



A Rasch Analysis of the Extraversion/Introversion Dimension of Personality

Paul Leeming

Abstract Researchers interested in individual differences have long sought to categorize personality and to develop a succinct model in order to achieve consistency within different academic fields. As a result of this, the Big Five-Model of Personality has emerged as the dominant model, describing personality in terms of five different traits. Various questionnaires have been developed to measure these traits and the International Personality Item Pool (IPIP) is a well-established and relatively concise example. The current study attempted to determine the reliability and dimensionality of the extraversion dimension of the IPIP by use of Rasch analysis to examine the results of 79 university students who completed the extraversion part of the IPIP measure. Results showed that generally the questions fit the Rasch model and support the unidimensionality of the extraversion dimension of the IPIP test.

Key words personality, Big Five Model, Rasch analysis
September 16, 2014 accepted

1. Introduction

Researchers in many different academic fields are interested in individual differences and how they affect behavior, from learning a language, to leading a group in a work setting (Forsyth, 2010). One of the most researched areas in individual differences is personality, and there have been numerous attempts to describe models of personality. The Big Five Model of personality (Goldberg 1992; 1993) has become widely accepted and is now the dominant model in psychology. This paper describes a Rasch analysis of the Extraversion/Introversion dimension from the IPIP test, which is currently one of the most popular questionnaires for measuring personality. After presenting a brief history of personality research, I describe the methodology for the study, and give a succinct summary of some of the key aspects of the Rasch model for measurement (Rasch, 1960). I then present the results of the analysis, before a brief discussion of how well the test conforms to the Rasch model. I close the paper by suggesting some possible modifications to the test that may enhance its validity in a Japanese context.

2. The Big Five Model of Personality

2.1 A Brief History

Personality has been of significant interest to researchers in all fields of psychology including those interested in leadership and groups, as it has been thought to influence the roles that members of the group will adopt and also how they interact with each other. Research in this area has generally striven for a parsimonious model to describe and measure personality across many different contexts and cultures, and several models have been suggested including Eysenck's three-part model (Eysenck & Eysenck, 1985). Despite the large number of models available, the model that has become ubiquitous in the field of personality research is the Big Five model, generally credited to Goldberg (1992; 1993), and McCrae and Costa (2003), and now used extensively. This model describes five different dimensions to personality

A Rasch Analysis of the Extraversion/Introversion Dimension of Personality (Leeming) labeled as Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. Neuroticism describes the extent of emotional adjustment or stability, while extraversion represents the scale from extroversion to introversion. Openness to experience measures how eager individuals are to experience new things and how imaginative they are, also incorporating broad-mindedness and tolerance of others' views. Agreeableness measures how well people get along with those around them and how well they adapt to social situations. The final dimension, conscientiousness, measures how hardworking individuals are and how much they strive to achieve in their daily lives. Each of these five domains is represented by six lower level dimensions, providing a more detailed model of personality. People are assumed to be along a scale for each of the different dimensions.

Although some researchers have criticized the model claiming that it oversimplifies the complexity of personality and therefore lacks comprehensiveness (Funder, 2001, p. 201), the model has been used extensively in varied contexts and cultures, with results showing generalizability of the dimensions. Verhoeven and Vermeer (2002) were the first to use the Big Five model in the field of second language acquisition, and the model has recently been used in a Japanese context to predict second language speaking anxiety (Apple, 2011).

Costa and McCrae (1992) developed the NEO-PI which is a 240-item self-report questionnaire designed to measure the five dimensions of personality. Although this instrument has been found to have good reliability, the length prohibits use in many situations and researchers have attempted to develop more parsimonious measures that can be used more readily. Goldberg produced a 100 item measure (Goldberg, 1992) and then subsequently developed the International Personality Item Pool (IPIP) (Goldberg, 1999), which is a 50-item five-factor measure with ten items designed to measure each of the Big Five factors. He called for collaboration and for researchers to pool resources in order to ensure that consistent and effective measures are developed. This measure has been widely used in the literature, and although some researchers have attempted to shorten it further with just four items for each dimension (Donnellan, Oswald, Baird, & Lucas, 2006) for the purposes of Rasch analysis it is beneficial to have at least six items measuring each construct.

The measure has been found to be reliable in a number of different contexts including Japan (Hendriks et al., 2003), where it has been translated into Japanese.

2.2 International Personality Item Pool

The International Personality Item Pool (IPIP) was used to measure the Big Five dimensions of personality in this study. This instrument was developed through the collective work of many researchers (Gow, Whiteman, Pattie, & Deary, 2005), has been used extensively (see website <http://ipip.ori.org/>), and is concise containing just 50 items. Unlike other lengthier measures of personality such as the NEO PI-R (Costa & McCrea, 1992), which has a total of 240 items, and the IPIP-NEO (Goldberg, 1992), which has 100 items, this questionnaire can be completed in a relatively short period of time. An example item designed to measure the introvert/extrovert dimension of personality is *I am always the life of the party* (see Appendix A for the full measure of introvert/extrovert in English and Japanese). The original measure adopts a five-point Likert scale but for the purposes of this study this was changed to four points by removing the middle option as it was felt that the middle option of “neither accurate nor inaccurate” effectively gave students a chance to avoid answering the question (1=Very inaccurate; 4=Very accurate). The Japanese instrument was administered in this study.

3. Methodology

3.1 Participants

The participants in the study were 79 students (56 male and 23 female students) in a first-year compulsory English communication course of a science department at a private university in western Japan. Age ranges of students were from 18 to 22, with 77 first year students and 2 students who were repeating the course and were in the third or fourth year. Participants were all native speakers of Japanese. The questionnaire was administered to the students online through Survey Monkey. Instructions were given in Japanese, and students were given a chance to ask questions. Students were given 10 minutes to complete the entire IPIP measure,

A Rasch Analysis of the Extraversion/Introversion Dimension of Personality (Leeming) and all students were able to do so in the allotted time. The survey and all online instructions were also in Japanese. Following the survey, Rasch analysis was performed to analyze the extraversion dimension of the IPIP measure.

3.2 Rasch Analysis

This study utilizes Rasch analysis for validation and refinement of the measure used, and I begin with a description of Rasch analysis, outlining the advantages it has over other methods of measurements, before explaining how to interpret the data produced during analyses.

Georg Rasch (1960) developed the Rasch model for measurement and described the basic principal of the model as follows:

a person having greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one. (Rasch, 1960, p. 117)

The two elements that are of importance are the ability of the person and the difficulty of the item. A person's ability is measured in relation to the difficulty of the items. The Rasch model places items in order of difficulty based on the answers of respondents, and people in order of ability based on their responses to the items. The raw score is converted using natural logarithm so that the distances between different item measures are on an interval scale. Person abilities and item difficulties are then represented in terms of logits (log odd units).

Interval scale data is a fundamental requirement for most statistical processes and many researchers take the ordinal data gathered from questionnaires and assume that it is scaled when conducting subsequent analysis. As Bond and Fox (2007, p. 24) explained, gaining extra points in the middle range of a test does not equate to the same increase in ability as gaining extra points at the top of the test. Often the data are not scaled and yet many researchers make this assumption by summing these raw scores from individual items to derive an overall score for

respondents with regard to a given latent construct. While ratio measures such as those adopted in the physical sciences might be beyond the possibilities of measurement in social science, it is essential that researchers attempt to produce measures that provide accurate interval scale data in order to allow for meaningful comparison between respondents.

3.2.1 Model Fit

Perhaps the most important consideration when using the Rasch model is that of model fit which addresses the issue of how closely the data gathered fit the idealized model of how the measurement should function. Rasch analysis does not explain the data but simply informs the researcher as to how closely the data gathered matches or fits with the properties of ideal measurement. The statistics of key relevance are the fit statistics, which indicate the extent to which our data obeys the principles of the Rasch model for measurement.

The two fit statistics most readily reported by researchers are the infit and outfit mean square. The infit mean square considers the response pattern of individuals around their level of ability, while outfit mean square describes how an individual responds to items that are either considerably above or below their actual ability.

The ideal value for both measures is 1 but they can range from zero to positive infinity. Infit or outfit values of less than 1 imply that the person or item is overfitting the model, while values over 1 suggest that the person or item is underfitting the model. Bond and Fox (2007) were quick to emphasize that the acceptable range of values for items depends on the testing situation, but gave benchmark ranges for infit and outfit mean square as greater than .75 and less than 1.30. In a language learning context McNamara (1996) recommended acceptable values as being within the range of double the standard deviation of fit for a given item, and this is the criteria that was adopted for the current study. Overfit means that the response pattern is too close the ideal response pattern and there is none of the ambiguity that is expected in the Rasch model, which takes into account the imperfect responses that can be expected when attempting measurement of latent constructs. Underfit means that the there is an erratic response pattern for a person or an item, and

A Rasch Analysis of the Extraversion/Introversion Dimension of Personality (Leeming)
that they cannot be accurately described by the Rasch model.

Generally outfit is not of as much concern for researchers as this is measuring performance far from actual ability or difficulty, and can simply indicate specialized knowledge or a lack of some simple knowledge (Bond & Fox, 2007). Infit is of far more interest as it is sensitive to the true ability of the individual or difficulty of the item, and underfit around the true ability means that there is a lack of sensitivity in measurement, or that the person is responding in an unpredictable manner.

Interpretation of underfitting and overfitting performance are quite different. Underfit means that the measures are simply not performing and not accurately measuring the performance of the people. As such, underfit is by far the most important concern for researchers using the Rasch model to verify their measures. Overfit might lead researchers to become overconfident in the measures they have developed, as it slightly inflates reliability estimates, but is unlikely to pose as serious a threat to the reliability of measures.

3.2.2 Principle Components Analysis of Item Residuals

Although factor analysis is often used to analyze dimensionality, I relied on Rasch principle components analysis of item residuals to check the dimensionality of the IPIP. Factor analysis requires a large number of participants and was therefore not suitable for this study. The Rasch model is based on a central assumption that the items are all attempting to measure the same latent construct. Misfit, as described above, is a strong indicator that the need for unidimensionality has not been met, but another strong indicator of the dimensionality of the measure is available through Principal Components Analysis (PCA) of item residuals.

PCA examines the amount of common variance in the items explained by the first dimension, in line with the Rasch model based on unidimensionality. Items that have large amounts of variance not explained by the first dimension contribute to the unexplained variance, and suggest that the measure has more than one component. Linacre (2007) stated that for a measure to be considered unidimensional it should account for greater than 50% of the variance explained by measures, while the unexplained variance explained by the first contrast should be less than

5 %, or have an eigenvalue lower than 3.0. Linacre (2007) did emphasize that these are general guidelines and that there are exceptions to these rules.

3.2.3 Rasch Reliability and Separation

Rasch analysis offers two indices of reliability. The person reliability index indicates the consistency of the person separation should the same respondents be given a different but equivalent measure hypothesized to measure the same construct. It addresses the issue of how replicable the results would be using the same participants with a different but equivalent measure. Person reliability also requires the items to separate the participants so that some perform at high levels and others at low, which means the measure must effectively distinguish between differences within the given population on the construct of interest. Values for person reliability range from 0 to 1 and are analogous to Cronbach's alpha, with values closer to 1 indicating higher reliability. Interpretation of reliability statistics follows that of Cronbach alpha with values over .90 indicating strong reliability, values over .80 indicating good reliability, and values over .70 indicating acceptable reliability (Sheridan & Puhl, 1996, p. 26).

The item reliability index indicates the replicability of item separation should the items be given to a different but equivalent group of respondents. That is, would the items still behave in the same way with a similar sample of the population? High reliability suggests that the difference in difficulty is consistent and the researcher can be confident that the item difficulties are reasonably stable. Again values range from 0 to 1 and follow the same principles as for person reliability.

The Rasch person separation index is a more sensitive measure of how well the items are able to differentiate between the respondents for the given variable. This index is not bound by 1, and values range from zero to infinity, with high values indicating that the measure is effective in distinguishing between the respondents. Generally values greater than 2.0 are regarded as providing acceptable separation. Measures with low person separation are not measuring differences in the sample and therefore are of limited use.

In a similar manner, the Rasch item separation index is an indicator of how

A Rasch Analysis of the Extraversion/Introversion Dimension of Personality (Leeming) well spread or separated the items are on the given variable. Again values range from zero to infinity and high values indicate that the items are well spaced in terms of difficulty, with values greater than 2.0 being acceptable. As with person separation, it is important that the measures developed cover a wide range of abilities to avoid floor and ceiling effects.

4. Results and Discussion

The next section presents the results of the Rasch analysis, focusing first on the items, and then on the overall dimensionality of the test.

4.1 Item analysis

The 10 items from the IPIP designed to measure extroversion were analyzed with the Rasch rating scale analysis. The Likert-scale category functioning was examined for the Extroversion measure (Table 1), in order to ensure that the scale was functioning according to the Rasch model. The measure adopted a four-point scale (1=Strongly disagree, 2=Disagree, 3=Agree, and 4=Strongly agree). The minimum of 10 observations per category was met, with the smallest number of observations being 70 in category 4. The outfit mean square statistic was below the criteria of 2.00 for all categories. Separation between categories was greater than the required 1.1 logits for a four-point scale. From this it was concluded that the Likert-scale was performing in a satisfactory manner, and none of the categories were collapsed.

Table 1. Category Structure Functioning for Extroversion

Category	Count	(%)	Infit MNSQ	Outfit MNSQ	Structure measure	Category measure
1 Strongly disagree	127	16	1.11	1.10	None	(-3.46)
2 Disagree	355	44	.86	.88	-2.30	-1.14
3 Agree	258	32	.92	.92	.08	1.18
4 Strongly agree	70	9	1.05	1.09	2.21	(3.39)

Table 2 shows the items fit statistics, which indicate that item E10 is not fitting

the Rasch model based on criteria provided by McNamara (1996), which would allow for items with infit mean square values ranging from .62 to 1.38. This item is negatively coded and states “*Am shy around strangers,*” seeming to fit in with the theoretical construct of extroversion. In order to further test the construct validity of this item, person measures with and without E10 were exported to SPSS, and a bivariate correlation analysis was performed. The correlation was significant ($p < .001$), and the Pearson correlation was .99, indicating that removal of the item does little to change the measure for people. The other items seem to be following the Rasch model for measurement.

Table 2. *Rasch Item Statistics for the Extrovert Items*

Item	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-Measure correlation
E01	.81	.18	.86	-1.0	.88	-.8	.71
E09	.59	.18	.97	-.1	.98	-.1	.57
E10	.46	.18	1.40	2.5	1.41	2.5	.58
E08	.00	.17	1.01	.2	1.01	.2	.57
E05	-.09	.17	.79	-1.4	.79	-1.5	.82
E06	-.09	.17	1.27	1.7	1.27	1.6	.49
E07	-.12	.17	.82	-1.2	.83	-1.2	.68
E04	-.21	.17	.84	-1.1	.83	-1.1	.74
E03	-.64	.17	1.04	.4	1.10	.7	.52
E02	-.70	.17	.91	-.6	.92	-.5	.67

The Wright map (Figure 1) shows persons on the left according to their extroversion measures, and items on the right according to their difficulty to endorse. A person on the left has a 50% chance of endorsing the item on the right side of the vertical line. Items increase in difficulty as they move up, and people increase in extraversion as they move up the scale. The average measure for persons was $-.38$, indicating that the items were well matched to the people, although the items were slightly difficult to endorse for this particular group.

The most difficult item to endorse was “am the life of the party” as this suggests someone who is very extroverted. The next most difficult item to endorse was “don’t mind being the center of attention” which again suggests someone who is at the core of any social gathering and is very extroverted. The item “don’t talk

A Rasch Analysis of the Extraversion/Introversion Dimension of Personality (Leeming)

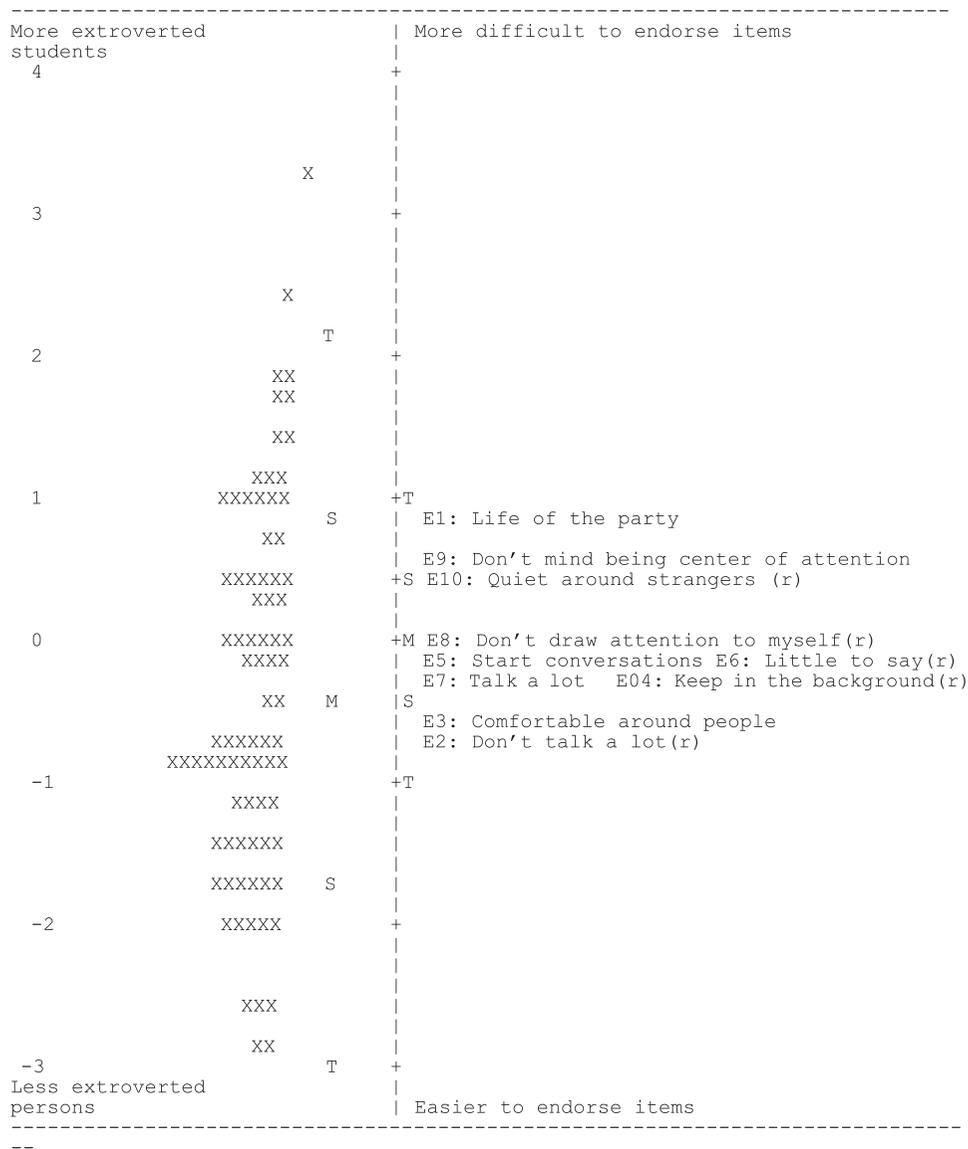


Figure 1. Wright map for Extrovert dimension of IPIP measure. M=Mean; S=1 SD; T=2 SDs. Each X=approximately 2 persons. (r) indicated negatively coded items.

a lot,” negatively coded, was easiest to endorse suggesting that most students consider themselves to be reasonably talkative. The second easiest item to endorse was “feel comfortable around people,” which again would suggest basic social skills. Items close to the mean for the people were generally related to being talkative, and clearly these students consider themselves to be comfortable in simple social

settings. The hierarchy of items is as expected, with easy to endorse items being accepted by most students, while the items that suggest high degrees of extraversion are rejected by most. There is a reasonable spread of items suggesting that they are measuring a range of extraversion. Based on the Wright map, the items are performing as expected.

4.2 Reliability and dimensionality

Person separation was 2.07 and item separation was 2.37, both above the benchmark of 2.0, indicating that the reliability of the measures was good, and that the items were reasonably well matched to the population in this study. Rasch person reliability was .81 and Rasch item reliability was .85, again indicating reasonable reliability.

The Rasch PCA of item residuals analysis showed that 44.1% of the variance (eigenvalue=7.9) was explained by the Rasch model, 19.4% of the variance (eigenvalue =3.5) was explained by the items, and 12.1% of the variance (eigenvalue=2.2) was explained by the first residual contrast. Although the variance explained by the Rasch model is slightly below the 50% criterion suggested by Linacre (2007), the unexplained variance in the first residual contrast with an eigenvalue of 2.2, meets the criterion of less than 3.0.

Table 3 shows the loading of items from the Extrovert measure. Items E02 and E06 have high positive loadings above .40, and items E09, E08, and E03 have strong negative loadings. Close examination of the items reveals that they are measuring slightly different aspects of what is commonly understood to be extroversion. Items E02 (Don't talk a lot) and E06 (Have little to say) both refer to talking, or having something to say, while items E03 (Feel comfortable around people), E08 (Don't like to draw attention to myself), which is a negatively coded equivalent of item E09 (Don't mind being center of attention), refer to being the center of attention or being around people. While these two dimensions differ slightly they are both construed to be part of extroversion, and therefore were deemed as being appropriate to include under the single dimension of extraversion. Overall, the items appear to form a fundamentally unidimensional construct.

Table 3. *Rasch PCA of Item Residuals for the Extrovert Items*

Item	Loading	Measure	Infit MNSQ	Outfit MNSQ
E02	.76	-.70	.91	.92
E06	.56	-.09	1.27	1.25
E10	.39	.46	1.40	1.41
E05	.15	-.09	.79	.79
E04	.13	-.21	.84	.83
E09	-.75	.59	.97	.98
E08	-.48	.00	1.01	1.01
E03	-.46	-.64	1.04	1.10
E07	-.21	-.12	.82	.83
E01	-.20	.81	.86	.88

5. Conclusion

The analysis showed that the 10 questions were effective in measuring the extraversion dimension of personality with this group of participants. All of the items seemed to fit the Rasch model, and the measure was unidimensional. The questions were generally effective in differentiating between the levels of extraversion in different students in this study, although the Wright Map suggests that the items are a little difficult to endorse for the majority of students, and therefore simpler items may be beneficial in a Japanese context. The results suggest that in this context the IPIP extraversion measure generally was suitable. It should be noted that the reliability of a test is concerned with the interaction of the items with the participants, and therefore, although I would recommend the IPIP to other researchers interested in measuring extraversion, extensive analysis should be conducted to ensure the measure remains valid with a different population.

This paper has also demonstrated how Rasch analysis can be used to assess quantitative questionnaire data to ensure that it is valid and can be used in subsequent analyses. Researchers in the social sciences, and particularly in my own field of second language acquisition, will often use a pre-established measure without considering its validity within their own research context. Rasch analysis not only gives us confidence that our measure is reliable, but also converts students' results to a true-scale score which is essential for most kinds of statistical analyses.

References

- [1] Apple, M. T. (2011). The Big Five personality traits and foreign language speaking confidence among Japanese EFL students. Unpublished doctoral dissertation. Temple University.
- [2] Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Routledge.
- [3] Costa, P. T., & McCrea, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment.
- [4] Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, 18(2), 192–203.
- [5] Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences*. New York: Plenum.
- [6] Forsyth, D. R. (2010). *Group dynamics*. Belmont, CA: Cengage Learning.
- [7] Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52(1), 197–221.
- [8] Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, 4(1), 26–42.
- [9] Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48(1), 26–34.
- [10] Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several Five-Factor models. In I. Mervielde, I. J. Deary, F. de Fruyt, & F. Ostendorf (Eds.). *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg: Tilburg University Press.
- [11] Gow, A., Whiteman, M., Pattie, A., & Deary, I. (2005). Goldberg's "IPIP" Big-Five factor markers: Internal consistency and concurrent validation in Scotland. *Personality and Individual Differences*, 39(2), 317–329.
- [12] Hendriks, J., Perugini, M., Angleitner, A., Ostendorf, F., Johnson, J., De Fruyt, F., Ruisel, I. (2003). The five-factor personality inventory: cross-cultural generalizability across 13 countries. *European Journal of Personality*, 17(5), 347–373.
- [13] Linacre, J. M. (2007). A user's guide to WINSTEPS: Rasch-model computer program. Chicago, IL: MESA Press.
- [14] Linacre, J. M., & Wright, B. D. (2007). WINSTEPS: Multiple-choice, rating scale, and partial credit Rasch analysis [Computer software]. Chicago, IL: MESA.
- [15] McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: a five-factor theory perspective*. New York: Guilford Press.
- [16] McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- [17] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Denmark Paedagogiske Institut.
- [18] Verhoeven, L., & Vermeer, A. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics*, 23(3), 361–374.

APPENDIX A

IPIP EXTROVERT/INTROVERT (ENGLISH AND JAPANESE VERSIONS)

Below there are phrases describing people's behaviors. Please use the rating scale provided to describe how accurately each statement describes you. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. Please read each statement carefully, choose the best response, and fill in your responses using the corresponding number.

Response Options:

- 1: Very Inaccurate 2: Moderately Inaccurate 3: Moderately Accurate
4: Very Accurate

Extroversion

1. Am the life of the party.
2. Don't talk a lot.
3. Feel comfortable around people.
4. Keep in the background.
5. Start conversations.
6. Have little to say.
7. Talk to a lot of different people at parties.
8. Don't like to draw attention to myself.
9. Don't mind being the center of attention.
10. Am quiet around strangers.

以下に人々の行動特性を述べた記述があります。あなた自身がそれぞれの記述にどの程度当てはまるかを、次の尺度値を用いて回答してください。

回答は現在のあなた自身についてで、将来の希望についてではありません。

あなた自身について正直に思ったように回答してください。あなたと同性で同年代の知っている誰かと比べて答えてください。

1. 全く当てはまらない 2. あまり当てはまらない 3. だいたい当てはまる

4. 良く当てはまる

Extroversion

1. 盛り上げ役である
2. おしゃべりではない
3. 人の中で気持ちよくいられる
4. 引っ込み思案である
5. 自分から話し掛ける方である
6. 多くを語らない
7. パーティでは色々な人と話す
8. 人の注意を引くようなことは好きではない
9. 注目の的になることを好む
10. 人見知りする