washington, D.C.: Hemisphere.
- Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Tenenbaum, G., Furst, D., & Weingarten, G. (1985). A statistical reevaluation of the STAI anxiety questionnaire. *Journal of Clinical Psychology*, 41(2), 239–244.
- Waugh, R. F., & Chapman, E. S. (2005). An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: What is the difference? Which method is better? *Journal of Applied Measurement*, 6(1), 80–99.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wolfe, E. W., & Smith, E. V, Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part I—Instrument development tools. *Journal of Applied Measurement*, 8(1), 97–123.
- Wright, B. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling*, 3(1), 3–24.
- Young, D. J. (1999). *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere*. Boston: McGraw Hill.

## **Author Biography**

**Matthew T. Apple** (MFA, MEd, EdD) is Associate Professor in the Department of Communication, Ritsumeikan University, Kyoto, Japan. Recent publications include *Language Learning Motivation in Japan* (2013, Multilingual Matters) and articles in *JALT Journal* and the *Journal of Applied Measurement*. His research interests include ESP, individual differences, and educational statistics.