Chapter 31

Applying the Mixed Rasch Model to Personality Questionnaires

Jürgen Rost, Claus Carstensen and Matthias von Davier

Institute for Science Education - IPN, Kiel

1. Introduction

Factor analytic research on the structure of personality traits continues to reveal five big factors, called the „Big Five“ (McCrae & Costa, 1987). Two of them, extraversion and neuroticism, are well known in classical theories of personality, in particular they are two of the three factors of the „type level“ in Eysenck's model (Eysenck, 1976). The remaining three of the Big Five are discussed under different labels, e.g., openness, agreeableness and conscientiousness.

This renaissance of the traditional factor analytic paradigm in personality research seems to ignore the current discussion on the consistency of human behavior that was initiated by Bem and Allen (1974) with their provocative paper „on predicting some of the people some of the time“. These authors demonstrated that personality traits may have more or less predictive power depending on how consistent people rate their own trait-related behavior. Incidentally, they did this empirical study by means of items that are strongly related to two factors of the Big Five, i.e., agreeableness and conscientiousness. Hence, the question arises, whether or not the Big Five personality factors turn out to be scalable traits for *all* individuals when analyzed with a latent trait model accounting for differential scalability in latent subpopulations.

Such a model is the *mixed Rasch* model identifying subpopulations of individuals in which the Rasch model holds, but with different sets of model parameters among these subpopulations (cf. section 2.3 of the introductory chapter). The present chapter applies this model in its generalized version for polytomous ordinal responses (Rost, 1991) to questionnaire data gained with the German version of the five-factor inventory by Costa and McCrae (1989; Borkenau and Ostendorf, 1991).

2. The data

The results of only two of the 5 traits are reported in detail here, i.e., one classical trait, extraversion, and one of the two traits more likely to be subject to differential scalability, conscientiousness. Two typical items of each trait are:

No 1: I keep my things proper and clean

No 7: I am working hard to reach my goals

for conscientiousness, and

No 4: I really like to talk with other people

No 8: I am a joyful and humorous person

for extraversion.

Each subscale of this questionnaire covers 12 items and the original response format is a 5-point scale ranging from „strongly disagree“ to „strongly agree“ including a neutral middle category („neither/nor“). This questionnaire has been applied in various samples by Borkenau and Ostendorf (1991) who were so kind as to leave the data to us for a reanalysis by means of mixture distribution models. The entire data set covers 2112 individuals, but we omitted all persons who did not answer at least one item in one of the scale, so that the reduced sample size is 1914.

In all analyses of these 5-category data it turned out that there is a strong tendency for people to avoid the middle category. This conclusion can be drawn from the threshold parameters of the ordinal Rasch model and also of the mixed Rasch model, when two or more latent classes are assumed. In these cases, the third threshold is easier than the second threshold for almost all items indicating that the middle category cannot be regarded as an intermediate stage between the second and the fourth category. As a consequence, we joined the second and third category by simply recoding the data. The three new thresholds then were in the correct order.

3. The identified subpopulations

For both cases, conscientiousness (C) and extraversion (E), we decided on the basis of the likelihood statistics to interprete the two-class solutions in a first step. Table 1 shows that this choice is justified for the C-scale, but not for the E-scale when the CAIC criterion (Bozdogan, 1987) is applied. The CAIC is defined as

$$\mathrm{CAIC} = -2\log \mathrm{L} + \mathrm{N}_{\mathrm{par}}\left(1+\log(\mathrm{N})\right) ,$$

where N is the sample size, N_{par} the number of independent parameters and L the likelihood of a particular model. According to this criterion, a model fits the data better than an alternative model if its CAIC value is smaller.

	1 class	*2 classes*	*3 classes*	*4 classes*	*5 classes*
E Scale	46689.41	46022.95	45778.67	*45740.12*	45853.29
C Scale	45905.77	*45344.81*	45373.21		

Table 1: The CAIC values for different numbers of classes of the mixed Rasch model

For the E-scale, we will end up with a 4-class model, which has the smallest CAIC. But it is easier to understand the different results for E and C when starting with both two-class solutions. Figures 1 and 2 show the profiles of the item parameters in both classes for both scales.

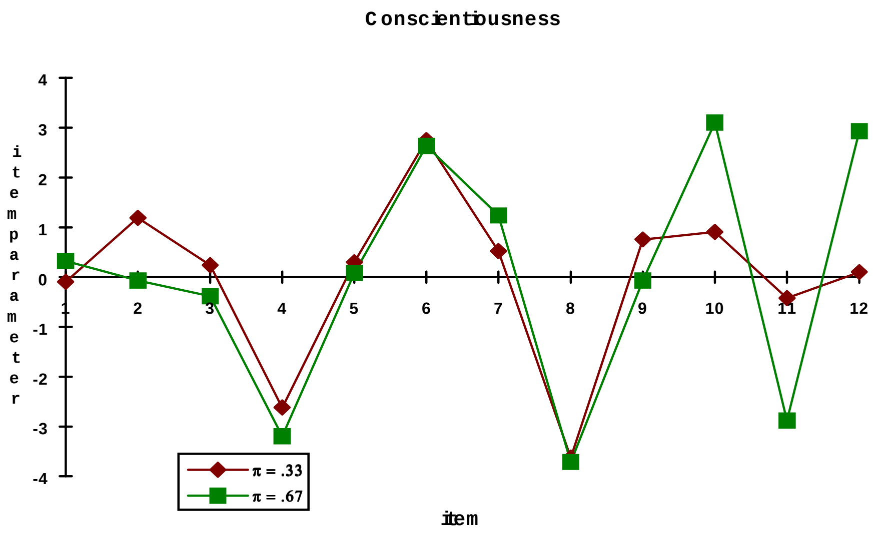

Figure 1: Item parameters for each item in both classes of conscientiousness

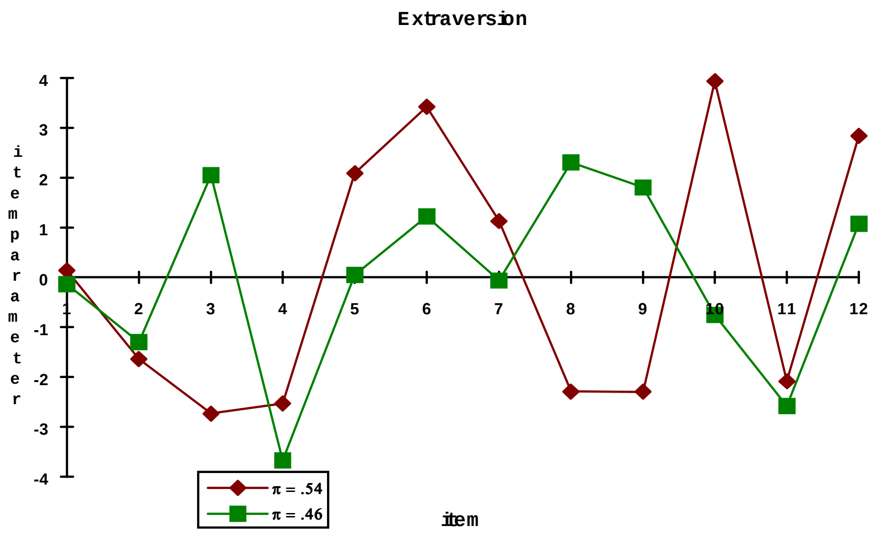

Figure 2: Item parameters for each item in both classes of extraversion

According to these graphs, there seems to be no difference at all between the two classes in the C-scale, whereas there seem to be strong differences in the item difficulties in the E-scale. Starting with the latter, items 3, 8, and 9 are much easier relatively in the first class than in the second. These items are:

No 3: I do not consider myself to be very cheerful

No 8: I am a joyful and humorous person

No 9: I am not a happy-go-lucky optimist

Items 3 and 9 are worded negatively and have been recoded accordingly (highest score for lowest category). These three items address a special part of extraversion, i.e., to be a cheerful and humorous person, that obviously is not shared by all extraverted people. The two classes of the mixed Rasch model divide the sample of individuals into two groups of persons, those who are more likely to have the happy-go-lucky feeling than to show other parts of extraverted behavior, and those who are extraverted without always having this feeling.

This result reflects the well-known heterogeneity of the extraversion construct, i.e., to have the two facets of *sociability* and *impulsivity*. Whereas in factor analysis the related items would have loadings on different factors, this heterogeneity is reflected by different item profiles in different classes in a mixed Rasch analysis.

In order to understand the two C-classes, it is necessary to look at the entire set of threshold parameters, not only at the item parameters, that are defined as the sum of all threshold parameters, i.e., they equal σ_{img} in section 2.3 of the introductory chapter. Figures 3 and 4 show the profiles of the threshold parameters τ_{ixg} for both C-classes.

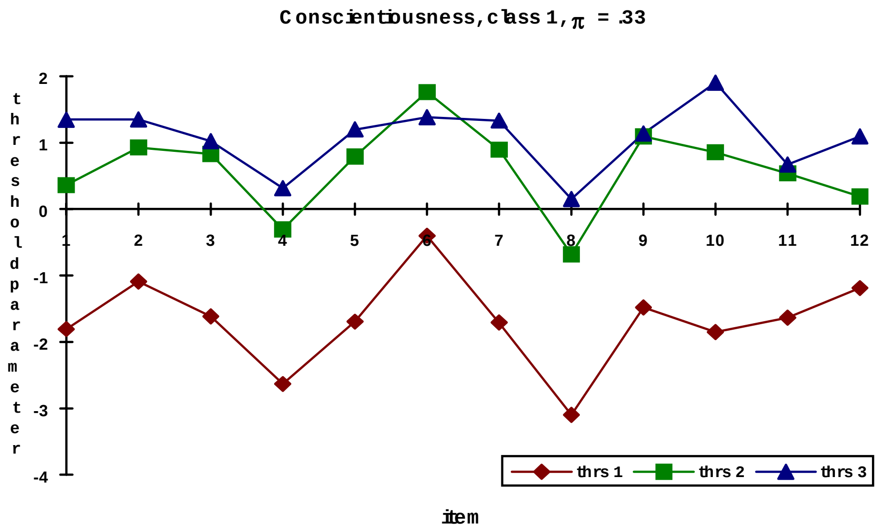

Figure 3 : Threshold parameters for each item in class 1, π_1 = 0.33, of conscientiousness

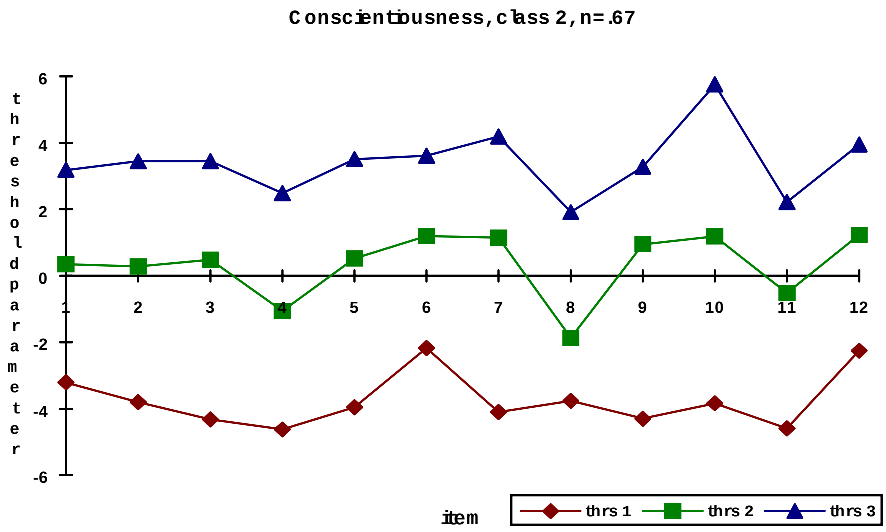

Figure 4: Threshold parameters for each item in class 2, π_2 = 0.67, of conscientiousness

It turns out that all items in both classes have ordered thresholds, except item 6 in the first class. In contrast to the first class, where the first threshold distance is about 2 units and the second only 0.5 units, the threshold distances in the second class are very large. That means, it is very easy to pass the first threshold and very hard to pass the last one. Hence, most responses in this class are in the two middle categories and the extreme categories are avoided.

Therefore, the difference between the two C-classes can be interpreted in terms of different *response sets*, i.e., the smaller class 1 has a tendency towards extreme ratings, the second, bigger class a tendency towards moderate ratings, i.e., towards the mean. In both classes, however, the items have - with few exceptions - the same difficulty parameters and, according to this criterion, measure the same dimension (see also section 4).

This interpretation of the C-classes as classes with different response sets raises the question of whether those response sets show up with the E-scale. In fact, the threshold distances of both E-classes shown in Figure 2 have nearly the same size, so that these classes do not reflect different response sets.

In order to further investigate the E-scale, we split the sample according to the results of the 2-class solution, i.e., each person was assigned to that class to which she or he belongs to with higher probability (cf. section 2.1 of chapter 1). The mean recruitment probabilities of being a member of these two classes given the response pattern is 0.917 for class 1 and 0.909 for class 2, respectively. These two subsamples were analyzed with the mixed Rasch model again, that is, subsequent analyses were carried out for the „impulsivity“ and the „sociability“ class separately.

This step is justified by the fact that the two class solution does not fit the data of the E-scale. According to the CAIC-values shown in table 1, only a 4-class solution would fit. However, instead of a 4-class solution of the mixed Rasch model (that produces nearly identical results), we present here the two-class solutions within both E-classes that already have been interpreted above.

E-scale subsample	*1 class*	*2 classes*	*3 classes*
impulsivity	23912,16	*23730,12*	23884,52
sociability	19765,54	*19631,46*	19758,65

Table 2: The CAIC-values for the mixed Rasch model within two manifest subsamples

Again, two-class solutions are indicated by the CAIC-statistics for both, the impulsivity and the sociability class. Figure 5 shows the item profiles of both classes for the impulsivity subsample, i.e., those people who regard humor and cheerfulness as part of being extraverted.

No major differences show up between these profiles. As in the case of the C-scale, the differences appear in the threshold distances that are smaller in the first class and extremely large in the second class, shown in Figures 6 and 7.

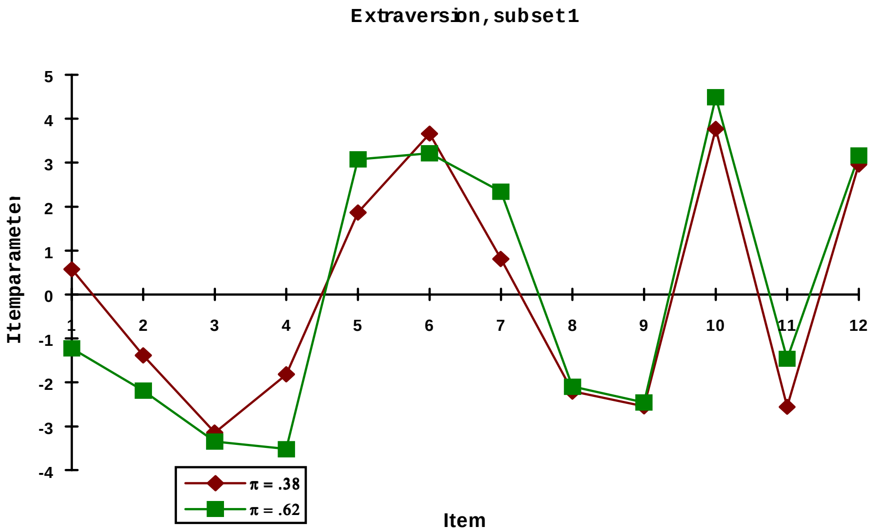

Figure 5: Item parameters for each item in both classes of subsample 1, extraversion, impulsivity

The second class obviously covers the persons with a response set of avoiding extreme ratings, as the threshold of passing to the second category is very easy and the threshold of passing to the last category is very difficult.

The results of the second subsample corresponding to class 2 in Figure 2, i.e., the sociability class, are similar as the item profiles are closer together and the threshold distances vary between the classes.

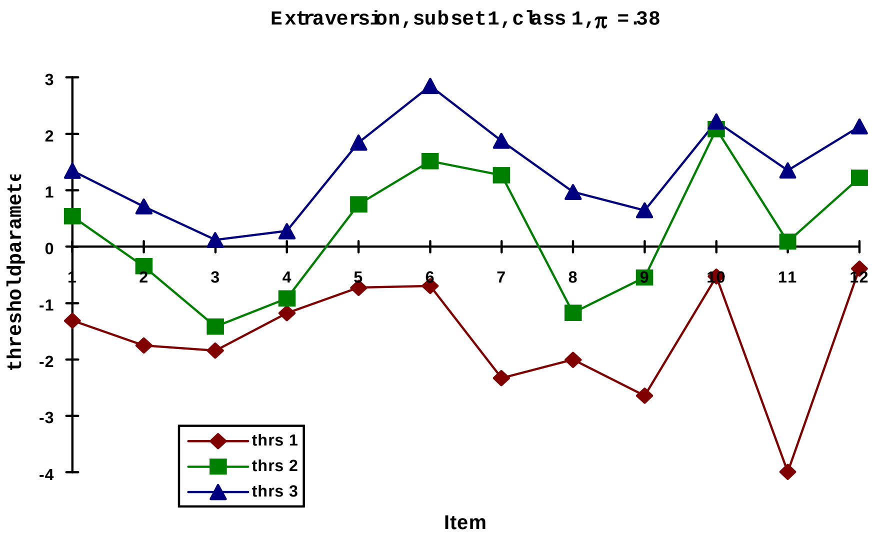

Figure 6: Threshold parameters for each item in class 1, π_1 = 0.38, of subsample 1, extraversion, impulsivity

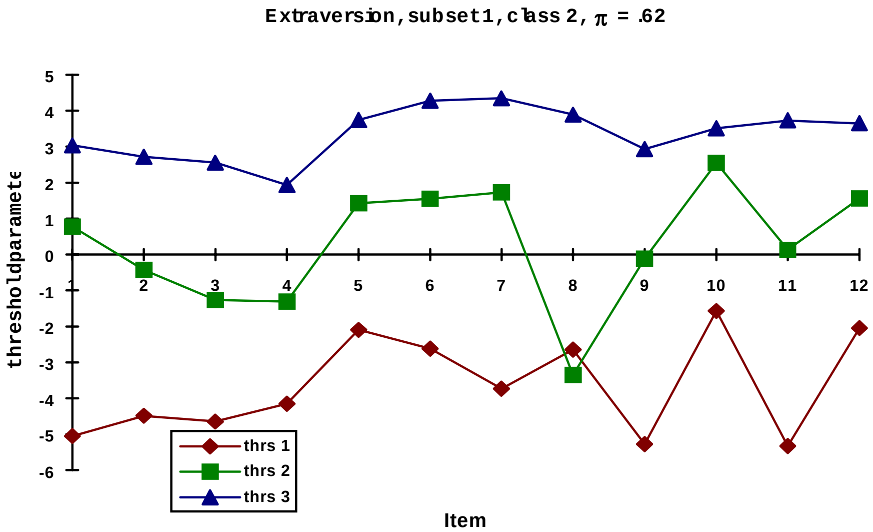

Figure 7: Threshold parameters for each item in class 2, $\pi_2 = 0.62$, of subsample 1, extraversion, impulsivity

In order to investigate whether the same individuals have the same response set for the C- and the E-scale, the cross tables between the class memberships of the C- and the E-scales were calculated.

	Imp 1	Imp 2	
C 1	160	122	282
C 2	211	554	765
	371	676	1047

	Soc 1	Soc 2	
C 1	138	151	289
C 2	96	482	578
	234	633	867

Table 3: Contingency tables for the class memberships of the C-scale and both classes of the E-scales

The tetrachoric correlation coefficients for these tables are 0.45 and 0.54, respectively. There is a strong relation between the response set classes of the C- and E-scales, but obviously this relation is not perfect: employing the response set of avoiding extreme ratings is partly a scale-unspecific phenomenon and partly depends on which items are considered.

4. Implications for estimating the trait parameters

The width of threshold distances has implications on the estimates of the trait parameters. When the threshold distances are large, i.e., the first threshold is easy and the last one is difficult, it is less likely for individuals to obtain a very low or a very high score in that subtest. As a consequence, people with the same sum score in a questionnaire get a different trait estimate depending on the response set they have. For people who *avoid* extreme ratings, a very low (high) score indicates an extremely low (high) trait level and, hence, get a much more extreme parameter estimate than people without such response set. Figure 8 shows this relationship by means of the parameter estimates of the two C-classes.

It can be inferred from this figure that e.g., a person with score 6 gets a trait estimate $\hat{\theta} = -2.0$ if she or he belongs to class 1, but an estimate $\hat{\theta} = -4.0$ if she or he belongs to class 2, i.e., usually avoids extreme ratings. In this sense, the trait estimates are corrected for the response set of the persons.

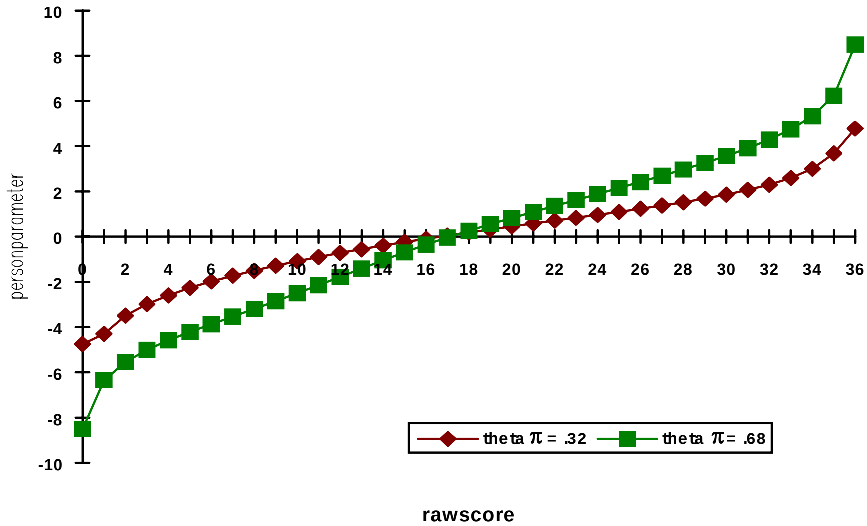

Figure 8: Person parameter θ as a function of the sum score r for both classes of conscientiousness

This is the most practical implication of employing the mixed Rasch model for taking account of different response sets: The trait parameter of the mixed Rasch model is automatically corrected for the effects of a response set on the sum scores. In Rasch models, the sum score is a sufficient statistic of the trait parameter. This is also true for the mixed Rasch model, but the same score represents a different trait level in classes with different response sets.

5. Discussion

This application demonstrates that a failure of fitting the ordinal Rasch model to questionnaire data does not nesessarily mean that the test is multidimensional. If identical item parameter estimates in different groups of persons is taken as a criterion for *unidimensionality*, a test may be unidimensional, but different response sets employed by different people lead the mixed Rasch model to identifying two or more latent classes. As long as there are relatively large groups of persons with the same response set, the mixed Rasch model can handle these individual differences and the trait estimates are corrected for the respective response set.

In general, it may be considered as problematical to compare trait values obtained from different classes. This is certainly true, if the item difficulties are substantially different in two classes indicating that the cognitive structure is different for people from both classes. In that case, it can be argued that the questionnaire measures different traits in different populations and, hence, trait values cannot be compared between the populations. In the present case, however, the item difficulties are very much the same in both classes (see Figures 1 and 5) and , following the same argument, the same trait is measured in both populations. It is only the dispersion of item responses that discriminates between the classes, not the difficulty of an item. In this particular situation, it is justified for us to compare trait values from different classes, as they only correct the effects of the class-specific dispersion of responses.

The results of the E-scale analysis show that substantial individual differences with respect to the measured construct may dominate those moderate response set effects. Here, the two-class solution revealed the substantial differences of sociability and impulsivity, whereas the different response sets only showed up in the second step. On a general level, these examples show not only that personality questionnaires can be analyzed with ordinal Rasch models. By applying the mixed Rasch model it is also possible to take account of different response sets when measuring a latent trait.

This research project was motivated by the question of scalability of the Big Five personality factors. As an answer to this question, we could not find classes of unscalable people. However, people are using the rating scale in different ways, and avoiding the two extreme categories out of four, may indicate that these people are less consistent with respect to the behaviour addressed by the items.

Acknowledgements

We are indebted to Peter Borkenau and Fritz Ostendorf who were so kind as to leave the data to us for a reanalysis. This research was part of the DFG-Project: Development and validation of psychometric mixture distribution models.

References

Bem, D. & Allen, A. (1974). On predicting some of the people some of the time: The search for cross situational consistency in behavior. *Psychological Review, 81*, 506-520.

Borkenau, P. & Ostendorf, F. (1991). Ein Fragebogen zur Erfassung fünf robuster Persönlichkeits-faktoren. *Diagnostica, 37*, 1, 29-41.

Costa, P.T. & McCrae, R.R. (1989). *NEO PI/FFI manual supplement*. Odessa, Florida: Psychological Assessment Resources.

Eysenck, H.J. (1976). *The measurement of personality*. Lancaster: MTP Press.

McCrae, R.R. & Costa, P.T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.

Rost, J. (1991). A logistic mixture distribution model for polytomous item responses. *British Journal of Mathematical and Statistical Psychology, 44*, 75-92.