Psychological Test and Assessment Modeling, Volume 56, 2014 (4), 332-347

The structural validity of the FPI Neuroticism scale revisited in the framework of the generalized linear model

Karl Schweizer[1] *& Siegbert Reiß*[2]

Abstract

The structural validity of the FPI Neuroticism scale that is composed of binary items is investigated by means of confirmatory factor analysis. Because of the binary nature of the items a link function is integrated into the model of measurement that turns it into a generalized linear model, and probability-based covariances serve as input. The structural investigation reveals that the scale shows a substructure that reflects the contents of the items originating from two different domains: the mental and physical domains. The weighted congeneric bifactor model shows that the general factor is the dominating factor besides two less prominent factors referring the mental and physical domains. A sufficient degree of homogeneity is indicated by McDonald's Omega coefficient. The use of factor scores is recommended for the representation of neuroticism.

Keywords: congeneric model, weighted congeneric model, neuroticism, link transformation, probability-based covariances

[1] *Correspondence concerning this article should be addressed to*: Karl Schweizer, PhD, Department of Psychology, Goethe University Frankfurt, Grüneburgplatz 1, 60323 Frankfurt a. M., Germany; email: K.Schweizer@psych.uni-frankfurt.de

[2] Goethe University Frankfurt, Germany

Introduction

Many personality scales constructed in the 50ies, 60ies and 70ies include binary items, as for example the neuroticism scale that is in the focus of this paper. At the time of the construction of these scales the use of factor-analytic methods was undisputed. But nowadays, other than in previous times special emphasis is given to the fit of the data and the model. The same properties are expected to characterize each one of them, as for example the same scale level (Raykov, 2012; Skrondal & Rabe-Hesketh, 2004). As a consequence of the change, structural investigations by means of confirmatory factor analysis must meet specific demands. In this paper the investigation of the structure of a personality scale is described within the framework of the generalized model of measurement with a coefficient of association as input to confirmatory factor analysis, which takes the binary nature of the data into consideration.

The FPI Neuroticism scale (Fahrenberg, Hampel, & Selg, 1984), the origin of which dates back to the 60ies, is one of these scales. The items of this scale are interspersed among the other items of a personality questionnaire so that it is not easy to identify the trait, which they represent in order to avoid acquiescence, reactance or other forms of bias in responding. At the time of the construction of this scale there were several reasons for giving preference to binary items. A major reason was the expectation that this way the respondents could be forced to come to a decision in cases in which they would prefer to avoid it. Another major reason was the expectation that scores computed from binary items would show interval level or would at least come closer to the interval level than scores computed from ordered-categorical items.

Since neuroticism is considered as a broad upper-level trait (Eysenck, 1952; McCrae & John, 1992), a scale for assessing it must meet a special demand: it must represent a number of specific facets simultaneously. Each one of these facets can give rise to an own scale that shows a high degree of homogeneity. However, because of the differences between the various facets the degree of homogeneity of a scale representing such an upper-level trait can be impaired. There is even the danger that subsets of items representing specific facets may show larger correlations among each other than with items representing other subsets. As a consequence, there may be inhomogeneity among the items that can lead to a low degree of homogeneity of the overall scale.

A special characteristic of the FPI Neuroticism Scale is that it is constructed according to Eysenck's PEN model (1952, 1967) that extends psychological concepts to the level of biological phenomena. Accordingly one half of the items of the scale refer to mental events or states, as for example anxiety and depression, and the other half to indicators of corresponding physical processes or states, such as the observation of unrest or nervousness. Therefore this scale is considered as especially useful for research focusing on the relationship between processes referring to the psychological and biological levels. Although this scale is expected to show a one-dimensional structure, the consideration of a subset of items representing physical processes and states besides another subset of items representing mental events and states can mean a deviation from uni-dimensionality. In the construction of the scale according to the guidelines of classical test theory (Novick,

1966) and by means of the factor-analytic methods of the time this potential impairment may not have been obvious.

Modeling binary data

Confirmatory factor analysis has become a preferred tool for the investigation of the structural validity of a scale. It is usually conducted on the basis of the congeneric model of measurement (Jöreskog, 1971) that is a linear model for continuous data. If binary data are to be investigated, data and model do not fit together. The formal representation of the model of measurement is helpful in revealing the discrepancy. This model describes the $n\times1$ vector of observations $\mathbf{y}$, which are also addressed as manifest variables, as the sum of the $n\times1$ vector of intercepts $\boldsymbol{\mu}$, of the product of the $n\times m$ matrix of factor loadings $\boldsymbol{\Lambda}$ and the $m\times1$ vector of latent variables (=latent factors) $\boldsymbol{\eta}$ and of the $n\times1$ vector of error components $\boldsymbol{\varepsilon}$:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} .$$

Since the scale level of the right-hand part of the equation is continuous, the left-hand part also has to be continuous. Binary data do not fit to this model since they show the categorical level. Another discrepancy is regarding the distributions. The binary data follow the binomial distribution, $Y_i \sim Bin(2, p)$ for i=1,…,n, where p denotes the probability of the selected binary event whereas the latent variables are assumed to be normally distributed, $\eta_j \sim N(0, 1)$ for j=1,…, m. Consequently, it is necessary to transform the scale level and to establish a relationship between the distributions before or as part of confirmatory factor analysis.

The switch from the binomial distribution to the normal distribution can be accomplished by means of a link transformation. It means the replacement of the linear model by a generalized linear model (McCullagh & Nelder, 1985; Nelder & Wedderburn, 1972). Such a generalized linear model includes a link function g( ) that related two random variables μ and η that follow different distributions to each other:

$$\eta = \mathrm{g}(\mu) .$$

Most link transformations apply to data characterizing individuals. But there is also the possibility of transforming sample statistics. The method of computing tetrachoric correlation (Pearson, 1900) includes a link transformation that applies to the probabilities of the binary events observed in the sample. The transformation is conducted by means of the normal distribution function. This correlation is suggested for the use as input to confirmatory factor analysis (Muthen, 1984). This correlation between binary variables is assumed to provide an estimate of the relationship between the continuous variables from which the binary ones originate. Because of its special sensitivity to skewness, it is recommended to use tetrachoric correlations as input to confirmatory factor analysis in combination with robust maximum likelihood estimation (Finney & DiStefano, 2013). However, the robust estimation only performs a correction of the fit statistics; it does not change the factor loadings in any way that may be important for the evaluation of the results of confirmatory factor analysis.

Furthermore, there is the possibility to transform variances by a link function (McCullagh & Nelder, 1985, p. 21). This possibility is especially useful for confirmatory factor analysis since confirmatory factor analysis is a method for analysing covariances in the first place (Jöreskog, 1970). Consequently, a link transformation of the variances and covariances can compensate for the difference between the distributions. Different link functions are used in combination with models including free and constrained factor loadings (Schweizer, Ren, & Wang, 2015). These link functions produce weights, which moderate the relationships between items and factors.

The link transformation can be achieved by integrating the $n \times n$ diagonal matrix $\mathbf{W}$ that includes the link function as weights into the congeneric model of measurement:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\left(\mathbf{W}\boldsymbol{\eta}\right) + \boldsymbol{\varepsilon}\ .$$

Since free factor loadings cannot be weighted, $\mathbf{W}$ is merged with $\boldsymbol{\eta}$ (=latent factors) but not with $\boldsymbol{\Lambda}$. As a consequence, some redefinitions are necessary: $\boldsymbol{\Lambda}$ becomes a $n \times n$ diagonal matrix and $\boldsymbol{\eta}$ a $n \times 1$ vector. Furthermore a second-order structure has to be added that serves the constraint of the first-order structure:

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}$$

where $\boldsymbol{\Gamma}$ is the $n \times m$ matrix of second-order factor loadings that are also addressed as gamma coefficients, $\boldsymbol{\xi}$ the $n \times 1$ second-order vector of latent variables and $\boldsymbol{\zeta}$ the $n \times 1$ vector of unique components. The matrix of second-order factor loadings is composed of zeros and ones in order to establish fixed relationships among the first-order factors. This way it is assured that the model is a one-factor model although there are several first-order factors.

Figure 1 provides an illustration of such a weighted congeneric model.

The arrows with dashed shafts represent parameters that are constrained whereas the arrows with solid shafts represent parameters that need to be estimated. Furthermore, there are squared weights. In the step from the model of measurement to the model of the covariance matrix that is finally investigated some parameters and also the weights are squared. These weights contribute as squares to the variances of the first-order latent variables.

The switch from the categorical level to the interval level can be accomplished in different ways. One way is included in the computation of tetrachoric correlations. It requires the computation of thresholds that are continuous and associated with the probabilities of the binary events. The achievement of these thresholds is an intermediary step in the computation of the tetrachoric correlation. Another option is the probability-based covariance. The computation of the probability-based covariance implicitly transforms the data level from binary to continuous since probabilities are computed in the first step and combined to give the probability-based covariance in the second step (Schweizer, 2013; Schweizer & Ren, 2013). This covariance can also be regarded as a pre-stage that is reached in the computation of Phi coefficient (McDonald & Ahlawat, 1974).

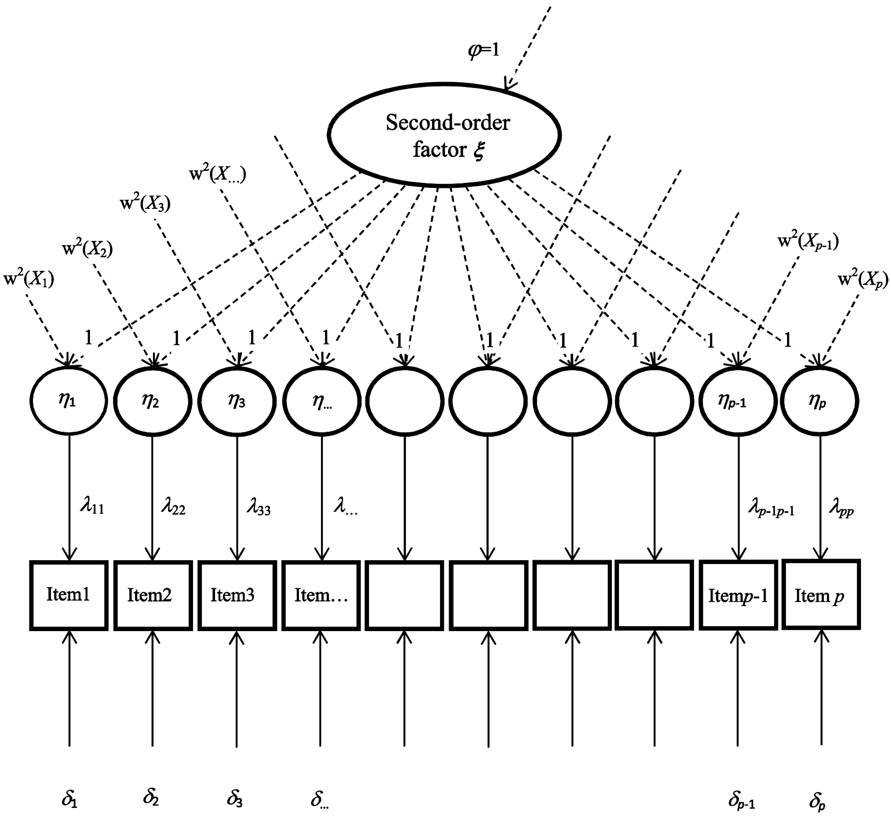

Figure 1:
Illustration of the weighted congeneric model of measurement

Modeling the structure of the FPI Neuroticism scale

The properties of FPI Neuroticism scale described in the previous paragraphs do not give reason to expect one specific structure of the items only. Instead there are several alternative structures that may apply to the data. First there is the unidimensional structure. This one is the structure that represents the authors' intention in constructing the FPI neuroticism scale. Second there is the possibility that the items from the mental and physical domains constitute two homogeneous subsets of items. These subsets may give rise to two unique dimensions in a structural investigation. Because of the common background that is the trait there may be a correlation between these dimensions. Third if there is strong commonality among all items besides the special degrees of homogeneity of the subsets of items, a complex structure including two levels may describe the data well.

The first hypothesis concerning structure can be assumed to be represented well by the congeneric model of measurement (Jöreskog, 1971). The second hypothesis requires the representation by the two-dimensional congeneric model. Two factors characterize this model, and each item loads on one of these factors depending on whether the contents of the item refer to the physical or mental domains. The third alternative can be realized as hierarchical model or bifactor model (Chen, West, & Sousa, 2006). It includes a general factor besides specific factors. Since in an investigation of the structural properties of a scale the focus is on the evaluation of the usefulness as a scale as a whole, the bifactor model is given preference over the hierarchical model (for an overview see Canivez, in press). The bifactor model that originates from work by Holzinger and Swineford (1937) enables the evaluation of neuroticism by the general factor.

The present study

The major objective of the present study is the investigation of the structural validity of the FPI Neuroticism scale by means of confirmatory factor analysis with probability-based covariances as input. The combination of probability-based covariances and the generalized linear model is given preference over confirmatory factor analysis as linear model with tetrachoric correlations as input because the sample size is too small for the computation of tetrachoric correlations but appropriate for probability-based covariances and because the accuracy of factor loadings is important. Whereas the model-data fit can be improved by robust estimation methods (Satorra & Bentler, 1994; Bryant & Satorra, 2012), there is no corresponding improvement of the quality of the factor loadings. The structural investigation has to reveal which one of the models discussed in the previous paragraph provides the best account of the data.

Method

Participants

The sample included 370 participants. It was randomly drawn from a big dataset that was representative of the German population according to gender (47 % males, 53 % females), age, educational level and some other characteristics.

The scale

The scale for the assessment of neuroticism was part of a personality inventory that was developed in several steps and widely used in German-speaking countries (Fahrenberg, Hampel, & Selg, 1984). This scale included 14 binary items. These items were interspersed among the other items of the inventory. Although the manual did not report the subdivision of the neuroticism items referring either to the mental and physical domains, the contents of the items suggested such a subdivision that was consistent with PEN theory. This theory was very influential during the time of the construction of the questionnaire. Therefore, the neuroticism items were rated according to their mental and physical contents in the first step and subdivided accordingly in the second step. The mental items were identified by the numbers 19, 42, 55, 82, 106, 110 and 112 while the numbers of the physical items were 28, 45, 49, 79, 115, 126 and 130.

Characteristics of the models

Several models were considered. The first model was the congeneric model of confirmatory factor analysis (Jöreskog, 1971) including weights based on the link function. This model represented the authors' assumption that the neuroticism scale was a homogeneous scale. It included one factor and the 14 items as manifest variables. There were 14 first-order latent factors and one second-order factor. In order to make sure that it was a one-factor model, all gamma coefficients relating the first-order factors to the second-order factor were set equal to one. Weights for accomplishing the transformation from the binomial distribution to the normal distribution were added to the first-order latent variables. This model of measurement that was a generalized linear model is denoted *weighted congeneric model*. The first part of Figure 2 provides a graphical representation of the core of the weighted congeneric model.

There are circles representing the first-order factors and ellipses the second-order factors. The arrows indicate that each first-order factor is tied to one second-order factor. Furthermore, there was the *original congeneric model*. It was characterized by the combination of one latent variable and 14 manifest variables referring to the 14 items. This model included no special weights. Although this original model was not appropriate for the data since it did not take the binary nature of the data into consideration, it was considered for demonstrating the effect of the link transformation by means of weights. The third model was the *weighted tau model*. This model assumed that the latent source contributed equally to each item. As a consequence, it was expected that the items showed equal factor loadings before the link transformation was conducted. It showed a structure according to the original congeneric model. However, it differed from this model because of the constraint of the factor loadings and the weighting. The fourth model was characterized by two instead of only one second-order factor. The first-order factors referring to items with mental contents loaded on one of the two second-order factors and the first-order factors referring to the other items on the other second-order factor. Another characteristic of this model was that the two second-order factors were allowed to correlate with each other. This model could be perceived as the combination of two weighted congeneric models. It was denoted *weighted congeneric two-factor model* and illustrated in the middle part of Figure 2. Finally, there was the *weighted congeneric bifactor model*. This model included three second-order factors and could be considered as the combination of the weighted congeneric model and the weighted congeneric two-factor model. There were the general second-order factor and the two specific second-order factors. The gamma coefficients relating all first-order factors to the general second-order factor were set equal to one. Furthermore, there were the two specific second-order factors that had factor loadings of the corresponding first-order factors. The three second-order factors were not allowed to correlate with each other in order to have a decomposition of the true variance and to make sure that the general second-order factor accounted for the common variance of the second-order latent variables. The last part of Figure 2 illustrates this model.

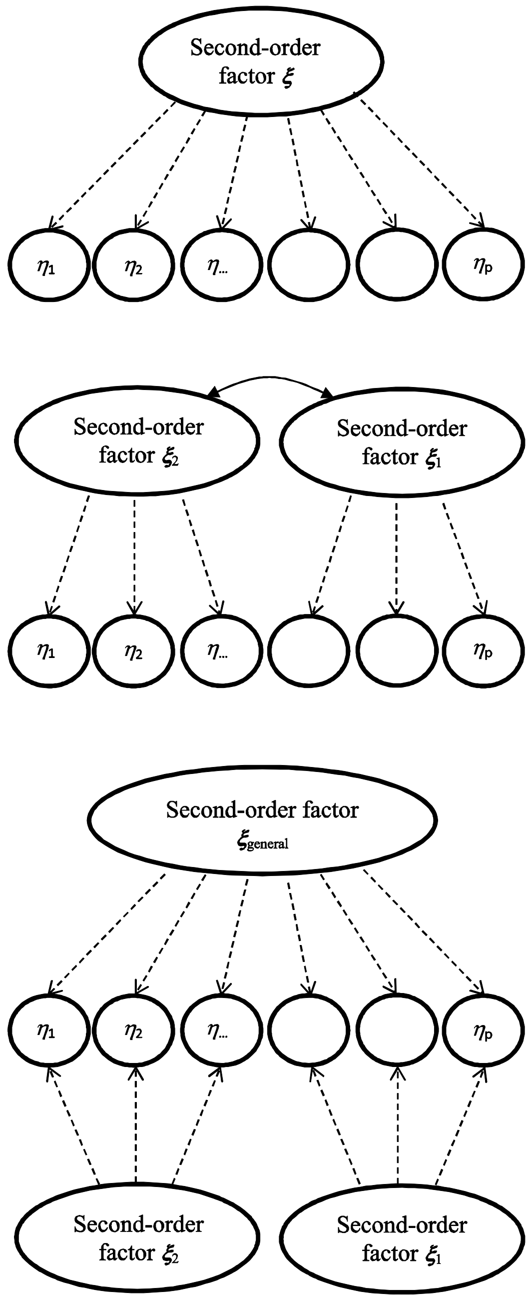

Figure 2:
Illustrations of the latent levels of the weighted congeneric one-factor model, the two-factor model and the bifactor models

The weight transformations were conducted by link functions used in other studies. The weights used in combination with the weighted congeneric model were computed by means of the following function (Schweizer, Ren, & Wang, 2015):

$$\mathrm{w}^2\left(X_i\right)=\left\{\sqrt{\frac{0.25}{\operatorname{var}\left(\Pr_i\right)\left[1-\operatorname{var}\left(\Pr_i\right)\right]}}+\frac{0.25}{\operatorname{var}\left(\Pr_i\right)\left[1-\operatorname{var}\left(\Pr_i\right)\right]}\right\}/2$$

where X_i (i=1, …, p) represented the binary random variable, $\mathrm{w}^2(\ )$ the squared weight for X_i and $\Pr_i$ the probability that X_i is equal to one. Since in this case the weight at Pr=.5 was 1.244, it was necessary to divide the estimates of the factor loadings by this number. It needs be added that in realizing the model in using LISREL the squared weights minus one are to be set equal to the elements of the diagonal of the PSI matrix. The link function of the weighted tau model was

$$\mathrm{w}^2\left(X_i\right)=\frac{\operatorname{var}\left(\Pr_i\right)}{0.25}.$$

Finally, the link transformation had to be completed by multiplying the factor loadings with the square root of two (Schweizer, 2013). This transformation could be expected to lead to factor loadings that corresponded to the factor loadings obtainable by investigating the continuous data from which the binary data originated by means of the confirmatory factor analysis.

Estimation and evaluation

The fit of the five models was investigated by means of LISREL (Jöreskog & Sörbom, 2006). The maximum likelihood method was used for parameter estimation since the link transformation could be expected to compensate for deviations from equal probabilities of the binary events. The evaluation of the outcomes was conducted by means of the following statistics: χ^2, normed χ^2 (=χ^2/df), RMSEA, SRMR, CFI, TLI, GFI. Criteria based on the work by Hu and Bentler (1999) and Kline (2005) were used (normed $\chi^2 \leq 3$, RMSEA≤.06, SRMR≤.08, CFI≥.95, TLI≥.95, GFI≥.90. Furthermore, the AIC that enables the comparison of non-nested models was included. Moreover, Kubinger's F test for the comparison of two non-nested models (Kubinger, Litzenberger, & Mrakotsky, 2006) was considered. The investigation was conducted on the basis of the covariance matrix including probability-based covariances.

Results

The results concerning model fit

The fit results obtained for the five models are provided in Table 1.

Table 1:
Fit Results for the Models Considered in Investigating the Neuroticism Data (N=370)

Model type	χ^2	df	Normed χ^2	RMSEA	SRMR	CFI	TLI	GFI
Model including one latent variable								
Weighted congeneric	271.1	77	3.1	.083	.064	0.91	0.89	0.91
Original congeneric	271.1	77	3.1	.083	.064	0.91	0.89	0.91
Weighted tau	317.9	90	3.5	.083	.091	0.89	0.89	0.89
Models including more than one latent variable								
Weighted congeneric two-factor	166.9	76	2.2	.057	.053	0.95	0.94	0.94
Weighted congeneric bifactor	166.7	77	2.2	.056	.054	0.95	0.95	0.94

The first row gives the results for the weighted congeneric model. According to these results the model-data fit was not generally good. There were only two statistics that were within the range for good results: SRMR and GFI. The second row includes the results for the original congeneric model. These results exactly corresponded to the results of the weighted congeneric model. Different results were observed for the weighted tau model presented in the third row. No one of the fit statistics was within the range for good results.

The second half of Table 1 provides the results for the models including more than one second-order factor. The results for the weighted congeneric two-factor model were generally good. There was only the TLI that was not within the range for good results. Furthermore, a high correlation between the two second-order latent factors was observed: r=.68 (t=14.02, p<.05). A correlation of this size could be considered as the basis for a third-order latent variable and as a justification for the general factor included in the weighted congeneric bifactor model. Virtually the same fit results were observed for the weighted congeneric bifactor model as for the weighted congeneric two-factor model. The only major difference was that in this model even the TLI was within the range of good results. In order to have a comparison between the weighted congeneric bifactor model and the weighted congeneric two-factor model the AICs that enabled the comparison of non-nested models were computed. The AICs were 222.7 and 224.9 for the weighted congeneric bifactor model and the weighted congeneric two-factor model indicating a slightly better model-data fit for the first one of the two models. In contrast, the CFI that was also recommended for such comparisons did not indicate a difference, and the results of Kubinger's *F* test for the comparison of two non-nested models ($F(75,76)=1.01$, ns) also indicated that there was no difference.

The evaluation of the factor loadings

This section reports the results of investigating the factor loadings that provide information on the appropriateness of the items serving as manifest variables of the model. The factor loadings and asymptotic *t* statistics obtained in investigating the one-factor model are listed in Table 2.

The completely standardized factor loadings of the weighted congeneric model are included in the first column of this Table. They varied between 0.37 and 0.75. The second column gives the corresponding asymptotic *t* statistics. They varied between 4.5 and 11.0. All of them indicated significance at the one-percent level. The completely standardized factor loadings of the original congeneric model presented in the third column were considerably lower than the ones presented in the first column. The corresponding asymptotic *t* statistics of the fourth column differed only to a minor degree from asymptotic *t* statistics of the second column. All of them reached the one-percent level of significance. Furthermore, there were the completely standardized factor loadings of the

Table 2:
Completely Standardized Factor Loadings Observed in the One-factor Models (N=370)

	Factor loadings				
Item (Number in questionnaire)	**Weighted cogeneric**	***t***	**Original congeneric**	***t***	**Weighted tau**
1 (19)	0.75	*11.0*	0.58	*11.0*	0.71
2 (42)	0.49	*6.9*	0.39	*7.0*	0.65
3 (55)	0.37	*4.5*	0.25	*4.5*	0.64
4 (82)	0.75	*10.8*	0.57	*10.8*	0.71
5 (106)	0.68	*10.0*	0.53	*10.0*	0.69
6 (110)	0.52	*7.2*	0.40	*7.1*	0.66
7 (112)	0.58	*8.3*	0.45	*8.3*	0.68
8 (28)	0.74	*11.0*	0.58	*11.0*	0.59
9 (45)	0.69	*10.3*	0.55	*10.3*	0.69
10 (49)	0.63	*8.9*	0.48	*9.0*	0.68
11 (79)	0.50	*7.0*	0.39	*7.0*	0.65
12 (115)	0.57	*7.9*	0.43	*7.9*	0.66
13 (126)	0.65	*8.8*	0.48	*8.8*	0.68
14 (130)	0.66	*9.5*	0.51	*9.5*	0.69

weighted tau model presented in the fifth column. These factor loadings varied between 0.59 and 0.71. In this case there were no asymptotic *t* statistics since these factor loadings were not estimated individually but showed variation because of the contribution of error in standardization.

The factor loadings and asymptotic *t* statistics obtained in investigating the weighted congeneric bifactor model are included in Table 3.

The first column of this Table comprises the completely standardized factor loadings that were estimated for the complete model. These factor loadings varied between 0.36 and 0.78 and the corresponding asymptotic *t* statistics given in the second column between 7.4 and 12.6. All asymptotic *t* statistics reached the one-percent level of significance.

Since the confirmatory factor models of the study did not provide factor loadings for the second-order factors directly, the relationships between the items and the three second-order factors were estimated from the standardized gamma coefficients that were made available by LISREL. The results of estimating the factor loadings as relationships between items and second-order factors this way are presented in the third to fifth columns.

Table 3:
Completely Standardized Factor Loadings Observed in the Weighted Congeneric Bifactor Model (N=370)

	Factor loadings				
Item (Number in questionnaire)	**Overall factor**	***t***	**Specific factor 1**[1]	**Specific factor2**[1]	**General factor**
1 (19)	0.74	*11.4*	0.43	0	0.60
2 (42)	0.50	*7.4*	0.29	0	0.40
3 (55)	0.36	*4.9*	0.21	0	0.29
4 (82)	0.73	*11.2*	0.43	0	0.59
5 (106)	0.71	*11.9*	0.41	0	0.57
6 (110)	0.56	*8.2*	0.33	0	0.50
7 (112)	0.63	*9.5*	0.37	0	0.51
8 (28)	0.78	*12.6*	0	0.46	0.63
9 (45)	0.63	*9.7*	0	0.37	0.51
10 (49)	0.69	*10.7*	0	0.40	0.56
11 (79)	0.49	*7.3*	0	0.29	0.40
12 (115)	0.52	*7.7*	0	0.30	0.42
13 (126)	0.59	*8.7*	0	0.34	0.47
14 (130)	0.73	*11.5*	0	0.43	0.59

[1] These factor loadings were obtained by weighting the factor loadings on the overall factor according to the second-order factor loadings.

In the third column these estimates of factor loadings are presented for the items with contents referring to the mental domain. They varied between 0.21 and 0.43. The results for the items with contents referring to the physical domain of the fourth column were between 0.29 and 0.46. The estimates of the factor loadings regarding the general second-order factor of the last column varied between 0.29 and 0.63. Out of the 14 coefficients there were 13 ones that were equal or larger than 0.40. In the specific second-order factors only three out of six coefficients were larger than 0.40.

The evaluation of the structural balance

Since the investigation of the structure revealed that FPI Neuroticism scale included a substructure composed of two major units, it was necessary to investigate whether the overall scale was biased in the direction of one of these units so that it was mainly representing one of these substructures. This additional investigation was conducted by comparing the variances of the second-order factors of the weighted congeneric two-factor model. The variances were 6.97 and 3.27 for the factors associated with the mental and physical contents in corresponding order. The comparison was conducted by means of Hartley's F_{max} test. This test indicated that there was no difference between the amounts of variance for which the two factors accounted ($F_{max}(2,6)=2.13$, n.s.). Consequently, the possibility of a structural imbalance could be excluded.

The evaluation of the homogeneity

The homogeneity of the FPI Neuroticism scale was investigated by means of McDonald's (1999) Omega coefficient. Since there was only one factor loading for each first-order factor and the second-order factors established fixed relationships between the first-order factors, the Omega was based on the factor loadings on the first-order factors that were expected to reproduce the covariance pattern. The Omega associated with weighted congeneric model was 0.81. The computation based on the results for the weighted congeneric bifactor model revealed an Omega coefficient of 0.86. Furthermore, the contributions of the specific second-order factors were removed from the factor loadings in order to obtain an estimate of the homogeneity of the general factor. This Omega coefficient was 0.80.

Discussion

This study that makes use of the generalized linear model and a special coefficient of association in order to comply with the binary nature of the items reveals that the FPI Neuroticism scale shows a substructure. This substructure reflects the theoretical background influencing the selection of item contents for representing neuroticism, which is provided by Eysenck's (1952) PEN theory. This substructure includes two major facets characterized by items from the mental and physical domains. It is this substructure that makes the scale unique since the more recent developments in the framework of the BIG FIVE theory give preference to the psychological level (McCrae & John, 1992).

Assuming that physical and mental phenomena are related to each other there is reason for including contents from both domains into one scale although there is the danger that specificities of the two domains may impair the homogeneity of the scale. Completely excluding such impairment complicates the construction of scales for the assessment of upper-level constructs. Only in the case that each facet is represented by one item solely an impairment of homogeneity is unlikely to occur. Another option is the possibility of computing an item parcel for each facet (Bandalos, 2008; Little, Cunningham, Shahar, & Widaman, 2002). However, in the case of FPI Neuroticism scale, which is investigated as an example, this provision does not work since there are only two major facets.

Consequently, it is necessary to investigate whether the degree of impairment due to different facets is tolerable. There are two results that suggest that this is the case. First the general factor of the weighted congeneric bifactor model turns out to be the dominating factor. After disentangling the effects of the second-order factors all factor loadings on the general factor surmount the factor loadings on the specific factors. Second there is a balance between the items reflecting mental contents and the items reflecting physical contents. Furthermore, there is the possibility to estimate participants' factor scores regarding the general factor. These factor scores could be expected to represent neuroticism as trait anchored in two different areas quite well.

The aim of the present study was not only to provide information on the Scale but also about the method used for investigating its quality. There are some observations that are in need of further interpretation. First of all, there is the observation that the change from the linear model to the generalized linear model does not have any effect on the model-data fit, which may appear to be strange. The explanation is that the two models only differ according to the weights so that the difference is compensated in the estimation of the factor loadings. In the linear model all weights are equal to one and in the other model the weights reflect the deviations of the distributions of the individual item from the expected distribution. In models with free parameter estimation the factor loadings simply compensate for the difference between the weights so that the model-data fit is simply retained.

Second the factor loadings of the linear and generalized linear models show a difference in the average level. This difference is because the factor loadings of the linear model reproduce the covariance matrix at hand whereas the factor loadings of the generalized linear model reproduce the covariance matrix of the continuous data from which the binary data originate. This is a characteristic that also characterizes confirmatory factor analysis with tetrachoric correlations as input.

Third ignoring the difference in the average level of the factor loadings the differences between the profiles of the two types of factor loadings are only minor. This similarity is a surprising observation since differences are expected. The reason is that the scale does not exhibit a broad range of item difficulties or include a few items showing extreme degrees of difficulty among a majority of item with medium degrees of item difficulty.

There is only one other method that enables the investigation of the structural validity of a scale composed of dichotomous items in the framework of confirmatory factor analysis: the other method is confirmatory factor analysis with tetrachoric correlations as input (Muthen, 1984). In this case the link transformation is part of the computation of the tetrachoric correlations. However, this method has occasionally been found to lead to problems regarding the model-data fit. Therefore, it is recommended to be applied in combination with robust estimation. Robust estimation improves the model-data fit but not the estimates of the factor loadings so that there is still a difference regarding the factor loadings. Furthermore, it needs to be mentioned that the method including the weighted congeneric model can be applied to covariances whereas confirmatory factor analysis with tetrachoric correlations as input to correlations only. In the first case the variances of the items contribute to the solution whereas in the second case they do not contribute because of the standardization.

Author note

The authors are very grateful to Jochen Fahrenberg (Freiburg University) for providing the data for the investigation.

References

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling, 19*, 373-398.

Canivez, G. L. (in press). Bifactor Modeling in Construct Validation of Multifactored Tests: Implications for Understanding Multidimensional Constructs and Test Interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: standards and recent advancements*. Göttingen: Hogrefe.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189-225.

Eysenck, H. J. (1952). *The Structure of Human Personality*. London.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, Ill.: Charles C. Thomas.

Fahrenberg, J., Hampel, R., & Selg, H. (1984). *Das Freiburger Persönlichkeitsinventar FPI* [Freiburg Personality Inventory]. 4th ed. Göttingen: Hogrefe.

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 439–492). Charlotte, NC: Information Age Publishing, Inc.

Jöreskog, K. G. (1970). A general method for analysis of covariance structure. *Biometrika, 57*, 239-257.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Jöreskog, K.G., & Sörbom, D. (2006). *LISREL 8.80*. Lincolnwood, IL: Scientific Software International Inc.

Kubinger, K. D., Litzenberger, M., & Mrakotsky, C. (2006). Practiced intelligence testing based on a modern test conceptualization and its reference to the common intelligence theories. *Learning and Individual Differences*, *16*, 175-193.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Mahwah,NJ: Erlbaum.

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology, 27,* 82–99.

McCrae, R. R. & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 175-215.

McCullagh, P., & Nelder, J. A. (1985). *Generalized linear models*. London: Chapman and Hall.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered, categorical, and continuous latent variables indicators. *Psychometrika, 32*, 1-13.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A, 135*, 370-384.

Novick, M.R. (1966) The axioms and principal results of classical test theory. Journal of Mathematical Psychology, 3, 1-18.

Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London Series A, 195*, 1-47.

Raykov, T. (2012). Scale construction and development using structural equation modelling. R. H. Hoyle (Ed.), *Handbook of structural equation modelling* (pp. 472-492). New York: Guilford Press.

Satorra, A., & Bentler, P. M. (1994). Corrections to the test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Glogg (Eds.), *Latent variable analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

Schweizer, K. (2013). A threshold-free approach to the study of the structure of binary data. *International Journal of Statistics and Probability, 2*, 67-75.

Schweizer, K. & Ren, X. (2013). The position effect in tests with a time limit: the consideration of interruption and working speed. *Psychological Test and Assessment Modeling, 55*, 62-78.

Schweizer, K., Ren, X., & Wang, T. (2015). A comparison of confirmatory factor analysis of binary data on the basis of tetrachoric correlations and of probability-based covariances: a simulation study. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W.-C. Wang (Eds.), *Quantitative Psychology Research* (pp. 273-292). Heidelberg: Springer.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modelling: multilevel, longitudinal and structural equation models*. Boca Raton, Fl: Chapman & Hall/CRC.