Investigating phenotypic heterogeneity of Trait Anger

Amsterdam, 2014
KLAASJAN G. OUWENS

Supervisors

dr. Marleen H.M. de Moor
Biological Psychology
VU University Amsterdam

dr. Gitta H. Lubke
Quantitative Psychology
University of Notre Dame

Abstract

The problem of *missing heritability* is mostly tackled from a genetic perspective, however, in complex traits, phenotypic heterogeneity also plays a role. This study examined phenotypic heterogeneity of Trait Anger, to reveal whether dispositions among the Dutch population reflect dinstinct subtypes, variants along a continuum of severity or severity differences within subtypes. Exploratory factor models, latent class models and factor mixture models were applied to questionnaire data collected by the Netherlands Twin Registry (NTR), using the Dutch translation of State-Trait Anger Scale (STAS). A comparison of the different models shows that a model that allows for multiple subtypes, each with their own three factors with severity differences provides the best fit.

Contents

Introduction

Anger has not been studied as intensively as other behavioral phenotypes such as anxiety and depression. It is however clear that Anger is related to health problems such as cardiovascular disease (e.g., Suls & Bunde, 2005), domestic violence (e.g., Leonard & Blane, 1992; Maiuro, Cahn, Vitaliano, Wagner, & Zegree, 1988; Pan, Neidig, & O'Leary, 1994), and substance abuse (e.g., Leibsohn, Oetting and Deffenbacher, 1994).

Just as for any other phenotype, it is critical to have an explicit, precise definition. However, the phenotype Anger has its complications. Anger has been used interchangeably with hostility or aggression, even though correlations between scales measuring anger and most hostility scales are low to moderate (Spielberger et al., 1983). While hostility can be described as a *cognitive* tendency to devaluate other people's worth and motives, and aggression as a variety of verbal and *psysical behavior*, Anger is more an unpleasant *emotion* ranging from annoyance to fury (Smith, 1994). Anger has often been assessed using ad hoc developed scales without much formal evidence of construct validity (e.g., Chang, Ford, Meoni, Wang, & Klag, 2002; Gallacher, Tarnell, Sweetnam, Elwood, & Stanfield, 1999), with insufficient differentiation between *expressing* Anger and the *emotional experience* of Anger (Deffenbacher et al., 1996) .

Spielberger et al. have made an effort to clarify the understanding of Anger by adapting state-trait personality theory to Anger (e.g., Spielberger,1988; Spielberger et al., 1983; Spielberger, Krasner, & Solomon, 1988; Spielberger, Reheiser, & Sydeman, 1995). In state-trait personality theory, the *State* refers to "*a transitory emotional-physiological condition consisting of subjective feelings and physiological activation*". The *Trait* refers to proneness or the tendency to experience State Anger. People scoring high on Trait Anger are presumed to experience more frequent and more intense State Anger (Spielberger, 1988; Spielberger et al., 1983, 1988). Both State and Trait Anger are experienced along a continuum from little or no Anger to highly emotional charged states such as fury and rage.

The psychometric instrument commonly used for measuring Trait Anger is part of the Spielberger State-Trait Anger Scale (STAS) (Spielberger et al., 1983) (table 3)(Spielberger et al., 1983). The ten items measuring Trait Anger are designed to assess the frequency and individual experiences state Anger over time and in response to a variety of situations.

There are four response options, ranging from *almost never* to *almost always*. The Trait Anger Scale (TAS) was developed to consist of two subscales, Angry Temperament (AT) and Angry Reaction (AR). The STAS is considered *"a strong measure of Anger, based on a solid theoretical model, with excellent psychometric properties across several normative groups. It has shown good discriminant and convergent validity, as well as clinical utility, and it has been administered across a wide range of subject populations and psychological domains"* (Eckhardt, Norlander and Deffenbacher, 2004). In the current study, this questionnaire was translated to Dutch by Van der Ploeg et al.(1982).

While Spielberger's work meant a great advance in Anger research, unclarity remains about what concept is actually measured with the TAS. In order to investigate whether the same construct is measured across different groups, it may be tested whether measurement invariance (MI) holds for the trait. If MI holds, the same construct is measured for every group. This means that no item will have a *bias* in a particular group of participants. Otherwise, people with the same latent score would score very differently on particular items. For example, an item with a gender bias would unjustly indicate a sex difference in Anger.

There are four different MI models (see table 1). In the context of factor analysis, intercepts, loadings, and residual variances, shouldn't differ in groups in order to reliably compare factor means and (co)variance differences. Absence of measurement invariance can be an indicator of population heterogeneity; when the same item has a different meaning for different populations, that item can not be used to distinguish populations with regard to the construct. In factor analysis terms, this means that the distribution of the observed variable conditional on the latent variables does not differ across groups.

MI model	intercepts	loadings	residual variances	factor means
M1	G	G	G	Z
M2	G	E	G	Z
M3	E	E	G	Z'
M4	E	E	Z	Z'

Table 1: Different Measurement Invariance Models. G = group specific, Z = fixed to zero, Z' = fixed to zero in one group, estimated in all other groups, E = fixed to be equal across groups

Recently, MI and age-related differences of Trait Anger have been investigated. Zimprich et al. (2012) found that strong measurement invariance for the Trait Anger scale did hold across all age groups, indicating that there are no substantial biases with regard to age. This means that the distribution of Trait Anger conditional on latent variables does not differ between age groups. In addition, it was found that the variance of Anger is lower at middle age as compared to young and old age. At old age, the covariance between factors AT and AR seemed to decrease, along with the mean level. This would mean that Trait Anger decreases with age, but that participants mainly score lower on items belonging to one factor. In their study, constraining the factor variances to be equal resulted in a significantly worse fit. They concluded that measurement of Trait Anger was unbiased across age groups. Strict MI, where residual variances are constrained to be equal across age groups, did not hold. Zimprich et al. state that this is not problematic with respect to comparing factor parameters.

Adding to the complexity of a trait is the possibility that relevant variables are omitted, causing a bias in regression coefficients. If a particular moderating variable only has an effect in one group, regression coefficients will reflect that. Including this variable as a covariate is therefore crucial to get reliable estimates of group differences.

Revealing the underlying complexity of a phenotype (e.g. Trait Anger) has a relevance for genetic analyses. Genome-wide analysis studies (GWAS) have located genetic variants associated with variation in psychiatric disorders. However, the estimated heritability obtained in family studies is much larger than the collective phenotypic variance explained by these genetic variants. While this problem of *missing heritability* is mostly tackled from a genetic perspective (for instance, investigating epistatis, epigenetics, and rare variants), methods of measuring complex traits are now also taken into consideration. Summing up scores on all items in a questionnaire ("sum-score") and using this as a phenotype for a GWAS is problematic given that such a sum-score might yield a conflated view of the phenotype. Neglecting phenotypic complexity, measurement bias, and phenotypic resolution in simulation studies have been found to contribute to missing heritability (Van der Sluis et al., 2010).

Van der Sluis et al. (2010) identified three conditions that should hold to declare sum-scores to be 'sufficient statistics' for a GWAS. These are 1) unidimensionality of the trait, with all items being statistically independent conditional on this latent trait; 2) having identical functional relations of item responses to the latent trait. In factor models, this holds if there are equal factor loadings for all items; and 3) equal residual variances for all items. This third condition does not hold for the effect of age on Trait Anger according to Zimprich, which might not be problematic for comparing factor parameters, but it does mean that carrying out a GWAS based on sum-scores would not be appropriate.

Comparing results from factor models, latent class models and factor mixture models can be used to further investigate heterogeneity of the trait in the population. While most of the time sum-scores are used for genetic association studies, under the assumption of homogeneity, population heterogeneity needs to be considered, and might increase the efficacy of genetic association studies.

More specifically, fitting different statistical models to the data might result in a more informative picture of the trait. These statistical models include factor mixture models, which combine factor analysis and latent class analysis (Arminger et al., 1999; Dolan and van der Maas, 1996; Heinen, 1996; Jedidi et al., 1997; Muthen and Shedden, 1999; Vermunt and Magidson, 2003; Yung, 1997). In this model type, individuals are divided into subtypes, while individuals within a subtype can still differ in severity or symptom patterns.

The purpose of this research is to to assess if differences in Trait Anger levels are attributable to severity differences, class differences, or both. In addition, the role of sex and age in the heterogeneity of Trait Anger will be investigated. Finally, it is discussed whether it would be appropriate to use a sum-score when analyzing Trait Anger, e.g. in genetic analyses.

Methods

Scale

The psychometric instrument that is widely used for measuring Trait Anger nowadays is the Trait Anger Scale (TAS) (Spielberger et al., 1983). There are four response options, ranging from *almost never* to *almost always*. Where applicable in this thesis, these response categories are consistently numbered 1 to 4, resulting in a sum-score ranging from 10 to 40. The Trait Anger Scale (TAS) has been developed to consist of two subscales, Angry Temperament and Angry Reaction. The manual for the STAXI, of which the TAS is part of, (Spielberger, 1988) reports alpha coefficients ranging from .84 to .93. Cronbach's Alpha was 0.865 in the dataset used in this study.

Both Spielberger and Van der Ploeg (who translated the STAS to Dutch) report a two-factor structure, with a distinction between Anger Temperament and Anger Reaction, see table 2. In the Dutch translation, items 1, 2, 3, 5 and 6 belong to Anger-Temperament, and items 7, 8 and 10 to Angry Reaction. Items 4 and 9 are reported to be cross-loading on both factors.

Scale	Characteristics of persons with high scores
Trait Anger	High Trait Anger individuals frequently experience angry feelings, especially when they feel they are treated unfairly by others.
Trait Anger *factor Angry Temperament*	Persons with high AT scores are quick tempered and readily express their Anger with little provocation. Such individuals are often impulsive and lacking in Anger control.
Trait Anger *factor Angry Reaction*	Persons with high AR scores are highly sensitive to criticism, perceived affronts, and negative evaluation by others. They frequently experience intense feelings of Anger under such circumstances.

Table 2: Guidelines for interpreting high Trait Anger scores. Adapted from Spielberger, C. D., & Sydeman, S. J. (1994). State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory.

While translating the STAS questionnaire to Dutch, Van der Ploeg already found that some items were difficult to translate, and therefore had to be replaced by new items. This affected the number of items per factor. One item that was deemed not translatable from Anger Reaction was replaced by an item that loads on factor Anger Temperament (item 5). Van der Ploeg suggests that for experimental versions of the questionnaire focusing on AR, two extra items could be included.

In the case of a general population of a city (Leiden, The Netherlands), these two items were loading mostly on AT. In a student population, these were loading mostly on AR. It was suggested that these items are differently interpreted by people depending on their Anger score. People with a high disposition for Trait Anger have these items loading on AT, while people with a low disposition for Trait Anger have these items loading on AR. It was hypothesized with these items address some hidden feeling of Anger. If people have a low Anger disposition, this frustration can be externalised faster and be counted as Anger Reaction. People with high Anger disposition are thought to hold back on their Anger externalizing, which is why these items are beloning to AT for these people. It is concluded that it is not possible to adequately classify these items.

Item	Content in Dutch *English equivalent*
1	Ik ben vlug driftig *I am quick-tempered*
2	Ik ben opvliegend van aard *I have a fiery temper*
3	Ik ben een driftkop *I am a hotheaded person*
4∗	Ik voel me gauw geergerd
5∗	Ik ben heetgebakerd
6	Ik verlies gauw mijn zelfbeheersing *I fly off the handle*
7	Als ik echt kwaad word, zeg ik hatelijke dingen *When I get mad, I say nasty things*
8	Ik word kwaad als ik in het bijzijn van anderen bekritiseerd word *It makes me furious when I am criticized in front of others*
9∗	Ik ben snel geirriteerd
10	Ik voel me woedend worden wanneer ik iets goeds doe en dit wordt negatief gewaardeerd *I feel infuriated when I do a good job and get a poor evaluation*
∗∗	*When I get frustrated, I feel like hitting someone*
∗∗	*I feel annoyed when I am not given recognition for doing good work*
∗∗	*I get angry when I'm slowed down by others' mistakes*

Table 3: The ten items constituting the translated Trait Anger Scale. ∗New items in Dutch translation, ∗∗not included in Dutch translation

Items 4 and 9, concerning the immediacy of an Anger response, are suggested to be part of either Anger Temperament for people with a high scores for Trait Anger, and part of Anger Reaction for people with low scores for Trait Anger (Van der Ploeg et al., 1982). Because not all Trait Anger items could be translated literally (partly due to cross-cultural differences), some items in the Dutch translation might be considered new items (see table 3).

Sample

Participants were registered with the Netherlands Twin Registry (NTR), kept by the Department of Biological Psychology at the Vrije Universiteit in Amsterdam. They are part of the adolescent and adult cohort that was recruited through city councils in 1990–1991, 1992–1993 and 2007-2008. They participate in longitudinal survey studies roughly every 2 years. (Boomsma et al., 2002). See table 4, and table 11 in the Appendix). Participants (N = 9213) from NTR were included in the analysis, by selecting the most recent questionnaire from any subject who participated at least once. The data contained familially related people. From all MZ twins, one was removed at random. To correct for other types of relatedness, the standard error was adjusted for clustering based on family number. For more summary statistics, see the Appendix.

Cohort	N	mean age (sd)	mean sum-score (sd)	%male
total	9213	39.03 (15.49)	15.92 (4.53)	46.12

Table 4: Summary statistics of the sample used in the study.

Analysis

All of the analyses were carried out using structural equation modeling software Mplus, version 6.12 (Muthen and Muthen, 1998-2006). 11 models were fitted to the data as shown in table 5.

The first models fitted were latent class models. In latent class analysis (LCA), subtypes of cases are identified by detecting unobserved *latent classes* from categorical data. With LCA, each case can be classified to their most likely latent class. Within each latent class, each variable is statistically independent of every other variable. The probability of obtaining a specific response pattern is a weighted average of the probabilities that are specific for a latent class (Vermunt and Magidson, 2003).

Next, exploratory factor analyses (EFA) were performed. Factor analysis is similar to latent class analysis, addressing patterns of association among observations. However, factor analysis identifies unobserved *continuous factors* among observed correlated variables. In a way, it summarizes the data in as few variables as possible. Each item has a *loading* on each factor, which expresses the amount of variance that is explained by that particular factor.

Finally, factor mixture models (FMM) were fitted to the data. These models are a combination of LCA and FA, combining the ability to identify subgroups (LCA) with the ability to investigate the common content of questionnaire items (FA). FMM is therefore especially suited to model heterogeneity. In contrast to LCA, variables are not assumed to be independent from each other, but variables are allowed to covary. Instead, this is modeled using underlying continuous factors (Lubke and Muthen, 2005).

Models 1f to 3f are exploratory factor analysis models with one, two, or three factors, models 1c to 6c are LCA models with one to six classes, and 3f,2c and 3f,3c are FMMs with three factors and two or three latent classes, respectively. All factor loadings were rotated using Yates's oblique Geomin method.

In the factor mixture models, factor loadings were constrained to be equal across classes, but factor variances and item thresholds were estimated for each class separately. All models were fitted to data of males and females of all ages combined. Finally, the covariate effects of sex and age on classes is tested using Vermunt's three-step approach. With this method, the model is first fitted without covariates, after which it is tested whether covariates influence class membership. With "naive" post-hoc covariate testing, parameter estimates can be biased since uncertainty in estimation of the mixture model and the class probabilities are ignored. With Vermunt's adjusted approach, this bias is corrected for by first extimating the mixture model, then obtaining class probabilities for the subjects, and then fitting a mixture model with these class probabilities as class indicator, and finally regressing the class variable on the covariate.

Multiple goodness-of-fit measures will be used to illustrate the results of fitting the different models. The Bayesian information criterion (BIC) and the root mean squared error of approximation (RMSEA) are used to evaluate the fit. The RMSEA is a measure of closeness of fit, and provides a measure of discrepancy per degree of freedom. A value of 0.05 or lower indicates a close fit, and values of 0.05 up to 0.08 represent reasonable errors of approximation in the population (Browne and Cudeck, 1993). Lower values of BIC indicate a better fit.

Results

Goodness-of-fit measures for LCA, FA and FMM models are shown in table 10.

Latent Class Models

When restricting the comparison to the six fitted latent class models models with a higher number of classes have higher log likelihood values. This is expected because models with more classes have more parameters and can therefore provide a better fit to the data. Among the LCA models, both log likelihood and BIC favor the the six-class model. See figure 1 how group proportions change when increasing the number of classes. When inspecting age and sex for different classes in this 6-class LCA solution, one class stands out with a low mean age and a high proportion of males (see figure 16 in the Appendix). Members of this small class (N = 125) often score in the highest response category and have a mean sum-score of 22.22. In contrast, most participants fall in a class with a mean sum-score below the mean of all classes together (15.92); either in the big low-scoring class (C2, N = 3569, mean sum-score

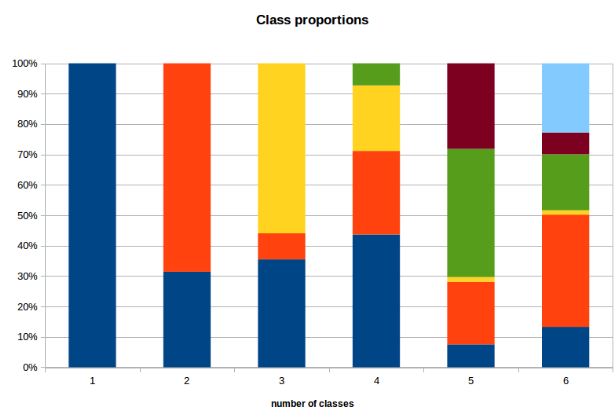


Figure 1: Relative sizes of classes. Colors do not indicate class transitions.

= 14.29), or in the extremely-low scoring class (C6, N = 2074, mean sum-score = 11.43). Possibly worth noticing is that of these participants, while answering *almost never* on nearly every item, items 10 and to a lesser extent items 7 and 8 are still answered with *sometimes*, while these items are not the highest-scoring items in the high-scoring class 3. Not all items are quantitatively ordered in the same way as mean score of the classes (see table 5). Class 1, a class of moderate size scoring slightly above average (N = 1036, mean sum-score = 18.15), deviates from this ordering for items 4, 7, 8, 9, and 10. This is visible in figure 2, where these items are scoring proportionally higher in class 1 compared to class 4, which has a higher mean sum-score.

class	sum-score	item 4	item 7	item 8	item 9	item 10
C3	32.22	3.44	3.23	3	3.20	3.30
C5	25.23	2.71	2.61	2.39	2.62	2.76
C4	18.95	2.00	1.94	1.82	1.92	2.14
C1	18.15	**2.13**	**2.29**	**2.33**	**2.08**	**2.86**
C2	14.29	1.73	1.67	1.58	1.56	1.86
C6	11.43	1.04	1.40	1.31	1.02	1.59

Table 5: Mean scores for different items 4, 7, 8, 9 and 10, for different classes. Classes are ordered from high to low sum-score. Deviations from quantitative ordering are in bold.

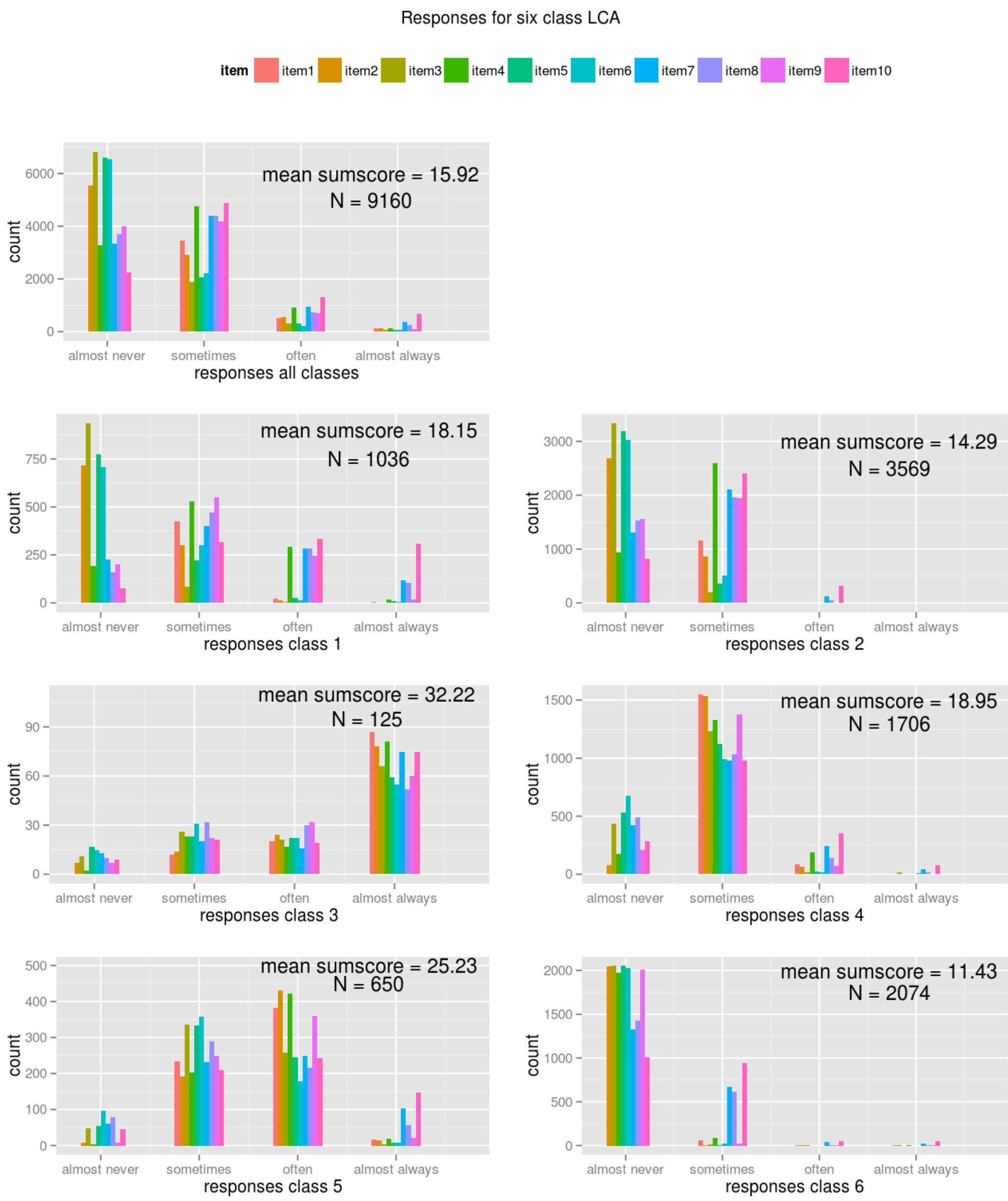
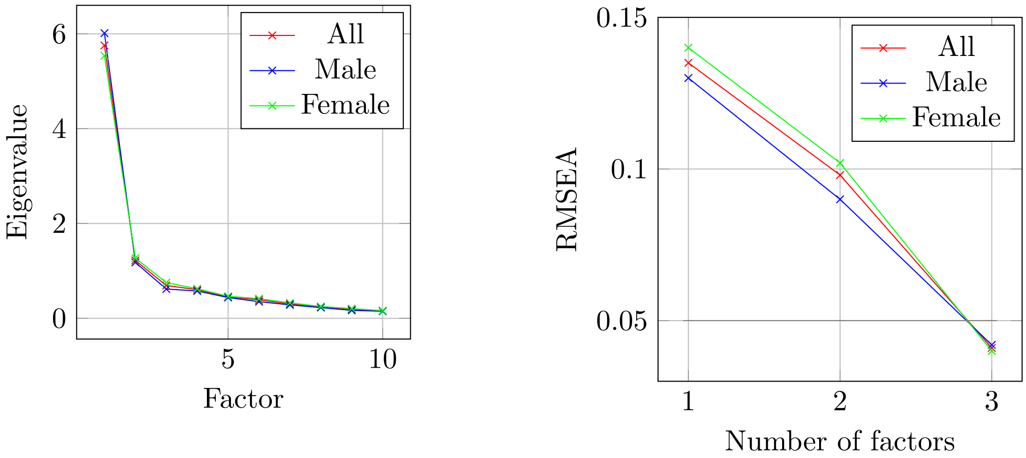


Figure 2: Response counts for all participants together (top left), and for all six latent classes separately. Each color represents an item. For example, class 2 mostly answers *almost never* on items 1, 2, 4 and 5, and mostly answers *sometimes* on the remaining items.

Exploratory Factor Analysis

Considering the FA models separately, the eigenvalues point to either a two-factor solution (based on Kaiser criterion) or one factor (based on the scree plot 'elbow'). See also figure 3a). However, both RMSEA and BIC are lowest for the three-factor model (see figure 3b). Moreover, the factor loading pattern points to a three-factor solution as well. While the two-factor solution has multiple crossloadings, the three-factor solution separates the items over factors a lot better (see table 7), consistent with the design of the STAS instrument. Items loading on the third factor, *Angry Reaction*, show high residual variances. Combined with a relatively low Cronbach's Alpha for these three items (even lower than the Cronbach's Alpha of the two item factor *Angry Other*), this factor is showing to be distinct from the other two factors, but not a very strong consistent measure in itself.



(a) Eigenvalues for the different numbers of factors

(b) RMSEA for the different numbers of factors

Figure 3: Criteria for number of factors

To further analyze the differences of the sum-score with one factor and three factor solutions, the correlations and Cronbach's alpha can be compared.

When performing a factor analysis, the factor score of each participant can be calculated. Letting only items 1, 2, 3, 5 and 6 load on a single factor, the average correlation of the factor score of each person with their sum-score is very high (0.952). Factor scores based on only items 1, 2 and 3 have an even higher correlation with the sum-score. Comparing the three factors separately to the sum-score, differences in correlation and internal consistency can be observed. Especially the factor *Angry Reaction* reveals a lower internal consistency (see also table 6).

This is a distinct difference from findings by Van der Ploeg et al., where heterogeneity leads to items 4 and 9 loading on different factors dependent on the factor score. In this analysis, these items do load strongly on a factor together. These findings already suggest that a Factor Mixture Model might provide a better fit to this data.

factor	correlation with sum-score	Cronbach's Alpha
1 factor	0.952	0.865
1 factor; only items 1, 2 & 3	0.985	0.850
3 factors - AT	0.928	0.872
3 factors - AO	0.904	0.765
3 factors - AR	0.895	0.655

Table 6: For every factor or combination of items, a Cronbach's Alpha as well as a correlation with the sum-score can be calculated. Note the relatively low value for Cronbach's alpha for the third factor in the three-factor solution, meaning that the items that were in included in *Angry Reaction* have a lower internal consistency. AT = factor Angry Temperament, AO = factor Angry Other, AR = factor Angry Reaction.)

	factor loadings			
	factor 1 (se) *Angry Temperament*	factor 2 (se) *Angry Other*	factor 3 (se) *Angry Reaction*	residual variances (se)
Item 1	**0.925** (0.020)	-0.056 (0.026)	-0.003 (0.010)	0.224 (0.008)
Item 2	**0.902** (0.014)	0.016 (0.011)	-0.033 (0.013)	0.196 (0.007)
Item 3	**0.996** (0.023)	-0.090 (0.032)	-0.011 (0.009)	0.149 (0.008)
Item 4	-0.001 (0.006)	**0.939** (0.049)	-0.109 (0.045)	0.238 (0.032)
Item 5	**0.647** (0.021)	0.209 (0.029)	0.043 (0.019)	0.287 (0.009)
Item 6	**0.578** (0.023)	0.094 (0.032)	0.202 (0.021)	0.387 (0.011)
Item 7	0.193 (0.027)	-0.002 (0.006)	**0.477** (0.017)	0.641 (0.011)
Item 8	-0.016 (0.014)	0.036 (0.037)	**0.748** (0.032)	0.419 (0.018)
Item 9	0.024 (0.051)	**0.830** (0.049)	0.017 (0.009)	0.263 (0.025)
Item 10	0.021 (0.020)	-0.031 (0.034)	**0.717** (0.031)	0.498 (0.017)

Table 7: Factor loadings for three factor model, resulting from exploratory factor analysis. For each item, the highest loading is in bold. Note the high residual variances for item 7, 8 and 10.

Factor Mixture Models

The two-class and three-class three-factor mixture models (3f,2c and 3f,3c) both provide a better fit than the LCA and EFA models. The BIC favors the model with three classes. Information criteria show that the FMMs provide a superior fit to the data than any of the LCA or exploratory FA models. The two-class solution splits the sample in a majority (79%) of low-scoring individuals, and a smaller high-scoring group. Interestingly, the covariances between the factors is higher for the lower scoring individuals. A third class takes participants from both classes. With item thresholds close to each other, this class seems to identify extreme-scoring participants, either very high or very low. However, the covariances between the factors in this class are even considerably higher then in the low-scoring class. The high-scoring class 1 is especially high for items loading on factor AR.

			Factor covariances												
			F1 with	F1 with	F2 with	Threshold of highest category per item									
		%	F2	F3	F3	1	2	3	4	5	6	7	8	9	10
						Without covariates predicting class membership									
C1	*high*	16	.39	.38	.19	2.27	2.36	2.66	2.14	2.69	2.68	1.22	1.65	2.16	0.96
C2	*low*	60	.67	.31	.38	‡	‡	‡	‡	‡	‡	3.13	2.73	3.50	2.08
C3	*extreme*	24	.82	.66	.80	1.82	1.86	2.09	1.83	2.01	2.02	1.31	1.55	1.92	1.09
						With covariates predicting class membership									
C1	*high*	16	.37	.41	.22	1.70	1.79	1.94	1.70	1.84	1.67	1.00	1.47	1.75	0.80
C2	*low*	64	.64	.21	.21	11.20	3.76	10.71	14.18	4.03	3.28	‡	2.80	3.66	2.18
C3	*medium*	20	.84	.68	.90	1.82	1.84	2.10	1.85	2.18	‡	1.29	1.48	1.91	1.00

Table 8: Class proportions, factor covariances and item thresholds of highest category. Note that item thresholds can be interpreted as Z-scores, corresponding to the probability of responding with the highest category. A higher thresholds means a lower probability. ‡ indicates a zero probability of scoring above the threshold.

Post-hoc testing covariate effects resulted in significant regression coefficients for both age and sex ($P < 0.001$). The size of C3 decreases, while the size of C1 increases, suggesting that the high-scoring participants from C3 move to C1, making C3 a *medium* scoring class (except for item 6, where the threshold for the highest category is even higher then C2). If age effects are not taken into account, a higher age decreases probability of membership of the high-scoring class 1, and increases probability of the low-scoring class 2 membership. This is because on average, Trait Anger decreases slightly with age. As a result, older people will be overrepresented in the low-scoring latent class. If sex-specific effects are not taken into account, women are overrepresented in class 3. Taking sex effects into account will increase female membership probability for both class 1 and 2. There are no significant regression sex covariate coefficients explaining a difference between class 1 and 2.

class		estimate (se)	P	odds ratio
		—Using reference class 3—		
C1	on Age	-0.013 (0.003)	0.000	0.987
	on Sex	0.284 (0.075)	0.000	1.328
	intercept	-0.236 (0.155)	0.128	
C2	on Age	0.010 (0.002)	0.000	1.010
	on Sex	0.266 (0.058)	0.000	1.305
	intercept	0.344 (0.125)	0.006	

Table 9: Results of regressing class membership on age and sex, using class 3 as a reference. For results using other classes as reference, see the Appendix.

	No. of Classes (c) or Factors (f)	No. of Estimated Parameters	RMSEA	Log Likelihood	BIC	Class Proportions
LCA model	1c	30		-83114.56	166502.8	1
	2c	61		-73006.10	146568.7	0.31, 0.69
	3c	92		-70379.63	141598.5	0.35, 0.08, 0.56
	4c	123		-69157.04	139436.2	0.44, 0.27 ,0.21 0.07
	5c	154		-68555.18	138515.2	0.07, 0.2, 0.01, 0.44, 0.27
	6c	185		-68092.81	137873.3	0.11, 0.39, 0.01, 0.19, 0.07, 0.23
EFA model	1f	40 (10∗)	0.135∗		139014.8	
	2f	49 (19∗)	0.098∗		136686.0	
	3f	57 (27∗)	0.041∗		135679.9	
FMM	3f, 2c	80		-67115.56	134960.9	0.79, 0.21
	3f, 3c	117		-66768.32	134604.0	0.16, 0.60, 0.24
	3f, 3c†	121		-70237.693	141579.2	0.16, 0.64, 0.20

Table 10: Results of fitting different statistical models to Trait Anger data. ∗Results when not using Mplus command *TYPE = COMPLEX*, instead of analysing a 1-class factor mixture model with command *TYPE = COMPLEX MIXTURE*. †Model estimation including age and sex covariates.

Discussion

Summary

The statistical approach used in the present study shows that factor mixture models provide the best fit to the behavioral measures of Trait Anger from questionnaire data. While latent class analysis and factor analysis are only capable of modeling individual differences based on subtypes or severity differences, respectively, the best fitting models incorporate *both* subtypes and severity differences. These best fitting models differentiate between a low-scoring majority, and minor medium- and high-scoring classes. As the NTR is a population sample, this division is not surprising. In our analysis, factor covariances differ considerably between classes and are especially high in the medium-scoring class (e.g. 0.90 between Angry Other and Angry Reaction). This correlation however does not exclude the possibility that individuals have opposite scores on the different dimensions. The fit of the mixture model could be approached with an LCA approach, using as much classes as possible. Continuous factor scores are then categorized, and individuals are classified into different subtypes. The finding that Trait Anger decreases with age and is lower in women might not be surprising, but ignoring this effect would lead to misclassification of individuals when using latent classes. The factor covariance differences for different classes that were found, are in line with earlier findings that some items may change factor depending on sum-score. Interesting to note is that the factors are the least differentiated in the medium scoring class.

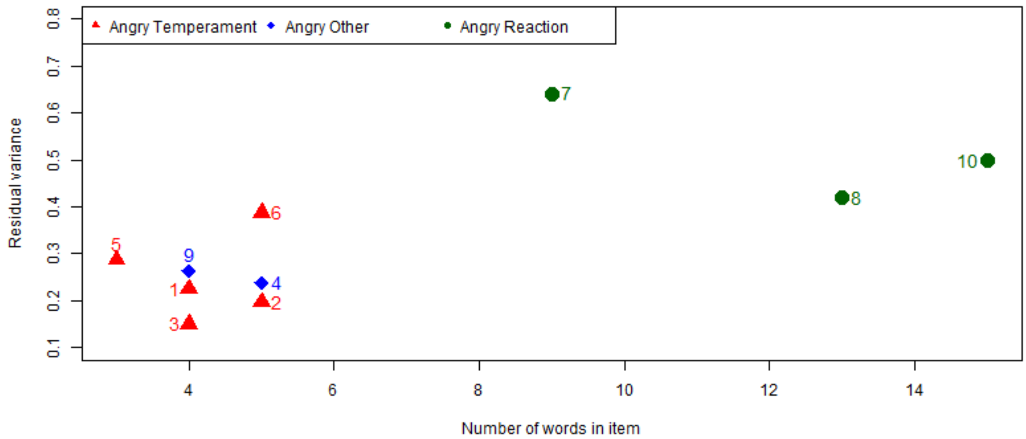


Figure 4: Item length vs. residual variance

The items belonging to the factor *Angry Reaction* stand out from the other items regarding residual variances, but also regarding the number of words in the item (see figure 4). It seems that the length of an item is relevant for the amount of residual variance. Not only are longer items more related to a specific *situation* that may be differently interpreted by different people, it is also possible that the increased time needed to read the item allows for more cognitive effort, and therefore introduce more variation in the response.

Limitations, implications and further research

This study indicates the problems arising from using a Trait Anger sum-score in a GWAS. It was already found that the condition of equal residual variances for all items did not hold. Since multiple factors can be identified, the condition of unidimensionality is violated as well. In addition, this research reported that item thresholds differ considerably between latent classes. Therefore, the condition of identical functional relations of item responses to the latent trait also does not hold. Considering this, power of a GWAS on Trait Anger would be severely limited by intrinsic features of the questionnaire.

Using a scale consisting of only ten items, with each item having only 4 response categories, seems very minimal to asses human variability of a behavioral trait. However, as shown in figure 5, a 30-point scale can approach a continuous variable quite closely. In addition, the data available were sufficient to reliably identify mixtures. In addition, the data available were sufficient to reliably identify mixtures. Using factor scores, or especially using factor mixture modeling, the phenotypic resolution can even be increased to some extent. In cases where mixture models are difficult to utilize, using as many latent classes as possible could approach the fit of a mixture model. However, results of these fits should be interpreted carefully, since class assignments are then based on categorizing continuous traits.

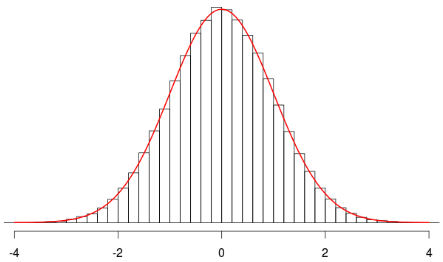


Figure 5: Precision of a 30-point scale vs. continuous variable

Relevant for meta-analysis is the fact that questionnaires can at least sometimes not be completely literally translated. This is in accordance with a recent phenotypic study of Neuroticism and Extraversion using Item Response Theory to harmonize different questionnaires (De Moor et al., 2014), where largest differences in item parameters where found in cohort with different spoken languages, which could be due to cultural and/or linguistic differences. It was however stated that these differences only slightly affect the results on a population level, since it was possible to achieve high correlations between different questionnaires using different item parameter values (*calibrations*).

Current behavioral genetic research focuses on increasing sample size or adding biological context to results through pathway analysis for example (or both, see Rietveld, 2013). However, analysing already present data might already result in a 'cleaner' phenotype with increased statistical power to detect genetic effects (for an example with Borderline personality features, see Lubke, 2013). It is highly likely that utilizing the information on phenotypic heterogeneity of Trait Anger would increase statistical power in a GWAS in a similar fashion.

References

[1] Arminger, G., Stein, P., & Wittenberg, J. *Mixtures of conditional mean-and covariance-structure models.* Psychometrika, 64(4), 475-494. 1999.

[2] van den Berg, S. M., de Moor, M. H., McGue, M., Pettersson, E., Terracciano, A., Verweij, K. J., ... & Pulkkinen, L. *Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory.* Behavior genetics, 1-19. 2014.

[3] Boomsma, D. I., Vink, J. M., Van Beijsterveldt, T. C., de Geus, E. J., Beem, A. L., Mulder, E. J., ... & van Baal, G. C. M. *Netherlands Twin Register: a focus on longitudinal research.* Twin Research, 5(05), 401-406. 2002.

[4] Browne, M. W., & Cudeck, R. *Alternative ways of assessing model fit.* Sage Focus Editions, 154, 136-136. 1993.

[5] Dolan, C. V., & van der Maas, H. L. *Fitting multivariage normal finite mixtures subject to structural equation modeling.* Psychometrika, 63(3), 227-253. 1998.

[6] Eckhardt, C., Norlander, B., & Deffenbacher, J. *The assessment of anger and hostility: A critical review.* Aggression and Violent Behavior, 9(1), 17-43. 2004. s

[7] Heinen, T. *Latent class and discrete latent trait models: Similarities and differences.* Sage Publications, Inc. 1996.

[8] Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. *Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity.* Marketing Science, 16(1), 39-59. 1997.

[9] Leonard, K. E. & Blane, J. T. *Alcohol and marital aggression in a national sample of young men.* Journal of Interpersonal Violence, 7, 19-30. 1992.

[10] Liebsohn, M. T., Oetting, E. R. & Deffenbacher, J. L. *Effects of trait anger on alcohol consumption and consequences.* Journal of Child and Adolescent Substance Abuse, 3, 17-32 1994.

[11] Lubke, G. H., Laurin, C., Amin, N., Hottenga, J. J., Willemsen, G., van Grootheest, G., ... & Boomsma, D. I. *Genome-wide analyses of borderline personality features.* Molecular psychiatry 2013.

[12] Lubke, G. H., & Muthén, B *Investigating population heterogeneity with factor mixture models.* Psychological methods, 10(1), 21. 2005.

[13] Maiuro, R. D., Cahn, T. S., Vitaliano, P. P., Wagner, B. C. & Zegree, J. B. *Anger, hostility, and depression in domestically violent versus generally assaultive men and nonviolent control subjects.* Journal of Consulting and Clinical Psychology, 56, 17-23 1998.

[14] Muthén, B., & Shedden, K. *Finite mixture modeling with mixture outcomes using the EM algorithm.* Biometrics, 55(2), 463-469. 1999.

[15] Muthén, L. K., & Muthén, B. O. *Mplus User's Guide Sixth Edition.* Los Angeles, CA: Muthén & Muthén 1998-2011. 1998-2011.

[16] Pan, H. S., Neidig, P. H., & O'Leary, K. D. *Predicting mild and severe husband-to-wife physical aggression.* Journal of consulting and clinical psychology, 62(5), 975. 1994.

[17] Van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. *Handleiding bij de Zelf-Analyse Vragenlijst ZAV. Een vragenlijst voor het meten van boosheid en woede, als toestand en als dispositie. Een nederlandstalige bewerking van de Spielberger State-Trait Anger Scale (The Dutch translation of the Speilberger State-Trait Anger Scale).* Lisse: Swets & Zeitlinger, 1982.

[18] Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & McMahon, G. *GWAS of 126,559 individuals identifies genetic variants associated with educational attainment.* science, 340(6139), 1467-1471., 2013.

[19] Van der Sluis, S., Verhage, M., Posthuma, D., & Dolan, C. V. *Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies.* PLoS One, 5(11), e13929., 2010.

[20] Spielberger, C.D. *The State -Trait Anger Scale (STAS).* In M. Hersen & A.S. Bellack (Eds.) Dictionary of behavioral assessment techniques (pp. 446-448). 1988.

[21] Spielberger, C. D., Jacobs, G., Russell, J. S., and Crane, R. S. *Assessment of anger: the state-trait anger scale.* In J.N. Butcher and C. D. Spielberger (eds.), Advances in Personality Assessment. Vol 2. Hillside, NJ: Erlbaum. 1983.

[22] Spielberger, C. D., & Sydeman, S. J. *State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory.* 1994.

[23] Suls, J., & Bunde, J. , *Anger, anxiety, and depression as risk factors for cardiovascular disease: the problems and implications of overlapping affective dispositions.* Psychological bulletin, 131(2), 260. 2005.

[24] Vermunt, J. K. *Latent Class Modeling with Covariates: Two Improved Three-Step Approaches.* Political Analysis, 18(4), 450-469. 2010.

[25] Vermunt, J. K., & Magidson, J. *Latent class models for classification.* Computational Statistics & Data Analysis, 41(3), 531-537. 2003.

[26] Yung, Y. F. *Finite mixtures in confirmatory factor-analysis models.* Psychometrika, 62(3), 297-330. 1997.

[27] Zimprich, D., & Mascherek, A. *Measurement invariance and age-related differences of trait anger across the adult lifespan.* Personality and Individual Differences, 52(3), 334-339, 2012.

Appendix

For readability issues, cohort 1990-1991, cohort 1992-1993 and cohort 2007-2008 are sometimes referred to as List 1, List 2 and List 7, respectively.

I Summary statistics of sample

See table 11 for cohort-specific information on sample size, sex and age for the data used in this study. Note that for these data, only one twin of a monozygotic twin pair and only the most recent questionnaire of each participant was included. Therefore, these numbers differ from the complete dataset as shown in table table 15.

Cohort	N	mean age (sd)	proportion male
1990-1991	1845	34.5 (14.9)	0.49
1992-1993	3511	35.0 (15.2)	0.52
2007-2008	9239	43.3 (14.7)	0.42
total	14595	39.0 (15.5)	0.46

Table 11: Summary statistics of the three cohorts included in the study.

a Sum-scores and sex differences

Sum-scores were highly skewed (see figure 6). Dividing the dataset in subgroups, small differences could be detected. For instance, female participants show a consistently lower mean. In addition, participants under the age of 20 show a higher mean and a higher standard deviation. See table 12 for mean sum-scores for different subgroups.

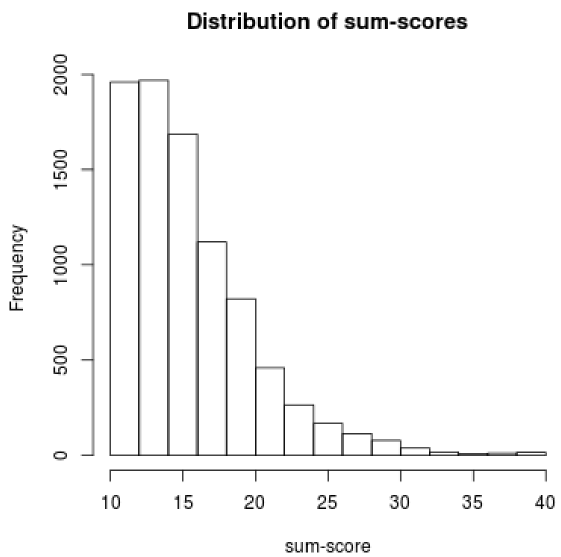


Figure 6: Distribution of sum-scores

Item	—*List*1—			—*List*2—			—*List*7—		
	n	mean	sd	n	mean	sd	n	mean	sd
all	6125	16.551	4.835	7510	16.576	4.834	3505	16.575	4.834
male	2823	16.737	4.782	4050	16.753	4.786	3302	16.753	4.782
female	3302	16.334	4.888	3460	16.368	4.882	1413	16.367	4.882
under 20	2632	16.973	5.231	2656	17.300	5.255	0	na	na
20 to 25	598	16.132	4.954	1045	16.371	4.944	118	17.136	5.547
over 25	2895	16.461	4.693	3809	16.080	4.418	3387	16.325	4.666

Table 12: Sum-scores and sex differences.

b Longitudinal effects and age

Age of the sample showed a bimodal distribution, with means around young adulthood and middle-aged adulthood. Mean age was for the three questionnaires was 30.89, 32.24 and 43.24 respectively (see table 11). Ages of participants included in cohort 2007-2008 were significantly higher than the other questionnaires. Item means were largely comparable for the three questionnaires. The last item has the highest mean in all questionnaires.

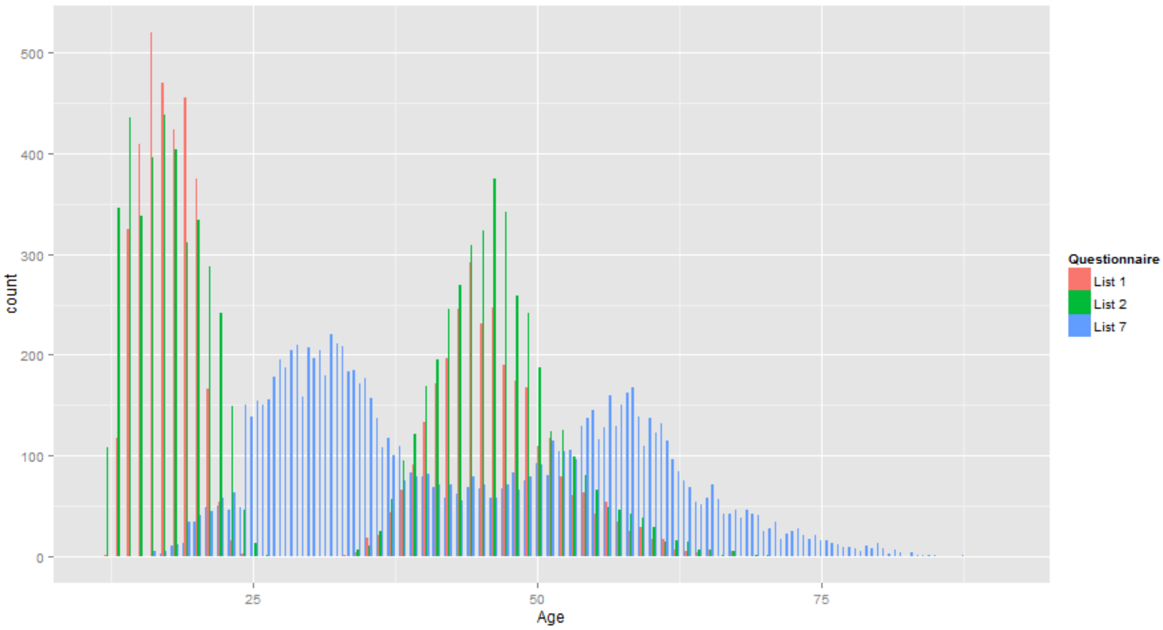


Figure 7: Age distributions for the three questionnaires

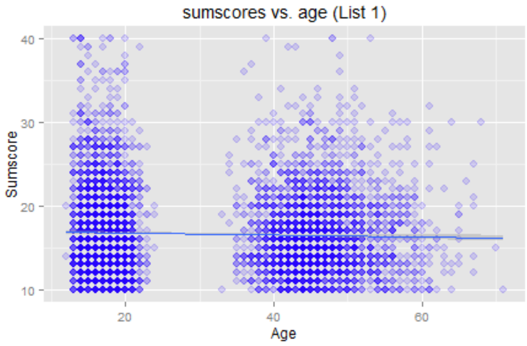
Of the 6322 participants in cohort 1990-1991, 3647 individuals also participated in cohort 1992-1993, and 2357 individuals also participanted in cohort 2007-2008. The overlap between cohort 1992-1993 and cohort 2007-2008 is 3641 (46.3%).

When selecting the participants with data on all items for all three questionnaires (n = 1608), the long-term changes of Anger can be investigated. See table 13 for correlation coefficient of item scores between questionnaires, ranging between 0.292 and 0.531.

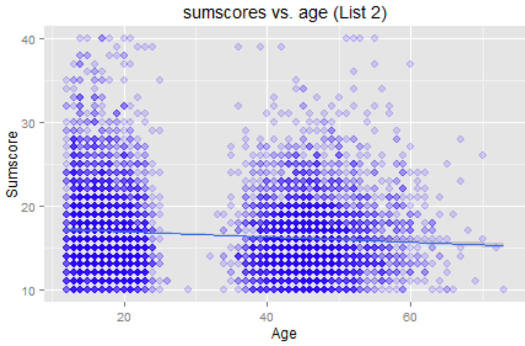
	List 1-2	List 1-7	List 2-7
Item 1	0.53	0.39	0.40
Item 2	0.52	0.39	0.38
Item 3	0.48	0.36	0.36
Item 4	0.46	0.35	0.37
Item 5	0.49	0.33	0.37
Item 6	0.42	0.29	0.31
Item 7	0.46	0.32	0.39
Item 8	0.37	0.28	0.34
Item 9	0.45	0.32	0.39
Item 10	0.38	0.29	0.31
sum-scores	0.65	0.48	0.54

Table 13: Correlation coefficients (Pearson's)

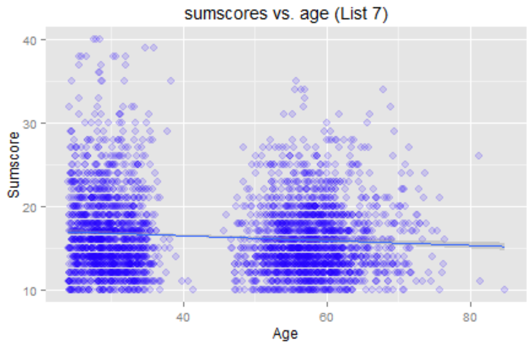
In the three separate datasets, mean sum-score decrease slightly with age (see figure 8).



(a) Intercept 17.02, slope -0.012, $R^2 = 0.00132$



(b) Intercept 17.56, slope -0.031, $R^2 = 0.009639$



(c) Intercept 17.64, slope -0.029, $R^2 = 0.008356$

Figure 8: Sum-scores per age for three separate questionnaires

c Individual trajectories

Plotting individual trajectories shows that Trait Anger is not a completely stable trait. See figure 9 for individual trajectories of sum-scores. A slight decrease in both item means and their standard deviations is visible

	—*List1*—		—*List2*—		—*List7*—	
Item	n	mean (sd)	n	mean (sd)	n	mean (sd)
Item 1	1818	1.52 (0.684)	1812	1.55 (0.688)	1786	1.40 (0.572)
Item 2	1823	1.49 (0.703)	1814	1.49 (0.689)	1782	1.38 (0.578)
Item 3	1822	1.31 (0.597)	1812	1.31 (0.586)	1781	1.23 (0.475)
Item 4	1821	1.74 (0.694)	1814	1.78 (0.694)	1782	1.74 (0.636)
Item 5	1820	1.36 (0.617)	1808	1.34 (0.590)	1780	1.25 (0.499)
Item 6	1820	1.37 (0.587)	1810	1.34 (0.563)	1782	1.22 (0.459)
Item 7	1821	1.93 (0.803)	1808	1.85 (0.778)	1778	1.63 (0.651)
Item 8	1817	1.77 (0.731)	1807	1.75 (0.720)	1779	1.64 (0.631)
Item 9	1820	1.68 (0.697)	1794	1.66 (0.682)	1756	1.60 (0.631)
Item 10	1818	2.26 (0.839)	1805	2.17 (0.860)	1782	1.85 (0.728)

Table 14: Sample sizes, means and standard deviations for different questionnaires, for individuals who participated in all three questionnaires

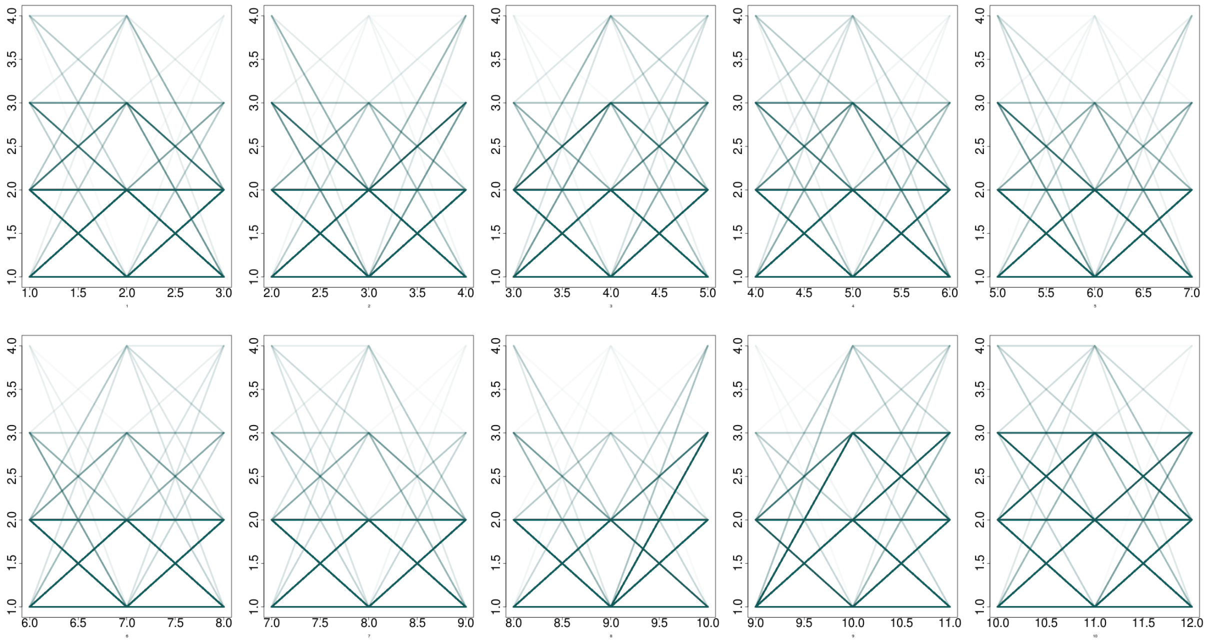


Figure 9: Item-specific longitudinal trajectories for people who participated in all three questionnaires.

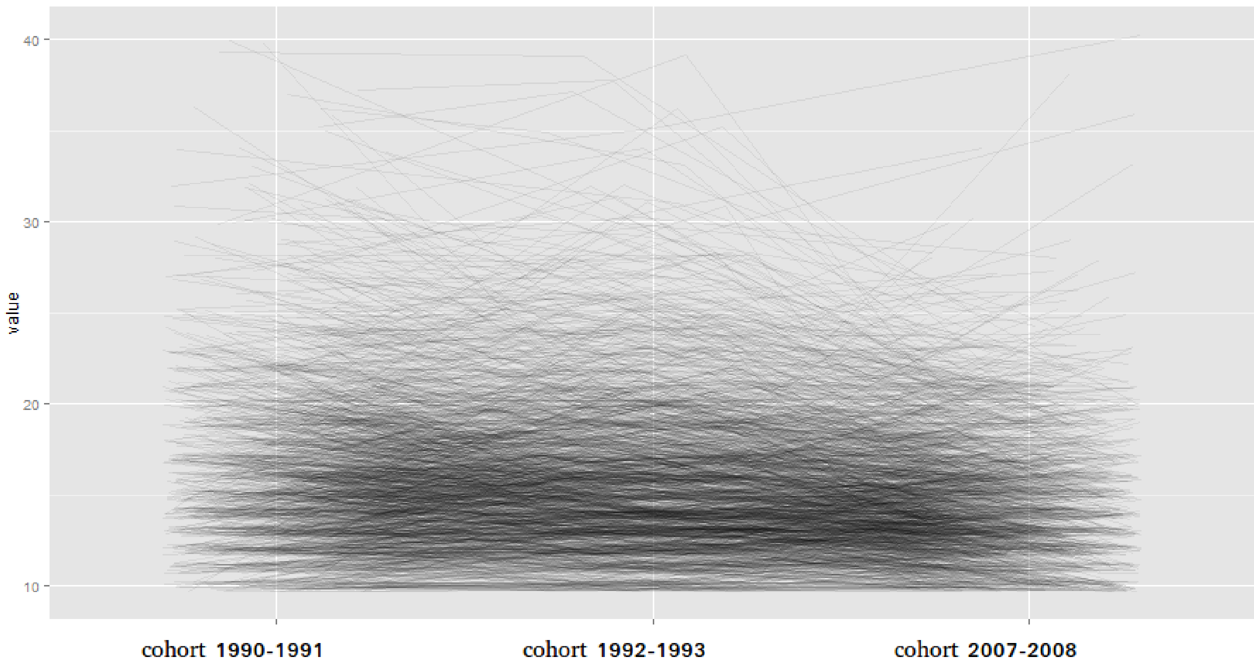


Figure 10: Individual trajectories of sum-scores. Some jitter was applied to x-coordinates to increase readability.

d Responses and prevalences per item and cohort

	n	mean age (sd)	min	max
List 1	6318	30.89 (15.04)	12	71
List 2	7859	32.24 (15.25)	12	73
List 7	11221	43.24 (14.72)	16	90

Table 15: Sample size and mean, minimum and maximum age for each questionnaire.

Mean age was for the three questionnaires was 30.89, 32.24 and 43.24 respectively (see also table 15). Ages of participants included in for List 7 were significantly higher than the other questionnaires. List 7 has almost twice as much participants as List 1. Note that there is overlap between questionnaires, since participants are enrolled in longitudinal research. The dataset used in the study consists of the the most recent questionnaire of each participant, with one twin of each twin pair excluded.

	—*List*1—			—*List*2—			—*List*7—		
Item	n	mean (sd)	age (sd)	n	mean (sd)	age (sd)	n	mean (sd)	age (sd)
1	6291	1.56 (0.706)	30.88 (15.04)	7790	1.60 (0.722)	32.20 (15.22)	11062	1.43 (0.579)	43.15 (14.68)
2	6292	1.52 (0.711)	30.90 (15.04)	7785	1.52 (0.719)	32.21 (15.21)	11064	1.40 (0.583)	43.14 (14.67)
3	6286	1.35 (0.623)	30.86 (15.03)	7783	1.34 (0.624)	32.20 (15.22)	11040	1.23 (0.480)	43.11 (14.67)
4	6289	1.74 (0.703)	30.87 (15.03)	7793	1.79 (0.717)	32.23 (15.23)	11059	1.77 (0.634)	43.12 (14.68)
5	6281	1.37 (0.625)	30.88 (15.03)	7767	1.35 (0.612)	32.22 (15.22)	11047	1.26 (0.507)	43.11 (14.66)
6	6277	1.37 (0.612)	30.89 (15.04)	7784	1.38 (0.608)	32.20 (15.22)	11025	1.22 (0.455)	43.12 (14.65)
7	6281	1.96 (0.830)	30.89 (15.04)	7769	1.95 (0.839)	32.19 (15.22)	11030	1.68 (0.686)	43.11 (14.66)
8	6265	1.81 (0.763)	30.89 (15.04)	7747	1.81 (0.777)	32.17 (15.21)	11040	1.63 (0.636)	43.11 (14.66)
9	6273	1.68 (0.704)	30.88 (15.03)	7705	1.68 (0.707)	32.11 (15.21)	10923	1.64 (0.634)	43.04 (14.62)
10	6281	2.30 (0.869)	30.90 (15.04)	7766	2.19 (0.882)	32.20 (15.22)	11065	1.85 (0.714)	43.13 (14.66)

Table 16: Item level sample sizes, score means, score sd, mean age and age sd for different questionnaires. Based on all valid items

For number of participants, mean score and mean age for each item individually, see table 16. Prevalences for the three cohorts separately are shown graphically in 11, 12 and 13, as well as in tables 17, 18, and 19.

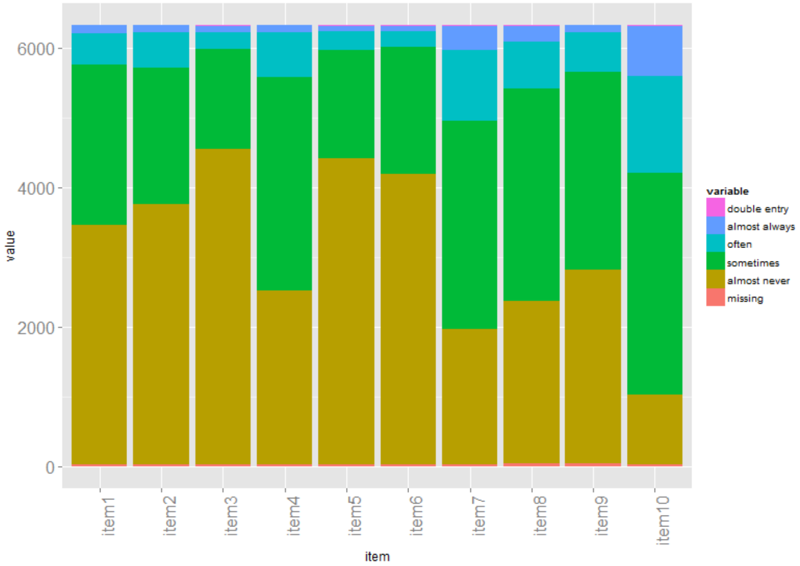


Figure 11: Relative item prevalences,List 1

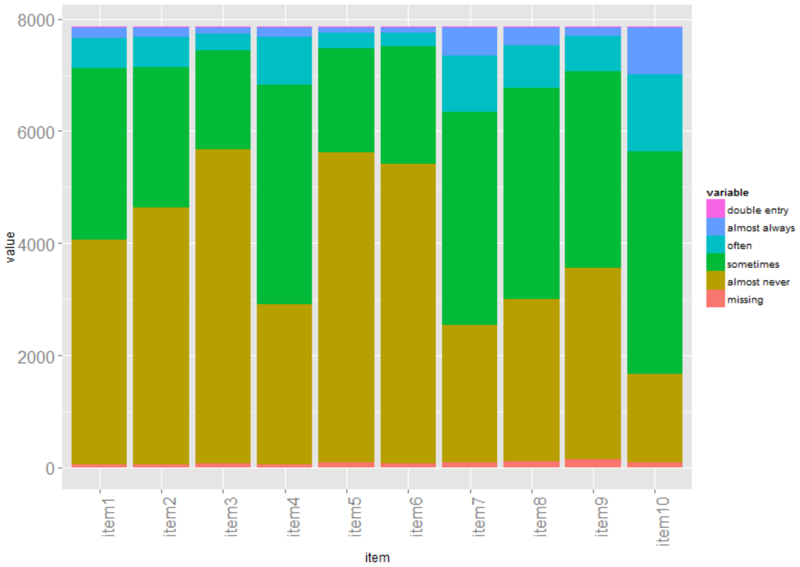


Figure 12: Relative item prevalences, List 2

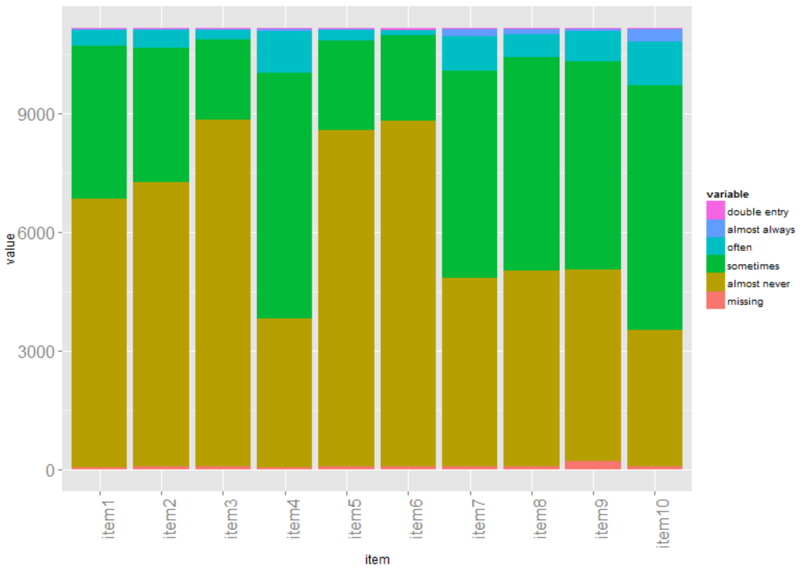


Figure 13: Relative item prevalences, List 7

	double (rel)	missing (rel)	almost never (rel)	sometimes (rel)	often (rel)	almost always (rel)
1	25 (0.00)	3437 (0.54)	2291 (0.36)	449 (0.07)	114 (0.02)	6 (0.00)
2	24 (0.00)	3735 (0.59)	1960 (0.31)	494 (0.08)	103 (0.02)	6 (0.00)
3	24 (0.00)	4521 (0.72)	1430 (0.23)	250 (0.04)	85 (0.01)	12 (0.00)
4	27 (0.00)	2493 (0.39)	3058 (0.48)	635 (0.10)	103 (0.02)	6 (0.00)
5	34 (0.01)	4382 (0.69)	1553 (0.25)	272 (0.04)	74 (0.01)	7 (0.00)
6	26 (0.00)	4166 (0.66)	1817 (0.29)	228 (0.04)	66 (0.01)	19 (0.00)
7	33 (0.01)	1939 (0.31)	2983 (0.47)	1015 (0.16)	344 (0.05)	8 (0.00)
8	48 (0.01)	2318 (0.37)	3051 (0.48)	674 (0.11)	222 (0.04)	9 (0.00)
9	43 (0.01)	2777 (0.44)	2825 (0.45)	571 (0.09)	100 (0.02)	6 (0.00)
10	33 (0.01)	987 (0.16)	3181 (0.50)	1384 (0.22)	729 (0.12)	8 (0.00)

Table 17: Item prevalences List 1

	double (rel)	missing (rel)	almost never (rel)	sometimes (rel)	often (rel)	almost always (rel)
1	61 (0.01)	3996 (0.51)	3065 (0.39)	544 (0.07)	185 (0.02)	12 (0.00)
2	58 (0.01)	4572 (0.58)	2515 (0.32)	525 (0.07)	173 (0.02)	20 (0.00)
3	62 (0.01)	5622 (0.71)	1755 (0.22)	290 (0.04)	116 (0.01)	18 (0.00)
4	55 (0.01)	2846 (0.36)	3920 (0.50)	866 (0.11)	161 (0.02)	15 (0.00)
5	88 (0.01)	5527 (0.70)	1868 (0.24)	273 (0.03)	99 (0.01)	8 (0.00)
6	79 (0.01)	5339 (0.68)	2088 (0.27)	248 (0.03)	94 (0.01)	15 (0.00)
7	84 (0.01)	2449 (0.31)	3801 (0.48)	1012 (0.13)	507 (0.06)	10 (0.00)
8	104 (0.01)	2901 (0.37)	3759 (0.48)	763 (0.10)	324 (0.04)	12 (0.00)
9	147 (0.02)	3407 (0.43)	3510 (0.45)	639 (0.08)	149 (0.02)	11 (0.00)
10	86 (0.01)	1585 (0.20)	3971 (0.51)	1369 (0.17)	841 (0.11)	11 (0.00)

Table 18: Item prevalences List 2

	double (rel)	missing (rel)	almost never (rel)	sometimes (rel)	often (rel)	almost always (rel)
1	25 (0.00)	59 (0.01)	6773 (0.61)	3861 (0.35)	394 (0.04)	34 (0.00)
2	19 (0.00)	63 (0.01)	7202 (0.65)	3381 (0.30)	444 (0.04)	37 (0.00)
3	36 (0.00)	70 (0.01)	8773 (0.79)	2018 (0.18)	227 (0.02)	22 (0.00)
4	26 (0.00)	61 (0.01)	3744 (0.34)	6201 (0.56)	1054 (0.09)	60 (0.01)
5	16 (0.00)	83 (0.01)	8498 (0.76)	2249 (0.20)	273 (0.02)	27 (0.00)
6	38 (0.00)	83 (0.01)	8714 (0.78)	2165 (0.19)	128 (0.01)	18 (0.00)
7	31 (0.00)	85 (0.01)	4745 (0.43)	5246 (0.47)	860 (0.08)	179 (0.02)
8	31 (0.00)	75 (0.01)	4942 (0.44)	5391 (0.48)	590 (0.05)	117 (0.01)
9	27 (0.00)	196 (0.02)	4853 (0.44)	5247 (0.47)	769 (0.07)	54 (0.00)
10	15 (0.00)	66 (0.01)	3448 (0.31)	6178 (0.55)	1102 (0.10)	337 (0.03)

Table 19: Item prevalences List 7

II Analyses

Details about the dataset used in the analyses (referred to as the *combined dataset* can be found is themethods section. Cronbach's Alpha was respectively 0.859, 0.852, and 0.840 in the three cohorts, and 0.865 in the combined dataset. Polychoric correlations indicate that items 1, 2, 3 and 5 are highly related. (table 20 and 21). Items 7, 8 and 10 stand out with relatively low polychoric correlations.

	1	2	3	4	5	6	7	8	9	10
1	1									
2	0.8	1								
3	0.82	0.82	1							
4	0.57	0.59	0.58	1						
5	0.68	0.75	0.77	0.64	1					
6	0.65	0.66	0.68	0.54	0.69	1				
7	0.38	0.37	0.39	0.36	0.39	0.46	1			
8	0.34	0.35	0.35	0.39	0.4	0.41	0.44	1		
9	0.57	0.6	0.58	0.75	0.62	0.59	0.39	0.43	1	
10	0.32	0.33	0.32	0.33	0.35	0.38	0.41	0.54	0.39	1

Table 20: Polychoric correlations (rho)

	1	2	3
V7	0.13	1.48	2.2
V8	0.28	1.46	2.2
V9	0.68	1.74	2.4
V10	-0.36	1.2	2.2
V11	0.61	1.71	2.4
V12	0.59	1.83	2.4
V13	-0.34	1.05	1.7
V14	-0.23	1.25	1.9
V15	-0.14	1.34	2.3
V16	-0.69	0.78	1.5

Table 21: Tau-values for polychoric correlations

III Exploratory Factor Analysis

a Overview of EFA results

For a scree-plot and table of the eigen-values resulting from the exploratory factor analysis, see figure 14 and table 22. RMSEA-values are shown in figure 15 and table 23.

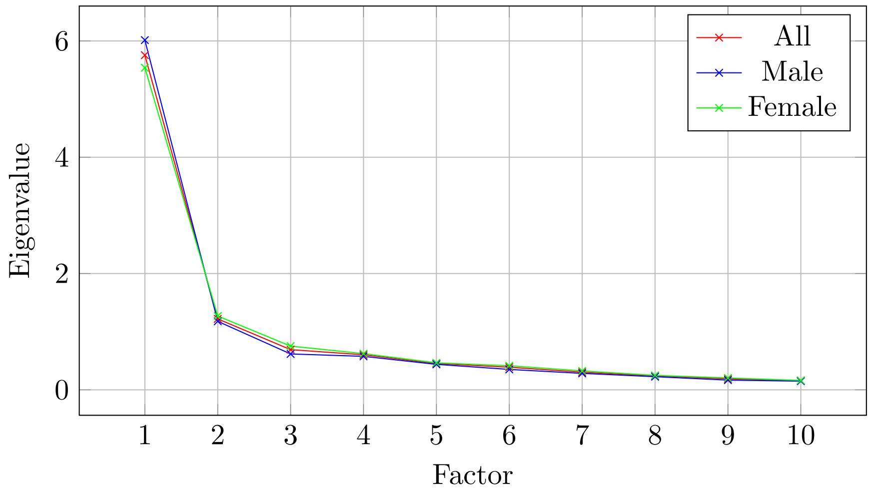


Figure 14: Eigenvalues for factor analysis on all both genders

	1	2	3	4	5	6	7	8	9	10
all	5.755	1.220	0.691	0.603	0.452	0.389	0.303	0.241	0.189	0.157
male	6.014	1.178	0.617	0.576	0.438	0.351	0.283	0.226	0.167	0.149
female	5.539	1.269	0.752	0.622	0.464	0.414	0.325	0.249	0.204	0.162

Table 22: Eigenvalues

Number of factors	Free parameters	*RMSEA* All	male	female
1	10	0.135	0.130	0.140
2	19	0.098	0.090	0.102
3	27	0.041	0.042	0.040

Table 23: Model fit information of exploratory factor analyses)

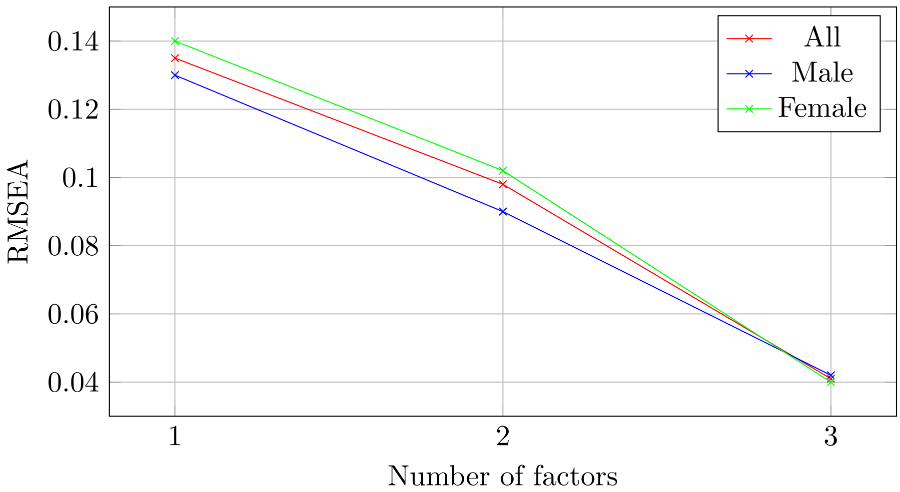


Figure 15: Eigenvalues for factor analysis on all both genders

b Results from one-factor EFA

Results of a one-factor explaratory factor analysis are shown below, for all participants together and for males and females separately.

Model fit information	
Number of Free Parameters	10
RMSEA estimate	0.135
RMSEA 90 Percent C.I.	0.132 - 0.138
Probability RMSEA <= 0.05	0.000

Table 24: Model fit information for one-factor model, all cases

	factor loadings	
	factor 1 (se)	*residual variances (se)*
Item1	0.853 (0.005)	0.272 (0.008)
Item2	0.873 (0.004)	0.238 (0.007)
Item3	0.893 (0.004)	0.202 (0.008)
Item4	0.762 (0.006)	0.419 (0.009)
Item5	0.835 (0.006)	0.303 (0.009)
Item6	0.768 (0.007)	0.410 (0.011)
Item7	0.518 (0.009)	0.732 (0.010)
Item8	0.542 (0.009)	0.706 (0.010)
Item9	0.783 (0.006)	0.386 (0.009)
Item10	0.500 (0.009)	0.750 (0.009)

Table 25: Factor loadings for one-factor model.

Model fit information	
Number of Free Parameters	10
RMSEA estimate	0.130
RMSEA 90 Percent C.I.	0.126 - 0.135
Probability RMSEA <= 0.05	0.000

Table 26: Model fit information for one-factor model, male cases

	factor loadings	
	factor 1 (se)	*residual variances (se)*
Item1	0.864 (0.006)	0.254 (0.010)
Item2	0.896 (0.006)	0.198 (0.010)
Item3	0.900 (0.006)	0.189 (0.011)
Item4	0.771 (0.008)	0.406 (0.012)
Item5	0.857 (0.007)	0.265 (0.013)
Item6	0.800 (0.010)	0.360 (0.016)
Item7	0.533 (0.013)	0.716 (0.014)
Item8	0.572 (0.013)	0.672 (0.015)
Item9	0.802 (0.008)	0.357 (0.013)
Item10	0.495 (0.014)	0.755 (0.013)

Table 27: Factor loadings for one-factor model, male cases

Model fit information	
Number of Free Parameters	10
RMSEA estimate	0.140
RMSEA 90 Percent C.I.	0.136 - 0.144
Probability RMSEA <= 0.05	0.000

Table 28: Model fit information for one-factor model, female cases

	factor loadings	
	factor 1 (se)	*residual variances (se)*
Item1	0.846 (0.007)	0.284 (0.011)
Item2	0.852 (0.006)	0.273 (0.010)
Item3	0.887 (0.007)	0.214 (0.012)
Item4	0.758 (0.008)	0.425 (0.012)
Item5	0.814 (0.008)	0.337 (0.014)
Item6	0.742 (0.010)	0.450 (0.015)
Item7	0.501 (0.013)	0.749 (0.013)
Item8	0.515 (0.013)	0.735 (0.013)
Item9	0.770 (0.008)	0.406 (0.013)
Item10	0.504 (0.013)	0.746 (0.013)

Table 29: Factor loadings for one-factor model, female cases.

c Results from two-factor EFA

Results of a two-factor explaratory factor analysis are shown below, for all participants together and for males and females separately.

Model fit information	
Number of Free Parameters	19
RMSEA estimate	0.098
RMSEA 90 Percent C.I.	0.094 - 0.101
Probability RMSEA <= 0.05	0.000

Table 30: Model fit information for two-factor model, all cases

	factor loadings		
	factor 1 (se)	*factor 2 (se)*	*residual variances (se)*
Item1	0.879 (0.007)	-0.006 (0.008)	0.232 (0.008)
Item2	0.894 (0.005)	0.003 (0.006)	0.198 (0.007)
Item3	0.932 (0.009)	-0.031 (0.014)	0.159 (0.008)
Item4	0.497 (0.013)	0.405 (0.013)	0.392 (0.009)
Item5	0.741 (0.010)	0.178 (0.014)	0.290 (0.009)
Item6	0.614 (0.012)	0.262 (0.015)	0.396 (0.011)
Item7	0.175 (0.015)	0.475 (0.014)	0.663 (0.011)
Item8	-0.003 (0.009)	0.733 (0.013)	0.465 (0.015)
Item9	0.475 (0.013)	0.464 (0.013)	0.342 (0.009)
Item10	-0.003 (0.011)	0.670 (0.013)	0.554 (0.014)

Table 31: Factor loadings for two factor model.

Model fit information	
Number of Free Parameters	19
RMSEA estimate	0.090
RMSEA 90 Percent C.I.	0.085 - 0.095
Probability RMSEA <= 0.05	0.000

Table 32: Model fit information for two-factor model, male cases

factor loadings			
	factor 1 (se)	*factor 2 (se)*	*residual variances (se)*
Item1	0.889 (0.011)	-0.013 (0.016)	0.222 (0.010)
Item2	0.913 (0.009)	-0.003 (0.011)	0.170 (0.010)
Item3	0.917 (0.009)	-0.004 (0.012)	0.164 (0.010)
Item4	0.575 (0.018)	0.298 (0.020)	0.394 (0.012)
Item5	0.773 (0.015)	0.149 (0.021)	0.256 (0.013)
Item6	0.632 (0.019)	0.266 (0.023)	0.347 (0.016)
Item7	0.158 (0.025)	0.507 (0.023)	0.631 (0.016)
Item8	0.007 (0.003)	0.753 (0.015)	0.427 (0.022)
Item9	0.551 (0.018)	0.376 (0.019)	0.330 (0.013)
Item10	-0.044 (0.029)	0.706 (0.027)	0.533 (0.022)

Table 33: Factor loadings for two factor model, male cases

Model fit information	
Number of Free Parameters	19
RMSEA estimate	0.102
RMSEA 90 Percent C.I.	0.098 - 0.107
Probability RMSEA <= 0.05	0.000

Table 34: Model fit information for two-factor model, female cases

factor loadings			
	factor 1 (se)	*factor 2 (se)*	*residual variances (se)*
Item1	0.873 (0.009)	0.001 (0.010)	0.237 (0.011)
Item2	0.883 (0.008)	-0.001 (0.009)	0.222 (0.010)
Item3	0.942 (0.012)	-0.050 (0.020)	0.155 (0.012)
Item4	0.422 (0.022)	0.490 (0.019)	0.383 (0.013)
Item5	0.714 (0.015)	0.194 (0.019)	0.319 (0.014)
Item6	0.596 (0.018)	0.253 (0.021)	0.435 (0.015)
Item7	0.167 (0.022)	0.460 (0.020)	0.687 (0.014)
Item8	-0.037 (0.027)	0.721 (0.022)	0.504 (0.019)
Item9	0.399 (0.022)	0.535 (0.018)	0.348 (0.013)
Item10	0.002 (0.002)	0.659 (0.014)	0.564 (0.018)

Table 35: Factor loadings for two factor model, female cases.

d Results from three-factor EFA

Results of a three-factor explaratory factor analysis are shown below, for all participants together and for males and females separately.

Model fit information	
Number of Free Parameters	27
RMSEA estimate	0.041
RMSEA 90 Percent C.I.	0.037 - 0.045
Probability RMSEA <= 0.05	1.000

Table 36: Model fit information for three-factor model, all cases

	factor loadings			
	factor 1 (se)	*factor 2 (se)*	*factor 3 (se)*	*residual variances (se)*
Item1	0.925 (0.020)	-0.056 (0.026)	-0.003 (0.010)	0.224 (0.008)
Item2	0.902 (0.014)	0.016 (0.011)	-0.033 (0.013)	0.196 (0.007)
Item3	0.996 (0.023)	-0.090 (0.032)	-0.011 (0.009)	0.149 (0.008)
Item4	-0.001 (0.006)	0.939 (0.049)	-0.109 (0.045)	0.238 (0.032)
Item5	0.647 (0.021)	0.209 (0.029)	0.043 (0.019)	0.287 (0.009)
Item6	0.578 (0.023)	0.094 (0.032)	0.202 (0.021)	0.387 (0.011)
Item7	0.193 (0.027)	-0.002 (0.006)	0.477 (0.017)	0.641 (0.011)
Item8	-0.016 (0.014)	0.036 (0.037)	0.748 (0.032)	0.419 (0.018)
Item9	0.024 (0.051)	0.830 (0.049)	0.017 (0.009)	0.263 (0.025)
Item10	0.021 (0.020)	-0.031 (0.034)	0.717 (0.031)	0.498 (0.017)

Table 37: Factor loadings for three factor model.

Model fit information	
Number of Free Parameters	27
RMSEA estimate	0.042
RMSEA 90 Percent C.I.	0.036 - 0.048
Probability RMSEA <= 0.05	0.986

Table 38: Model fit information for three-factor model, male cases

	factor loadings			
	factor 1 (se)	*factor 2 (se)*	*factor 3 (se)*	*residual variances (se)*
Item1	0.913 (0.019)	-0.022 (0.025)	-0.014 (0.020)	0.212 (0.011)
Item2	0.914 (0.016)	0.013 (0.017)	-0.017 (0.019)	0.164 (0.010)
Item3	0.942 (0.021)	-0.023 (0.028)	-0.006 (0.017)	0.154 (0.010)
Item4	0.000 (0.003)	0.981 (0.061)	-0.160 (0.063)	0.229 (0.042)
Item5	0.639 (0.031)	0.256 (0.037)	0.013 (0.017)	0.252 (0.013)
Item6	0.568 (0.033)	0.122 (0.043)	0.205 (0.032)	0.342 (0.016)
Item7	0.163 (0.040)	0.000 (0.007)	0.507 (0.031)	0.620 (0.017)
Item8	-0.020 (0.034)	0.049 (0.069)	0.738 (0.060)	0.422 (0.025)
Item9	0.074 (0.081)	0.786 (0.078)	0.016 (0.012)	0.265 (0.029)
Item10	0.010 (0.014)	-0.113 (0.077)	0.789 (0.062)	0.481 (0.027)

Table 39: Factor loadings for three factor model, male cases

Model fit information	
Number of Free Parameters	27
RMSEA estimate	0.040
RMSEA 90 Percent C.I.	0.034 - 0.045
Probability RMSEA <= 0.05	0.999

Table 40: Model fit information for three-factor model, female cases

	factor loadings			
	factor 1 (se)	*factor 2 (se)*	*factor 3 (se)*	*residual variances (se)*
Item1	0.922 (0.023)	-0.068 (0.030)	0.010 (0.014)	0.231 (0.012)
Item2	0.899 (0.013)	0.002 (0.010)	-0.038 (0.020)	0.220 (0.011)
Item3	1.022 (0.026)	-0.131 (0.036)	-0.013 (0.011)	0.147 (0.012)
Item4	0.004 (0.035)	0.902 (0.077)	-0.060 (0.052)	0.241 (0.051)
Item5	0.657 (0.024)	0.160 (0.038)	0.079 (0.033)	0.316 (0.014)
Item6	0.594 (0.027)	0.044 (0.035)	0.218 (0.027)	0.422 (0.016)
Item7	0.195 (0.029)	-0.001 (0.021)	0.468 (0.022)	0.656 (0.016)
Item8	-0.018 (0.013)	-0.007 (0.028)	0.781 (0.026)	0.409 (0.027)
Item9	0.003 (0.048)	0.835 (0.070)	0.035 (0.037)	0.261 (0.042)
Item10	0.032 (0.029)	0.002 (0.026)	0.682 (0.029)	0.512 (0.023)

Table 41: Factor loadings for three factor model, female cases

IV Latent Class Analysis

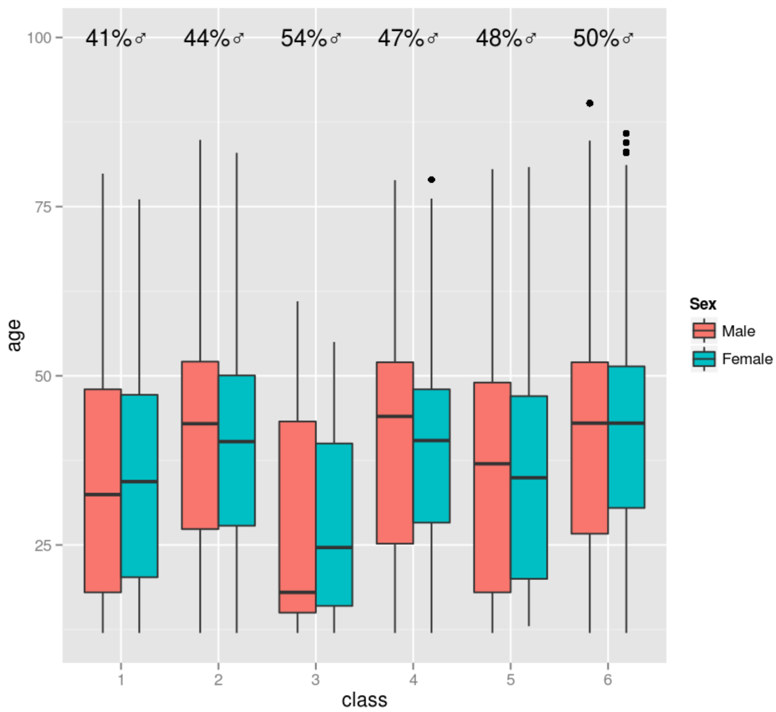


Figure 16: Age and sex differences for different classes

V Factor Mixture Modeling

class		estimate (se)	P
	—Using reference class 1—		
C2	on Age	0.023 (0.002)	0.000
	on Sex	-0.017 (0.063)	0.780
	intercept	0.58 (0.128)	0.000
C3	on Age	0.013 (0.003)	0.000
	on Sex	-0.284 (0.075)	0.000
	intercept	0.236 (0.155)	0.128
	—Using reference class 2—		
C1	on Age	-0.023 (0.002)	0.000
	on Sex	0.017 (0.063)	0.780
	intercept	-0.580 (0.128)	0.000
C3	on Age	-0.010 (0.002)	0.000
	on Sex	-0.266 (0.058)	0.000
	intercept	-0.344 (0.125)	0.006
	—Using reference class 3—		
C1	on Age	-0.013 (0.003)	0.000
	on Sex	0.284 (0.075)	0.000
	intercept	-0.236 (0.155)	0.128
C2	on Age	0.010 (0.002)	0.000
	on Sex	0.266 (0.058)	0.000
	intercept	0.344 (0.125)	0.006

Table 42: ∗Results when not using Mplus command *TYPE = COMPLEX*, instead of analysing a 1-class factor mixture model with command *TYPE = COMPLEX MIXTURE*. †Model estimation including age and sex covariates.

List of Figures

List of Tables

List of abbreviations

AT	Angry Temperament
AE	Angry Expression
BIC	Bayesian information criterion
C	(Latent) Class
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
F	Factor
FA	Factor Analysis
FMM	Factor Mixture Model
GWAS	Genome-wide association study
LCA	Latent Class Analysis
MI	Measurement Invariance
MZ	Monozygotic
N	Sample size
NTR	Netherlands Twin Registry
RMSEA	Root Mean Square Error of Approximation
SD	Standard deviation
SE	Standard error
STAS	State-Trait Anger Scale
TA	Trait Anger
TAS	Trait Anger Scale