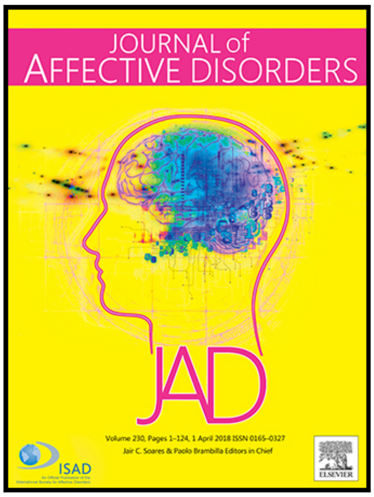
Accepted Manuscript



A reliability generalization meta-analysis of the Child and Adolescent Perfectionism Scale

María Vicent , María Rubio-Aparicio , Julio Sánchez-Meca ,
Carolina Gonzálvez

PII: S0165-0327(18)31781-6
DOI: https://doi.org/10.1016/j.jad.2018.11.049
Reference: JAD 10267

To appear in: *Journal of Affective Disorders*

Received date: 14 August 2018
Revised date: 20 September 2018
Accepted date: 3 November 2018

Please cite this article as: María Vicent , María Rubio-Aparicio , Julio Sánchez-Meca , Carolina Gonzálvez , A reliability generalization meta-analysis of the Child and Adolescent Perfectionism Scale, *Journal of Affective Disorders* (2018), doi: https://doi.org/10.1016/j.jad.2018.11.049

Highlights:

The average reliability was .87 for the original CAPS total score.

The mean alpha values were .84 and .83, respectively, for SPP and SOP subscales.

The original version of the CAPS can be employed with general research purposes.

The O'Connor's version of the CAPS must be used only for explanatory research.

The reliability induction rate was 29.8%.

A reliability generalization meta-analysis of the Child and Adolescent Perfectionism Scale

Abstract

Background: Perfectionism is a prevalent disposition of personality involved in the development and maintenance of a wide range of psychological disorders. The Child and Adolescent Perfectionism Scale (CAPS) is the most usually applied test to assess perfectionism in children and adolescents. This study aimed: (a) to conduct a reliability generalization meta-analysis to estimate the average reliability of the CAPS scores and to search for characteristics of the studies that may explain the variability among reliability estimates, and (b) to estimate the reliability induction rate of the CAPS.

Method: An exhaustive search allowed to select 56 studies that reported alpha coefficients with the data at hand for the CAPS.

Results: The average alpha coefficients were .87, .84 and .83, respectively for the CAPS total score and its two subscales, Socially Prescribed Perfectionism (SPP) and Self-Oriented Perfectionism (SOP). Regarding O'Connor's version, the average reliability coefficients were .82, .74 and .73, respectively, for SPP, SOP-Critical and SOP-Strivings. Some study characteristics (ethnicity, language, mean age and standard deviation of the scores, psychometric vs applied) showed a statistical association with the reliability coefficients of SPP and SOP. The reliability induction rate was 29.8%.

Limitations: Due to the scarcity of studies, we could not examine the reliability scores of other versions of the CAPS and test-retest reliability.

Conclusions: In terms of reliability, the original version of the CAPS present better results than O'Connor's version. The original version of the CAPS is a reliable instrument to be employed with general research purposes, but not for clinical practice.

Keywords: Meta-analysis, reliability generalization, Child and Adolescent Perfectionism Scale.

Perfectionism can be defined as "a multidimensional personality disposition characterized by striving for flawlessness and setting exceeding high standards of performance accompanied by overly critical evaluations of one's behavior" (Stoeber, 2018a, p. 3). It is a stormy worldview that constitute a psychological vulnerability factor of clinical relevance, predisposing to the development and maintenance of lot of problems (Hewitt, Flett and Mikail, 2017). Likewise, far from being an exclusive disposition of adulthood, perfectionism is closely related with several disorders, such as anxiety, depression, Obsessive Compulsive Disorder and Eating Behavior Disorders, in child and adolescent population (Morris and Lomax, 2014). In fact, it is deemed that three out of ten young people present maladaptive forms of perfectionism; a rate that increases considerably when other more moderate forms are taken into account (Sironic and Reeve, 2015). On the other hand, Hong et al. (2017) concluded that maladaptive perfectionist trajectories emerge at the beginning of formal education, reflecting children's reactions to a prevalent culture that excessively values academic excellence. It is not surprising, therefore, the growing interest in research about perfectionism in samples of children and adolescents.

Leone and Wade (2017) conducted a systematic review on the psychometric properties of the scales used to measure perfectionism in the population under 15 years old. Concretely, four specific measures of child perfectionism were identified: (a) *The Adaptive-Maladaptive Perfectionism Scale* (AMPS; Rice and Preusser, 2002), (b) *The Children's Disfunctional Attitudes Scale* (CDAS; Allessandro and Abela, unpublished results), (c) *The Perfectionistic Self-Presentation Scale-Junior Form* (PSPS-JR; Hewitt et al., 2011), and (d) *The Child and Adolescent Perfectionism Scale* (CAPS; Flett et al., 2016). Authors concluded that the CAPS was the most advisable scale of the four, not only because it has relatively strong psychometric properties, but also because of its wide use and comparative data availability. In effect, the CAPS is currently the most used instrument of child and adolescent perfectionism (García-Fernández et al., 2016), having being applied in children and adolescents age 8 and over from several countries, mostly English-speaking, such as Canada (Flett et al., 2016), United States (e.g., Affrunti and Woodruff-Borden, 2017), United Kingdom (e.g., Kerr et al., 2016) and Australia (e.g., Ferrari et al., 2018), but also in population from Spain (e.g., Vicent, Inglés, Sanmartín et al., 2017b), Israel (e.g., Freudenstein et al., 2012), Portugal (e.g., Bento et al., 2017), Romania (e.g., Damian et al., 2017), Turkey (e.g., Uz-Bas and Siyez, 2010), France (e.g., Douilliez and Hénot, 2013), China (e.g., Yang et al., 2015), Ecuador (e.g., Vicent, Inglés, Gonzálvez et al., 2017a), etc. This fact may create some confusion, since the validation of the CAPS was not definitively published until a few years ago, despite having been used for almost two decades since it was cited for the first time as an unpublished manuscript by Hewitt et al. (1997).

The relevance of the CAPS is partly due to the fact that it was developed by one of the research groups with the greatest impact in the field of perfectionism on the bases of the scale for adults of these same authors (i.e., *Multidimensional Perfectionism Scale*, Hewitt and Flett, 2004). The original version of the test consists of a 5-point Likert response scale and 22 items structured around two dimensions: Self-Oriented Perfectionism (SOP; 12 items) which measures the motivation and efforts to be a perfectionist as well as the tendency to self-criticize; and Socially Prescribed Perfectionism (SPP; 10 items) that captures the belief about the perfectionist demands of the environment. The authors also estimated the reliability of the scale across different populations, finding fluctuations between $\alpha = .68$ and .82 for SOP and between .68 and .89 for SPP. Test-retest reliability was also calculated for intervals of one, three and five years, ranging these values between $r = .65$ and .40 for SOP and between .35 and .59 for SPP. From our knowledge, seven additional psychometric studies on CAPS have been published (Bento et al., 2014; Douilliez and Hénot, 2013; McCreary et al., 2004; Nobel et al., 2012; O'Connor et al., 2009a; Uz-Baş and Siyez, 2010; Yang et al., 2015).

All of them eliminated some items, with the exception of the Portuguese validation (Bento et al., 2014) that keeps the original scale intact. However, the studies of McCreary et al. (2004), O'Connor et al. (2009a) and Nobel et al. (2012), not only dispense with certain items but they also question the two-dimensional structure of the scale when considering that SOP dimension is better conceptualized by dividing its items into two independent dimensions called Self-Oriented Perfectionism Critical (SOP-C) and Self-Oriented Perfectionism-Striving (SOP-S). These two dimensions refers to self-criticism perfectionism and strivings to reach perfection, respectively. In this way, a new three-dimensional structure of the CAPS is proposed (i.e., SPP, SOP-C and SOP-S). Lastly, there is a Chinese validation of the CAPS consisting of 16 items of the original 22 and three items newly created, structuring all of them in four dimensions: Socially Prescribed Perfectionism Positive, Socially Prescribed Perfectionism Negative, Self-Oriented Perfectionism Positive and Self-Oriented Perfectionism Negative.

In terms of internal consistency, these additional psychometric studies offered good levels of reliability, Cronbach's alpha, for the SPP dimension, ranging between .82 and .86. Nevertheless, taken into account the Nunnally's criterion (1987), who established a minimum value of .70 to consider that a reliability coefficient is acceptable, not all psychometric studies obtained adequate levels of reliability for SOP, SOP-C and SOP-S. Specifically, values ranged from .64 to .83, from .66 to .74, and from .58 to .78, respectively. In contrast, regarding the temporal reliability, those studies that provided data on the test-retest obtained acceptable values, higher than .60 in all cases, with the exception of the Portuguese validation, whose test-retest level was .59 for the SOP dimension. These data show the existence of considerable fluctuations in the reliability levels depending on the characteristics of the employed sample. Meyer defines internal consistency reliability as "the extent to which test scores are consistent with another set of test scores produced from a similar process" (2010, p. 9). It is a psychometric property that must be taken into account in any study because it determines the validity of the conclusions obtained (Nunnally, 1982). However, there is a fairly widespread belief that reliability is an inherent property of an instrument (Sánchez-Meca et al., 2013). Thus, it is common in research to find studies in which either reliability estimates of the measures used are not provided, or the reliability coefficients obtained in previous studies are cited; generally the original validation of the scale (Sánchez-Meca et al., 2013). It has been coined with the name of reliability induction (Vacha-Haase et al., 2000), and it is an erroneous practice because, as mentioned, reliability is a property of the scores of a test for a particular sample of participants. Therefore, it is not an immutable property, but it can vary depending on different factors, such as the characteristics of the sample, the version of the test used, etc. According to Shields and Caruso (2004), and Sánchez-Meca et al. (2017), it is possible to distinguish two types of reliability induction: (a) by omission, that is, when the authors make no reference to the reliability of the test, or (b) by report, when reliability estimates from previous studies are mentioned. In turn, the induction by report may be exact or vague, respectively, depending on whether or not accurate estimates of reliability are provided.

The Reliability Generalization (RG) is a meta-analytical approach that emerges as a criticism of the widespread practice of induction of reliability. The purpose of this method is to estimate the average reliability of the scores of a given test, as well as to determine the variability of the reliability coefficients reported by the different studies that have used this test. Moreover, if the variability is very high, another aim is to explore which characteristics of the studies may be statistically associated to the reliability estimates (Henson and Thompson, 2002; Rodríguez and Maeda, 2006; Sánchez-Meca et al., 2013).

The purpose of this research was to conduct an RG meta-analysis of those empirical studies that have applied the CAPS. The specific aims of this study were: (a) to calculate the average reliability of the CAPS dimensions scores to have an approximate estimate of their overall reliability; (b) to identify which characteristics of the studies may influence the variability of the reliability coefficients; and (c) to propose a predictive model to estimate the expected reliability of the CAPS according to the characteristics of the studies. Likewise, (d) the reliability induction rate of the CAPS was also estimated. Finally, in order to assess the extent to which the results of our RG meta-analysis can be generalized, we compared the characteristics of the studies that induced the reliability with those that provided some reliability coefficient with the data at hand.

Method

Selection criteria

The following criteria were considered to include each study in the meta-analysis: (a) being an empirical research where the original version of the CAPS (Flett et al., 2016) or any of its adaptations or versions were applied; (b) being written in English, Spanish or French; (c) being published and evaluated by experts; (d); employing any type of target population (community or clinical); (e) using a sample of at least 10 participants; (f) and reporting any reliability estimate of the CAPS or any of its subscales (internal consistency, test-retest) with the data at hand. The same criteria were considered for selecting studies that induced reliability, with the exception of (e) and (f).

Searching for the studies and selection process

The following data bases were consulted: Web of Science, Scopus, PsycINFO and ProQuest. The research strategy employed was: *"Child-Adolescent Perfectionism Scale" or "Child and Adolescent Perfectionism Scale" or (CAPS and perfectionism).* The search period covered from 1997 (date of publication of the first study that have used the CAPS) to march 2018.

Figure 1 shows a flowchart describing the selection process of the studies. A total of 214 references were obtained, out of which 130 were removed for different reasons. Of the remaining 84 empirical studies, 59 reported some reliability coefficient whereas the other 25 induced the reliability.

INSERT HERE FIGURE 1

Data extraction

The following characteristics of the studies were extracted: (a) mean and standard deviation of CAPS (for total score and subscales), (b) CAPS adaptation (original, O'Connor, Portuguese adaptation), (c) language of the scale/adaptation, (d) study focus (psychometric vs. applied), (e) continent where the study was carried out, (f) target population (community, clinical), (g) type of disorder (in case of clinical sample), (h) mean age of the sample, (i) gender (% male), (j) ethnicity (% Caucasian), (k) financial source of the study, (l) year of the study, and (m) conflict of interest declaration. These characteristics were extracted from studies that reported any reliability estimate with the data at hand. In addition, such characteristics as the target population, mean and standard deviation of the CAPS and subscales, mean age, gender, and ethnicity were also extracted from the studies that induced reliability. This enabled us to compare the characteristics of the studies that induced and reported reliability estimates, with the purpose of examining the extent to which our meta-analytic results could be generalized to the total population of studies that applied the CAPS, regardless of whether they induced or reported reliability estimates.

To assess the reliability of the coding process of the study characteristics, all studies were doubly coded by two independent coders, both psychologists with PhD in psychology. Results were highly satisfactory, with kappa coefficients for qualitative characteristics ranging between .82 and 1 (M = .93), and intra-class correlations for continuous variables yielding values between .88 and 1 (M = .96).

Reliability estimates

In this RG study, the alpha coefficients were taken into account to assess internal consistency of the measures. Although, we intended to include in our meta-analysis test-retest temporal stability coefficients, the scarce references (e.g., Bento et al., 2014; Flett et al., 2016; O'connor et al., 2009a) that reported this type of reliability did not allow us to carry out this analysis. Therefore, only alpha coefficients were extracted for the CAPS score and for each one of their subscales. In order to normalize their distribution and stabilizing their sampling variances, alpha coefficients, $\hat{\alpha}$, were transformed by means of Bonett's (2002) formula: $L_i = Ln(1 - |\hat{\alpha}_i|)$, with *Ln* being the natural logarithm. The sampling variances were obtained by (Bonett, 2002):

$$V(L_i) = \frac{2J}{(J-1)(n_i-2)}, \qquad (1)$$

with *J* being the number of items of the scale and n_i being the sample size of the study.

Statistical analyses

Separate meta-analyses were conducted for the alpha coefficients obtained from the total scale and for each of the two subscales of the original version of the CAPS.

To obtain summary statistics of alpha coefficients, a random-effects model was assumed (Borenstein et al., 2010). Thus, the alpha coefficients were weighted by the inverse variance, this defined as the sum of the within-study (Equation 1) and the between-studies variance, estimated by restricted maximum likelihood (López-López et al., 2013). In each meta-analysis, an average alpha coefficient and a 95% confidence interval were computed using the method proposed by Hartung (1999; see also Sánchez-Meca and Marín-Martínez, 2008; Sánchez-Meca et al., 2013). The heterogeneity exhibited by the alpha coefficients was assessed by constructing a forest plot and by calculating the *Q* statistic and the I^2 index. The *Q* statistic can be applied to test the homogeneity assumption among the alpha coefficients and I^2 values about 25%, 50%, and 75% can be considered as reflecting low, moderate, and large heterogeneity, respectively (Huedo-Medina et al., 2006).

For meta-analyses with at least 30 coefficients where evidence of heterogeneity was found, moderator analyses were performed through weighted ANOVAs for qualitative variables and meta-regressions for continuous variables. Mixed-effects models were assumed for these analyses, using the improved method proposed by Knapp and Hartung to test the statistical significance of the moderator variable (Knapp and Hartung, 2003; Rubio-Aparicio et al., 2017; Viechtbauer et al., 2015). In addition, the proportion of variance accounted for by the moderator variables was estimated with R^2 (López-López et al., 2014). Q_W and Q_E statistics were applied for testing the model misspecification of ANOVAs and meta-regressions, respectively.

To facilitate the interpretation of the results, the average alpha coefficients, their confidence limits, and the slope estimates obtained with Bonett´s transformation were back-transformed to the original metric of alpha coefficient.

Last, the risk of publication bias was assessed applying the Egger test and constructing funnel plots with the trim-and-fill method (Duval & Tweedie, 2000).

All statistical analyses were carried out with *metafor* package in *R* (Viechtbauer, 2010).

Results

Mean reliability and heterogeneity

The present RG study was focused on the 59 studies that reported alpha coefficients with the data at hand. Of the 59 studies, three of them could not be included in our RG meta-analysis because they reported a range of alpha coefficients (Fairweather-Schmidt and Wade, 2015; Flett et al., 2012c; Vekas and Wade, 2017), or they employed other versions of the CAPS with not enough studies to be compared, this is the case of the French (Douilliez & Hénot, 2013) and Chinese (Yang et al., 2015) versions of the scale, or due to other reasons. Thus, the remaining 56 studies that reported alpha coefficients were included in our RG meta-analysis.

As several studies reported alpha coefficients for two or more different samples, the dataset of our RG meta-analysis was composed by a total of 64 independent samples.[1] The total number of participants was N = 28483 (min. = 37; max. = 2142), with a mean of 445 participants per sample (Median = 257; SD = 489). Out of the 64 independent samples, 59 (92.2%) were written in English, and the 5 remaining samples (7.8%) were written in Spanish. Regarding the location of the studies, five continents were represented in our RG study: North America with 26 samples (40.6%), Europe with 23 samples (35.9%), Asia with 8 samples (12.5%), Oceania with 5 samples (7.8%), and South America with 2 samples (3.1%). Finally, we found that 54 samples (84.4%) used the CAPS original version, 8 samples (12.5%) used the O'Connor version, and 2 samples (3.1%) used the Portuguese version. Separate meta-analyses for each one of these versions of the CAPS were carried out.

[1] The database with the 64 independent samples can be obtained from the corresponding author on request.

Table 1 presents the average alpha coefficients obtained for the total scores as well as for the two subscales of the original CAPS version. The 11 samples that reported alpha coefficients for the total score yielded a mean coefficient of .87 (95%CI: .84 and .90; 95% prediction interval (PI): .73 and .94). For the subscales, alpha coefficients were computed in 51 samples, yielding an overall estimate of .84 (95%CI: .82 and .85; 95%PI: .72 and .91) for the SPP subscale, and for the SOP subscale the average coefficient calculated with the 47 samples was of .83 (95%CI: .81 and .84; 95%PI: .66 and .91). The number of samples that applied the subscales was greatly larger than those that applied the total scale. For this reason, forest plots were only constructed for the SPP and SOP subscale scores (see Figures 2 and 3, respectively). Alpha coefficients for the total scale and subscales presented a statistically significant heterogeneity, with I^2 above 90%.

INSERT HERE TABLE 1, FIGURES 2 AND 3

Table 1 also presents the average alpha coefficients obtained for the three subscales of the O´Connor version. The reason for not including the total scale in the analyses was that only one study (Wojtowicz and Von Ranson, 2012) reported an alpha coefficient (α = .91) for that. The 7 estimates reported for SPP yielded a mean coefficient of .82 (95%CI: .76 and .86; 95%PI: .62 and .92). SOP-C and SOP-S showed lower average reliability coefficients than the SPP subscale above described. Concretely, the 6 samples that reported an alpha coefficient for SOP-C yielded an overall estimate of .74 (95%CI: .65 and .80; 95%PI: .52 and .86) and the 6 estimates for SOP-S presented a mean of .73 (95%CI: .67 and .77; 95%PI: .59 and .82).

Finally, only two studies reported reliability coefficients for the total scale of the Portuguese version: α = .81 (Bento et al., 2014) and .88 (Bento et al., 2010).

Analysis of moderator variables

As alpha coefficients for the SPP and SOP subscales of the original version presented more than 30 reliability estimates, the analyses of moderator variables were carried out only for these subscales. Meta-regressions and ANOVAs were conducted for continuous and categorical variables, respectively, on transformed alpha coefficients separately for SPP and SOP.

Table 1

Average alpha coefficients, 95% confidence and prediction intervals, and heterogeneity

•š š]•š] • (}Œ šZ }Œ]P]v o W^ v K[}vv}Œ À Œ•]}v

	k	+	95% CI LL ; UL	95% PI LL ; LU	Q	I^2	^
Original CAPS:							
Total	11	.87	.84 ; .90	.73 ; .94	174.970****	96.8	.108
SPP	51	.84	.82 ; .85	.72 ; .91	851.738****	93.4	.079
SOP	47	.83	.81 ; .84	.66 ; .91	1010.134****	95.0	.109
K[}vv}Œ À							
SPP	7	.82	.76 ; .86	.62 ; .92	34.585****	91.1	.080
SOP_C	6	.74	.65 ; .80	.52 ; .86	14.554**	78.7	.045
SOP_S	6	.73	.67 ; .77	.59 ; .82	12.078*	57.5	.018

SPP = Socially Prescribed Perfectionism. SOP = Self-Oriented Perfectionism. SOP_C = Self-Oriented Perfectionism-Critical. SOP_S = Self-Oriented Perfectionism tStriving. k = number of studies. + = mean coefficient alpha. CI = confidence interval. PI = prediction interval. LL and UL= lower and upper limits of the 95% confidence and prediction intervals for +. Q = } ZŒ v[• Z š Œ}P statistic; Q statistic has k t 1 degrees of freedom. I^2 = heterogeneity index. ^ = between-studies variance estimated using restricted maximum likelihood. *p < .05. **p < .01. ****p < .0001.

Table 2

Results of the simple meta-regressions applied on alpha coefficients for the SPP original CAPS version, taking continuous moderator variables as predictors

Predictor variable	k	b_j	F	p	Q_E	R^2
Mean SPP score	40	-0.001	0.03	.971	403.99****	0.0
SD of SPP score	37	0.029	3.61	.066	323.96****	.08

Mean age (years)	43	0.019	1.37	.249	700.26****	0.0
Gender (% male)	46	-0.002	1.28	.264	689.69****	0.0
Ethnicity (% Caucasian)	43	0.005	10.68	.002	666.18****	.20
Year of the study	51	-0.012	1.40	.243	829.20****	0.0

k = number of studies. b_j = regression coefficient of each predictor. F = Knapp-, Œ š µ v P [• statistic for testing the significance of the predictor (the degrees of freedom for this statistic are 1 for the numerator and k t 2 for the denominator). p = probability level for the F statistic. Q_E = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the predictor. ****p < .0001.

Table 3

Results of the weighted ANOVAs applied on alpha coefficients for the SPP original CAPS version, taking qualitative moderator variables as independent variables

Variable	k	α_+	95% CI LL	95% CI LU	ANOVA results
Language:					
English	30	.86	.84	.87	$F(6,44) = 5.20$, $p < .001$
Hebrew	3	.87	.83	.90	$R^2 = .39$
Spanish	9	.82	.79	.85	$Q_W(44) = 408.31$, $p <.0001$
French	1	.84	.73	.90	
Romanian	3	.82	.77	.86	
Chinese	4	.72	.64	.78	
Russian	1	.77	.62	.86	

Language (dich.):					$F(1,49) = 9.58$, $p = .003$
English	30	.85	.84	.87	$R^2 = .20$
Other	21	.81	.79	.84	$Q_W(49) = 509.29$, $p < .0001$
Study focus:					$F(1,49) = 2.52$, $p = .119$
Applied	9	.81	.77	.85	$R^2 = .03$
Psychometric	42	.84	.83	.86	$Q_W(49) = 842.01$, $p < .0001$
Continent:					
Europe	17	.83	.81	.85	$F(4,46) = 2.32$, $p = .071$
N. America	21	.85	.83	.87	$R^2 = .13$
Asia	7	.80	.75	.84	$Q_W(46) = 590.70$, $p < .0001$
Oceania	4	.87	.82	.90	
S. America	2	.80	.70	.86	
Target population:					
Community	36	.83	.82	.85	$F(3,47) = 1.47$, $p = .239$
Clinical	10	.87	.84	.89	$R^2 = .02$
Comm.+Clinical	4	.83	.78	.88	$Q_W(47) = 828.52$, $p < .0001$
Athletes	1	.80	.64	.89	
Type of disorder:					$F(3,6) = 3.71$, $p = .081$
Anxiety/depression	4	.86	.83	.89	$R^2 = .99$
Eating disorder	2	.90	.86	.93	$Q_W(6) = 5.32$, $p = .514$

Mixed psychiatric sample	3	.87	.85	.89	
Other	1	.79	.67	.87	
Financial source:					$F(1,49) = 1.54$, $p = .220$
Public funding	34	.83	.81	.85	$R^2 = .02$
No funding	17	.85	.83	.87	$Q_W(49) = 789.00$, $p < .0001$
Conflict of interest:					$F(1,49) = 0.16$, $p = .689$
No reported	44	.84	.82	.85	$R^2 = 0.0$
No conflict	7	.85	.81	.88	$Q_W(49) = 801.20$, $p < .0001$

k = number of studies. $_{+}$ = mean coefficient alpha. LL and LU = lower and upper 95% confidence limits for $_{+}$. F = Knapp- , Œ š µ v P [• • š š] • š] (} Œ š • š] v P š Z •] P v] ( ] v moderator variable. Q_W = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the moderator.

Table 4

Results of the simple meta-regressions applied on alpha coefficients for the SOP original CAPS version, taking continuous moderator variables as predictors

Predictor variable	k	b_j	F	p	Q_E	R^2
Mean SOP score	38	0.007	2.46	.125	362.45****	.04
SD of SOP score	35	0.052	13.27	.001	283.17****	.29
Mean age (years)	39	0.045	5.10	.029	627.83****	.11
Gender (% male)	42	-0.004	1.75	.193	820.08****	.03

Ethnicity (% Caucasian)	39	0.003	1.71	.199	802.20****	.02
Year of the study	47	-0.042	16.97	<.001	708.37****	.30

k = number of studies. b_j = regression coefficient of each predictor. F = Knapp- , Œ š µ v P [• statistic for testing the significance of the predictor (the degrees of freedom for this statistic are 1 for the numerator and k t 2 for the denominator). p = probability level for the F statistic. Q_E = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the predictor. ****p < .0001.

Table 5

Results of the weighted ANOVAs applied on alpha coefficients for the SOP original CAPS version, taking qualitative moderator variables as independent variables

Variable	k	+	95% CI LL	95% CI LU	ANOVA results
Language:					
English	28	.84	.82	.86	$F(6,40) = 3.94$, $p = .003$
Hebrew	3	.89	.84	.92	$R^2 = .33$
Spanish	7	.79	.74	.83	$Q_W(40) = 416.26$, $p < .0001$
French	1	.82	.67	.90	
Romanian	3	.79	.71	.85	
Chinese	4	.75	.67	.82	
Russian	1	.68	.42	.82	
Language (dich.):					$F(1,45) = 5.56$, $p = .023$
English	28	.84	.82	.86	$R^2 = .12$
Other	19	.80	.77	.83	$Q_W(45) = 634.80$, $p < .0001$

Study focus:					$F(1,45) = 4.84$, $p = .033$
Applied	9	.78	.73	.83	$R^2 = .08$
Psychometric	38	.84	.82	.85	$Q_W(45) = 1004.07$, $p < .0001$
Continent:					
Europe	15	.79	.75	.82	$F(4,42) = 3.09$, $p = .026$
N. America	20	.85	.83	.87	$R^2 = .19$
Asia	7	.83	.78	.86	$Q_W(42) = 533.98$, $p < .0001$
Oceania	3	.85	.78	.90	
S. America	2	.79	.68	.87	
Target population:					
Community	32	.82	.80	.84	$F(3,47) = 1.46$, $p = .239$
Clinical	10	.85	.82	.89	$R^2 = .02$
Comm.+Clinical	4	.81	.73	.87	$Q_W(47) = 828.52$, $p < .0001$
Athletes	1	.76	.52	.88	
Type of disorder:					$F(3,6) = 0.07$, $p = .976$
Anxiety/depression	4	.85	.74	.91	$R^2 = 0.0$
Eating disorder	2	.85	.68	.93	$Q_W(6) = 34.66$, $p < .0001$
Mixed psychiatric sample	3	.87	.76	.92	
Other	1	.86	.59	.95	
Financial sources:					$F(1,45) = 0.04$, $p = .842$

Public funding	31	.83	.80	.85	R^2 = 0.0
No funding	16	.83	.80	.86	Q_W(45) = 965.99, p <.0001
Conflict of interests:					
No reported	40	.83	.81	.85	F(1,45) = 0.77, p = .384
No conflict	7	.81	.75	.85	R^2 = 0.0
					Q_W(45) = 878.47, p <.0001

k = number of studies. $_+$ = mean coefficient alpha. LL and LU = lower and upper 95% confidence limits for $_+$. F= Knapp-, Œ š µ v P [• •š š] • š] (} Œ š •š] v P š Z •] P v] ( ] v moderator variable. Q_W = statistic for testing the model misspecification. R^2 = proportion of variance accounted for by the moderator.

Table 6

Results of the multiple meta-regression applied on alpha coefficients for the SPP original CAPS version, taking as predictors the percentage of Caucasian and the language dichotomized (k = 43)

Predictor variable	b_j	t	p	Model fit
Intercept	1.330	9.56	< .0001	F(2, 40) = 7.69, p = .002
Ethnicity (% Caucasian)	0.005	3.02	.004	R^2 = .30
Language (dich.)	0.165	1.99	.053	Q_E(40) = 431.62, p < .0001

Model	F	p	R^2	ΔR^2
Full model	7.69	.002	.30	-
Ethnicity (% Caucasian)	10.68	.002	.20	.11
Language (dich.)	9.58	.003	.19	.10

b_j = regression coefficient of each predictor. t = statistic for testing the significance of the predictor (with 40 degrees of freedom). p = probability level for the t statistic. F = Knapp-, Œšµv P[• •š š]•š] (}Œ š •š]v P šZ Q_E = statistic for testing the (µoo u} oX model misspecification. R^2 = proportion of variance accounted for by the predictors. ΔR^2 = increase in R^2 as consequence of including in the model a predictor once the other predictors had already been introduced.

Table 7

Results of the multiple meta-regression applied on alpha coefficients for the SOP original CAPS version, taking as predictors the SD of PAO scores, the mean age, the year of the study, the language dichotomized and the study focus (k = 28)

Predictor variable	b_j	t	p	Model fit
Intercept	31.017	1.07	.295	$F(5, 22) = 10.94, p < .0001$
SD of PAO score	0.063	5.02	<.0001	$R^2 = .78$
Mean age (years)	0.036	2.43	.024	$Q_E(22) = 61.75, p < .0001$
Year of the study	-0.015	-1.07	.296	
Language (dich.)	0.083	1.02	.319	
Study focus	0.347	3.67	.001	

Model	F	p	R^2	ΔR^2
Full model	10.94	<.0001	.78	-
SD of PAO score	13.27	.001	.29	.27
Mean age (years)	5.10	.029	.11	.22
Year of the study	16.97	.000	.30	0
Language (dich.)	5.56	.023	.12	0

Study focus	4.84	.033	.08	.24

b_j = regression coefficient of each predictor. t = statistic for testing the significance of the predictor (with 22 degrees of freedom). p = probability level for the t statistic. F= Knapp-, Œšµv P[• •š š]•š] (}Œ š •š]vP šZ Q_E = statistic for testing the }(šZ (µoo u} oX model misspecification. R^2 = proportion of variance accounted for by the predictors. ΔR^2 = increase in R^2 as consequence of including in the model a predictor once the other predictors had already been introduced.