

Table S1

*Factor Loadings for the exploratory (EFA) and confirmatory (CFA) Factor Analyses of the Personality Subscale in the two Subsets from Study 1(S1-1, S1-2) and the Data from Study 2 (S2) for the Items from the shorter Version of the GERAS after first Reduction of the Item-pool from Study 1*

Masculine personality items					Feminine personality items				
Factor	EFA		CFA		Factor	EFA		CFA	
<b>Risk-taking</b>	S1-1	S1-1	S1-2	S2	<b>Expressivity</b>	S1-1	S1-1	S1-2	S2
Reckless	.78	.77	.74	.70	Warm-hearted	.78	.81	.84	.82
Willing to take risks	.76	.80	.84	.85	Loving	.74	.85	.89	.78
Courageous	.74	.83	.67	.63	Caring	.71	.73	.80	.75
Adventurous	.67	.73	.71	.72	Compassionate	.71	.81	.78	.73
<i>Bold/Daring*</i>	.53				Delicate	.69	.76	.69	.77
<i>Brave*</i>	.52				Tender	.65	.72	.70	.65
<i>Strong*</i>	.43				Family-oriented	.46	.39	.50	.50
					<i>Affectionate*</i>	.69			
					<i>Sensitive*</i>	.56			
					<i>Open to feelings*</i>	.33			
<b>Assertiveness</b>	EFA	CFA			<b>Neuroticism</b>	EFA	CFA		
	S1-1	S1-1	S1-2	S2		S1-1	S1-1	S1-2	S2
Dominant	.67	.75	.72	.72	Anxious	.68	.70	.59	.62
Controlling	.67	.67	.67	.69	Thin-skinned	.53	.63	.53	.51
Boastful	.38	.52	.49	.38	Careful	.45	.59	.76	.63
					<i>Insecure*</i>	.74			
					<i>Vulnerable*</i>	.56			
<b>Rationality</b>	EFA	CFA							
	S1-1	S1-1	S1-2	S2					
Rational	.81	.86	.83	.70					
Analytical	.66	.70	.72	.84					
Pragmatic	.59	.63	.56	.45					

*Note.* Items in *italics\** were dropped in accordance with the modification indices; EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis; S1-1/2 = Subsets 1/2 from Study 1; S2 = Data from Study 2.

Table S2

*Factor Loadings for the Exploratory (EFA) and confirmatory (CFA) Factor Analyses of the Cognition Subscale in the two Subsets from Study 1(S1-1, S1-2) and the Data from Study 2 (S2) for the Items from the shorter Version of the GERAS after first Reduction of the Item-pool from Study 1*

Masculine cognition items					Feminine cognition items				
Factor	EFA		CFA		Factor	EFA		CFA	
<b>Spatial</b>	S1-1	S1-1	S1-2	S2	<b>Verbal</b>	S1-1	S1-1	S1-2	S2
To find an address for the first time	.85	.86	.89	.91	To explain foreign words	.67	.73	.78	.50
To find a way again	.68	.75	.78	.76	To find the right words to express certain content	.72	.78	.74	.74
To follow directions	.82	.91	.88	.89	To find synonyms for a word in order to avoid repetitions	.71	.83	.85	.92
					To phrase a text	.75	.74	.81	.80
					<i>To understand texts *</i>	.56			
					<i>To learn foreign languages *</i>	.39			
<b>Numerical</b>	EFA	S1-1	CFA	S1-2	<b>Memory</b>	EFA	S1-1	CFA	S1-2
	S1-1	S1-1	S1-2	S2		S1-1	S1-1	S1-2	S2
To solve equations	.87	.88	.82	.95	Remembering events from your own life	.66	.72	.95	.52
To understand formulas	.89	.95	.88	.72	To notice small changes	.59	.59	.52	.65
Day-to-day calculations	.65	.74	.83	.79	To remember names and faces	.37	.47	.38	.46
To write a computer program	.49	.56	.40	.42	<i>To remember the exact wording of a statement *</i>	.47			
<i>To recognize regularities *</i>	.45								

*Note.* Items in *italics*\* were dropped in accordance with the modification indices; EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis; S1-1/2 = Subsets 1/2 from Study 1; S2 = Data from Study 2.

Table S3

*Factor Loadings for the exploratory (EFA) and confirmatory (CFA) Factor Analyses from the Interests Subscale in two Subsets from Study 1 (S1-1, S1-2) and the Data from Study 2 (S2) for the Items from the shorter Version of the GERAS after first Reduction of the Item-pool from Study 1*

Factor	Masculine interest items				Factor	Feminine interest items			
	EFA		CFA			EFA		CFA	
<b>ISoM</b>	S1-1	S1-1	S1-2	S2	<b>ISoF</b>	S1-1	S1-1	S1-2	S2
Paintball	.73	.76	.76	.71	Shopping	.78	.64	.64	.57
Driving go-carts	.51	.74	.76	.80	To gossip	.60	.70	.60	.61
Drinking beer	.41	.49	.45	.49	Watching a romantic movie	.45	.70	.74	.73
Watching action movies	.37	.51	.59	.55	Talking on the phone with a friend	.45	.66	.60	.74
Playing cards	.36	.48	.46	.52	<i>Spa vacation*</i>	.43			
<i>Listening to “hard” music*</i>	.39				<i>Beauty products and make-up*</i>	.68			
<i>Computer and video games*</i>	.52								
<b>ISpM</b>	EFA		CFA		<b>ISpF</b>	EFA		CFA	
	S1-1	S1-1	S1-2	S2		S1-1	S1-1	S1-2	S2
Watching sports on TV (boxing, Formula 1, ball games...)	.60	.75	.69	.70	Yoga	.56	.61	.59	.63
Doing certain sports (soccer, basketball, hand ball, etc.)	.50	.71	.73	.67	Rhythmic gymnastics	.45	.73	.72	.72
Gym (weightlifting)	.31	.42	.38	.55	Going for a walk	.44	.42	.40	.50
					Dancing	.43	.59	.61	.71
					<i>Decorating an apartment*</i>	.40			
					<i>Baking*</i>	.39			
					<i>Crocheting/Sewing/Knitting*</i>	.48			
					<i>Drawing/</i>	.47			
					<i>Painting/Pottery*</i>				
<i>Cars*</i>	EFA								
	S1-1								
<i>Interest in fast vehicles*</i>	.75								
<i>Car driving*</i>	.51								
<i>Tinkering with cars*</i>	.49								

*Note.* Items in *italics\** were dropped in accordance with the modification indices; EFA = Exploratory Factor Analysis; CFA = Confirmatory Factor Analysis; S1-1/2 = Subsets 1/2 from Study 1; S2 = Data from Study 2; ISoM = Interests Social Masculine; ISoF = Interests Social Feminine; ISpM = Interests Sport Masculine; ISpF = Interests Sport Feminine.

Table S4

*Model Fit of the Solution Suggested by EFA and the Confirmatory Factor Analysis (CFA) of the three GERAS Subscales Cross-Validated in the Second Subset from Study 1 (S1-2) and the Data from Study 2 (S2)*

		<b>Model fit</b>					
	<i>Data set</i>	$\chi^2$	<i>df</i>	<i>CFI</i>	<i>RMSEA</i>	<i>SRMR</i>	$\chi^2/df$
<i>Personality</i>							
EFA	S1-1	1876.08	340	.920	.079	.079	5.52
	S1-1	620.54	160	.951	.063	.064	3.88
CFA	S1-2	631.01	160	.949	.063	.064	3.94
	S2	583.15	160	.923	.075	.077	3.64
<i>Cognition</i>							
EFA	S1-1	597.15	129	.947	.070	.069	4.63
	S1-1	178.87	71	.980	.046	.048	2.52
CFA	S1-2	249.90	71	.969	.059	.057	3.52
	S2	204.04	71	.957	.063	.064	2.87
<i>Interests</i>							
EFA	S1-1	1528.55	314	.904	.073	.074	4.87
	S1-1	283.25	98	.956	.051	.053	2.89
CFA	S1-2	405.83	98	.931	.066	.064	4.14
	S2	438.52	98	.906	.086	.083	4.47

*Note.* EFA = exploratory factor analysis; CFA = confirmatory factor analysis; S1-1/2 = Subsets 1/2 from Study 1; S2 = Data from Study 2.

Table S5

*Results for the Test of the Second-/Third-order Models in Three Subsamples: Study 1 Subset 1 (S1-1), Study 1 Subset 2 (S1-2), and Study 2 (S2) for the Factor Solution for the GERAS Items*

	Data subset	<i>df</i>	$X^2$	CFI	RMSEA	SRMR	$\chi^2/df$	Decision
Model 1		61						
	S1-1		90.64	.927	.026	.092	1.49	A
	S1-2		127.60	.843	.039	.101	2.09	R
	S2		49.70	1.00	<.001	.102	0.81	R
Model 2		60						
	S1-1		84.83*	.939	.024	.089	1.41	A
	S1-2		125.97	.845	.039	.103	2.10	R
	S2		39.37*	1.00	<.001	.093	0.66	A
Model 3		66						
	S1-1		88.83	.944	.022	.092	1.35	A
	S1-2		98.26*	.924	.026	.092	1.49	A
	S2		35.59*	1.00	<.001	.088	0.54	A

*Note.* S1-1/2 = Subsets 1/2 from Study 1; S2 = Data from Study 2; \*significantly better compared with the previous Model; R = reject model; A = accept model.

## Factor Structure

### First-Order Models

**Personality subscale.** For the personality subscale, the EFA suggested an optimal number of six factors. However, only five of these factors were defined by at least three items with high loadings ( $> .30$ ). After dropping the items with low or multiple loadings, a total of 28 items remained in the model (13 masculine, 15 feminine). The solution is presented in Table 1 (ESM 4). It suggested that all positively connoted feminine items should be grouped into one factor called “expressivity” and all negative feminine items into another factor called “neuroticism.” Masculine items were grouped into three factors, termed “risk-taking,” “assertiveness,” and “rationality,” respectively. A CFA that rebuilt this structure showed a moderate fit ( $\chi^2[340, N = 733] = 1876.08$ ,  $CFI = .920$ ,  $RMSEA = .079$ ,  $SRMR = .079$ ) and did not meet the criteria for the exploratory approach. In accordance with the modification indices, we additionally dropped three masculine and five feminine items for which multiple loadings on different factors were suggested. This resulted in a final five-factor model including 10 masculine and 10 feminine items with good model fit ( $\chi^2[160, N = 733] = 620.54$ ,  $CFI = .951$ ,  $RMSEA = .063$ ,  $SRMR = .064$ ), which was confirmed twice, once in Subset 2 from Study 1 ( $\chi^2[160, N = 733] = 631.01$ ,  $CFI = .949$ ,  $RMSEA = .063$ ,  $SRMR = .064$ ) and once in the data from Study 2 ( $\chi^2[160, N = 471] = 583.15$ ,  $CFI = .923$ ,  $RMSEA = .075$ ,  $SRMR = .077$ ; see Table 4 of ESM 4).

**Cognition subscale.** For the cognition subscale, the EFA suggested an optimal number of four factors. After dropping items with low loadings ( $< .30$ ), a total of 16 items remained in the model (eight masculine, eight feminine). None of these items loaded on more than one factor. The solution is presented in Table 2 of ESM 4 and suggests that masculine items be grouped into “spatial abilities” and “numerical abilities” and that feminine items be grouped into “verbal abilities” and “memory functions.” A CFA that rebuilt this structure had an acceptable fit ( $\chi^2[129, N = 733] = 597.15$ ,  $CFI = .947$ ,  $RMSEA = .070$ ,  $SRMR = .069$ ) but

did not meet the criteria for the exploratory approach. In accordance with the modification indices, one masculine and three feminine items were dropped, resulting in a final four-factor model that included seven masculine and seven feminine items. This model showed very good fit ( $\chi^2[71, N = 733] = 178.87, CFI = .980, RMSEA = .046, SRMR = .048$ ) and was confirmed in both Subset 2 from Study 1 ( $\chi^2[71, N = 733] = 249.90, CFI = .969, RMSEA = .059, SRMR = .057$ ) and the data from Study 2 ( $\chi^2[71, N = 471] = 204.04, CFI = .957, RMSEA = .063, SRMR = .064$ ; see Table 4 of ESM 4).

**Interests and activity subscale.** For the interests subscale, the EFA suggested an optimal number of five factors. After dropping items with low or multiple loadings, a total of 27 items remained in the model (13 masculine, 14 feminine). The solution is presented in Table 3 of ESM 4. It suggested that items be grouped according to masculine and feminine social activities (SocialM, SocialF), masculine and feminine Sports (SportM, SportF), as well as an Interest in Cars. A CFA that rebuilt this structure had a moderate fit ( $\chi^2[314, N = 733] = 1528.55, CFI = .904, RMSEA = .073, SRMR = .074$ ). In accordance with the modification indices, five masculine and six feminine items, including the whole “car factor” were dropped on the basis of multiple loadings on different factors. The resulting model included eight masculine and eight feminine items and had a good model fit ( $\chi^2[98, N = 733] = 283.25, CFI = .956, RMSEA = .051, SRMR = .053$ ), which was confirmed in both Subset 2 from Study 1 ( $\chi^2[98, N = 733] = 405.83, CFI = .931, RMSEA = .066, SRMR = .064$ ) and the data from Study 2 ( $\chi^2[98, N = 471] = 438.52, CFI = .906, RMSEA = .086, SRMR = .083$ ; see Table 4 of ESM 4).

For all first-order models, measurement invariance between men and women was established for all three data sets (Tables S1-S3, ESM 5).

## **Second-Order Models**

In all second-order models, we allowed error covariances between the two masculine cognitions, the two feminine cognitions, the two masculine activities, and the two feminine activities.

In the exploratory data set from Study 1, the second-order model summarizing all first-order factors into one gender role factor (Model 1) displayed a good fit but did not reach the cutoff set for exploratory analyses. A second-order model in which the masculine first-order factors were subsumed under a Masculinity factor and the feminine first-order factors under a Femininity factor (Model 2) displayed significantly better fit but also did not meet the criteria set for exploratory analyses. In this model, the loadings for verbal abilities and memory functions on the Femininity factor were low ( $< .10$ ). Therefore, we additionally tested a third-order model that also better displays the theoretically intended factor structure of our measurement: the facets served as first-order factors, subsumed by PS, CS and AIS second-order factors which were overarched by the third-order factors Masculinity and Femininity, respectively. In this third-order model, we allowed error covariances between the first-order facets risk-taking and neuroticism and between the second-order factors of masculine and feminine cognitions. Model 3 displayed a better fit than Model 2 and met the criteria for exploratory analyses. Therefore, Model 3 was kept as the final third-order model ( $\chi^2[66, N = 733] = 88.83, CFI = .944, RMSEA = .022, SRMR = .092$ ) and confirmed in both Subset 2 from Study 1 ( $\chi^2[66, N = 733] = 98.26, CFI = .924, RMSEA = .026, SRMR = .092$ ) and the data from Study 2 ( $\chi^2[66, N = 471] = 35.59, CFI = 1.00, RMSEA < .001, SRMR = .088$ ). Loadings are presented in Figure 1, model fits for all higher-order models is summarized in Table S5 of ESM 2.