

1 A DSM-5 AMPD and ICD-11 compatible measure for an early identification of  
2 personality disorders in adolescence – LoPF-Q 12-18 latent structure and short  
3 form

4  
5 Ronan Zimmermann <sup>1,2, \*</sup>, Martin Steppan <sup>1,3, \*</sup>, Johannes Zimmermann <sup>4</sup>, Lara Oeltjen <sup>5</sup>, Marc  
6 Birkhölzer <sup>1</sup>, Klaus Schmeck <sup>1</sup>, Kirstin Goth <sup>1</sup>

7  
8  
9 <sup>1</sup> Child and Adolescent Psychiatric Research Department, Psychiatric University Hospitals of the  
10 University of Basel, Basel, Switzerland

11 <sup>2</sup> Division of Clinical Psychology and Psychotherapy, Faculty of Psychology, University of  
12 Basel, Basel, Switzerland

13 <sup>3</sup> Division of Developmental and Personality Psychology, Faculty of Psychology, University of  
14 Basel, Basel, Switzerland

15 <sup>4</sup> Department of Psychology, University of Kassel, Kassel, Germany

16 <sup>5</sup> Department of Psychological Methods, Institute of Psychology, Friedrich-Schiller-Universität  
17 Jena, Jena, Germany

18  
19 \* These authors contributed equally to this work.

20  
21 Corresponding author:

22 Ronan Zimmermann

23 [ronan.zimmermann@unibas.ch](mailto:ronan.zimmermann@unibas.ch)

24

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

## Abstract

The LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire) was designed for clinical use and to promote early detection of personality disorder (PD). It is a self-report measure with 97 items to assess personality functioning in adolescents from 12 years up. It operationalizes the dimensional concept of personality disorder (PD) severity used in the Alternative DSM-5 Model for Personality Disorders and the ICD-11. In this study, we investigated the factorial structure of the LoPF-Q 12-18. Additionally, a short version was developed to meet the need of efficient screening for PD in clinical and research applications.

To investigate the factorial structure, several confirmatory factor analysis models were compared. A bifactor model with a strong general factor and four specific factors showed the best nominal fit (CFI = .91, RMSEA = .04, SRMR = .07).

The short version was derived using the ant colony optimization algorithm. This procedure resulted in a 20-item version with excellent fit for a hierarchical model with four first order factors to represent the domains and a secondary higher order factor to represent personality functioning (CFI = .98, RMSEA = .05, SRMR = .04). Clinical validity (effect size  $d = 3.1$  between PD patients and controls) and clinical utility (cutoff  $\geq 36$  providing 87.5% specificity and 80.2% sensitivity) for detecting patients with PD were high for the short version. Both, the long and short LoPF-Q 12-18 version are ready to be used for research and diagnostic purposes.

*Keywords:* Personality functioning, self-report, ant optimization algorithm, adolescents, youth, early detection, personality disorder, bifactor model

46 DSM-5 AMPD and ICD-11 compatible measure for an early identification of personality  
47 disorders in adolescence – LoPF-Q 12-18 latent structure and short form

## 48 **Introduction**

### 49 **Recent changes in the conceptualisation of personality disorders (PD)**

50 The conceptualisation of personality disorder (PD) and, subsequently, the diagnostic system for  
51 PDs is currently transitioning from a categorical system (e.g. narcissistic or avoidant PD) which is still the  
52 official system in the DSM-5 [1] to a dimensional approach. The dimensional approach is used both in the  
53 ICD-11 [2] as well as in the Alternative DSM-5 Model for Personality Disorders (AMPD). The AMPD is  
54 described in the section “emerging measures and models” of the DSM-5. The reason for the fundamental  
55 change of the guidelines for diagnosing PD are well documented shortcomings of the categorical system:  
56 For instance, individual differences in the expression of PD characteristics are not dichotomous but  
57 appear continuously distributed, thresholds (i.e. number of present symptoms required to assign a  
58 diagnosis) for categorical PD diagnoses have been criticised as largely arbitrary, and the empirical  
59 covariation of the individual criteria does not fully correspond with their assignment to the ten distinct PD  
60 categories in the diagnostic manual [3]. As a consequence, the categorical approach is no longer regarded  
61 as the only valid taxonomy and has been criticised as a hindrance to research and practice [4]. A growing  
62 number of publications in the field now argue in favour of the dimensional approach.

63 The AMPD and the ICD-11 dimensional PD models are conceptually similar. They each contain  
64 two assessment modules to characterise PDs: A first diagnostic module is the evaluation of the self- and  
65 the interpersonal functioning of the patients to represent general features and the severity of the PD. This  
66 is referred to as ‘criterion A’ in the AMPD. Criterion A is constructed from four domains: identity, self-  
67 direction (those account for self-related functioning), empathy and intimacy (accounting for interpersonal  
68 functioning). The second module is used to evaluate maladaptive personality traits to represent stylistic  
69 differences in the expression of PD [5]. This assessment of ‘trait domain qualifiers’ in the ICD-11 is not

70 mandatory for a diagnosis but helps to refine the assessment. In contrast, in the AMPD, the usage of this  
71 second module, ‘Criterion B’, is required.

72         As a further major shift in the PD diagnosis paradigm, experts now broadly agree on the  
73 importance of early detection and treatment of PD [6]. Strong evidence has been delivered showing that  
74 PD is a valid diagnosis in youth [7]. PDs can have their origin during childhood and can emerge in early  
75 adolescence [8,9]. Early detection and treatment are important as adolescence is a critical and formative  
76 period which lays the foundation in terms of psychosocial functioning for the wellbeing and productivity  
77 of the adult [10]. Additionally, from a neurocognitive perspective, adolescence also represents a window  
78 of opportunity to effectively and efficiently treat mental disorders [8]. By providing early interventions in  
79 adolescents, clinicians are trying to avert the harmful psychosocial consequences of a developing disorder  
80 and prevent chronification. This is important as PDs can have a heavily incapacitating impact on the  
81 patients and their environment, including the somatic health and life expectancy of those affected [11].  
82 Additionally, societal costs of untreated PD are high (e.g. direct healthcare costs and loss of productivity)  
83 [12,13]. To account for a perspective of PD across lifetime, both the ICD-11 and the AMPD have  
84 abolished the age limit for PD diagnoses. Multiple manualised psychotherapies are available for young  
85 PD patients [14–17]. However, to allow for early treatment, age-adequate assessment procedures to  
86 detect PDs in adolescents according to the dimensional PD concept are required.

### 87 **Need for psychometric instruments**

88         These two-fold changes (dimensional approach and earlier diagnosis) in the diagnostic  
89 systems of PDs pose a challenge for mental health care services on a global level. The World  
90 Health Organization emphasises the ICD-11 system for PDs needs to be useful and usable also  
91 for health care workers in lower-resource settings who are not highly trained specialists [18]. To  
92 overcome this challenge, evidence-based assessments are of critical importance (19).  
93 Zimmermann et al. [5] provide a brief but comprehensive review on research regarding the

94 dimensional PD models summarizing currently available measures. Birkhölzer et al. [19] provide  
95 an updated review for instruments to measure criterion A.

96 The ICD-11 model for PD is relatively new. Tools specifically targeting the ICD-11  
97 operationalisation of PD diagnosing are currently being developed and validated. Based on  
98 strong similarities of ICD-11 and the AMPD regarding personality functioning, Bach & First  
99 [20] propose that assessment tools developed for the DSM-5 AMPD model can also be used to  
100 support an ICD-11 dimensional PD diagnosis. To comply with the new ICD-11 lifetime  
101 perspective on mental disorders in general, all psychometric instruments will in principle have to  
102 be adapted for younger ages.

103 To date, the Levels of Personality Functioning Questionnaire for Adolescents (LoPF-Q  
104 12-18) [21] is the only available self-report questionnaire to assess personality functioning  
105 according to the AMPD that was developed specifically for adolescents from 12 years up. The  
106 items of the LoPF-Q 12-18 have been carefully designed to take into account the developmental  
107 stage and life situation of adolescents [22]. It has been optimised for use in clinical practice,  
108 providing several descriptive subscales matching classical psychological concepts in addition to  
109 the total score and the four domain scores. This is supposed to inform differentiated diagnoses  
110 and therapy planning and to facilitate the upcoming fundamental changes in diagnostic  
111 guidelines for PD. First developed in German language, it has been translated and culturally  
112 adapted by expert teams for English [23], Spanish [24], Turkish [25,26] and Lithuanian [27].  
113 Adaptations for Slovenian, Russian, French, Danish, Swahili (Tanzania), Hebrew, Chinese and  
114 Romanian are currently under development, showing that this instrument is supported by an  
115 international clinical and research community including low- and middle-income countries. The  
116 LoPF-Q 12-18 shows excellent scale reliability and accurately detects patients with personality

117 disorders [21]. It can be requested for free for research purposes and is also available in  
118 electronic format at the project website ([academic-tests.com](http://academic-tests.com)).

### 119 **Test construction and psychometric properties of the LoPF-Q 12-18**

120 The LoPF-Q 12-18 is a 97-item self-report measure for adolescents between 12 and 18  
121 years (+/- 2 years) to assess the dimensions of personality functioning: Identity, Self-direction,  
122 Empathy, and Intimacy. It is designed to enable a dimensional differentiation between healthy  
123 and impaired personality functioning to promote early detection of PD (criterion A). The  
124 construction was inspired by the AMPD [28] and the ICD-11 beta draft capturing the full scope  
125 of self- and interpersonal functioning. To operationalize the LoPF-Q 12-18, all descriptors of the  
126 four AMPD domains were carefully analyzed and enriched with available concepts from child  
127 and adolescent psychology with focus on clinical validity. This led to a detailed structure for  
128 operationalizing the domains of functioning (see S1 Table), building the basis for a deductive  
129 item formulation. The derived item pool was then revised in an empirically informed iterative  
130 process to make them appropriate for a self-rating instrument for adolescents. Accordingly, the  
131 four resulting primary scales identity, self-direction, empathy, and intimacy are composed of two  
132 subscales per scale. These subscales are reported in addition to the total score and scale scores to  
133 support detailed clinical decision making. Because they represent less abstract and more  
134 commonly shared concepts (like e.g. Purposefulness or Prosociality), they may be helpful to  
135 better understand a patient's situation or to trace developments over time.

136 The process of test construction as well as psychometric properties have been described  
137 in detail in [22]. The main psychometric targets were clinical validity, good applicability for  
138 older and younger adolescents, and good scale reliabilities. The LoPF-Q 12-18 shows good scale  
139 reliability (Cronbach's alpha of .96 for the total scale, .92, .94, .87, and .92 for the primary scales

140 and between .76 and .96 for the subscales), good construct validity and substantial clinical  
141 validity. The LoPF-Q 12-18 total score distinguished between adolescents from the general  
142 population and  $n = 96$  SCID-II diagnosed PD patients at a highly significant level and with a  
143 large effect size of 2.1 standard deviations [21].

144 As all four dimensions of personality functioning were designed to build upon the joint  
145 construct of PD severity, and since the AMPD defines a current PD as the presence of  
146 impairments in two or more of them, scales were expected and found to be highly intercorrelated  
147 (Pearson correlation coefficients ranged between .41 and .83). Exploratory factor analysis on  
148 item level supported a one-factor solution (i.e., strong first factor and a ratio of first to second  
149 factors' eigenvalue of 5.1) speaking for a common factor of "personality pathology". This is  
150 intended and in line with the goal of creating an assessment of the generalised severity of  
151 personality pathology. However, all four domains of functioning had been operationalized  
152 independently and in careful contrast to each other to make sure that each domain only covers  
153 one of the described aspects of PD-related impairments with minimum overlap. Each item had to  
154 show: sufficient item-total correlation as part of the assigned a) subscale, b) primary scale, and c)  
155 total scale, respecting an internal consistent structure on all scale levels, and a reasonable effect  
156 size for discriminating the school population and the PD patient sample as a sign of clinical  
157 validity. Factor analytic approaches were not used to empirically select the final item set.  
158 However, in an exploratory factor analysis on item level, a model with four factors accounted for  
159 39.9% of the variance, and 72.2% of the items showed a loading  $> .30$  on the factor that  
160 corresponded to the theoretically assumed domain. This was interpreted as preliminary evidence  
161 for the appropriateness of using the four domain scores [22]. However, with a Turkish translation

162 of the LoPF-Q 12-18, a four-factor model did not show adequate fit in a CFA [26]. Therefore,  
163 the factor-analytical basis of the four domain scores has not yet been fully clarified.

### 164 **The Current Study**

165 The first goal of this study was an in-depth investigation of the factorial structure of the  
166 LoPF-Q 12-18 items. Based on the preliminary analyses reported above, we expected that the  
167 LoPF-Q 12-18 is essentially unidimensional, in the sense that most of the reliable variance of the  
168 total score is due to a general factor. This is in line with research showing that different measures  
169 of PD severity capture a strong common factor and can therefore be scaled along a single latent  
170 continuum [29]. Nevertheless, previous research also indicates that specific factors might still  
171 play a role even when a strong first factor is present [30–33]. Hence, Goth et al., (23, p. 687)  
172 hypothesized that a bifactor structure might be suitable for the LoPF-Q 12-18, taking into  
173 account a strong general factor as well as four empirically distinguishable domains.

174 Our second goal for the current study was to achieve a considerably shorter version  
175 which maintains the structure of the questionnaire in terms of the four domains and the high  
176 clinical validity of the original long version. With 97 5-point likert scale items, the LOPF-Q 12-  
177 18 can be considered a somewhat long measure, at least for many research and clinical  
178 applications with a focus on fast and efficient screening. Length can, therefore, be considered a  
179 barrier for its usage. For instance, individuals with mental health problems often present in non-  
180 specialised settings like primary care, school psychologist offices, or emergency departments  
181 [34]. A shorter version would allow for administration of the instrument in a resource saving  
182 manner. This is important, as the LoPF-Q 12-18 might not be the only instrument that needs to  
183 be administered at a certain time. A short version can reduce burden for the patients and,

184 additionally, it reduces resources required for the scoring of the questionnaire. Taken together,  
185 we expect that a short version will have a high impact on the practicability of the instrument.

## 186 **Materials and methods**

### 187 **Participants and procedures**

188 The current analyses were conducted on the same samples previously described in Goth  
189 et al. 2018 [22]. In short, a school sample of  $n = 351$  students was assessed at three public  
190 schools. The BPFSC-11 (Borderline Personality Features Scale for Children, 11 Item Version;  
191 [35]) was used to screen for the PD related health status,  $n = 337$  were below the Cut-Off  $\geq 34$   
192 and was taken as healthy control group. The study was reviewed and approved by the ethics  
193 committee "Ethikkommission Beider Basel" which is now "Ethikkommission Nordwest- und  
194 Zentralschweiz". Written informed consent has been obtained from all participants. A clinical  
195 sample of  $n = 415$  patients was recruited at inpatient and outpatient units of six child and  
196 adolescent psychiatric hospitals in Basel, Innsbruck, Berlin, Mainz, Idar-Oberstein, and  
197 Heidelberg. Inclusion criteria were age of 12 to 20 years, sufficient language and cognitive skills,  
198 no autistic disorder, and no current psychotic episodes. Diagnoses were based on the results of  
199 the clinical interviews Structured Clinical Interview for DSM-IV Axis II (SCID-II; [36]), the  
200 Children's Diagnostic Interview for Psychiatric Diseases (K-DIPS; [37]) and a classification  
201 conference. Patients with a PD diagnosis were assigned to the PD group independently from  
202 Axis I diagnoses. Of the total clinical sample,  $n = 96$  patients (23.1%) met the DSM-IV criteria  
203 of one or more PDs (44.8% BPD). The total sample of  $n = 766$  adolescents consists of 44.4%  
204 boys and 55.6% girls, the age range was 12-20 years ( $M = 15.5$ ,  $SD = 1.9$ ). For details, please  
205 see the description of the full study [22].

### 206 **Measures**

207           The LoPF-Q 12-18 [21] has been described above. It contains 97 items to be answered on  
208 a 5-point scale ranging from 0 (no), 1 (more no), 2 (part/part), 3 (more yes) to 4 (yes). The  
209 resulting four scales Identity, Self-Direction, Empathy, and Intimacy are coded towards  
210 pathology and add up to a total score Personality Functioning, ranging from no impairment to  
211 severe impairment. For descriptive reasons, two subscales per scale are included, matching  
212 classical psychological concepts to facilitate interpretation. The test is available on the self-  
213 publishing project website ([academic-tests.com](http://academic-tests.com)).

#### 214 **Investigation of the latent structure**

215           For the investigation of the latent structure of the LoPF-Q 12-18, confirmatory factor  
216 analyses on item level were used. Scale reliabilities were evaluated using McDonald's Omega.  
217 The analyses were conducted with the software 'R' [38] and the package 'lavaan' [39]. Fig 1  
218 illustrates models representing different factorial assumptions that were tested in order to  
219 compare their fit.

220

221           *--- Fig 1: Different configural assumptions tested for the long version ---*

222

223           The following fit indices are reported: Comparative Fit Index (CFI), Root Mean Squared Error of  
224 Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) [40]. The  
225 following combination of indices was used as the cut-off to determine acceptable models: CFI > .  
226 90 acceptable and CFI > .95 good, RMSEA < .05, SRMR < .08 [41,42]. The model fit of the short  
227 version created by ACO method (see below) was investigated using the same criteria. Scale  
228 reliabilities were estimated using the package 'semTools' [43]. We report an ordinal version of  
229 coefficient alpha according to Zumbo et. al [44], as well as omega hierarchical according to

230 McDonald [45]. Note that in bifactor models, omega hierarchical corresponds to OmegaH for the  
231 total score and to OmegaHS for the subscale scores [46].

232

### 233 **Ant Colony Optimization to create a short version**

234 For small item pools it can be an option to iterate through all possible item combinations  
235 or to apply simple iterative methods, e.g. a Stepwise Confirmatory Factor Analysis Approach  
236 (SCOFA) in order to find a well-suited combination of items that can be used as a short version  
237 of a test [47]. Considering that the LoPF-Q 12-18 consists of 97 items, the number of possible  
238 item combinations reaches a level where this is no longer possible. To illustrate that such an  
239 attempt is impossible, we calculated the number of possible combinations based on “n over k”,  
240 as suggested previously [48]. The S2 Table illustrates the estimated computation time, memory,  
241 and amount of energy that would be required to execute these iterations. Results show that the  
242 algorithm would need to run for billions of years while using multiple times the global estimated  
243 yearly energy consumption. Consequently, it is inevitable to run an optimization algorithm which  
244 approximates a close-to-optimal short version of the questionnaire in a shorter period of time.

245 The Ant Colony Optimization (ACO) meta-heuristic [49] was used to select a set of items  
246 for the short version. For our use case, this algorithm ran for less than 5 hours CPU time (see S2  
247 Table). The ACO consists of virtual ‘ants’ who explore the selection of sets of items and  
248 attribute a ‘pheromone’ level to the items of the selection according to a statistical criterion  
249 (defined below). Items with a higher pheromone level have a higher probability to be re-selected  
250 in future item sets to be explored. The pheromone level fades over time (“evaporation”). The  
251 ACO was started with 30 ants, and the algorithm stopped after 20 runs without improvement.

252 This method has already proven useful in the construction of the 34-item “Personality Inventory  
253 for DSM-5, Brief Form Plus” (PID5BF+) [50].

254 Our goal was to generate a short version of the LoPF-Q 12-18 with a total of 20 items,  
255 including 5 items for each domain (Identity, Self-direction, Intimacy, Empathy). For this purpose,  
256 the ACO method was set up to select a subset of 5 items from each domain of the original  
257 version. The criterion to calculate the pheromone used by the ants was a combination of model  
258 fit, reliabilities of the domain scales and clinical validity i.e. the capacity to discriminate between  
259 patients with personality disorder (n = 96) and students without signs of personality disorder  
260 according to the BPFSC-11 (n = 337). Model fit was based on a confirmatory factor analysis  
261 with 4 first order factors to represent the domains and a secondary higher order factor to  
262 represent generalised severity. The loadings of the domains on the higher order factor were  
263 constrained to be equal, thereby ensuring a balanced interpretation of the general severity  
264 continuum. Pheromones were calculated based on logistic transformations ( $\phi$ ) of fit measures  
265 (CFI, RMSEA), measures of reliability (McDonald’s Omega and minimum factor loading) and  
266 criterion validity (adjusted R<sup>2</sup>). The ultimately optimised (i.e. maximised) pheromone was based  
267 on the sum of all three  $\phi$ -values. Please refer to S3 Materials for the formulas used for  
268 calculating pheromone levels.

269 Finally, to show the advantage of the applied ACO-algorithm over the iterative approach,  
270 we compared reliability, CFA model fit as well as criterion validity with 100,000 random  
271 combinations of items. Due to the high computational load, the calculation of these 100,000  
272 models was performed at the sciCORE (<http://scicore.unibas.ch/>), the scientific computing centre  
273 of the University of Basel. All analyses have been conducted using R (> version 4.0.1), as well  
274 as the R packages ‘lavaan’ [39] and ‘semTools’ [43] for Confirmatory Factor Analyses. Receiver

275 operating characteristic (ROC) analysis was used to determine the clinical utility of the LoPF-Q  
276 short version and to derive empirical cut-off scores for defining clinically relevant thresholds.

## 277 **Results**

### 278 **Investigation of the factorial structure of the original (97-item) version**

279 Table 1 shows the parameters of different confirmatory factor analyses (CFA) testing  
280 different factorial assumptions.

281

282 Table 1: *Confirmatory Factor Analyses (CFA) testing different factorial assumptions (long*  
283 *version)*

Model (id)	Factors	par	$\chi^2$	CFI	RMSEA	SRMR
1-dim (1)	1	485	12341.2	0.851	0.057	0.090
2-dim (2)	2	486	11584.6	0.865	0.054	0.087
4-dim (3)	4	491	11019.2	0.876	0.052	0.084
2+bifactor (4)	2	582	9286.3	0.907	0.046	0.070*
4+bifactor (5)	4	582	9099.5*	0.911*	0.045*	0.072
2-dim hierarchical (6a) <sup>i</sup>	(2)	(487)	(21882.8)	(0.667)	(0.085)	(0.087)
2-dim hierarchical (6b)	2	486	11584.6	0.865	0.054	0.087
4-dim hierarchical (7a)	4	489	11089.0	0.875	0.052	0.085
4-dim hierarchical (7b)	4	486	11397.4	0.869	0.054	0.086

284 Note. *npar*: number of estimated parameters; *df*: degrees of freedom; *CFI*: Comparative Fit

285 *Index*; *RMSEA*: Root Mean Squared Error of Approximation; *SRMR*: Standard Root Mean

286 *Residual*. Best fit indices are highlighted with asterisks. <sup>i</sup> For model 6a standard errors could not

287 be computed and the information matrix could not be inverted. This may be a symptom of a non-

288 identifiable model. Therefore, the parameters of this model are shown in parenthesis.

289

290  
 291 All factorial assumptions are related to the basic personality functioning concept, highlighting  
 292 either the joint construct of PD severity, the two areas of Self-related and Interpersonal  
 293 functioning, or the four domains Identity, Self-Direction, Empathy, and Intimacy according to  
 294 the AMPD. Overall, the bifactor models performed slightly better than all other correlated or  
 295 hierarchical factor models. A four-dimensional bifactor was the only model to show acceptable  
 296 fit based on all three fit measures, RMSEA (<.05) and SRMR (<.08) and CFI (>.90). The two  
 297 best fitting bifactor models (“two-dimensional bifactor” and “four-dimensional bifactor”)  
 298 showed very similar fit indices with only a subtle difference on RMSEA. Table 2 summarizes the  
 299 model-based scale reliabilities. Based on the best fitting model (model 5), ordinal alpha was  
 300 excellent for the general factor as well as all four domains (>.90). However, while OmegaH  
 301 showed excellent reliability of the total score (.94), OmegaHS was substantially lower for the  
 302 four domains scales (.07, .11, .50, .20) (see S4 Tables – sheet 1 for additional details). Factor  
 303 loadings of model 5 indicated that several items from the domain of empathy did not  
 304 substantially (> .30) load on the general factor, and several items from the domain of identity had  
 305 even negative loadings on the respective specific factor (see S4 Tables – sheet 2). S4 Tables –  
 306 sheet 3 shows the factor intercorrelations for the bi-factor models.

307  
 308 Table 2: *Factor reliabilities for the long and short version*

	Long version (models)									Short Version <sup>i</sup>
	1	2	3	4	5 <sup>ii</sup>	6a	6b	7a	7b	
<i>Ordinal Alpha</i>										
Total score	0.98	0.98	0.98	0.97	0.98	0.98	0.97	0.97	0.97	.91
Identity		0.97	0.94	0.97	0.94	0.97	0.97	0.94	0.94	.80
Self-direction			0.96		0.96			0.96	0.96	.84

	Long version (models)								Short	
Empathy	0.95	0.90	0.95	0.90	0.95	0.95	0.90	0.90	.71	
Intimacy		0.94		0.94			0.94	0.94	.78	
<i>Hierarchical Omega</i>										
Total score	1.00	1.00	1.00	0.99	0.94	1.00	1.00	1.00	1.00	.93
Identity		0.99	0.96	0.17	0.07	0.99	0.99	0.96	0.99	.78
Self-direction			0.97		0.11			0.97	0.99	.84
Empathy		0.92	0.77	0.22	0.50	0.92	0.92	0.77	0.61	.68
Intimacy			0.95		0.20			0.95	0.93	.75

309 *Note. Alpha = ordinal alpha; Omega = hierarchical Omega*

310 <sup>i</sup> *the optimized short version corresponds to the factorial assumptions of model 7b.*

311 <sup>ii</sup> *best fitting model in Confirmatory Factor Analysis (see Table 1)*

312

313 Taken together, this suggests that although model 5 has the best fit, it is not a very satisfying  
 314 representation of the structure of the LoPF-Q 12-18 [51]. In consequence, we have chosen a  
 315 hierarchical model with four lower-order factors (i.e., model 7b) for developing the short version.

316

### 317 **Creating a short version**

318 As intended, a 20-item version with 5 items for each of the domain scales was obtained.

319 Fit indices presented in Table 3 show that the optimised short version had a very good fit on all  
 320 fit indices (CFI = .980, RMSEA = .046, SRMR = .038).

321

322 *Table 3: Model fit indices and external validity for the Ant Colony Optimised Short version*  
 323 *compared to 100,000 random combinations of items*

324

Model (id)	Factors	par	$\chi^2$	CFI	RMSEA	SRMR	Adj. R <sup>2</sup>
ACO short version <sup>a</sup>	101	169	252.3*	0.980*	0.046*	0.038*	0.425*
100,000 random combinations <sup>a</sup>	101	169	716.7	0.918	0.080	0.067	0.389
				(0.02)	(0.01)	(0.01)	(0.02)

325 Note. *npar*: number of estimated parameters; *df*: degrees of freedom; *CFI*: Comparative Fit  
326 Index; *RMSEA*: Root Mean Squared Error of Approximation; *SRMR*: Standard Root Mean  
327 Residual. Best fit indices are highlighted with asterisks. *Adj. R<sup>2</sup>* = external validity (variance  
328 explained). <sup>a</sup> modelling corresponds to the factorial assumptions of model 7b.

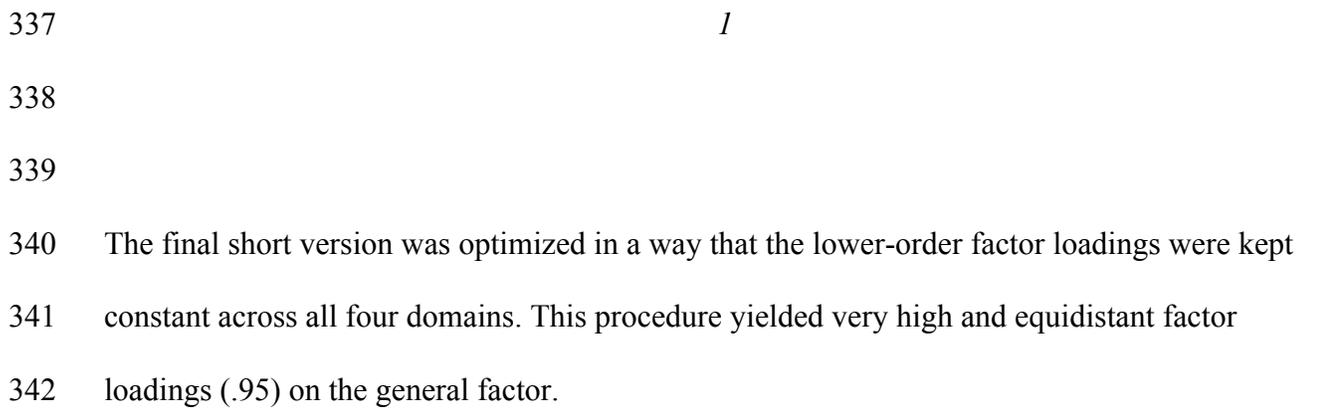
329

330 Due to the smaller number of items, ordinal alpha was slightly lower in comparison to the long  
331 version (.91 total scale, .71 - .84 domains). Omega hierarchical of the total score was similar  
332 (.93). Factor loadings of the optimised short version are depicted in Fig 2.

333

334 --- Fig 2: Factor loadings of short version ---

335 Fig 2 - legend: Optimized model using ant colony optimization to develop a short version to  
336 identify personality disorders in adolescence. The configuration corresponds to model 7b in Fig

337   
338 1

339

340

341 The final short version was optimized in a way that the lower-order factor loadings were kept  
342 constant across all four domains. This procedure yielded very high and equidistant factor  
343 loadings (.95) on the general factor.

344

345 For comparison, Table 2 additionally shows the average model fits (and standard  
346 deviations) for 100,000 randomly selected item combinations testing the same hierarchical factor

345 model. Fig 3 visually compares the 20-item solution that was generated with the ACO with  
346 100'000 random combinations of items regarding external validity and model fit. Compared to  
347 the random combinations, the combination of model fit and external validity (the ability to  
348 differentiate between healthy controls and PD patients) of the short ACO version are excellent  
349 with an adjusted R square of .425 (i.e. 42,5% explained variance).

350

351 *--- Fig 3: Fit and external validity of short version ---*

352 *Fig 3 - legend: Model fit and external validity of the optimized short version in comparison to*  
353 *100,000 random item combinations.*

354

355

356 Expressed with the more traditional effect size Cohens d, the LoPF-Q Short total score differs  
357 between the PD patients and the healthy controls with  $d = 3.1$  standard deviations. ROC analysis  
358 showed an area under the curve (AUC) of .92 ( $p < .001$ ; 95 % confidence interval .89-.95). A  
359 preliminary cut-off score for the LoPF-Q Short total score was defined to be  $\geq 36$  using Youden  
360 index, corresponding to a T-score of 74. Specificity for detecting patients with personality  
361 disorder compared to healthy controls was 87.5% and sensitivity was 80.2%. Reliability  
362 coefficients for the short version can be seen in Table 2.

363

364

## Discussion

365 The current study had two aims: First, to investigate the latent structure of the LoPF-Q  
366 12-18 original 97-item version. We assumed that a bifactor model with a general factor –  
367 representing a joint construct of PD severity – with four specific factors matching the four

368 domains of functioning according the AMPD (DSM-5) would perform well. The second goal  
369 was the construction of an optimised short version to meet the needs for an efficient screening  
370 instrument for PD in adolescents.

### 371 **Factorial structure of LoPF-Q 12-18 long version**

372 As hypothesised, the nominally best fitting models were bifactor models when compared  
373 to correlated factor or hierarchical factor models (see Fig 1). The model fit of both bifactor  
374 models (including either two broad dimensions or four narrow domains) was acceptable when  
375 considering all evaluated criteria. However, other aspects besides model fit should be considered  
376 when interpreting bifactor models [51] due i.a. to their less restrictive nature which results in a  
377 higher overall chance of good fit, even when using random data. In addition to an acceptable fit,  
378 the items should show substantial loadings on the general factor, and the specific scales should  
379 have sufficient reliability after controlling for variance of the general factor (i.e., Omega HS). In  
380 both these respects the estimates of the bi-factor models were lacking. Conclusively, bi-factor  
381 models did not satisfactorily represent the structure of the questionnaire despite the acceptable fit.  
382 Nevertheless, the following important conclusions can be drawn from the performed analyses.  
383 First, as expected, the item level data collected with the LoPF-Q 12-18 contain a very strong  
384 general factor. This can be seen, for example, in the fact that model fit was only moderately  
385 improved by extracting more than one factor, that the four domains in model 3 were very highly  
386 correlated (S4 Tables – sheet 3), and that the reliable variance in the total score in model 5 (i.e.,  
387 Omega H) was almost entirely attributable to the general factor (S4 Tables – sheet 1). This  
388 support for a general factor of personality functioning is very much in line with the usage of the  
389 LoPF-Q 12-18 in the framework of diagnostic procedures of both the AMPD (DSM-5) and the  
390 ICD-11. In both diagnostic models, personality functioning is seen as an overarching construct

391 important to establish a PD diagnosis and to judge its severity. Importantly, according to [52],  
392 the general factor of personality pathology has been primarily described in adult populations.  
393 The current study might be one of the first to describe this general factor in a sample of younger  
394 patients.

395                 Second, the four domain subscales, with the possible exception of empathy,  
396 contain hardly any reliable variance beyond general severity. In other words: Although the four  
397 domain scores were reliable in their own right (i.e., ordinal alpha > .90), their very high  
398 correlation in the underlying sample makes it seem unlikely that distinctive and clinically  
399 interpretable profiles will emerge in individual cases. This contrasts with PD criteria from DSM-  
400 IV [32,53] or items of the Inventory of Personality Organization [31], which tend to warrant  
401 scoring of subscales in adult samples. At least on a group level, [22] found first evidence of  
402 distinctive profiles, for example, the empathy scale was severely impaired only in patients with  
403 narcissistic and antisocial PD, whereas the identity scale was particularly impaired in patients  
404 with Borderline or anxious-avoidant PD. The specific clinical variation of the empathy scale may  
405 be an explanation for why only this one showed an independent variance beyond the general  
406 factor. In sum, whether the use of each of the four domain scores is clinically meaningful needs  
407 to be investigated in clinical trials with different types of PD patients and optimally with  
408 different therapeutic approaches in a longitudinal design.

409                 The debate on the meaningfulness of the domain scales is important as mental health care  
410 workers tend to find the primary scales and subscales of the LoPF-Q 12-18 useful for the  
411 interpretation of the assessments regarding clinical decision making and therapy planning. This  
412 is comprehensible as the less abstract denomination of the subscales appear to be closer to  
413 commonly shared concepts and can be used to find a shared language with the patients and their

414 families. According to the authors [22] the LoPF-Q 12 -18 has been primarily developed to meet  
415 the needs of clinical practitioners and to cover a wide range of symptoms related to the four  
416 domains, because often specific aspects of functioning (identity pathology, problems with self-  
417 regulation or problems with social interaction etc.) are the primary target for psychotherapy. This  
418 discrepancy between the authors' experiences and intentions and our current findings cannot be  
419 conclusively clarified. The currently investigated sample consisted mainly of subjects without  
420 signs of PD (351 from schools and 319 patients without PD vs 96 patients with PD). The general  
421 factor might turn out being less pronounced and the domain scales more independent from each  
422 other when investigating clinical samples of PD patients [54]. Similarly, Watts et al. found that  
423 the inclusion of undiagnosed individuals causes more positive correlations in psychopathological  
424 data, leading to a stronger  $p$ -factor [55]. For a further optimization of the structure in a short  
425 form of the LoPF-Q 12-18, it seemed reasonable to keep the four domains in terms of content  
426 validity, but to put the focus of the optimization on the general factor.

#### 427 **LoPF-Q Screener (20-item version)**

428 The short version was optimised for clinical validity and internal consistency accounting  
429 for a structure with four first order factors to represent the personality functioning domains and a  
430 secondary higher order factor to represent the general personality functioning denoting PD  
431 severity. Thanks to the optimisation, the short version performed excellently regarding both  
432 external validity and internal consistency. The optimisation was done with the ACO heuristic  
433 which had already proven useful in previous studies for creating short personality assessments  
434 [31,50] and performed very well in the current study (see Fig 3). The derived short version  
435 "LoPF-Q Screener" contains 20 items and preserves the four scales Identity, Self-direction,  
436 Empathy and Intimacy as well as the total scale Personality Functioning. It showed an excellent

437 model fit concerning all parameters and good scale reliabilities. Most importantly, it showed  
438 excellent clinical validity, with the total scale differentiating significantly and with an effect size  
439 of 3.1 standard deviations between PD patients and healthy adolescents.

440         The LoPF-Q Screener can be used in contexts where employing the longer version is not  
441 feasible or inconvenient. This flexibility cannot be overestimated in the presence of a general  
442 global mental health gap [56] in adolescents and a specific gap regarding personality disorders in  
443 youth [6,7]. Tools that can help address these gaps are required, and while diagnostic tools  
444 cannot solve this issue alone, they are one of the cornerstones to advance research and  
445 interventions. The results on psychometric properties of this short version are still preliminary  
446 and need to be verified with test data that were not used for its construction. The data ideally  
447 needs to be collected with this short version in order to validate it, since using a subsample of  
448 items of data collected with the long version might potentially introduce bias (e.g. memory  
449 effects, effects of the sequential order of items, attention span of the subject etc.). Finally, the  
450 question arises whether an even shorter version wouldn't be better in terms of practicality of the  
451 assessment and, thus, versatility in clinical contexts. However, an even shorter version may come  
452 at the expense of inferior measurement precision and diagnostic validity, both of which are  
453 highly relevant for clinical usage [57]. The 20-item version of the LoPF Q 12-18 is likely to  
454 present a solid compromise between psychometric precision and practicality.

#### 455 **Research recommendations**

456         Research on the usefulness of the levels of personality functioning model for clinical  
457 decision making such as selection of appropriate treatment and treatment customisation is needed.  
458 The long and short version need to be compared in future studies regarding their usability and  
459 user experience of the different stakeholders. For instance, do users benefit from the more

460 comprehensive data collection of the long version or are they looking for more efficient tools? A  
461 further question is the preparation of a pathway towards shorter versions for different cultural  
462 settings. The authors of the LoPF-Q 12-18 pursue a strategy in which they emphasise the  
463 importance of the same set of items for all cultural settings and actively support the development  
464 of cultural adaptations and networking among interested colleagues. A shared set of items across  
465 culturally adapted versions is necessary because it facilitates scientific exchange and  
466 management of the different versions and enables joint data analyses in cross-cultural settings.  
467 This possibility is particularly important because the development of PD in early adolescence is  
468 an under-researched area and data pooling is key. In addition, LoPF-Q versions for informant  
469 report and for even younger age groups (from 6 years up) are under development, and the  
470 seamless and clear transferability of the assessed scales in all cultural adaptations is crucial,  
471 especially for longitudinal studies. Future research will show whether the optimised short version  
472 LoPF-Q Screener will provide measurement invariance across different cultural settings and  
473 translations.

474         The current study highlights the usefulness of a more detailed and more time-efficient  
475 assessment of personality functioning in adolescence. Whereas there is no doubt about a  
476 common core, i.e. a general latent construct, there is somewhat mixed evidence regarding the  
477 usefulness of the lower-order domains (identity, self-direction, empathy, intimacy). Earlier  
478 research on alcohol use disorders has shown that determining the factor structure in a sample  
479 including individuals with no clinical symptoms may have a debilitating impact on the  
480 discrimination of sub-factors [54]. Future research on the LoPF-Q 12-18 and the introduced  
481 LoPF-Q Screener short version might provide more comprehensive insights by comparing the  
482 factor structure between clinical and non-clinical samples.

484

## Literature

- 485 1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders.  
486 5th ed. Arlington (USA): American Psychiatric Association; 2013.
- 487 2. World Health Organization. International Statistical Classification of Diseases and Related  
488 Health Problems. 11th ed. 2019.
- 489 3. Hengartner MP, Zimmermann J, Wright AGC. Personality pathology. In: Virgil Zeigler-  
490 Hill, Shackelford TK, editors. The SAGE Handbook of Personality and Individual  
491 Differences: Volume III: Applications of Personality and Individual Differences. SAGE;  
492 2018.
- 493 4. HOPWOOD CJ, KOTOV R, KRUEGER RF, WATSON D, WIDIGER TA, ALTHOFF  
494 RR, et al. The time has come for dimensional personality disorder diagnosis. *Personal Ment*  
495 *Health*. 2018;12: 82–86. doi:10.1002/pmh.1408
- 496 5. Zimmermann J, Kerber A, Rek K, Hopwood CJ, Krueger RF. A Brief but Comprehensive  
497 Review of Research on the Alternative DSM-5 Model for Personality Disorders. *Curr*  
498 *Psychiatry Rep*. 2019;21: 92. doi:10.1007/s11920-019-1079-z
- 499 6. Chanen AM, Sharp C, Hoffman P, Global Alliance for Prevention and Early Intervention  
500 for Borderline Personality Disorder. Prevention and early intervention for borderline  
501 personality disorder: A novel public health priority. *World Psychiatry*. 2017;16: 215–216.  
502 doi:10.1002/wps.20429
- 503 7. Winsper C, Bilgin A, Thompson A, Marwaha S, Chanen AM, Singh SP, et al. The  
504 prevalence of personality disorders in the community: a global systematic review and meta-  
505 analysis. *The British Journal of Psychiatry*. 2020;216: 69–78. doi:10.1192/bjp.2019.166
- 506 8. Dahl RE, Allen NB, Wilbrecht L, Suleiman AB. Importance of investing in adolescence  
507 from a developmental science perspective. *Nature*. 2018;554: 441–450.  
508 doi:10.1038/nature25770
- 509 9. Kaess M, Brunner R, Chanen A. Borderline Personality Disorder in Adolescence.  
510 *Pediatrics*. 2014;134: 782–793. doi:10.1542/peds.2013-3677
- 511 10. Kramer U, Temes CM, Magni LR, Fitzmaurice GM, Aguirre BA, Goodman M, et al.  
512 Psychosocial functioning in adolescents with and without borderline personality disorder.  
513 *Personality and Mental Health*. 2017;11: 164–170. doi:10.1002/pmh.1377
- 514 11. Tyrer P, Tyrer H, Yang M. Premature mortality of people with personality disorder in the  
515 Nottingham Study of Neurotic Disorder. *Personality and Mental Health*. 2021;15: 32–39.  
516 doi:10.1002/pmh.1466
- 517 12. Bloom DE, Cafiero E, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, et al. The  
518 global economic burden of noncommunicable diseases. PGDA Working Papers. Geneva:

- 519 Program on the Global Demography of Aging; 2012 Jan. Report No.: 8712. Available:  
520 <https://ideas.repec.org/p/gdm/wpaper/8712.html>
- 521 13. Hastrup LH, Jennum P, Ibsen R, Kjellberg J, Simonsen E. Societal costs of Borderline  
522 Personality Disorders: a matched-controlled nationwide study of patients and spouses. *Acta*  
523 *Psychiatrica Scandinavica*. 2019;140: 458–467. doi:10.1111/acps.13094
- 524 14. Fleischhaker C, Böhme R, Sixt B, Brück C, Schneider C, Schulz E. Dialectical Behavioral  
525 Therapy for Adolescents (DBT-A): a clinical Trial for Patients with suicidal and self-  
526 injurious Behavior and Borderline Symptoms with a one-year Follow-up. *Child and*  
527 *Adolescent Psychiatry and Mental Health*. 2011;5: 3. doi:10.1186/1753-2000-5-3
- 528 15. Foelsch PA, Schlüter-Müller S, Odom AE, Arena HT, H AB, Schmeck K. *Adolescent*  
529 *Identity Treatment: An Integrative Approach for Personality Pathology*. Springer  
530 International Publishing; 2014. Available:  
531 <https://books.google.ch/books?id=DTkqBAAAQBAJ>
- 532 16. Loose C. Die Schematherapie bei adoleszenten Patienten mit Borderline-Störung. *PTT:*  
533 *Persönlichkeitsstörungen Theorie und Therapie*. 2015;19: 44–54.
- 534 17. Rossouw TI, Fonagy P. Mentalization-Based Treatment for Self-Harm in Adolescents: A  
535 Randomized Controlled Trial. *J Am Acad Child Adol Psychiatry*. 2012;51: 1304-1313.e3.  
536 doi:10.1016/j.jaac.2012.09.018
- 537 18. Reed GM. Progress in developing a classification of personality disorders for ICD-11.  
538 *World Psychiatry*. 2018;17: 227–229. doi:10.1002/wps.20533
- 539 19. Birkhölzer M, Schmeck K, Goth K. Assessment of Criterion A. *Current Opinion in*  
540 *Psychology*. 2021;37: 98–103. doi:10.1016/j.copsyc.2020.09.009
- 541 20. Bach B, First MB. Application of the ICD-11 classification of personality disorders. *BMC*  
542 *Psychiatry*. 2018;18: 351. doi:10.1186/s12888-018-1908-3
- 543 21. Kirstin Goth, Marc Birkhölzer, Klaus Schmeck. LoPF-Q 12-18 (Levels of Personality  
544 Functioning Questionnaire) German Version: A self-report questionnaire for measuring  
545 personality functioning in adolescence - Short manual. Offenbach: academic-tests; 2018.
- 546 22. Goth K, Birkhölzer M, Schmeck K. Assessment of Personality Functioning in Adolescents  
547 With the LoPF-Q 12–18 Self-Report Questionnaire. *Journal of Personality Assessment*.  
548 2018;100: 680–690. doi:10.1080/00223891.2018.1489258
- 549 23. Sharp C, Vanwoerden S. Culture-adapted version English USA of the self-report  
550 questionnaire LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire; authors  
551 Goth & Schmeck) - Short manual. academic-tests. 2018. Available: [https://academic-](https://academic-tests.com)  
552 [tests.com](https://academic-tests.com)

- 553 24. Kassin M, Hackradt J. Culture-adapted version Spanish Mexico of the self-report  
554 questionnaire LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire; authors  
555 Goth & Schmeck) - Short manual. academic-tests. Available: <https://academic-tests.com/>
- 556 25. Cosgun S, Cakiroglu S. Culture-adapted version Turkish of the self-report questionnaire  
557 LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire; authors Goth & Schmeck)  
558 - Short manual. academic-tests. Available: <https://academic-tests.com/>
- 559 26. Cosgun S, Goth K, Cakiroglu S. Levels of Personality Functioning Questionnaire (LoPF-Q)  
560 12–18 Turkish Version: Reliability, Validity, Factor Structure and Relationship with  
561 Comorbid Psychopathology in a Turkish Adolescent Sample. *J Psychopathol Behav Assess.*  
562 2021;43: 620–631. doi:10.1007/s10862-021-09867-2
- 563 27. Barkauskienė R, Skabeikytė G. Culture-adapted version Lithuanian of the self-report  
564 questionnaire LoPF-Q 12-18 (Levels of Personality Functioning Questionnaire; authors  
565 Goth & Schmeck) - Short manual. academic-tests. Available: <https://academic-tests.com/>
- 566 28. Bender DS, Morey LC, Skodol AE. Toward a Model for Assessing Level of Personality  
567 Functioning in DSM–5, Part I: A Review of Theory and Methods. *Journal of Personality*  
568 *Assessment.* 2011;93: 332–346. doi:10.1080/00223891.2011.583808
- 569 29. Zimmermann J, Müller S, Bach B, Hutsebaut J, Hummelen B, Fischer F. A Common  
570 Metric for Self-Reported Severity of Personality Disorder. *PSP.* 2020;53: 168–178.  
571 doi:10.1159/000507377
- 572 30. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p  
573 factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical*  
574 *Psychological Science.* 2014;2: 119–137.
- 575 31. Hörz-Sagstetter S, Volkert J, Rentrop M, Benecke C, Gremaud-Heitz DJ, Unterrainer H-F,  
576 et al. A Bifactor Model of Personality Organization. *Journal of Personality Assessment.*  
577 2021;103: 149–160. doi:10.1080/00223891.2019.1705463
- 578 32. Sharp C, Wright AGC, Fowler JC, Frueh BC, Allen JG, Oldham J, et al. The structure of  
579 personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal*  
580 *Psychology.* 2015;124: 387–398. doi:10.1037/abn0000033
- 581 33. Sharp C, Wall K. DSM-5 Level of Personality Functioning: Refocusing Personality  
582 Disorder on What It Means to Be Human. *Annu Rev Clin Psychol.* 2021;17: 313–337.  
583 doi:10.1146/annurev-clinpsy-081219-105402
- 584 34. Costello EJ, He J, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with  
585 psychiatric disorders: 12-month data from the national comorbidity survey–adolescent. *PS.*  
586 2014;65: 359–366. doi:10.1176/appi.ps.201100518
- 587 35. Sharp C, Steinberg L, Temple J, Newlin E. An 11-item measure to assess borderline traits  
588 in adolescents: refinement of the BPFSC using IRT. *Personal Disord.* 2014;5: 70–78.  
589 doi:10.1037/per0000057

- 590 36. Wittchen H-U, Zaudig M, Fydrich T. Strukturiertes Klinisches Interview für DSM-IV.  
591 Göttingen: Hogrefe. 1997.
- 592 37. Schneider S, Unnewehr S, Margraf, J., editors. Diagnostisches Interview bei psychischen  
593 Störungen im Kindesund Jugendalter (Kinder-DIPS). 2. aktualisierte und erweiterte Aufl.  
594 Heidelberg: Springer; 2009.
- 595 38. R Core Team. R: A Language and Environment for Statistical Computing. Vienna (AU): R  
596 Foundation for Statistical Computing; 2020. Available: <https://www.R-project.org/>
- 597 39. Rosseel Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical*  
598 *Software*. 2012;48: 1–36.
- 599 40. Kline RB. *Principles and Practice of Structural Equation Modeling, Fourth Edition*.  
600 Guilford Publications; 2015.
- 601 41. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis:  
602 Conventional criteria versus new alternatives. *Structural Equation Modeling: A*  
603 *Multidisciplinary Journal*. 1999;6: 1–55. doi:10.1080/10705519909540118
- 604 42. Fan X, Thompson B, Wang L. Effects of sample size, estimation methods, and model  
605 specification on structural equation modeling fit indexes. *Structural Equation Modeling: A*  
606 *Multidisciplinary Journal*. 1999;6: 56–83. doi:10.1080/10705519909540119
- 607 43. Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y. semTools: Useful tools for  
608 structural equation modeling. 2021. Available: [https://CRAN.R-](https://CRAN.R-project.org/package=semTools)  
609 [project.org/package=semTools](https://CRAN.R-project.org/package=semTools)
- 610 44. Zumbo B, Gadermann A, Zeisser C. Ordinal Versions of Coefficients Alpha and Theta for  
611 Likert Rating Scales. *Journal of Modern Applied Statistical Methods*. 2007;6.  
612 doi:10.22237/jmasm/1177992180
- 613 45. McDonald RP. *Test Theory: A Unified Treatment*. New York: Psychology Press; 1999.  
614 doi:10.4324/9781410601087
- 615 46. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and  
616 interpreting statistical indices. *Psychol Methods*. 2016;21: 137–150.  
617 doi:10.1037/met0000045
- 618 47. Fung S, Chow EO, Cheung C. Development and Evaluation of the Psychometric Properties  
619 of a Brief Wisdom Development Scale. *International Journal of Environmental Research*  
620 *and Public Health*. 2020;17: 2717. doi:10.3390/ijerph17082717
- 621 48. Schroeders U, Wilhelm O, Oлару G. Meta-Heuristics in Short Scale Construction: Ant  
622 Colony Optimization and Genetic Algorithm. *PLOS ONE*. 2016;11: e0167110.  
623 doi:10.1371/journal.pone.0167110

- 624 49. Leite WL, Huang I-C, Marcoulides GA. Item Selection for the Development of Short  
625 Forms of Scales Using an Ant Colony Optimization Algorithm. *Multivariate Behav Res.*  
626 2008;43: 411–431. doi:10.1080/00273170802285743
- 627 50. Kerber A, Schultze M, Müller S, Rühling RM, Wright AGC, Spitzer C, et al. Development  
628 of a Short and ICD-11 Compatible Measure for DSM-5 Maladaptive Personality Traits  
629 Using Ant Colony Optimization Algorithms. *Assessment.* 2020; 1073191120971848.  
630 doi:10.1177/1073191120971848
- 631 51. Watts AL, Poore HE, Waldman ID. Riskier Tests of the Validity of the Bifactor Model of  
632 Psychopathology. *Clinical Psychological Science.* 2019;7: 1285–1303.  
633 doi:10.1177/2167702619855035
- 634 52. Shields AN, Giljen M, España RA, Tackett JL. The p factor and dimensional structural  
635 models of youth personality pathology and psychopathology. *Current Opinion in*  
636 *Psychology.* 2021;37: 21–25. doi:10.1016/j.copsyc.2020.06.005
- 637 53. Paap MCS, Braeken J, Pedersen G, Urnes Ø, Karterud S, Wilberg T, et al. A psychometric  
638 evaluation of the DSM-IV criteria for antisocial personality disorder: Dimensionality, local  
639 reliability, and differential item functioning across gender. *Assessment.* 2020;27: 89–101.
- 640 54. Steppan M, Piontek D, Kraus L. The effect of sample selection on the distinction between  
641 alcohol abuse and dependence. *International Journal of Alcohol and Drug Research.*  
642 2014;3: 159–168. doi:10.7895/ijadr.v3i2.100
- 643 55. Watts AL, Meyer FA, Greene AL, Wood PK, Trull TJ, Steinley D, et al. Spurious empirical  
644 support for the p-factor arises with the inclusion of undiagnosed cases. 2021.
- 645 56. Keynejad R, Spagnolo J, Thornicroft G. WHO mental health gap action programme  
646 (mhGAP) intervention guide: updated systematic review on evidence and impact.  
647 *Evidence-Based Mental Health.* 2021;24: 124–130. doi:10.1136/ebmental-2021-300254
- 648 57. Kemper CJ, Trapp S, Kathmann N, Samuel DB, Ziegler M. Short Versus Long Scales in  
649 Clinical Assessment: Exploring the Trade-Off Between Resources Saved and Psychometric  
650 Quality Lost Using Two Measures of Obsessive–Compulsive Symptoms. *Assessment.*  
651 2019;26: 767–782. doi:10.1177/1073191118810057

652

653

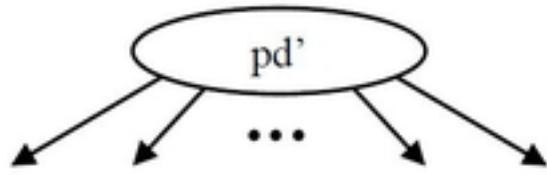
654

655

656

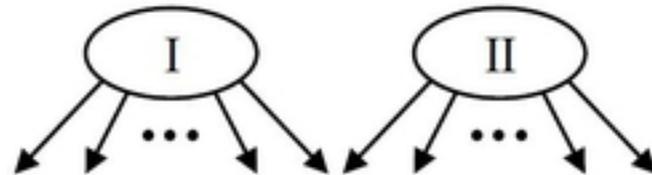
**Model 1**

**1-dimensional**



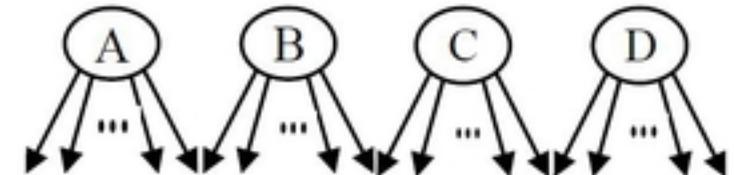
**Model 2**

**2-dimensional**  
(factors allowed to correlate)



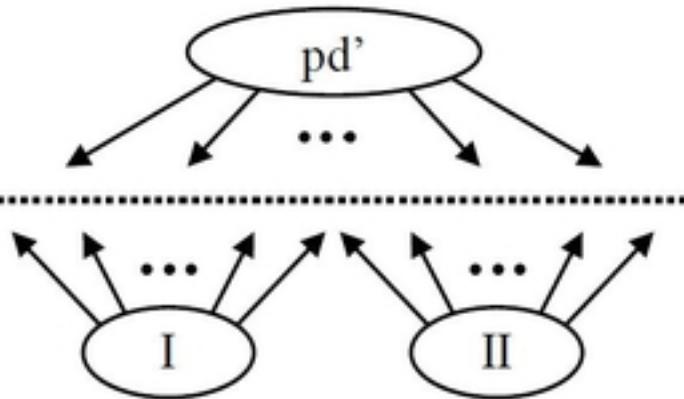
**Model 3**

**4-dimensional**  
(factors allowed to correlate)



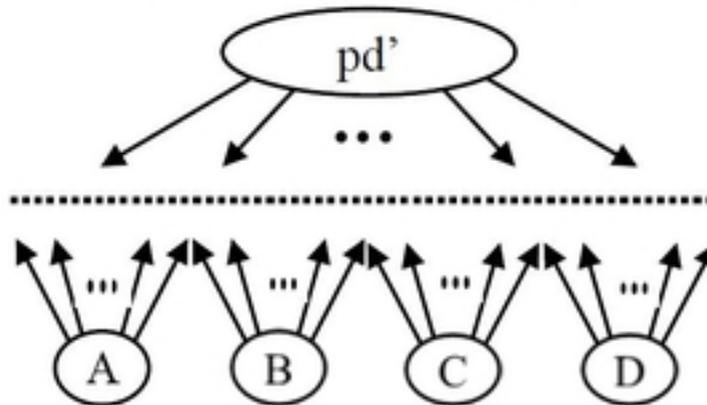
**Model 4**

**2-dimensional + bifactor**  
(all factors orthogonal)



**Model 5**

**4-dimensional + bifactor**  
(all factors orthogonal)



**Factors:**

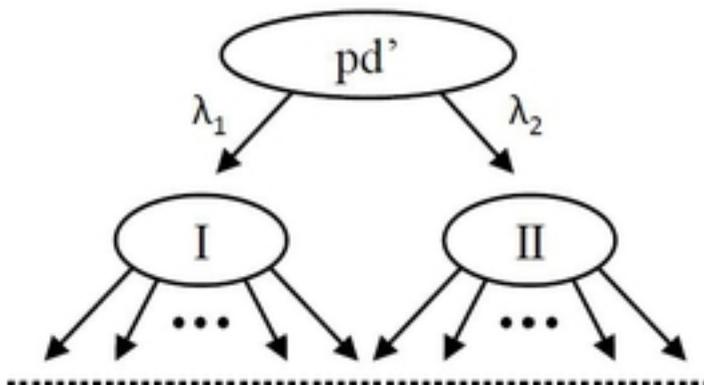
pd': General factor of personality functioning

I: Interpersonal  
II: Intrapersonal

A: Identity  
B: Self-direction  
C: Intimacy  
D: Empathy

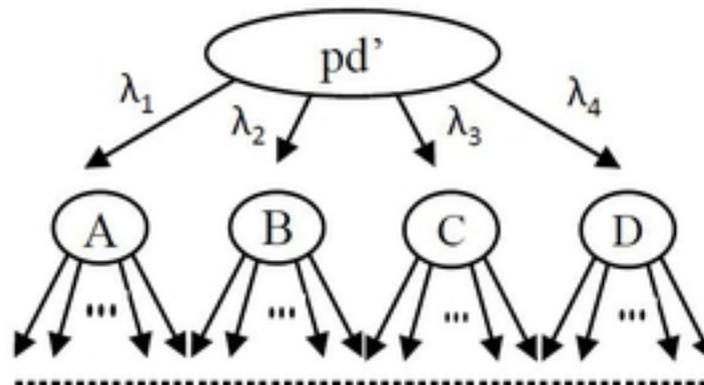
**Models 6a, 6b**

**2-dimensional**  
(hierarchical)



**Models 7a, 7b**

**4-dimensional**  
(hierarchical)



**Model restrictions:**

Model 6b:  $\lambda_1 = \lambda_2$

Model 7b:  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$

**Indicators:**

97 items

Figure 1

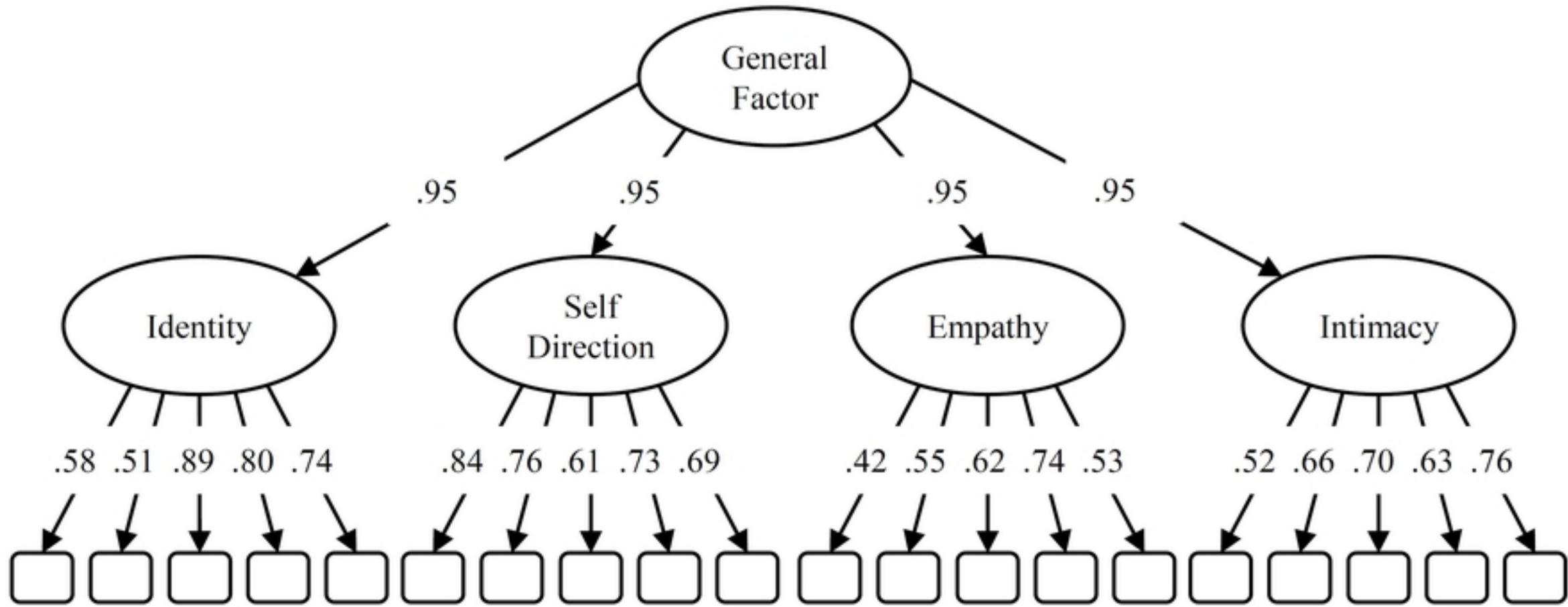


Figure 2

● 100k random combinations ● Short version

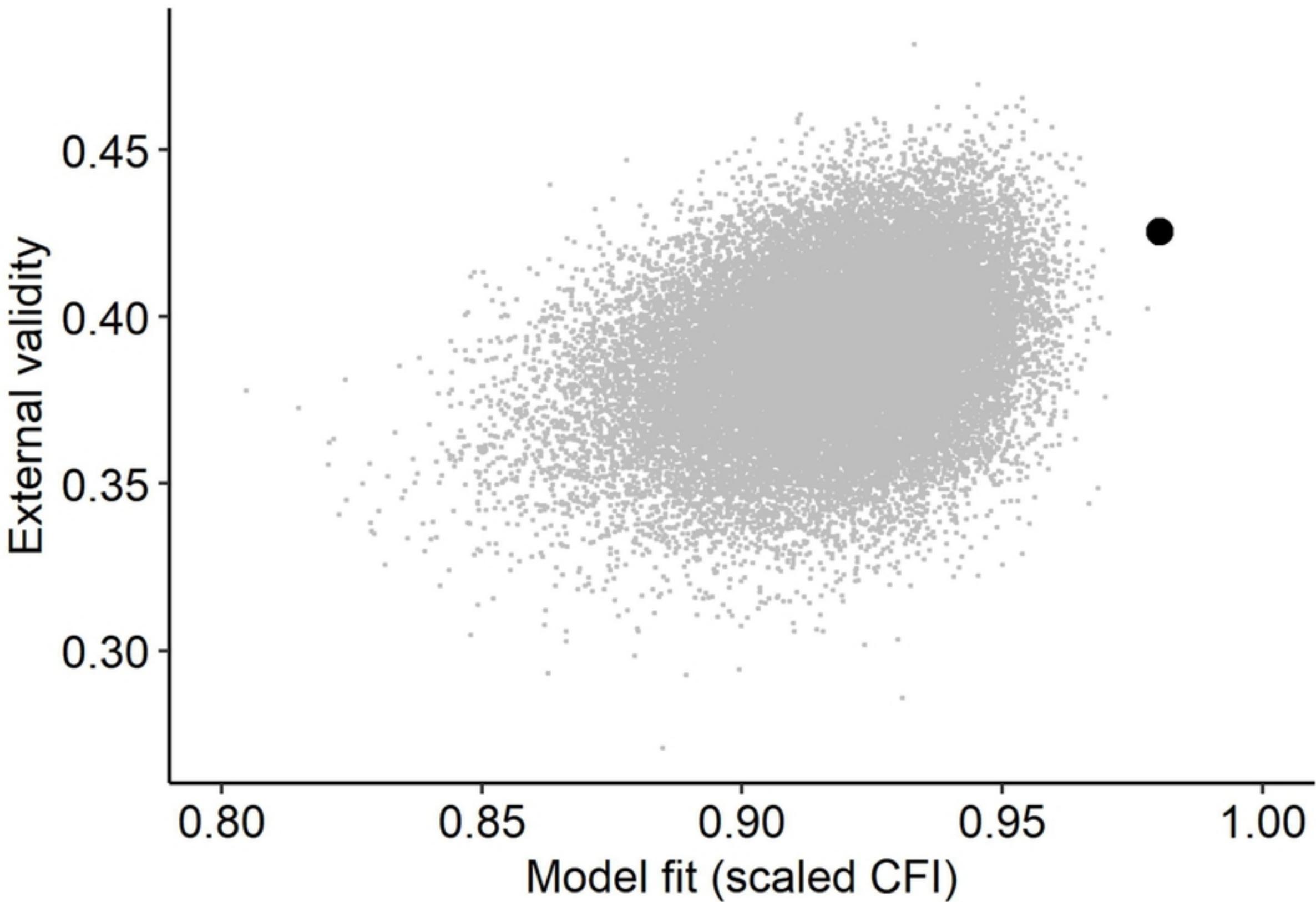


Figure 3