Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327
مجلة اللغة الانكليزية في العالم العربي

AWEJ Volume.3 Number.3, 2012 pp. 18 - 34

Developing an Instrument that Tests for English-Language Personality

Omar Karlin
Faculty of Global Japanese Studies
Meiji University
Tokyo, Japan

Abstract

Studying a second-language involves several cultural and environmental pressures that can shift behaviour and alter personality. A person may act one way in their native surroundings, but act completely differently when travelling, studying a foreign language, or talking with people from a different culture. Existing personality instruments do no capture this shift in personality, and thus a specialized personality instrument is required. This study surveyed 287 Japanese university students with 263 specialized second-language personality items. Items focused on situations involving travel, being in English class, meeting native English-speakers, watching English movies, and other activities that involve the study of English. A factor analysis, a correlational analysis and a Rasch principal components analysis was conducted on this new instrument, as well as an established personality instrument to determine their effectiveness at capturing the participants' personality. The ultimate result was a 50 item personality instrument intended for second-language contexts, which performed better than the established personality instrument. The correlation analysis also revealed that the underlying constructs of personality were not highly correlated across instruments, indicating the presence of a native language personality and a second-language personality for participants.

Keywords Personality, TESOL, Rasch, correlations

Introduction

There have been several attempts to research personality within a TESOL-context, however, in all instances the instrument used to measure personality was intended for an overarching conception of personality, situated within a person's native-language and culture, indifferent to many situational considerations. One only needs to talk to a fluent bilingual to hear how, at least in some instances; people develop different personalities based on the situation. In Japan, it is not uncommon to hear how a person can feel rigid and reserved when ensconced in the cultural, social, and language trappings of Japan, yet this same person may feel free and unhindered when speaking English in a culture less tied to rules of social conduct. However, as of yet, a personality instrument has not been developed to handle these situational complications. If such an instrument could be developed, one that could capture personality shifts that arise from contextual pressures, it may have a profound effect on TESOL research involving personality, yielding new theories and relationships between personality dimensions and second language acquisition that have not yet revealed them.

Literature Review

Through lexical studies of the five-factor adjective descriptors of the Big Five model of personality (i.e. personality is comprised of the five dimensions of extraversion, neuroticism, conscientiousness, agreeableness, and openness), it has been suggested that personality factors may be affected by the surrounding culture. By analyzing the English lexicon, some researchers have claimed that each of the five personality dimensions in the Big Five lexically evolved at a different rate, a finding that suggests that within society as a whole, personality develops and changes in response to social developments (Piedmont & Aycock 2006). Specifically in Western society, openness, the dimension which covers how open a person may be to new concepts and ideas, appears to be the most lexically recent of the five factors, possibly being catalyzed by major social events such as the Age of Enlightenment and the Industrial Revolution, when new ways of thinking and conducting business were regarded more highly than at other times in Western history. An ANOVA based on the mean lexical origins of descriptors (by year) for each of the five personality dimensions revealed that extraversion, neuroticism, agreeableness, and conscientiousness were the first four personality dimensions to be represented in the English lexicon, with openness's lexical origins coming significantly later. The implication is that personality dimensions, in this case openness, are not stable across different historical periods, and by extension, are not stable across different environments and cultures. To confirm this, researchers examined a non-industrial, agrarian sub-Saharan African community to determine if all five personality dimensions were as strongly represented as in present-day Western society. They discovered that openness was more difficult to isolate, probably because the lack of vocabulary representing openness in the culture resulted in a lack of self-identification with that construct for the participants in the study (Heuchert, Parker, Stumpf, & Myburgh 2000; Horn 2000). Further, while some cultures might display fewer than five personality factors, other cultures might display more. In other non-English lexical studies (Ashton, Lee, & Goldberg 2006), a humility/humble factor was discovered in some Asian cultures, resulting in a six-factor personality model termed *HEXACO* (Lee & Ashton 2004; Lee & Ashton 2006). A subsequent modification of the open-source personality instrument *IPIP Big Five Factor Markers* (to hereby be referred to as the *IPIP BFFM*), resulting in the *IPIP-HEXACO*, reflects the possible future sphere of influence that the *HEXACO* model might someday wield (Ashton, Lee, & Goldberg 2006; Goldberg 2007).

Suffice it to say, whether four, five, six, or more factors, environment and culture need to be considered in any study discussing personality. Language classrooms, particularly those taught by native speakers, have affiliations with the target language's culture, different to the one in which the classroom is situated, whether through direct influence from the native teacher's conduct or through cultural perceptions tied to the target language that seep into the classroom. As a result, inclusion in a second language classroom must be considered, to some degree, as being in another culture. Unfortunately the wording of items used in the prevalent personality instruments does not recognize respondents situated in another culture, where personality might shift because of environmental and cultural pressures. When an item asks whether a respondent is shy, respondents might confidently respond "no," yet that same respondent might respond differently when asked if they are shy when speaking a second language, when travelling abroad, or in the presence of a native speaker of the target language. Certainly some of this might be tied to other variables such as heightened anxiety or lack of motivation, but it seems reasonable to assume that some people just change in the presence of another culture (although maybe not consciously). It is important to modify questionnaire items to effectively capture shifts in personality brought on by these environmental and cultural pressures.

One emerging personality model that has challenged the notion of personality as a static trait is the *Personality and Role Identity Structural Modal*, or *PRISM* (Wood & Roberts 2006). PRISM is a three-tiered hierarchical structure with general personality traits, such as extraversion and neuroticism situated at the top. General traits are an average of the second tier, role identities. For example, most people occupy a variety of roles in their life, such as work, with friends, with family, in clubs, and romantically. Each of these usually involves a unique identity that is tied to that specific context. If a person is relatively talkative with family, with friends, and in romantic situations, but not talkative at work, those four role identities taken together establish their general trait as relatively extraverted (three extraverted roles and one introverted role would average to be extraverted). The bottom tier of this bottom-up personality model is referred to as role experiences. It is the most dynamic of these tiers, and represents the positive or negative associations a person may be having in their role at any given time, such as satisfaction from completing a project at work or angst from an acrimonious romantic relationship. Just as the first tier of general traits is an amalgam of the second tier, each role identity from the second tier is an amalgam of third-tier experiences. In sum, the third-tier can gradually change the second-tier, which can in turn alter the first-tier, challenging the notion that personality is fixed and offering support to a trait-state compromise of personality theory. Existing trait-based personality instruments such as the *NEO-PiR, IPIP Big Five Factor Markers*, and even the *MBTI* attempt to capture personality on this first tier, yet because of the aforementioned peculiarities with language learning, it is likely that language learning anomalies are occurring on the second and third tiers, and are not strong enough to overcome conflicting second- and third-tier identities and experiences that occur in native language situations.

Evidence of this need to specialize items to reflect cultural pressures can be found in immigration studies designed to investigate how personality and identity become destabilized in foreign culture contexts. In a study examining Chinese immigrants to Canada and how they behaved relative to ethnic-Chinese born in Canada, (McCrae, Yik, Trapnell, Bond, & Paulhus 1998) found that despite similar genetic compositions because of their shared ethnicity, the two populations displayed different personality characteristics. The Canadian-born group scored significantly higher on extraversion, openness, and agreeableness dimensions, likely because of behavioral demands created by Canadian society. Over time, the immigrant community

acculturated to their surroundings and began to display some of the same characteristics as the Canadian-born group. Perhaps language students function in the same way, not displaying certain personality characteristics in their native culture, resulting in the lack of an identifiable relationship in many TESOL research studies between personality (as measured by existing instruments) and second language acquisition. However, when in a language classroom, or better yet, a new country, their personality might shift, like the Chinese immigrants, or entire personality dimensions might be gained or lost, like the missing openness of the aforementioned non-industrial African Bushmen or the added humility dimension in some Asian cultures.

Lastly, there have been some recent efforts by educational psychologists to create a taxonomy of *which* situations can alter personality (Saucier, Bel-Bahar, & Fernandez 2007), yet this effort is only at the preliminary stages. It is important that changes in personality spurred by environmental and cultural pressures be addressed by existing instruments, something that could greatly enhance future research.

In a study that demonstrated the relative malleability of the personality construct, and the need to adapt it to different contexts, (Bowler, Bowler, & Phillips 2009) demonstrate that the five-factor personality model fluctuates in response to the individual. This study, involving 718 undergraduate university students, split participants into three groups, a low cognitive complexity group, an average cognitive complexity group, and a high cognitive complexity group. Cognitive complexity was measured by the Computer-Administered Rep Test (Woehr, Miller, & Lane 1998). Results indicated that a 3-factor personality model was best-suited for the low cognitive complexity group, a 5-factor personality model was best-suited for the average cognitive complexity group (as well as the entire sample as a whole), and a 7-factor personality model was best-suited for the high cognitive complexity group. The personality instrument administered to students was the Unipolar 100 Big Five Markers (Goldberg 1992), which is intriguing because one wonders which two additional factors might be embedded within this 5-factor instrument, in the case of the high cognitive complexity group. One also wonders which two factors were dropped (or merged) in the low cognitive complexity group (the authors do not specify). While the Bowler, Bowler, & Phillips study (2009) questions the consistency of the Five Factor Model (*FFM*) across different groups, this paper does not question the general validity of the FFM. The convergent validity of the FFM has been repeatedly endorsed (McCrae & Costa 1999; Garcia, Aluja, & Garcia 2006), and even the Bowler, Bowler and Phillips study (2009) confirmed the FFM as appropriate when viewing their sample as a whole. However, the study does underscore the need for scrutiny of personality instruments in varying contexts, rather than acceptance across all contexts. Just as personality factors may disappear or appear in the face of varying degrees of cognitive complexity, personality instrument items may need to be dropped or added in the face of cultural and environmental pressures.

Method

The construction and validation of the Questionnaire of English Environment Personality (*QuEEP*) involved several steps. First, a factor analysis was conducted with the initial 263 items across five personality dimensions. The number of items was reduced to the 50 items that loaded most strongly on their respective personality dimensions, 10 for each of the big five personality dimensions. Items that loaded on two or more dimensions simultaneously, or loaded on an unexpected dimension were not included amongst the final 50 items. A factor analysis was also conducted on the 50-item International Personality Item Pool Big Five Factor Markers (*IPIP BFFM*) to provide context to the *QuEEP* results. The *IPIP BFFM* was developed to provide an

alternative to existing personality instruments, which had more items and were under copyright protection. The *IPIP BFFM* is intended to provide personality researchers with an open-source instrument that is shorter, easier to procure, and more accurate than competing instruments like the NEO-Pi-R (Goldberg 1999).

Second, a correlational analysis was conducted to determine how closely the constructs in each instrument were situated. Specifically, if like-factors had a high correlation with each other (for example, if the *QuEEP* extraversion factor was highly correlated with the *IPIP BFFM* extraversion factor), it would suggest that personality is in fact stable across different contexts. However, a relatively low correlation between like-factors would suggest a situational dissonance between the native-language extraversion of the *IPIP BFFM* and the second-language extraversion of the *QuEEP*, thereby justifying the construction of the *QuEEP* as a second-language situation-specific personality instrument.

Finally, a Rasch analysis was conducted using the data obtained for the 50-item version of the *QuEEP* in order to confirm the suitability of the 50 items. A particular focus was placed upon item fit, explained variance, unexplained variance, and the Rasch reliability and separation indices to determine item quality. Alongside this Rasch analysis of the *QuEEP*, a parallel Rasch analysis was conducted on the 50-item *IPIP BFFM* to compare the two instruments and ascertain the effectiveness of the *QuEEP* in relation to a more established personality instrument.

Participants

The Questionnaire of English Environment Personality was completed by 287 Japanese university students; 188 first-year students, 92 second-year students, 6 third-year students, and 1 fourth-year student. The participants completed the questionnaire in May of 2012, with almost all students in 13 classes at five universities completing the survey. The students were taking a variety of courses and came from several different faculties (see Table 1), but were all taught by the same native English teacher.

Table 1. ***Background of the Participants***

School	Location	*n*	Male	Female	Faculty
1	South of Tokyo	3	2	1	Social welfare
2	Central Tokyo	99	83	16	Engineering, Architecture
3	West Tokyo	70	35	35	Law
4	West-Central Tokyo	72	16	56	Global Japanese Studies
5	North-West Tokyo	43	6	37	Music

The settings for the online survey collection website, www.surveymonkey.com, were set so that students had to make a selection (questions could not be skipped) for one of the six answer choices for each question on the *QuEEP* (*Strongly disagree, Disagree, Slightly disagree, Slightly agree, Agree,* and *Strongly agree*). As a result, no data were missing for any of the 263 items across 287 students (75,481 completed items in total). However, for the *IPIP BFFM*'s 50 items, students were asked to enter a value from one to five (ascending from one, answer choices were *Very inaccurate, Moderately inaccurate, Neither inaccurate nor accurate, Moderately accurate*, and *Very accurate*) into a box. Ultimately there were four instances of a student not entering a number, as well as two instances of a student entering an incorrect number (40 and 1555). In the case of the missing values, an asterisk was entered in the place of the missing numerical values so that the Winsteps program could conduct the Rasch analysis without disruption. In the case of the incorrect values, the first digit of the value ("4" and "1") was taken as the intended response,

while the remaining digits were deleted ("0" and "555"). There were no instances in which a participant provided implausible data, such as repeatedly entering in the same value for all responses. As a result, no participants were deleted from the sample.

Results – Factor Analysis

The first phase of this study involved conducting a factor analysis with 262 of the items. After selecting several different extraction and rotation methods, it appeared that the data most closely followed the expected five-construct pattern with maximum likelihood extraction and equamax rotation. Only 262 items were included in the factor analysis, instead of 263, because the data for item E35 (the 35th item within the extraversion set of items) was not collected properly by www.surveymonkey.com. It is not known why the data could not be collected properly, or how to remedy the situation so the data could be saved, so the item was eliminated from the first factor analysis.

The Kaiser-Meyer-Olkin measure indicated poor sampling for the analysis, KMO = .33 (Hutcheson & Sofronion 1999), with a guideline of 0.5-0.7 as mediocre, 0.7-0.8 as good, 0.8-0.9 as very good, and greater than 0.9 as excellent. Bartlett's test of sphericity was significant at 67324.35, p = .00. There were 67 factors with eigenvalues over Kaiser's criterion of 1 (Kaiser 1974), however the first five factors had eigenvalues of 48.72, 17.32, 8.00, 6.71, and 5.27, accounting for 32.95% of the variance. The rotated sums of squared loadings indicated five factors with eigenvalues of 20.80, 18.56, 16.50, 13.89, and 12.93 accounting for 31.68% of variance. The scree plot did not indicate a clear inflection point to justify five factors, as many items and factors obscured visual interpretation. The items that clustered on the same factors suggested that factor 1 represented neuroticism, factor 2 openness, factor 3 conscientiousness, factor 4 extraversion, and factor 5 agreeableness.

A standard of +/-.40 was used to screen items. Items that had a factor loading below +/- .40 were deleted from the initial set of 262 items on the questionnaire. However, because so many non-performing items were to be deleted, it was expected that the remaining items would show a significant increase in their factor loadings. As a result, the +/- .40 loading criteria was only applied to items from the neuroticism, conscientiousness, and openness personality factors because there was an abundance of items that met the criteria. With the extraversion and agreeableness factors, far fewer items exceeded the +/- .40 criterion, so the cut-off threshold was lowered to +/- .35. Also, items in the neuroticism, conscientiousness, and openness factors that loaded onto multiple factors or onto an unintended factor were deleted from the questionnaire, while items in the extraversion or agreeableness factors were granted an exception to these selection criteria.

The first screening of items resulted in 14 remaining neuroticism items (25 items deleted), 23 remaining conscientiousness items (24 items deleted), 17 remaining openness items (37 items deleted), 17 remaining extraversion items (41 items deleted), and 19 remaining agreeableness items (45 items deleted).

A factor analysis (FA) with maximum likelihood extraction and orthogonal rotation (equamax) was conducted on the remaining 90 items. The Kaiser-Meyer-Olkin measure verified the sampling adequacy of the analysis, KMO = .89, which should be considered very good (Kaiser 1974). Bartlett's test of sphericity was significant at 14766.72, p = .00. There were 20 factors with eigenvalues over Kaiser's criterion of 1, however the first five factors had eigenvalues of 20.02, 8.56, 4.41, 3.26, and 2.93, accounting for 43.05% of the variance. The rotated sums of squared loadings indicated five factors with eigenvalues of 8.46, 8.28, 7.33, 6.45, and 5.64 accounting for 39.73% of variance. The scree plot indicated an inflection point after the

fifth factor, supporting the Big Five factor model of personality. The items that clustered on the same factors suggested the following: Factor 1 represented conscientiousness, Factor 2 openness, Factor 3 neuroticism, Factor 4 extraversion, and Factor 5 agreeableness.

For the final screening of items, 40 items were removed to create a 50-item instrument. Only 10 items per factor were selected, and preference was given to items with the strongest factor loading on the intended factor. However, some consideration was also given to factor loading on unintended factors, in that if there were two items with a similar factor loading on the intended factor, priority was given to the item with the weaker factor loading on the secondary factor. For example, on the conscientiousness factor, item C43 had a stronger factor loading than item C36 (.54 vs. .53), yet it was not selected for the final version of the instrument because it had a much stronger loading on a secondary factor (.31 on the openness factor) than item C36 did (.17 on the extraversion factor). A factor analysis (FA) with maximum likelihood extraction and orthogonal rotation (equamax) was conducted on the remaining 50 items (see Table 2).

Table 2. ***Final Factor Analysis of the 50 QuEEP Items***

	Factor						Factor				
Item	1	2	3	4	5	Item	1	2	3	4	5
N1	**.76**	.15	.05	.17	.18	*O30*	.10	**.64**	.22	.25	.02
N2	**.58**	.12	.18	.13	.09	*O36*	.02	**.67**	.19	.15	-.09
N3	**.57**	.12	.12	.29	.06	*O37*	.06	**.69**	.30	.06	.14
N5	**.67**	.06	.07	-.03	.13	*O42*	.15	**.64**	.27	.05	.24
N6	**.71**	.12	.12	.11	.19	*O53*	.25	**.53**	.07	.17	.19
N18	**.76**	.07	.14	.12	.14	*E3*	.24	.03	.22	.01	**.43**
N21	**.68**	.09	-.03	.02	.28	*E4*	.33	.24	-.02	.09	**.49**
N23	**.50**	.27	-.05	.13	.31	*E6*	.17	.08	.24	.14	**.55**
N32	**.50**	-.02	-.07	-.08	.22	*E7*	.12	-.05	.14	.04	**.59**
N39	**.61**	.02	.15	-.10	.29	*E8*	.23	.03	-.03	-.05	**.62**
C3	.00	.21	**.57**	.36	.02	*E12*	.31	.24	-.02	.08	**.49**
C12	-.05	.04	**.57**	.23	.15	*E18*	.11	-.05	.11	-.04	**.53**
C14	.05	.16	**.61**	.12	.12	*E28*	.09	.02	-.06	.04	**.60**
C16	.12	.06	**.57**	.21	.21	*N22*	.26	.12	-.01	.04	**.57**
C17	.09	.22	**.62**	.10	-9.9E5	*C25*	.18	.12	.29	.06	**.46**
C24	.03	.17	**.61**	.20	-.14	*A5*	.01	.20	.21	**.59**	-.15
C36	.09	.04	**.59**	.11	.12	*A6*	.00	.33	.22	**.51**	-.19
C39	.11	.27	**.60**	.24	.08	*A7*	.10	.27	.10	**.60**	-.09
C40	.08	.15	**.49**	.21	.09	*A11*	.12	.25	.31	**.64**	.10
C44	.08	.12	**.54**	.22	.03	*A13*	.15	.37	.17	**.40**	.16
O1	-.02	**.56**	.17	.24	-.04	*A22*	.07	.22	.22	**.66**	.15
O2	-.06	**.53**	.28	.33	-.22	*A33*	.08	.17	.14	**.52**	.24
O3	.14	**.53**	.12	.35	-.01	*A35*	.21	.15	.32	**.56**	.29
O20	.18	**.68**	.00	.13	.17	*A43*	.02	.05	.15	**.40**	-.12

O22	.15	**.55**	-.00	.22	.11	*A61*	.01	.040	.15	**.52**	.27

Note. Underlined items indicate items that were included in the final 50-item version of the QuEEP. Extraction and rotation: Maximum likelihood and equamax.

The Kaiser-Meyer-Olkin measure verified the sampling adequacy of the analysis, KMO = .90, which can be considered excellent (Kaiser 1974). Bartlett's test of sphericity was significant at 6863.55, p = .00. There were nine factors with eigenvalues over Kaiser's criterion of 1, however the first five factors had eigenvalues of 12.06, 5.66, 2.96, 2.30, and 2.93, accounting for 49.97% of the variance. The rotated sums of squared loadings indicated five factors with eigenvalues of 4.90, 4.76, 4.49, 4.14, and 3.96 accounting for 44.50% of variance. The scree plot indicated a clear inflection point after the fifth factor, further justifying the Big Five factor model of personality. The items that clustered on the same factors suggested the following: Factor 1 represented neuroticism, Factor 2 openness, Factor 3 conscientiousness, Factor 4 agreeableness, and Factor 5 extraversion.

It was thought that conducting a FA on a native-language based personality instrument, such as the International Personality Item Pool Big Five Factor Markers test, would provide valuable context to the results obtained for the *QuEEP*. Results of the factor analysis on the *IPIP BFFM* can be seen in Table 3.

Table 3. ***Factor Analysis of the 50 IPIP BFFM Items***

Item	Factor					Item	Factor				
	1	2	3	4	5		1	2	3	4	5
E1	**.66**	.02	.27	-.05	.09	*C6*	.04	**-.42**	-.11	.18	.07
E2	**-.52**	.09	-.13	.06	-.15	*C7*	.00	**.43**	.06	-.16	-.01
E3	.38	.14	.18	-.11	.37	*C8*	.11	-.39	-.01	.04	-.26
E4	**-.54**	.01	-.24	.25	-.03	*C9*	-.09	**.54**	-.05	.00	.16
E5	**.62**	.16	.17	-.01	.22	*C10*	.20	.39	.10	.09	.21
E6	**-.40**	.07	-.09	.06	-.09	*N1*	-.12	-.02	.01	**.62**	-.05
E7	**.64**	.13	.20	.02	.19	*N2*	.27	.15	.12	-.25	.10
E8	**-.50**	.02	-.30	.01	-.07	*N3*	-.28	.34	-.14	**.41**	.26
E9	**.54**	-.01	.24	.03	.00	*N4*	.18	.07	.10	**-.50**	-.18
E10	**-.49**	.09	-.11	.21	-.07	*N5*	-.28	.08	-.16	**.49**	.14
A1	-.21	-.01	.10	.00	**-.72**	*N6*	-.14	.06	-.11	**.44**	.12
A2	.30	.08	.05	.05	**.65**	*N7*	.17	-.18	.01	**.53**	-.05
A3	.12	-.08	-.04	.07	**-.53**	*N8*	.16	-.13	.10	**.58**	-.04
A4	.16	.32	.08	.10	.38	*N9*	-.02	-.10	-.07	**.54**	-.16
A5	-.26	-.09	-.02	-.06	**-.62**	*N10*	-.19	.02	-.04	**.67**	.07
A6	.21	.38	.27	.07	.25	*O1*	.26	.21	.22	-.03	-.10
A7	-.05	-.07	-.04	.11	**-.51**	*O2*	-.05	-.01	-.25	.14	.13
A8	.12	**.45**	.15	.13	.32	*O3*	.07	.06	**.81**	-.02	.23
A9	.11	**.44**	.02	.08	.15	*O4*	.01	-.08	-.03	-.07	-.18
A10	.27	.38	.20	-.04	.25	*O5*	.20	.14	**.60**	-.05	-.06

C1	-.06	**.55**	.05	-.17	.04	*O6*	-.12	.06	**-.73**	.09	-.21
C2	.11	-.39	.08	.35	-.15	*O7*	.24	**.44**	.30	-.05	-.08
C3	.11	**.54**	.30	.05	.06	*O8*	.15	.05	.15	.05	-.15
C4	.19	.00	.23	.32	-.08	*O9*	-.19	**.43**	.12	.03	.15
C5	-.01	**.47**	-.06	-.12	.13	*O10*	.20	.10	**.81**	.04	.07

Note. Extraction and rotation: Maximum likelihood and equamax.

A factor analysis using maximum likelihood extraction and orthogonal rotation (equamax) was conducted on the 50-item *IPIP BFFM*. The Kaiser-Meyer-Olkin measure verified the sampling adequacy of the analysis, KMO = .79, which can be considered good (Kaiser 1974). Bartlett's test of sphericity was significant at 5498.18, p = .00. There were 13 factors with eigenvalues over Kaiser's criterion of 1, however the first five factors had initial eigenvalues of 7.57, 4.38, 3.74, 2.75, and 1.93, accounting for 40.74% of the variance. The rotated sums of squared loadings indicated five factors with eigenvalues of 4.07, 3.49, 3.29, 3.22, and 3.10 accounting for 34.32% of variance. The scree plot indicated a clear inflection point after the fourth factor, which was perhaps indicative of the inadequate factor loading of +/- .30 (or less) for 6 of the 10 openness items. The items that clustered on the same factors suggested that factor 1 represented extraversion, factor 2 conscientiousness, factor 3 openness, factor 4 neuroticism, and factor 5 agreeableness. Only 33 out of 50 items had a loading of +/- .40 on their intended factor (9 for extraversion, 6 for conscientiousness, 4 for openness, 9 for neuroticism, and 5 for agreeableness). Interestingly, the extraversion and neuroticism factors seemed to be the most fully realized and they are also the two factors in the Big Five personality model with the longest history of theory, both appearing in Eysenck's Big Two model of personality (predating the Big Five and the Big Three models).

Results – Correlational Analysis

To get a clear picture of how strong the correlations were within each instrument, Rasch person ability measures for each of the five personality factors were obtained for the 287 participants involved in the study and a correlational analysis was performed. The correlations for the person measures for the five personality factors for both the *QuEEP* and the *IPIP BFFM* are displayed in Table 4. As seen in the results of the Rasch analysis that follows, the personality factors that had the highest correlations with each other were the ones that appeared to cohere when item dimensionality was tested, most notably the extraversion and neuroticism factors for the *QuEEP* at .61 and the extraversion and openness factors for the *IPIP BFFM* at .39. However, this was not always the case, as the conscientiousness and neuroticism factors on the *IPIP BFFM* did cohere, yet did not display a significant correlation. Interestingly, the correlations between the personality factors of each instrument were relatively low, with the correlation between the *QuEEP* agreeableness factor and the *IPIP BFFM* agreeableness factor at .33. The other correlations between the two instruments were .27 for the conscientiousness factor, .32 for the extraversion factor, .23 for the neuroticism factor, and an insignificant correlation for the openness factor. This would seem to suggest that there is a dissonance between native language personality, as measured by the *IPIP BFFM* and a second language personality as measured by the *QuEEP*. If the two instruments were tapping into the same constructs, then the correlation between like factors would presumably be much higher. Interestingly, in the case of conscientiousness, neuroticism, and openness, there are actually different personality factors

with higher correlations across the two instruments than like correlations across the two instruments.

Table 4. ***Correlations among Factors for the QuEEP and IPIP BFFM***

	1	2	3	4	5	6	7	8	9	10
1. A QuEEP	-									
2. C QuEEP	.57*	-								
3. E QuEEP	.31*	.29*	-							
4. N QuEEP	.31*	.22*	.61*	-						
5. O QuEEP	.59*	.44*	.30*	.31*	-					
6. A IPIP BFFM	.33*	.34*	.06	.13	.22*	-				
7. C IPIP BFFM	.10	.27*	-.06	.02	.05	.39*	-			
8. E IPIP BFFM	.28*	.22*	.32*	.41*	.30*	.37*	.06	-		
9. N IPIP BFFM	-.06	-.01	.12	.23*	.00	.06	.11	.32*	-	
10. O IPIP BFFM	.20*	.27*	.24*	.25*	.17	.33*	.18*	.37*	.14	-

Note. A = Agreeableness; C = Conscientiousness; E = Extraversion; N = Neuroticism; O = Openness.
* $p = 0.01$ (two-tailed)

When comparing the strength of the correlations among factors within each instrument, the correlations among *QuEEP* factors are much higher than the correlations among the *IPIP BFFM* factors. The implication to this difference in correlational strength suggests that the *IPIP BFFM* factors are more independently realized than their *QuEEP* counterparts, likely because the premise of the *IPIP BFFM* (and other native language personality tests) are not limited to specific situations or experiences. When items can be drawn from any situation or experience, there is less likelihood of overlap that might yield similarities. For instance, with a second language personality instrument such as the *QuEEP*, the extraversion and openness scales both include items focused on travel. Despite being situated in different personality scales, these items might yield some synchronicities because of the common theme of travel, in this particular case, someone who is welcoming of travel is likely to be both extraverted and open in travel situations, resulting in a correlation. Despite the higher correlation among *QuEEP* factors, this should not diminish the veracity of the instrument as unidimensionality was evident, and is outlined below. Rather, the higher correlations among *QuEEP* factors should be recognized as a by-product of situational limits imposed on items in order to capture a specific environment, in this case a second language learning environment.

Results – Rasch Analysis

To confirm the dimensionality of the instrument, a Rasch analysis was conducted on the *QuEEP*. Testing unidimensioanlity can be accomplished by conducting a Rasch principal component analysis of item residuals. If the first residual contrast has an eigenvalue under 2.0, it can be assumed that the construct is fundamentally unidimensional. As mentioned above, it was

believed that the *QuEEP* would follow a Big Five model of personality, comprising the constructs of neuroticism, conscientiousness, openness, extraversion, and agreeableness.

The first step in confirming unidimensionality involved entering the data for all 50 QuEEP items into Winsteps, and then checking the item dimensionality. If items within a single construct clump together, it would indicate unidimensionality, which would need to be clarified with further analysis. If items from two constructs are cohering together, this may again be an indication of unidimensionality, provided the two constructs can be separated in a further analysis. Results indicated that it may be possible to parse out the constructs of extraversion and neuroticism into single constructs, as the negative loading items were either extraversion or neuroticism items, with the sole exception of C25. However, item C25 was earlier shifted to the extraversion scale, thereby keeping intact the extraversion and neuroticism cohering of items.

Taking only the items from the neuroticism and extraversion scales and re-entering them into Winsteps, the explained variance for the two sets of items was revealed to be 42.3%, while the unexplained variance after the first contrast had an eigenvalue of 3.2, or 9.3% of the unexplained variance. The items clustered perfectly, with neuroticism items positively loading, and extraversion items negatively loading, indicating two distinct dimensions. Again, the N22 and C25 items had negative loadings, but had earlier been converted to the extraversion set of items, preserving the interpretation of two distinct dimensions existing within this data set.

Since these two sets of items cohered, a further analysis of their structure was conducted, in particular, an examination of their item fit, item separation, item reliability, person separation, person reliability, and Wright maps (to see how well the items were spread across the study's participants). The explained variance for the neuroticism set of items was 52.6%, while the unexplained variance after the first contrast had an eigenvalue of 1.7, or 8.1% of the unexplained variance. The explained variance for the extraversion set of items was 40.0%, while the unexplained variance after the first contrast had an eigenvalue of 1.6, or 9.7% of the unexplained variance, as shown in Table 5.

The remaining three QuEEP factors that needed to be tested for unidimensionality included the conscientiousness, openness, and agreeableness factors. When the items from all three factors were entered into Winsteps and tested for unidimensionality, the variance explained by the measures was 43.6%, while the first contrast had an eigenvalue of 4.0 and unexplained variance of 7.6%. The openness items separated themselves as the items with the most positive loadings, while the conscientiousness items all had negative loadings and the agreeableness items straddled the middle ground with either slightly positive or slightly negative loadings. It was clear that the openness items were unidimensional, and it appeared there was a good chance that the conscientiousness and agreeableness items would be as well since there was only one conscientiousness item, C39, that was even mixing with the agreeableness items at this early stage. Nevertheless, openness appeared to be unidimensional, and a further examination would be required with the openness items isolated. The explained variance for the openness set of items was 54.5%, while the unexplained variance after the first contrast had an eigenvalue of 1.8, or 8.1% of the unexplained variance.

The remaining two *QuEEP* factors that needed to be tested for unidimensionality after removing the openness items were conscientiousness and agreeableness. When the items from both factors were entered into Winsteps and tested for unidimensionality, the variance explained by the measures was 45.7%, while the first contrast had an eigenvalue of 3.0 and unexplained variance of 8.2%. The agreeableness items all loaded positively, while the conscientiousness items all loaded negatively, suggesting two distinct dimensions. The only item from the previous

dimensionality analysis that did not cohere with items from its own scale, C39, now cohered with other items from the conscientiousness scale.

The explained variance for the agreeableness set of items was 53.3%, while the unexplained variance after the first contrast had an eigenvalue of 1.9, or 8.9% of the unexplained variance. The explained variance for the conscientiousness set of items was 52.5%, while the unexplained variance after the first contrast had an eigenvalue of 1.6, or 7.7% of the unexplained variance, as can be seen in Table 5.

While examining the dimensionality of the *IPIP BFFM*, it was expected that the Rasch analysis would similarly reveal an underlying five factor construct to the instrument. After processing all of the items through Winsteps, the explained variance was 24.7%, while the unexplained variance after the first contrast had an eigenvalue of 5.0, or 7.5% of the unexplained variance. Results suggested that the factors of conscientiousness and neuroticism needed to be isolated to determine if they could be separated into defined factors, as the positive loading portion of the table was dominated by either conscientiousness or neuroticism items. There were several openness and agreeableness items interspersed among some of the weaker loading conscientiousness and neuroticism items, extending to the negative loading scale, however, for the most part, the conscientiousness and neuroticism items cohered. Recall that with the *QuEEP*, it was the extraversion and neuroticism items, not conscientiousness items, that cohered in the initial PCA, signalling an immediate divergence between these two instruments.

Taking only the items from the conscientiousness and neuroticism scales and re-entering them into Winsteps, the explained variance for the two sets of items was revealed to be 28.4%, while the unexplained variance after the first contrast had an eigenvalue of 3.7, or 13.3% of the unexplained variance. The items clustered perfectly, with neuroticism items positively loading, and conscientiousness items negatively loading, except for one conscientiousness item, C4, which had the weakest positive loading (still clustering with the negatively loading conscientiousness items as it was weaker than *all* of the neuroticism items). This indicated two clear dimensions within the data set.

A further analysis of each set of items was conducted, in particular, an examination of item fit, item separation, item reliability, person separation, person reliability, and Wright maps to see how well the items were spread across the study's participants. The explained variance for the conscientiousness set of items was 34.0%, while the unexplained variance after the first contrast had an eigenvalue of 1.8, or 12.2% of the unexplained variance. The explained variance for the neuroticism set of items was 35.9%, while the unexplained variance after the first contrast had an eigenvalue of 2.2, or 13.9% of the unexplained variance.

The remaining three constructs that needed to be tested for unidimensionality included extraversion, openness, and agreeableness. When the items from all three constructs were entered into Winsteps and tested for unidimensionality, the variance explained by the measures was 31.0%, while the first contrast had an eigenvalue of 3.6 and unexplained variance of 8.4%. The agreeableness items separated themselves as the items with the most positive loadings, while the extraversion and openness items all had negative loadings or very low positive loadings. It is clear that the agreeableness items were unidimensional, and it appeared that the extraversion and openness items could be, but would require further analysis. Nevertheless, agreeableness appeared to be unidimensional, and a further analysis would be required with the agreeableness items isolated. The explained variance for the agreeableness set of items was 35.8%, while the unexplained variance after the first contrast had an eigenvalue of 2.3, or 14.8% of the unexplained variance.

Taking only the items from the *IPIP BFFM* openness and extraversion scales and re-entering them into Winsteps, the explained variance for the two sets of items was revealed to be 32.8%, while the unexplained variance after the first contrast had an eigenvalue of 3.0, or 10.0% of the unexplained variance. The items clustered perfectly, with openness items positively loading, and extraversion items negatively loading, indicating two clear dimensions within the data set. The explained variance for the extraversion set of items was 41.3%, while the unexplained variance after the first contrast had an eigenvalue of 1.8, or 10.4% of the unexplained variance. The explained variance for the openness set of items was 41.8%, while the unexplained variance after the first contrast had an eigenvalue of 2.6, or 15.0% of the unexplained variance.

In Table 5, the aforementioned statistics for the Rasch PCA for both the *QuEEP* and the *IPIP BFFM* have been arranged to allow for a quick comparison, and it would appear that the *QuEEP* performed better than the *IPIP BFFM* in this particular study, exhibiting superior variance, separation, reliability, and item fit in all but a few instances.

Factor	Variance accounted (eigenvalue)	Variance accounted (%)	Variance unaccounted (eigenvelue)	Variance unaccounted (%)	Item sep.	Item rel.	Person sep.	Person rel.	Infit range	Outfit range
C QuEEP	11.1	52.5%	1.6	7.7%	8.27	.99	2.31	.84	.88-1.10	.87-1.13
C IPIP BFFM	5.2	34.0%	1.8	12.2%	5.55	.97	1.65	.73	.70-1.56	.70-1.75
N QuEEP	11.1	52.6%	1.7	8.1%	5.89	.97	2.64	.87	.73-1.30	.72-1.35
N IPIP BFFM	5.6	35.9%	2.2	13.9%	5.04	.96	1.85	.77	.78-1.15	.76-1.12
A QuEEP	11.4	53.3%	1.9	8.9%	9.63	.99	2.22	.83	.72-1.29	.73-1.30
A IPIP BFFM	5.6	35.8%	2.3	14.8%	3.20	.91	1.76	.76	.85-1.32	.83-1.26
E QuEEP	6.7	40.0%	1.6	9.7%	3.55	.93	1.88	.78	.84-1.26	.85-1.24
E IPIP BFFM	7.0	41.3%	1.8	10.4%	3.60	.93	2.06	.81	.77-1.38	.78-1.36
O QuEEP	12.0	54.5%	1.8	8.1%	8.79	.99	2.31	.84	.85-1.34	.91-1.16
O IPIP BFFM	7.2	41.8%	2.6	15.0%	8.60	.99	1.54	.70	.74-1.37	.73-1.47

Note, Variance unaccounted for refers to the 1st contrast

Discussion

With regard to the factor analysis results, it is difficult to conclusively say that the items on the *QuEEP* are better at capturing personality than those of the *IPIP BFFM* because of the superior factor loading of the *QuEEP* items. It is difficult because the *QuEEP* items in this study were cherry-picked *because* of their superior factor loading. However, if the results of this study can be replicated, then it would seem the *QuEEP* is indeed a superior instrument, at least in this particular context. Both instruments seem to do a fairly good job at capturing temperament-dimensions like extraversion and neuroticism, with all 20 of the *QuEEP* items and 18 of the *IPIP BFFM* items with a factor loading above +/- .40.Yet these two instruments diverge when measuring the agreeableness, conscientiousness, and openness items, with all 30 of the *QuEEP* items and only 15 of the *IPIP BFFM* items achieving an acceptable factor loading. It is a bit surprising to see conscientiousness items in this group as it would seem conscientious behaviour would be fairly consistent across different contexts, cultures, and relatively easy to identify. Generally, items that signify hard-work do not get lost in translation and conscientiousness is easily quantifiable in the minds of respondents when they evaluate their own behaviour, and so the divide between these two instruments probably should be minimal, yet the *IPIP BFFM* had four items that did not reach an acceptable minimum factor loading.

However, in all fairness, of the four that did not achieve the minimum factor loading, three had a factor loading of +/- .39, which should probably be considered close enough. The real weakness in the *IPIP BFFM* is seen in the agreeableness and openness factor, where only 9 items (of 20) achieved acceptable factor loadings and four items even loaded on the wrong personality factor (which was the conscientiousness factor in all four cases). Only four openness items actually loaded on the openness factor, and this was probably just a result of redundancy in the items, not a connection to an underlying construct of openness. Items O5 and O10 are both about ideas ("am full of ideas" and "have excellent ideas") while items O3 and O6 are both about imagination ("have a vivid imagination" and "do not have a good imagination"). With four of the ten items about the related concepts of ideas and imagination, their similarity probably formed the basis of the construct, not an actual insight into openness. Redundancy in items can skew a factor towards that redundancy, making non-redundant items ineffective, weakening the cohesiveness of the set of items. In this respect, the *QuEEP* succeeds in there is less redundancy in the wording of items, and underlying constructs are based on the collective influence of all ten items, not a dominating inner group of four items that skew the result.

Interestingly, the results of the correlational analysis indicate stronger correlations between the *QuEEP* factors than with the *IPIP BFFM* factors, which may also stem from redundancy, in this case the re-use of common themes in the items of different factors within the *QuEEP*. With the *QuEEP* limited to specific situations in order to capture a "language personality", there is a re-use of certain themes that must occur, such as travelling, studying in English class, and talking with a native speaker. As a result, the higher inter-factor correlations seems almost unavoidable, and a necessary consequence of this highly situational instrument. For example, travel themed items in the extraversion factor may be eliciting similar responses as travel themed items in the openness factor, resulting in higher correlations. This should not diminish the validity of the responses as the factor analysis and the Rasch analysis proved each set of *QuEEP* items to be sufficiently unidimensional.

Looking at the results of the Rasch analysis, four of the five *QuEEP* personality factors achieved superior statistical performance over their *IPIP BFFM* counterparts. The explained variance was higher, the unexplained variance was lower, the item and person separation was

greater, the item and person reliability was higher, and the item fit spreads narrower for the neuroticism, agreeableness, conscientiousness, and openness factors. Only the *QuEEP* extraversion factor underperformed the *IPIP BFFM* extraversion factor, yet even in this case the two sets of items performed relatively similarly.

Viewed holistically, it would seem that the *QuEEP* outperformed the *IPIP BFFM*. Further, if you look at the correlations among like factors between the two instruments, correlations were relatively weak. This affirms the main thesis of this article in that native-language personality instruments are not effectively measuring students' second-language personality, because if they were, these correlations would be much higher. This opens up the exciting possibility that much of the existing TESOL research involving personality may not have accurately captured second-language personality. If this is indeed the case, it would necessitate a critical re-examination of previous TESOL research involving personality, possibly invalidating several studies.

Conclusion

The research in this article showed, at the most basic level, the *QuEEP* and the *IPIP BFFM* were fundamentally measuring different things. Very limited correlations between the two instruments indicate that a person may indeed have two distinct personalities that are affected by culture, society, and language. At the very least, this research should open a dialogue on the need for a specialized TESOL personality instrument, and how one can be developed. A more effective personality instrument could change how existing TESOL personality research is perceived, and open up several avenues of future research.

About the author

Omar Karlin is a non-tenured Associate Professor at Meiji University in Tokyo. He earned his M.Ed in TESOL from Temple University, and is a doctoral candidate in Temple University's Doctorate of Education program. His primary research interest is how personality can affect second language acquisition.

References

Ashton, M. C., Lee, K., & Goldberg, L. R. (2006). The IPIP-HEXACO scales: an alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences, 42*(8), 1515-1526. doi:10.1016/j.paid.2006.10.027

Bowler, M. C., Bowler, J. L., & Phillips, B. C. (2009). The Big-5+2? The impact of cognitive complexity on the factor structure of the five factor model. *Personality and Individual Differences, 47,* 979-984. doi:10.1016/j.paid.2009.08.002

Garcia, L. F., Aluja, A., Garcia, O., & Cuevas, L. (2005). Is openness to experience an independent personality dimension? Convergent and discriminant validity of the openness domain and its NEO-PI-R facets. *Journal of Individual Differences*, *26*(3), 132-138. doi:10.1027/1614-0001.26.3

Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment, 4*, 26-42. doi:10.1037/1040-3590.4.1.26

Goldberg, L. R. (2007). A Comparison between the 24 facet scales in Lee and Ashton's HEXACO personality inventory and the corresponding 24 preliminary IPIP scales measuring similar constructs. Retrieved April 15, 2012, from International Personality Item Pool Web site: http://ipip.ori.org/newHEXACO_PI_table.htm

Heuchert, J. W. P., Parker, W. D., Stumpf, H., & Myburgh, C. P. H. (2000). The five-factor model of personality in South African college students. *American Behavioral Scientist, 44(*1), 112-125. doi:10.1177/00027640021956125

Horn, B. S. (2000). *A Xhosa translation of the revised NEO personality inventory: a pilot study.* Unpublished master's thesis. Port Elizabeth, South Africa: University of Port Elizabeth.

Hutcheson, G., & Sofronion, N. (1999). *The multivariate social scientist*. London: Sage.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 35*, 401-415. doi:10.1007/BF02291575

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioural Research, 39*(2), 329-358. doi:10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO personality inventory: Two new facet scales and an observer report form. *Psychological Assessment, 18(*2), 182-191.

McCrae, R. R., & Costa, P. T. (1999). A five factor theory of personality. In L. A. Pervin & O. P. John (Eds.) *Handbook of personality: Theory and research* (pp. 102-138). New York: The Guilford Press.

McCrae, R. R., Yik, M. S. M., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. *Journal of Personality and Social Psychology, 74*(4), 1041-1055.

Piedmont, R. L., & Aycock, W. (2006). An historical analysis of the lexical emergence of the Big Five personality adjective descriptors. *Personality and Individual Differences, 42(*6), 1059-1068. doi:10.1016/j.paid.2006.09.015

Saucier, G., Bel-Bahar, & Fernandez, C. (2007). What modifies the expression of personality tendencies? Defining basic domains of situation variables. *Journal of Personality*, *75*(3), 479-504. doi:10.1111/j.1467-6494.2007.00446.x

Woehr, D. J., Miller, M. J., & Lane, J. A. (1998). The development and evaluation of a computer administered measure of cognitive complexity. *Personality and Individual Differences, 25*, 1037-1049. doi:10.1016/S0191-8869(98)00068-3

Wood, D., & Roberts, B. W. (2006). Cross-sectional and longitudinal tests of the personality and role identity structural model (PRISM). *Journal of Personality*, *74(*3), 779-810. doi:10.1111/j.1467-6494.2006.00392.x