**Appendix A** REGEMA checklist.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TITLE |  | Yes | No  | Unclear | NA |
| *1. Title* | The title must include: (a) the term “reliability generalization” or “meta-analysis” together with some explicit indication to reliability (internal consistency, test-retest, inter- or intra- rater) and (b) the name of the scale or, if more than one scale, the attribute/outcome measure that the scales are assessing. | X |  |  |  |
| ABSTRACT |  | Yes | No | Unclear | NA |
| *2. Abstract* | The abstract must state explicitly: (a) that the objective was to carry out a reliability generalization (RG) meta-analysis of one or several scales; (b) eligibility criteria of the studies; (c) data sources with the temporal range covered; (d) types of reliability coefficients analyzed; (e) statistical model applied; (f) main results (e.g., pooled reliability coefficient and 95% CI, moderator variables related to reliability); and (g) main conclusions. In case of space limitation, (b) and (c) criteria can be omitted.  | X |  |  |  |
| INTRODUCTION |  | Yes | No | Unclear | NA |
| *3.* *Background* | The background must include: (a) a conceptual definition of the attribute/outcome measure assessed by the scale/s; (b) description of the target population/s to which the scale/s is/are applied and its/their purposes (e.g., screening, clinical diagnosis); (c) a complete description of the scale/s (length, number of categories), including the versions and adaptations to other languages/cultures; and (d) a brief presentation of reliability estimates obtained in previous psychometric studies of the scale/s. Optionally, a brief review of validation studies of the scale/s (e.g., exploratory/confirmatory factor analyses, concurrent/convergent/discriminant validity, responsiveness) could be included. | X |  |  |  |
| *4. Objectives* | State whether the purpose of the meta-analysis was to obtain a more precise overall reliability coefficient estimate and/or investigate how reliability coefficients vary among different applications of the scales. Optionally, specify whether one objective of the meta-analysis is to estimate the reliability induction rates of the scale/s.  | X |  |  |  |
| METHOD |  | Yes | No | Unclear | NA |
| 5. *Selection criteria* | Specify inclusion criteria: (a) name/s of the scale/s analyzed in the RG meta-analysis, as well as the versions and/or adaptations included; (b) years considered; (c) language restrictions; (d) publication status; (e) to report any reliability estimate based on the study-specific sample/s; (f) type/s of reliability considered (e.g., internal consistency, temporal stability, inter-/intra-rater reliability); and (g) target population/s (e.g., community, clinical, subclinical/analogue, university).  | X |  |  |  |
| 6. *Search strategies* | Specify how the studies were located: (a) electronic databases consulted; (b) other formal search procedures (e.g., manual search in specific journals, backward search from references listed in selected studies); and (c) informal search procedures (e.g., internet searches, contacting study authors to identify additional studies). For electronic searches, describe the search strategy, including the keywords used and how they were combined, and the search limits (e.g., fields where the keywords were searched - title, abstract, full-text -, temporal range, language).  | X |  |  |  |
| 7. *Data extraction* | Describe the characteristics extracted from the studies, including: (a) sample size/s, mean/s and standard deviation/s of total test scores and subscales (if applicable); (b) sample characteristics (e.g., target population, country, mean age, standard deviation of the age, gender distribution, ethnic distribution, disorder history −mean and SD in years); (c) test version (e.g., adaptation/version, number of items, reporting format −self-report, clinician); (d) methods (e.g., study design, purpose of the study −psychometric versus applied−, quality checklist); (e) extrinsic characteristics (e.g., publication status, researchers’ affiliations, funding source). | X |  |  |  |
| *8. Reported reliability* | Identify the types of reliability coefficients included in the RG meta-analysis: internal consistency (e.g., Cronbach’s alpha, KR-21, parallel forms), temporal stability (test-retest), inter- and intra-rater reliability (e.g., intraclass correlation, kappa coefficient). | X |  |  |  |
| *9. Estimating the reliability induction* | In case that the meta-analysis intends to estimate the reliability induction, identify the types of reliability induction: induction by omission (no mention of test reliability whatsoever) or reporting induction (vague or precise reporting).  | X |  |  |  |
| *10. Data extraction of inducing studies* | Declare whether characteristics of inducing studies were also extracted or if, on the contrary, only characteristics of studies that reported reliability were extracted. | X |  |  |  |
| *11. Reliability of data extraction* | Describe how the reliability of data extraction process was appraised: how many coders, which agreement coefficients were applied (e.g., kappa coefficient, intraclass correlation), which values were obtained, and how disagreements were dealt with.  | X |  |  |  |
| *12. Transformation method* | State whether or not the reliability coefficients were transformed for the meta-analytic integration. If relevant, specify the transformation methods: Fisher´s Z for correlation coefficients (e.g., test-retest coefficients), Bonett’s and Hakstian and Whallen’s transformation for internal consistency coefficients (e.g. Cronbach’s alpha), reliability index, measurement error (e.g., standard error of measurement), or other (specify). | X |  |  |  |
| *13. Statistical model* | Describe the statistical model(s) assumed in the meta-analytic integration for estimating the average reliability coefficient and for analysing the influence of moderator variables (e.g. fixed-effect(s), random-effects, mixed-effects, varying-coefficient models, generalized linear models), as well as the analysis framework (frequentist or Bayesian).  | X |  |  |  |
| *14. Weighting method* | Specify the weighting method applied in the meta-analytic integration: unweighted, weighting by sample size, weighting by inverse variance, or other weighting methods. | X |  |  |  |
| *15. Heterogeneity assessment* | Describe how heterogeneity among reliability coefficients was assessed (e.g., standard deviation, *Q* statistic, *I*2 index, between-studies variance). If relevant, specify the between-studies variance estimator (DerSimonian and Laird, Maximum Likelihood, Restricted Maximum Likelihood, Empirical Mayes, Paule and Mandel), as well as how confidence intervals and/or credibility intervals were calculated.  | X |  |  |  |
| 16. *Moderator analyses* | If relevant, describe how the influence of moderator variables was assessed (e.g., subgroup analyses, meta-regression analyses).  | X |  |  |  |
| *17. Additional analyses* | Describe other additional analyses accomplished, such as sensitivity analyses (e.g., statistical analyses with transformed and untransformed reliability coefficients, one-to-one deleting of reliability coefficients, assessment of publication bias and other reporting biases). | X |  |  |  |
| *18. Software* | Mention the software and version used to carry out the statistical analyses (e.g., metafor in R, Proc MIXED in SAS, Comprehensive Meta-analysis). | X |  |  |  |
| RESULTS |  | Yes | No | Unclear | NA |
| *19. Results of the study selection process* | Describe, ideally with a flow chart, the selection process of the studies, specifying the number of studies identified from each search source, excluded studies and reasons why, and the number of studies that reported and induced reliability of test scores. Regarding reliability induction, report induction rates, distinguishing between induction “by omission” and “by report” (see e.g., REGEMA flowchart). Furthermore, it is advisable to compare the reliability induction rates as a function of variables such as publication year, country/continent and study purpose (psychometric vs. applied).  | X |  |  |  |
| *20. Mean reliability and heterogeneity* | Present pooled reliability coefficients and confidence/credibility intervals for the scale (and subscales, if applicable) and for each type of reliability (e.g., internal consistency, temporal stability, inter- and intra-rater agreement). In case of applying any transformation of the reliability coefficients, results should be back-transformed to the original metric to facilitate interpretation. Illustrate the distribution of reliability coefficients with graphical techniques (e.g., forest plots, box plots, stem and leaf displays, histograms) and describe the degree of heterogeneity by one or more heterogeneity measures (see Item 15).  | X |  |  |  |
| *21. Moderator analyses* | For categorical moderators, provide the pooled reliability coefficient, confidence interval and other heterogeneity measures for each category of the moderator. For continuous moderators, include the regression coefficients, standard errors and confidence limits. For both types of moderators, report results of the statistical significance tests, misspecification tests, and proportion of variance accounted for. As a further step, it is advisable to fit a predictive/explanatory model including the most relevant moderator variables.  | X |  |  |  |
| *22. Sensitivity analyses* | Report or describe the results of any sensitivity analyses conducted (see Item 17).  | X |  |  |  |
| *23. Comparison of inducing and reporting studies* | A comparison of the characteristics of inducing and reporting studies (e.g., sociodemographic and clinical characteristics of the samples) can be insightful. If performed, results must be reported.  | X |  |  |  |
| *24. Data set* | Tabulate the characteristics of the individual studies that reported reliability (see Item 7). Tables can be presented as appendices or supplementary files. In addition, list of all studies included in the RG meta-analysis, either in the reference section or as a supplementary file.  | X |  |  |  |
| DISCUSSION |  | Yes | No | Unclear | NA |
| *25. Summary of results* | Present the main results, such as mean reliability exhibited by the scale/test and moderators of the reliability coefficients. If available, discuss the results in the light of previous evidence.  | X |  |  |  |
| *26. Limitations* | Discuss the limitations of the meta-analysis. Include an explicit statement of the reliability induction rates and the extent to which inducing and reporting studies are comparable in terms of samples characteristics.  | X |  |  |  |
| *27. Implications for practice* | Provide guidelines for professional practice regarding the usefulness of the scale/test in different settings and target populations.  | X |  |  |  |
| *28. Implications for future research* | Include recommendations for researchers regarding the conditions under which the scale/test should be applied.  | X |  |  |  |
| FUNDING |  | Yes | No | Unclear | NA |
| *29. Funding* | State the financial sources of the meta-analysis, as well as potential conflict of interests of the authors. | X |  |  |  |

*Note*. NA: Not Applicable

**Appendix B** Studies included in the meta-analysis.

\*Aardema, F., Trihey, M., Kleijer, T. M., O'connor, K., & Emmelkamp, P. M. (2006). Processes of inference, schizotypal thinking, and obsessive-compulsive behaviour in a normal sample. *Psychological reports*, *99*, 213-220. <https://doi.org/10.2466/PR0.99.1.213-220>

\*Aardema, F., Wu, K.D., Moulding, R., Jean-Sebastien, A., & Louis-Philippe, B. (2018). The relationship of inferential confusion and obsessive beliefs with specific obsessive-compulsive symptoms. *Journal of Obsessive-compulsive and Related Disorders, 18,* 98-105. https://doi.org/[10.1016/j.jocrd.2018.03.004](https://doi.org/10.1016/j.jocrd.2018.03.004)

\*Anholt, G. E., van Oppen, P., Emmelkamp, P. M., Cath, D. C., Smit, J. H., van Dyck, R., & van Balkom, A. J. (2009). Measuring obsessive-compulsive symptoms: Padua inventory-revised vs. Yale-Brown obsessive compulsive scale. *Journal of Anxiety Disorders*, *23*, 830-835. <https://doi.org/10.1016/j.janxdis.2009.04.004>

\*Beşiroğlu, L., Ağargün, M. Y., Boysan, M., Eryonucu, B., Güleç, M., & Selvi, Y. (2005). The assessment of obsessive-compulsive symptoms: the reliability and validity of the Padua inventory in a Turkish population. *Turkish Journal of Psychiatry*, *16*, 179-189.

\*Bogaerts, S., Spreen, M., & Palermo, G. B. (2011). Obsessive-compulsive and schizotypal characteristics in a group of male child molesters. *Crimen et Delictum: International Journal of Criminological and Investigative Sciences*, *2*, 15-26.

\*Boysan, M. (2014). Dissociative experiences are associated with obsessive-compulsive symptoms in a non-clinical sample: a latent profile analysis. *Archives of Neuropsychiatry*,*51*, 253-262. https://doi.org/10.4274/npa.y6884

\*Boysan, M., & Çam, Z. (2016). An investigation into the role of attachment insecurities in obsessive-compulsive symptoms. *British Journal of Guidance & Counselling*, 1-16. <https://doi.org/10.1080/03069885.2016.1262533>

\*Boysan, M., Gulec, M., Deveci, E., & Barut, Y. (2015). Diagnostic Performance of the Turkish Version of the Vancouver Obsessional Compulsive Inventory (VOCI) Versus Padua Inventory-Revised (PI-R): A Validation Study. *Bulletin of Clinical Psychopharmacology*, *25*, 44-56. <https://doi.org/10.5455/bcp.20141103123307>

\*Boysan, M., Yildirim, A., Besiroglu, L., Kefeli, M. C., & Kagan, M. (2018). Development and preliminary psychometric properties of an instrument for the measurement of obsessional dissociative experiences: The Van Obsessional Dissociation Questionnaire (VOD-Q). *Psychiatric Quarterly, 89*, 549-568. <https://doi.org/10.1007/s11126-017-9555-2>

\*Carter, J. C., & Bewell‐Weiss, C. V. (2011). Nonfat phobic anorexia nervosa: clinical characteristics and response to inpatient treatment. *International Journal of Eating Disorders*, *44*, 220-224. <https://doi.org/10.1002/eat.20820>

\*Cosentino, T., Pellegrini, V., Gacomantonio, M., Saliani, A., Basile, B., Saettoni, M., Gragnani, A., Buonanno, C., & Mancini, F. (2020). Validation and psychometric properties of the Italian version of the Fear of Guilt Scale. *Rassegna di Psicologia, 37*, 59-70. https://doi.org/[10.13133/1974-4854/16724](https://doi.org/10.13133/1974-4854/16724)

\*De Berardis, D., D'Albenzio, A., Gambi, F., Sepede, G., Valchera, A., Conti, C. M., Fulcheri, M., Cavuto, M., Ortolani, C, Salerno, R. M., Serroni, N., & Ferro, F. M. (2009). Alexithymia and its relationships with dissociative experiences and Internet addiction in a nonclinical sample. *CyberPsychology & Behavior*, *12*, 67-69. <https://doi.org/10.1089/cpb.2008.0108>

\*de Bruin, G. O., Rassin, E., & Muris, P. (2005). Cognitive self-consciousness and meta-worry and their relations to symptoms of worry and obsessional thoughts. *Psychological Reports*, *96*, 222-224.

\*de Bruin, G. O., Muris, P., & Rassin, E. (2007). Are there specific meta-cognitions associated with vulnerability to symptoms of worry and obsessional thoughts? *Personality and Individual Differences*, *42*, 689-699. <https://doi.org/10.1016/j.paid.2006.08.015>

De Putter, L.M.S., & Koster E.H.W. (2018). Can selective attention and inhibition (interactively) predict future obsessive compulsive symptoms? A prospective study*. Journal of Behavior Therapy and Experimental Psychiatry, 61*, 150-157. https://doi.org/[10.1016/j.jbtep.2018.07.007](https://doi.org/10.1016/j.jbtep.2018.07.007)

\*Denys, D., de Geus, F., van Megen, H. J., & Westenberg, H. G. (2004). Symptom dimensions in obsessive-compulsive disorder: Factor analysis on a clinician-rated scale and a self-report measure. *Psychopathology*, *37*, 181-189. <https://doi.org/10.1159/000079509>

\*Gangemi, A., Mancini, F., & Dar, R. (2015). An experimental re-examination of the inferential confusion hypothesis of obsessive–compulsive doubt. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 90-97. <https://doi.org/10.1016/j.jbtep.2015.02.008>

\*Gönner, S., Ecker, W., & Leonhart, R. (2010). The Padua Inventory: Do revisions need revision? *Assessment*, *17*, 89-106. <https://doi.org/10.1177/1073191109342189>

\*Mancini, F., D’Olimpio, F., Del Genio, M., Didonna, F., & Prunetti, E. (2002). Obsessions and compulsions and intolerance for uncertainty in a non-clinical sample. *Journal of Anxiety Disorders*, *16*, 401-411. [https://doi.org/10.1016/S0887-6185(02)00133-0](https://doi.org/10.1016/S0887-6185%2802%2900133-0)

\*McFarlane, T., MacDonald, D. E., Trottier, K., & Olmsted, M. P. (2015). The effectiveness of an individualized form of day hospital treatment. *Eating disorders*, *23*, 191-205. https://doi.org/10.1080/10640266.2014.981430

\*Rassin, E., Cougle, J. R., & Muris, P. (2007). Content difference between normal and abnormal obsessions. *Behaviour Research and Therapy*, *45*, 2800-2803. <https://doi.org/10.1016/j.brat.2007.07.006>

\*Rassin, E., Merckelbach, H., Muris, P., & Schmidt, H. (2001). The thought-action fusion scale: further evidence for its reliability and validity. *Behaviour Research and Therapy*, *39*, 537-544. [https://doi.org/10.1016/S0005-7967(00)00031-0](https://doi.org/10.1016/S0005-7967%2800%2900031-0)

\*Rassin, E., & Muris, P. (2005). To be or not to be… indecisive: Gender differences, correlations with obsessive–compulsive complaints, and behavioural manifestation. *Personality and Individual Differences*, *38*, 1175-1181. <https://doi.org/10.1016/j.paid.2004.07.014>

\*Riskind, J. H., & Williams, N. L. (2005). The looming cognitive style and generalized anxiety disorder: Distinctive danger schemas and cognitive phenomenology. *Cognitive Therapy and Research*, *29*, 7-27. <https://doi.org/10.1007/s10608-005-1645-z>

\*Schulte-van Maaren, Y. W., Giltay, E. J., van Hemert, A. M., Zitman, F. G., de Waal, M. W., & Carlier, I. V. (2013). Reference values for anxiety questionnaires: the Leiden routine outcome monitoring study. *Journal of Affective Disorders*, *150*, 1008-1018. <https://doi.org/10.1016/j.jad.2013.05.031>

\*Steketee, G., Siev, J., Fama, J. M., Keshaviah, A., Chosak, A., & Wilhelm, S. (2011). Predictors of treatment outcome in modular cognitive therapy for obsessive–compulsive disorder. *Depression and Anxiety*, *28*, 333-341. <https://doi.org/10.1002/da.20785>

\*Slof‐Op't Landt, M. C., Claes, L., & van Furth, E. F. (2016). Classifying eating disorders based on “healthy” and “unhealthy” perfectionism and impulsivity. *International Journal of Eating Disorders*, *49*, 673-680. <https://doi.org/10.1002/eat.22557>

\*Van Oppen, P., Hoekstra, R. J., & Emmelkamp, P. M. (1995). The structure of obsessive-compulsive symptoms. *Behaviour Research and Therapy*, *33*, 15-23. [https://doi.org/10.1016/0005-7967(94)E0010-G](https://doi.org/10.1016/0005-7967%2894%29E0010-G)

\*Verwoerd, J., de Jong, P. J., Wessel, I., & van Hout, W. J. (2013). “If I feel disgusted, I must be getting ill”: Emotional reasoning in the context of contamination fear. *Behaviour Research and Therapy*, *51*, 122-127. <https://doi.org/10.1016/j.brat.2012.11.005>

\*Vriend, C., de Wit, S. J., Remijnse, P. L., van Balkom, A. J., Veltman, D. J., & van den Heuvel, O. A. (2013). Switch the itch: a naturalistic follow-up study on the neural correlates of cognitive flexibility in obsessive-compulsive disorder. *Psychiatry Research: Neuroimaging*, *213*, 31-38. https://doi.org/10.1016/j.pscychresns.2012.12.006

\*Wahl, K., Ertle, A., Bohne, A., Zurowski, B., & Kordon, A. (2011). Relations between a ruminative thinking style and obsessive–compulsive symptoms in non-clinical samples. *Anxiety, Stress, & Coping*, *24*, 217-225. <https://doi.org/10.1080/10615806.2010.482985>

**Appendix C** Supplementary figures



Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Impulses subscale scores. The outer edges of the bottom polygon indicate the confidence interval limits and the dotted line indicates the bounds of the 95% prediction interval. Tau = between-study standard deviation.



Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Washing subscale scores. The outer edges of the bottom polygon indicate the confidence interval limits and the dotted line indicates the bounds of the 95% prediction interval. Tau = between-study standard deviation.



Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Checking subscale scores. The outer edges of the bottom polygon indicate the confidence interval limits and the dotted line indicates the bounds of the 95% prediction interval. Tau = between-study standard deviation.



Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Rumination subscale scores. The outer edges of the bottom polygon indicate the confidence interval limits and the dotted line indicates the bounds of the 95% prediction interval. Tau = between-study standard deviation.



Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Precision subscale scores. The outer edges of the bottom polygon indicate the confidence interval limits and the dotted line indicates the bounds of the 95% prediction interval. Tau = between-study standard deviation.