Optimising the measurement of anxious-depressive, compulsivity and intrusive thought and social withdrawal transdiagnostic symptom dimensions

Alexandra Kathryn Hopkins[1], Claire Gillan[2], Jonathan Roiser[3], Toby Wise[4] & Nura Sidarus[1]

[1] Department of Psychology, Royal Holloway University of London
[2] School of Psychology & Global Brain Health Institute, Trinity College Dublin
[3] Institute of Cognitive Neuroscience, University College London
[4] Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, Kings College London

Corresponding author: Nura Sidarus, nura.sidarus@rhul.ac.uk
Department of Psychology, Royal Holloway University of London, Egham, TW20 0EX

Running title: Optimising transdiagnostic symptom measurement

Keywords: mental health, transdiagnostic, questionnaire, computational psychiatry

Abstract

Transdiagnostic mental health factors measuring anxious-depressive, compulsivity and intrusive thought and social withdrawal symptom dimensions have been identified as useful constructs that relate to a variety of cognitive constructs. However, the factor measurement relies on the administration of hundreds of questionnaire items which is both costly and a burden to participants. Additionally, it is unclear to what extent these factors generalise across datasets. Therefore, this study sought to optimise the measurement as well as assess the stability of these three factors. Using exploratory factor analysis on 209 questionnaire items, we replicated the same three-factor structure across a pooled dataset of n = 4782 participants, as well as within four independent subsets of the data. Using a machine learning approach, we reduced the number of items to 71 while still measuring the factors with high accuracy. We externally validated these factor scores by replicating previously observed associations with behaviour on a metacognition task. Our results support the generalisability of these symptom dimensions and provide a useful approach to optimising their measurement.

Introduction

Mental health problems are a significant source of distress for millions of people across the world and are amongst the largest contributors to disability (Ritchie & Roser, 2018). Traditionally, the classification of mental health problems has distilled symptoms down into discrete disorder categories, reflected in manuals such as the DSM-5 (American Psychiatric Association, 2013) or ICD-10 (World Health Organization, 1992). However, symptom overlap between disorder categories and heterogeneity within disorders present issues for classification. Dimensional alternatives to classification, such as the Research Domain Criteria (RDoC) project (Insel et al., 2010) and the Hierarchical Taxonomy of Psychopathology (HiTOP) (Kotov et al., 2017) support a move away from this type of categorisation and towards a more neurobiologically plausible and mechanistic understanding of mental health issues, spurring novel research questions and methodologies. For example, the general psychopathology (p)-factor was discovered through the attempt to understand the high comorbidity between disorders using questionnaires that assess multiple mental health domains (Caspi et al., 2014; Lahey et al., 2012). Of particular interest are transdiagnostic mechanisms that are shared across traditionally separate symptom domains. For example, the ability to adapt to volatile environments has been found to associate with internalising psychopathology including depression and anxiety, but not each disorder separately (Gagne et al., 2020).

Notably, research by Gillan et al. (2016) has linked transdiagnostic symptom dimensions representing ‘anxious-depression’ (AD), ‘compulsive behaviour and intrusive thought’ (CIT) and ‘social withdrawal’ (SW) to different aspects of cognition, including goal-directed behaviour. Subsequent research has used these same factors, based on the same set of questionnaires administered in the original study, to investigate a number of potential transdiagnostic cognitive mechanisms, including metacognition (Rouault et al., 2018), model-based learning (Patzelt et al., 2019) and aversive learning (Wise & Dolan, 2020). A significant drawback of this approach is the substantial time, effort and monetary cost arising from administering a large questionnaire set, with 209 items in total. Long questionnaires may increase response burden for participants, resulting in a reduction in data quality (Diehr et al., 2005) and may be especially burdensome for participants who suffer from disrupted cognitive function (Atkinson et al., 2019). Reducing the set of questions but keeping the specificity and span of the symptomatology represented would be greatly beneficial for mental health researchers interested in these transdiagnostic factors, and would reduce the cognitive and emotional burden for participants. In a first step towards this aim, Wise & Dolan (2020) utilised a machine learning approach in order to reduce the number of items. This yielded a set of 63 questions that showed strong factor score predictability, could be administered quickly, and which showed associations with aversive learning processes. However, the analysis was based on a relatively small sample size limited to US-based participants recruited through Amazon Mechanical Turk, and thus may lack generalisability. Additionally, it is not known how stable the original factor structure is across samples. Many of the studies that have made use of these transdiagnostic factors simply used the published item loadings from Gillan et al. (2016) to generate factor scores using a confirmatory approach, rather than estimating loadings from their own data (Patzelt et al., 2019; Seow et al., 2020, 2021; Seow & Gillan, 2020). Thus, the main goal of this work was to identify a reduced set of items to measure these 3 transdiagnostic factors, based on a large and varied dataset, with the secondary aim of verifying that the factor structure is consistent in different samples.

Materials and Methods

Sample

The sample consisted of four separate datasets in which the exact same set of questionnaires (described below) was administered, with a combined sample size of 4782 participants (mean age = 38.43(13.92), 35% men, 63% women). Two of the datasets were derived from published work by Rouault et al. (2018) and Patzelt et al. (2019) and consisted of participants (*n* = 496 and *n* = 820 respectively) recruited from Amazon Mechanical Turk https://www.mturk.com/. A further dataset was obtained from Kelley et al. (2021), *n* = 1006, with participants recruited through Clickworker https://www.clickworker.com/. The final dataset (*n* = 2460) was unpublished data obtained from the smartphone app Neureka https://www.neureka.ie/. The Rouault et al. (2018) and Patzelt et al. (2019) data were collected before the COVID-19 pandemic, the Kelley et al. (2021) and Neureka datasets were obtained during the COVID-19 pandemic.

Questionnaires

All datasets included the original questionnaire set reported in Gillan et al. (2016), consisting of measures assessing alcohol use (Alcohol Use Disorders Identification Test; AUDIT) (Saunders et al., 1993), apathy (Apathy Evaluation Scale; AES) (Marin et al., 1991), impulsiveness (Barrett Impulsiveness Scale; BIS) (Patton et al., 1995), eating disorders (Eating Attitudes Test; EAT-26) (Garner et al., 1982), social anxiety (Leibowitz Social Anxiety Scale; LSAS) (Liebowitz, 1987), OCD (Obsessive Compulsive Inventory-Revised; OCI-R) (Foa et al., 2002), schizotypy (Short Scales for Measuring Schizotypy; SSMS) (Mason et al., 2005), anxiety (State Trait Anxiety Index; STAI) (Spielberger et al., 1983) and depression (Zung Depression Scale; SDS) (Zung, 1965). All questionnaires had the same administration across datasets, except for the LSAS. Two datasets (Patzelt, Rouault) took an average of the fear/avoidance dimensions, whereas the two other datasets (Kelley, Neureka) asked participants to rate both fear and avoidance using a single response measure. Despite this difference, scores were treated equally here. Three of the datasets (Kelley, Neurka, Patzelt) had used a linear coding system for the EAT-26 questionnaire ranging from 0-5, instead of the questionnaire coding system, which ranged from 0-3, therefore scores were transformed back to the EAT-26 coding system.

Analysis

Factor Analysis

In order to determine whether the factor structure of the combined dataset was analogous to the original factor structure reported in Gillan et al. (2016), we initially performed exploratory factor analysis (EFA) on the combined dataset. We performed the factor analysis in R on the 209 questionnaire items, using the factanal() function, and Maximum Likelihood Estimation. We used an oblique rotation, which allows for correlations between factors. In order to assess the appropriate number of factors, we defined an empirical elbow using Cattell's criterion (Cattell, 1966), obtaining an empirical estimate of the elbow using the Cattell-Nelson-Gorsuch (CNG) test (Gorsuch & Nelson, 1981), implemented in R using the nCng() function. We then also performed factor analyses on each dataset separately in the same manner, in order to determine whether the factor structure replicates across the separate samples. Despite the CNG test being the main method used to determine the factor structure in Gillan et al. (2016), we further looked at Parallel Analysis as a method for assessing the number of factors. Here, an exploratory factor analysis was performed on 50% of the data, whilst a confirmatory factor analysis with increasing number of factors was performed on the other 50%. We compared fit metrics of Chi-squared, AIC, BIC, RMSEA and CFI.

Item reduction

In order to reduce the number of items used to measure the three factors of interest, we first split the dataset into a 75% training and 25% independent test set of data. We used lasso regularised regression to predict each participant's factor score from responses to individual questions. Performing this analysis with a range of values of the hyperparameter C, which governs the degree of regularisation, produced a model that included varying numbers of questions as predictors. We selected the best model based on a trade-off between model predictive ability and number of questions. The ability of these models to predict the true factor scores was assessed using five-fold cross validation, whereby the model was trained on 80% of the training data and tested on the remaining 20%, with this procedure repeated across combinations of training and test data and the prediction R^2 averaged across these five folds. We then used the resulting item weights on an independent test set of data, to examine the predictive accuracy on the held out set. We note that alternative approaches to item reduction exist, such as item response theory, however we chose to use a simple, data-driven approach that avoids the complexities associated with specifying appropriate statistical models of the underlying processes driving participants' responses.

External Validation

In order to test external validation of the reduced set, we took cognitive variables obtained from one of the sample datasets, (Rouault et al., 2018) which reported the results of a metacognition task. The cognitive variables consisted of accuracy, confidence level and metacognitive efficiency. We repeated the same regression analyses on these task variables using the factors derived from the reduced set of questions, plus the covariates age, gender and IQ.

Results

Descriptive Statistics

To assess whether the subsamples had comparable sum scores for each symptom domain, an ANOVA was conducted for each questionnaire, with substudy as between-subjects variable. Skewed variables, such as the OCIR, AUDIT and EAT scales were transformed using square root transformation. All questionnaires had significant substudy differences, with follow-up Tukey HSD tests indicating that the Kelley et al. (2021) dataset was generally significantly higher in symptoms than others, except for social anxiety (Table S1).

Exploratory Factor Analysis (EFA)

Based on the difference in slopes of the eigenvalues, the CNG test gave a three-factor solution to our data. These three factors were the same as those originally reported in Gillan et al., (2016); the first factor represented ‘anxious-depression’, consisting mainly of depression, apathy and anxiety symptoms; the second factor represented ‘compulsive behaviour and intrusive thought’, consisting mainly of OCD, eating disorder, schizotypy, alcohol use and impulsivity symptoms; and the third factor was a ‘social withdrawal’ factor, primarily comprising social anxiety. The EFAs conducted on each substudy independently also all gave a 3-factor solution to the data, following the same structure.

	AD	CIT	SW
	Factor 1	*Factor 2*	*Factor 3*
Depression (SDS)	**0.39 (0.23)**	0.19 (0.17)	0.07 (0.05)
Trait Anxiety (STAI)	**0.47 (0.18)**	0.21 (0.16)	0.15 (0.08)
OCD (OCIR)	-0.07 (0.12)	**0.49 (0.07)**	0.14 (0.05)
Social Anxiety (LSAS)	0.03 (0.03)	0.06 (0.09)	**0.63 (0.12)**
Impulsivity (BIS)	0.24 (0.21)	0.22 (0.20)	-0.09 (0.11)
Schizotypy (SSMS)	0.15 (0.13)	0.24 (0.18)	0.07 (0.13)
Alcohol Use (AUDIT)	0.08 (0.06)	**0.27 (0.07)**	-0.15 (0.03)
Eating Disorders (EAT-26)	-0.06 (0.09)	**0.26 (0.12)**	0.05 (0.04)
Apathy (AES)	**0.53 (0.15)**	-0.05 (0.16)	0.03 (0.09)

Table 1. Mean (SD) factor loadings for 3-factor solution for each questionnaire. Bold represents the dominant construct (above 0.25) for each factor. AD mostly represents depression, anxiety and apathy. CIT mostly represents OCD, alcohol addiction and eating disorders. SW is substantively representative of social anxiety specifically.

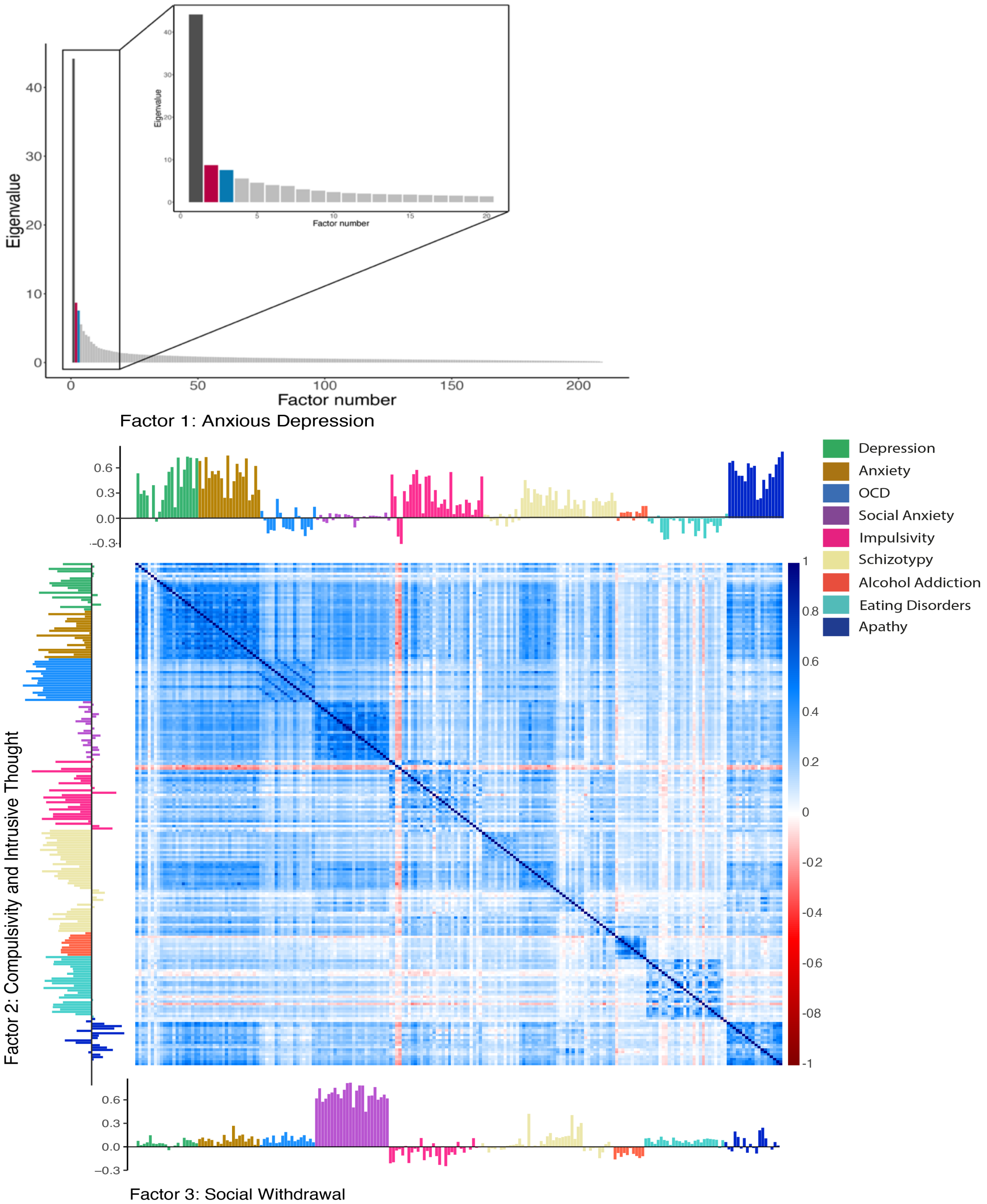

Figure 1. Trans-diagnostic factor structure for 3-factor solution suggested by Cattell's criterion. A) Eigenvalues for three factor structure, factor 1 (grey): AD (Anxious Depression), factor 2 (red): CIT (Compulsivity and Intrusive Thought); factor 3 (blue): SW (Social Withdrawal). B) Item loadings for each factor are shown above the correlation matrix, with the colour legend indicating which questionnaire they belong to.

Item Reduction

Testing a range of values for the hyperparameter tuning gave a range of prediction R^2 values (Figure 2A) for each of the three factors. We selected a hyperparameter value of 0.125 as a good trade-off between number of items and prediction accuracy, resulting in 72 items. Upon inspection of the item content, one of the items from the OCI-R ('I feel there are good and bad numbers') was deemed too culturally specific (Daniel-Watanabe et al., 2020) and therefore we removed this item from the set, resulting in 71 items in total. The cross-validated predictive accuracy scores on the training set of data were 0.92, 0.84 and 0.93 for the AD, CIT and SW factors respectively (Figure 2C). These items also had good predictive accuracy when applied to the test set (0.92, 0.84 and 0.94 for the AD, CIT and SW factors respectively).

Measure	Item Numbers	Total
Zung Depression Scale	10, 11, 12, 13, 14, 16, 17, 18, 20	9
State Trait Anxiety Inventory	1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 16, 17, 19, 20	14
Obsessive Compulsive Inventory	2, 3, 4, 6, 7, 9, 11, 12, 13, 15, 16, 18	11
Liebowitz Social Anxiety Scale	2, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24	19
Barratt Impulsivity Scale	6, 9, 13, 14, 17, 19, 20, 22, 25	9
Eating Attitudes	1, 11	2
Apathy Evaluation Scale	1, 2, 7, 8, 16, 17, 18	7
		71

Table 2. Questionnaire item numbers for the reduced set of items (71) used to approximate the three transdiagnostic factors identified from our EFA. No items were taken from the Alcohol Use Disorder or Schizotypy scale.

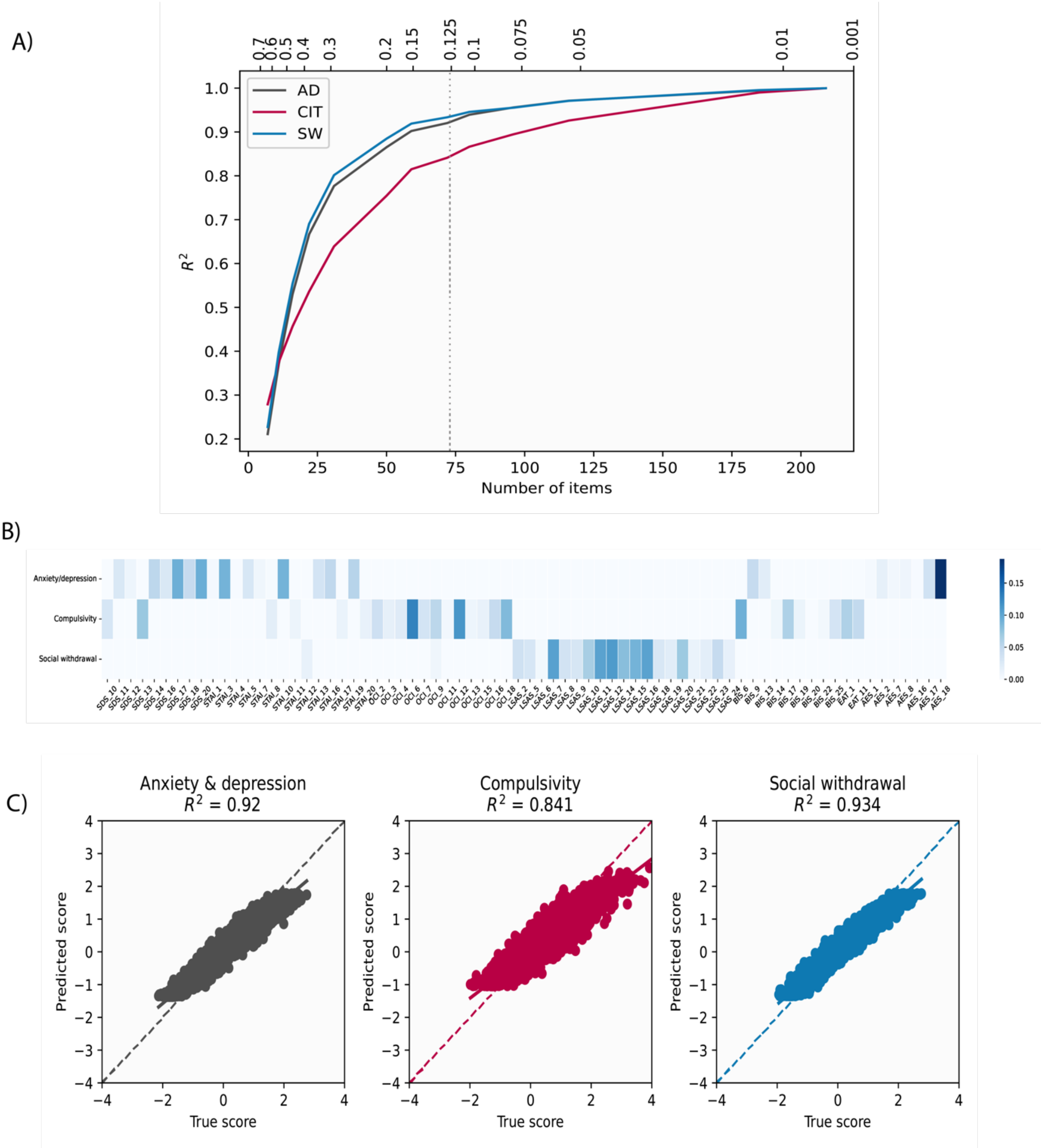

Figure 2. A) Number of items (bottom) plotted against the predictive accuracy (R^2) of each model with different hyperparameter (*C*) settings (top). A good trade-off between accuracy and number of items is at hyperparameter value 0.125, resulting in 71 items (after the exclusion of one culturally-specific item). B) Selected 71 items and regression coefficients for each factor. C) True vs predicted factor scores for the test set of data using 5-fold CV. The classifier had strong predictive accuracy overall, but showed some evidence of bias at the extremes.

Of the 71 items, no items were taken from the AUDIT (alcohol use) or SSMS (schizotypy). Inspecting the item content for each factor (Table S1, S2 and S3) suggested that the content of the AD and CIT factors could be considered slightly differently to the original factors. Although the AD factor contained some items from the STAI (thought to measure anxiety), many of the items appeared to be mood and apathy related, for example, STAI items such as 'I feel like a failure' (STAI item 5), 'I feel satisfied with myself' (STAI item 3, negatively scored) and 'I am content' (STAI item 16, negatively scored) could be considered more representative of mood than of anxiety. However, items such as 'I feel secure' (STAI item 13, negatively scored) and 'I am a steady person' (STAI item 19, negatively scored) may more closely reflect anxiety. Similarly the CIT factor contains items that more closely reflect anxiety, such as 'I am restless and can't keep still' (SDS item 13), 'I get into a state of tension or turmoil as I think over my recent concerns and interests' (STAI item 20) and 'I feel that difficulties are piling up so I cannot overcome them' (STAI item 8). Despite these considerations, we have decided to use the same factor labelling as before, as it was felt that that it is still largely appropriate.

External validation of reduced set of items

Similar to the original analysis we identified a significant relationship between confidence level and CIT trans-diagnostic factor (Figure 3, Table 3). Also similar to the original analysis, there was no relationship between any of the factors and accuracy. The original results for metacognitive efficiency showed a significant relationship to AD and CIT factors at the uncorrected level, however not when Bonferroni corrected. Here, we only find trend levels of significance at the uncorrected level and replicate the non-significant Bonferroni corrected results.

		Predicted Factors			Original		
		β	***t***	***p* (unc.)**	***β***	***t***	***p* (unc.)**
Accuracy	AD	0.04	0.64	0.52	0.03	0.57	0.57
	CIT	0.04	0.72	0.47	0.03	0.58	0.56
	SW	-0.08	-1.44	0.15	-0.06	-1.2	0.22
Confidence Level	AD	-0.23	-4.12	<.001***	-0.20	-3.91	<.001***
	CIT	0.24	4.50	<.001***	0.23	4.98	<.001***
	SW	-0.04	-4.47	0.43	-0.05	-0.93	0.35
Metacognitive Efficiency	AD	0.10	1.77	0.08	0.11	2.02	<.05
	CIT	-0.10	-1.82	0.07	-0.12	-2.43	<.05
	SW	-0.01	-0.06	0.98	-0.01	-0.16	0.87

Table 3. Regression coefficients and *p* values for original (Rouault et al., 2018) and replicated results using the predicted factor scores.

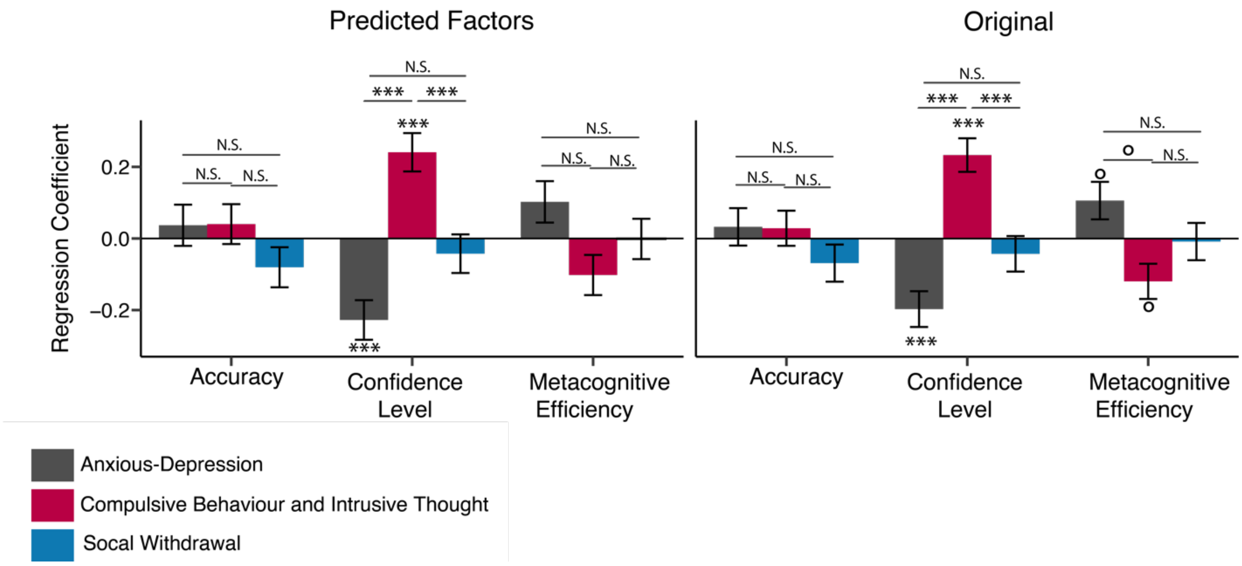

Figure 3. Results of regression model with accuracy, confidence level and metacognitive efficiency from a metacognition task (Rouault et al., 2018) predicted by the three trans-diagnostic factors. Results revealed a replication of the significant relationship between confidence level and trans-diagnostic factors. N.S. not significant, °p <.05 uncorrected, ***p <.001 corrected for multiple comparisons.

Discussion

The primary aims of this study were to optimise the measurement of three transdiagnostic factors that broadly represent anxious-depressive, compulsivity and intrusive thought and social withdrawal mental health symptoms, as well as assess their replicability across large and independent datasets. Using a large dataset of questionnaire items from four independent studies, we replicated the three-factor structure originally reported in Gillan et al. (2016). This was observed both for the pooled dataset and the independent datasets when analysed separately, supporting the replicability of these factors. Our item reduction analysis resulted in a reduced set of items that had high predictive validity for all three factors. Notably, the content validity of the AD and CIT factors may differ slightly from the original factors, particularly with respect to anxiety, as discussed below. Using our reduced set of items, we were able to reproduce published results relating the factor scores to variables from a metacognition task (Rouault et al., 2018).

The three-factor structure was the best fitting factor structure for both the pooled and subset datasets. The fact that this specific three-factor structure replicates is reassuring, as it reinforces their reliability and generalisability across datasets with different demographic characteristics and symptom score severities. The three-factor structure was obtained using Cattell's criterion (Cattell, 1966), which is not the only metric for assessing optimal factor structure. Using different metrics, we observed alternative factor structures; however the three-factor structure was more parsimonious and relative fit metrics were not deemed substantive enough to warrant the inclusion of more factors.

The reduced set of items has high predictive validity, meaning that it permits the calculation of factor scores that closely match those scores obtained from the full questionnaire set. Upon examination of the items that contribute to the calculation of each factor score (see supplementary tables 3-5) , one could consider whether the construct validity is slightly different to the original factor structure reported. The AD factor was largely represented by mood and apathy symptoms, with fewer items representing anxiety. Anxiety symptoms also featured in the CIT dimension. The SW factor remained strongly related to LSAS items. Additionally, the reduced set did not contain any items from the AUDIT (alcohol use) or SSMS (schizotypy) questionnaires, meaning those items were not particularly predictive of these 3 factor scores.

Although there are slight differences in representation of symptoms in our reduced items compared to the original factor structure, the factor names are arguably still indicative of the factor content. It has been suggested that the STAI is a non-specific measure of negative affect, rather than a specific measure of trait anxiety (Knowles & Olatunji, 2020). In a meta-analysis investigating the specificity of the STAI (Knowles & Olatunji, 2020), it was shown that both depressive and anxiety symptoms correlated with STAI-T symptom severity. Moreover, people with a depressive disorder had significantly higher STAI-T scores than individuals with an anxiety disorder, reinforcing its role in assessing mood. The STAI-T items within the anxious-depressive factor seem to plausibly assay both depressive and anxiety symptoms. Likewise, the ‘compulsivity and intrusive thought’ factor includes many items from the OCI-R and BIS scales examining intrusive thought, although it no longer contains questions related to schizotypy and alcohol use (which previously had relatively low weights on the factor). The few items from the SDS and the STAI seem to represent somatic symptoms that could be linked to anxiety. Interestingly it was previously found that cognitive and somatic anxiety symptoms can be dissociated, with opposing relations to learning from punishment vs. safety (Wise & Dolan, 2020). OCD, depression and anxiety are highly comorbid, so it can be difficult to fully disentangle these constructs (Overbeek et al., 2002), though different symptoms may differently co-occur. Therefore, although the item content for these two factors is slightly different to that originally reported, the names still reflect the dominant constructs.

In summary, we show that a three-factor solution replicates across four independent samples that administered a common set of psychopathology questionnaires. Using a machine learning approach, we extracted a set of 71 items that optimise measurement of ‘anxious-depression’, ‘compulsive and intrusive thought’ and ‘social withdrawal’ transdiagnostic dimensions.

Data and Code Availability

All data and code to reproduce the analysis from this manuscript are available on GitHub https://github.com/alexKhopkins/optimising-transdiagnostic-measurement

Acknowledgements

Alexandra Hopkins and Nura Sidarus were supported by an ESRC grant (CAMCODEP). Toby Wise is supported by a fellowship from the Anthony and Elizabeth Mellows Charitable Settlement. Claire M Gillan is funded by a fellowship from MQ: transforming mental health (MQ16IP13), a project award from Science Foundation Ireland's Frontiers for the Future Scheme (19/FFP/6418), and a European Research Council (ERC) Starting Grant (ERC-H2020-HABIT).

We would like to acknowledge Edward Patzelt for providing assistance with information about the dataset and Oliver Robinson for advice about the analysis and manuscript.

Disclosures

The authors declare no conflict of interest.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5.* (5th edition). Washington, D.C.: American Psychiatric Publishing.

Atkinson, T. M., Schwartz, C. E., Goldstein, L., Garcia, I., Storfer, D. F., Li, Y., Zhang, J., Bochner, B. H., & Rapkin, B. D. (2019). Perceptions of Response Burden Associated with Completion of Patient-Reported Outcome Assessments in Oncology. *Value in Health*, *22*(2), 225–230. https://doi.org/10.1016/j.jval.2018.07.875

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science : A Journal of the Association for Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Daniel-Watanabe, L., McLaughlin, M., Gormley, S., & Robinson, O. J. (2020). Association Between a Directly Translated Cognitive Measure of Negative Bias and Self-reported Psychiatric Symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2020.02.010

Diehr, P., Chen, L., Patrick, D., Feng, Z., & Yasui, Y. (2005). Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials. *Contemporary Clinical Trials*, *26*(1), 45–58. https://doi.org/10.1016/j.cct.2004.11.014

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485–496.

Gagne, C., Zika, O., Dayan, P., & Bishop, S. J. (2020). Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *ELife*, *9*, e61387. https://doi.org/10.7554/eLife.61387

Garner, D. M., Olmsted, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*(4), 871–878. https://doi.org/10.1017/S0033291700049163

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *ELife*, *5*, e11305. https://doi.org/10.7554/eLife.11305

Gorsuch, R. L., & Nelson, J. (1981). CNG scree test: An objective procedure for determining the number of factors. *In Annual Meeting of the Society for Multivariate Experimental Psychology.*

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Kelley, S., Mhaonaigh, C. N., Burke, L., Whelan, R., & Gillan, C. (2021). *Can the internet tell that you’re depressed? Machine learning of language use on Twitter reveals weak and non-specific predictions*. PsyArXiv. https://doi.org/10.31234/osf.io/rzx73

Knowles, K. A., & Olatunji, B. O. (2020). Specificity of trait anxiety in anxiety and depression: Meta-analysis of the State-Trait Anxiety Inventory. *Clinical Psychology Review*, *82*, 101928. https://doi.org/10.1016/j.cpr.2020.101928

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is There a General Factor of Prevalent Psychopathology during Adulthood? *Journal of Abnormal Psychology*, *121*(4), 971–977. https://doi.org/10.1037/a0028355

Liebowitz, M. R. (1987). Social Phobia. *Anxiety*, *22*, 141–173. https://doi.org/10.1159/000414022

Marin, R. S., Biedrzycki, R. C., & Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Research*, *38*(2), 143–162. https://doi.org/10.1016/0165-1781(91)90040-V

Mason, O., Linney, Y., & Claridge, G. (2005). Short scales for measuring schizotypy. *Schizophrenia Research*, *78*(2), 293–296. https://doi.org/10.1016/j.schres.2005.06.020

Overbeek, T., Schruers, K., Vermetten, E., & Griez, E. (2002). Comorbidity of obsessive-compulsive disorder and depression: Prevalence, symptom severity, and treatment effect. *The Journal of Clinical Psychiatry*, *63*(12), 1106–1112. https://doi.org/10.4088/jcp.v63n1204

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768–774.

Patzelt, E. H., Kool, W., Millner, A. J., & Gershman, S. J. (2019). The transdiagnostic structure of mental effort avoidance. *Scientific Reports*, *9*(1), 1689. https://doi.org/10.1038/s41598-018-37802-1

Ritchie, H., & Roser, M. (2018). Mental Health. *Our World in Data*. https://ourworldindata.org/mental-health

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biological Psychiatry*, *84*(6), 443–451. https://doi.org/10.1016/j.biopsych.2017.12.017

Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., McDonough, M., & Gillan, C. M. (2020). A dimensional investigation of error-related negativity (ERN) and self-reported psychiatric symptoms. *International Journal of Psychophysiology*, *158*, 340–348. https://doi.org/10.1016/j.ijpsycho.2020.09.019

Seow, T. X. F., Benoit, E., Dempsey, C., Jennings, M., Maxwell, A., O'Connell, R., & Gillan, C. M. (2021). Model-Based Planning Deficits in Compulsivity Are Linked to Faulty Neural Representations of Task Structure. *Journal of Neuroscience*, *41*(30), 6539–6550. https://doi.org/10.1523/JNEUROSCI.0031-21.2021

Seow, T. X. F., & Gillan, C. M. (2020). Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. *Scientific Reports*, *10*(1), 2883. https://doi.org/10.1038/s41598-020-59646-4

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the State-Trait Anxiety Inventory. *Palo Alto, CA: Consulting Psychologists Press*.

Wise, T., & Dolan, R. J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nature Communications*, *11*(1), 4179. https://doi.org/10.1038/s41467-020-17977-w

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines.*

Zung, W. W. K. (1965). A Self-Rating Depression Scale. *Archives of General Psychiatry*, *12*(1), 63–70. https://doi.org/10.1001/archpsyc.1965.01720310065008

Supplementary Information

	Kelley	Patzelt	Rouault	Neureka
Age (years)	30.49 (10.06)	34.95 (10.10)	35.59 (10.57)	43.44 (14.92)
Gender				
Man	312 (31%)	421 (51%)	240 (48%)	729 (29%)
Woman	668 (66%)	399 (49%)	256 (52%)	1682 (68%)
Transgender Man	6 (1%)	-	-	11 (<1%)
Transgender Woman	1 (<1%)	-	-	2 (<1%)
Non-binary	16 (1.5%)	-	-	32 (1.3%)
Other	3 (<1%)	-	-	4 (<1%)

Table S1. Demographic information for each sub-study.

	Kelley[(a)]	Patzelt[(b)]	Rouault[(c)]	Neureka[(d)]
Depression (SDS)	47.2 (11.9)[*bcd]	37.5 (10.5)[*ad]	37.5 (11.3)[*ad]	42 (11.8)[*abc]
Trait Anxiety (STAI)	51.7 (13.6)[*bcd]	39.3 (14.5)[*acd]	41.5 (13.6)[*abd]	45.7 (14.6)[*abc]
OCD (OCIR)	19.5 (13.4)[*bcd]	10 (10.9)[*acd]	13.4 (11.3)[*abd]	15.8 (12.6)[*abc]
Social Anxiety (LSAS)	30.7 (16.8)[*bc]	20 (16.6)[*acd]	22.7 (14.5)[*abd]	30.3 (16.4)[*bc]
Impulsivity (BIS)	65.8 (12.2)[*bcd]	57 (12.4)[*ad]	58.2 (12.2)[*ad]	61.8 (12.4)[*abc]
Schizotypy (SSMS)	17.4 (8.46)[*bcd]	10.4 (7.71)[*acd]	12.4 (8.07)[*abd]	13.7 (8.60)[*abc]
Alcohol Addiction (AUDIT)	7.19 (5.24)[*bcd]	4.04 (4.70)[*ad]	3.96 (4.92)[*ad]	4.6 (5.00)[*abc]
Eating Disorders (EAT-26)	11.3 (10.8)[*bcd]	8.75 (8.93)[*a]	8.37 (8.53)[*a]	8.33 (8.88) [*a]
Apathy (AES)	37.4 (9.97)[*bcd]	32.5 (9.93)[*ad]	32.3 (9.74)[*abd]	34 (9.48)[*abc]

Table S2. Mean (SD) sum scores for each questionnaire across datasets. [abcd] Refer to Tukey HSD tests for significance across datasets.

Anxious-Depression	Item Number
Zung Depression Scale	
My mind is as clear as it used to be (R)	11
I find it easy to do the things I used to (R)	12
I feel hopeful about the future (R)	14
I find it easy to make decisions (R)	15
I feel that I am useful and needed (R)	17
My life is pretty full (R)	18
I still enjoy the things I used to do (R)	20
State Trait Anxiety Inventory-Trait	
I feel pleasant (R)	1
I feel satisfied with myself (R)	3
I wish I could be as happy as others seem to be	4
I feel like a failure	5
I am calm cool and collected (R)	7
I am happy (R)	10
I feel secure (R)	13
I am content (R)	16
I am a steady person (R)	19
Barratt Impulsivity Scale	
I concentrate easily (R)	9
I plan for job security (R)	13
I am a steady thinker (R)	20
Apathy Evaluation Scale	
I am interested in things (R)	1
I get things done during the day (R)	2
I approach life with intensity (R)	7
Seeing a job through to the end is important to me (R)	8
Getting things done during the day is important to me (R)	16
I have initiative (R)	17
I have motivation (R)	18

Table S3. Items and item numbers for anxious-depression factor. Note that items marked with (R) are reverse coded.

Compulsivity and Intrusive Thought	**Item Number**
Zung Depression Scale	
I get tired for no reason	10
I am restless and can't keep still	13
State Trait Anxiety Inventory	
I feel that difficulties are piling up so I cannot overcome them	8
I have disturbing thoughts	11
Some unimportant thought runs through my mind and bothers me	17
I get in a state of tension or turmoil as I think over my recent concerns and interests	20
Obsessive Compulsive Inventory	
I check things more often than necessary	2
I get upset if objects are not arranged properly	3
I feel compelled to count while I am doing things	4
I find it difficult to control my own thoughts	5
I collect things I don't need	7
I get upset if others change the way I have arranged things	9
I am upset by unpleasant thoughts that come into my mind against my will	12
I avoid throwing things away because I am afraid I might need them later	13
I need things to be arranged in a particular way	15
I frequently get nasty thoughts and have difficulty getting rid of them	18
Barratt Impulsivity Scale	
I have racing thoughts	6
I say things without thinking	14
I act on impulse	17
I act on spur of the moment	19
I buy things on impulse	22
I spend or charge more than I earn	25
EAT	
I am terrified about being overweight	1
I am preoccupied with a desire to be thinner	11

Table S4. Items and item numbers for compulsivity and intrusive thoughts factor.

Social Withdrawal	Item Number
State Trait Anxiety Inventory	
I lack self-confidence	12
Liebowitz Social Anxiety Scale	
Participating in small groups	2
Talking to people in authority	5
Acting, performing or giving a talk in front of an audience	6
Going to a party	7
Working while being observed	8
Writing while being observed	9
Calling someone you don’t know very well	10
Talking with people you don’t know very well	11
Meeting strangers	12
Entering a room when others are already seated	14
Being the centre of attention	15
Speaking up at a meeting	16
Expressing a disagreement or disapproval to people you don’t know very well	18
Looking at people you don’t know well in the eyes	19
Giving a report to a group	20
Trying to pick up someone	21
Returning goods to a store	22
Giving a party	23
Resisting a high pressure salesperson	24

Table S5. Items and item numbers for social withdrawal factor.

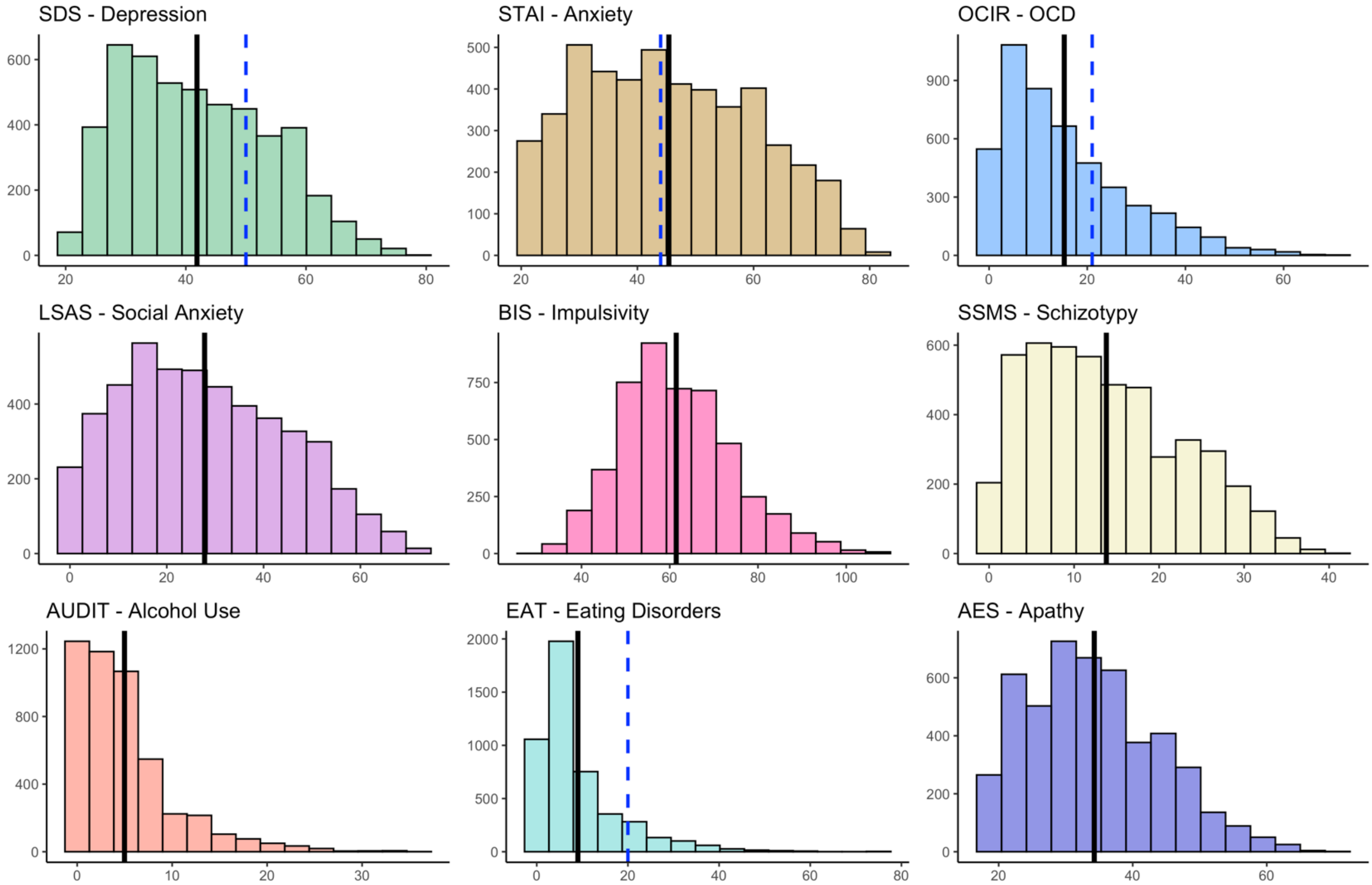

Figure S1: Distribution of symptom scores for each questionnaire for the pooled dataset. Black line indicates mean. Blue dashed line indicates suggested significant 'clinical' cut-off for highly symptomatic individuals according to questionnaire criteria. >50 is suggested clinical depression cut-off (Zung, 1965), >44 for STAI-T is suggested clinical anxiety cut-off (Ercan et al., 2015), >21 for OCI-R indicates likely presence of OCD (Foa et al., 2002), >20 for EAT-26 suggests presence of an eating disorder (Garner et al., 1982). Cut-off scores for the LSAS could not be calculated due to the use of an averaged metric rather than summing the ratings of fear and avoidance subscales.

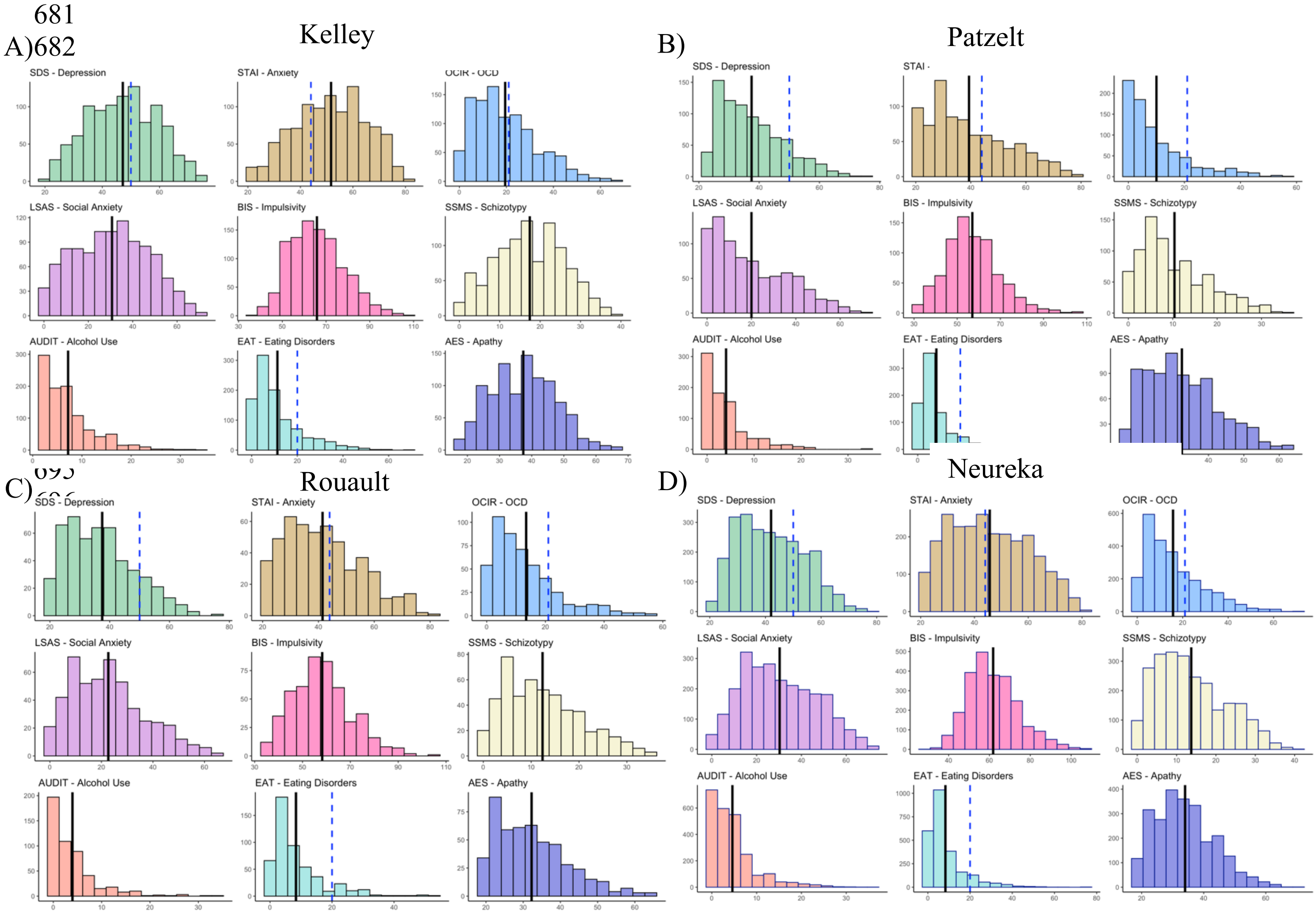

Figure S2. Distribution of symptom scores for each questionnaire for A) Kelley B) Patzelt C) Rouault D) Neureka. Black line indicates mean. Blue dashed line indicates suggested significant 'clinical' cut-off for highly symptomatic individuals according to questionnaire criteria. >50 is suggested clinical depression cut-off (Zung, 1965), >44 for STAI-T is suggested clinical anxiety cut-off (Ercan et al., 2015), >21 for OCI-R indicates likely presence of OCD (Foa et al., 2002), >20 for EAT-26 suggests presence of an eating disorder (Garner et al., 1982). Cut-off scores for the LSAS could not be calculated due to the use of an averaged metric rather than summing the ratings of fear and avoidance subscales.

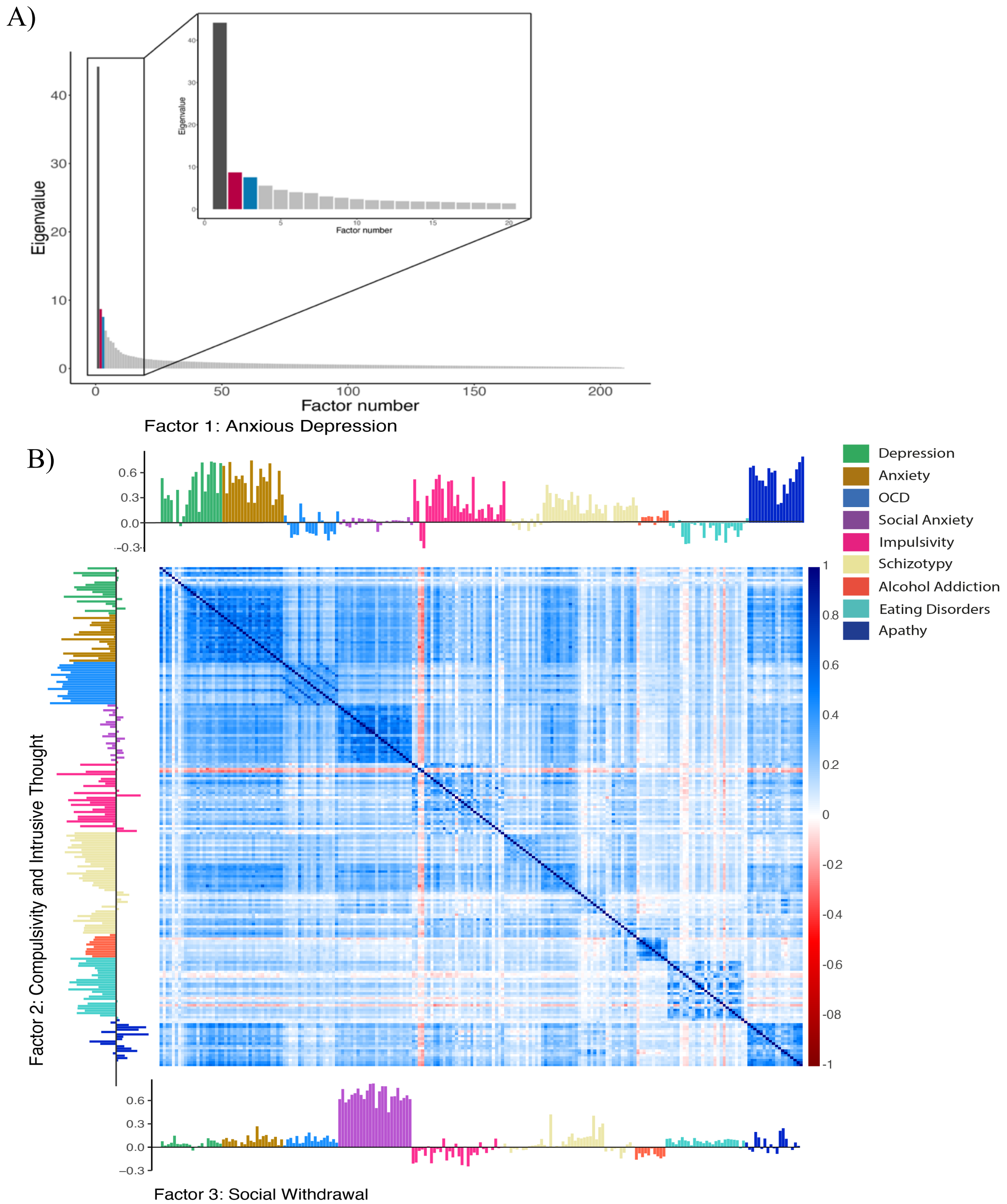

Figure 1. Trans-diagnostic 3-factor structure for the combined dataset suggested by Cattell's criterion. A) Eigenvalues for three factor structure, factor 1 (grey): AD, factor 2 (red): CIT factor 3 (blue): SW. B) Item loadings for each factor are shown above the correlation matrix, with the colour legend indicating which questionnaire they belong to.

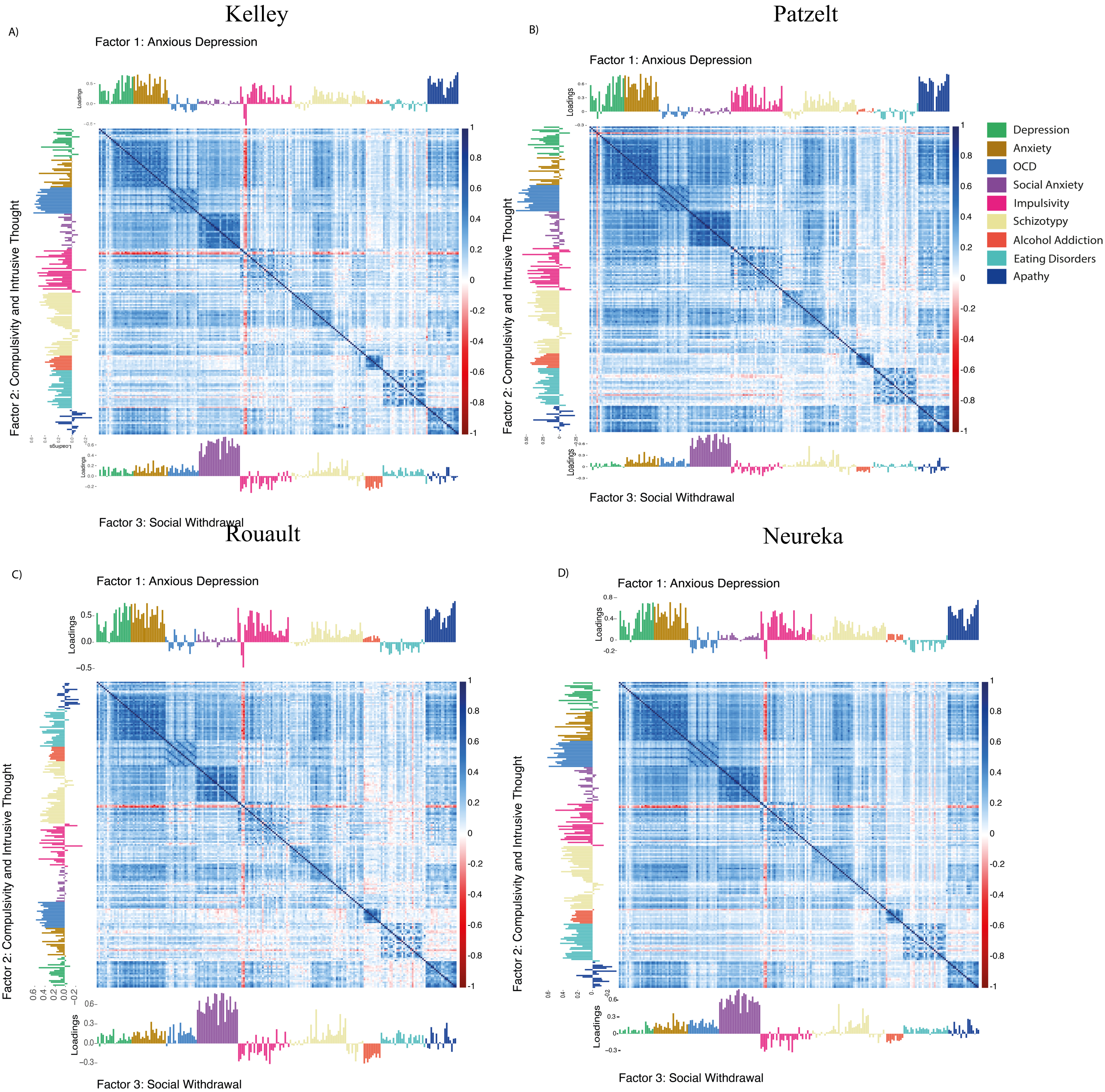

Figure S3. Trans-diagnostic factor structure for 3-factor solution suggested by Cattell's criterion for A) Kelley B) Patzelt C) Rouault D) Neureka datasets. Item loadings for each factor are shown above the correlation matrix, with the colour legend indicating which questionnaire they belong to.

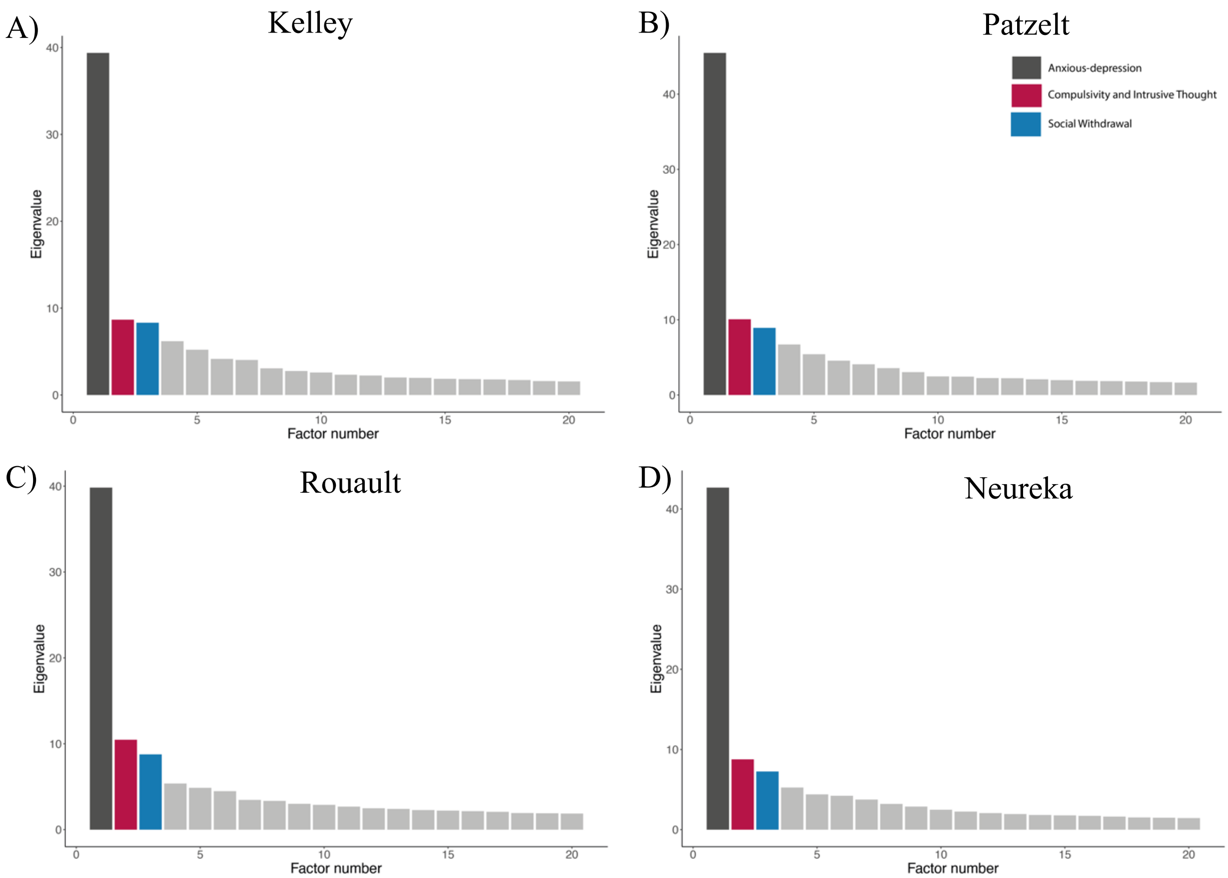

Figure S4. Eigenvalues for first 20 factor solution for A) Kelley B) Patzelt C) Rouault D) Neureka datasets. Three-factor solutions were the best solution according to Cattell's criterion. Across datasets, anxious-depression was always the factor with the largest eigenvalues, meaning it explained the largest amount of variance, followed by compulsivity and intrusive thought then social withdrawal.

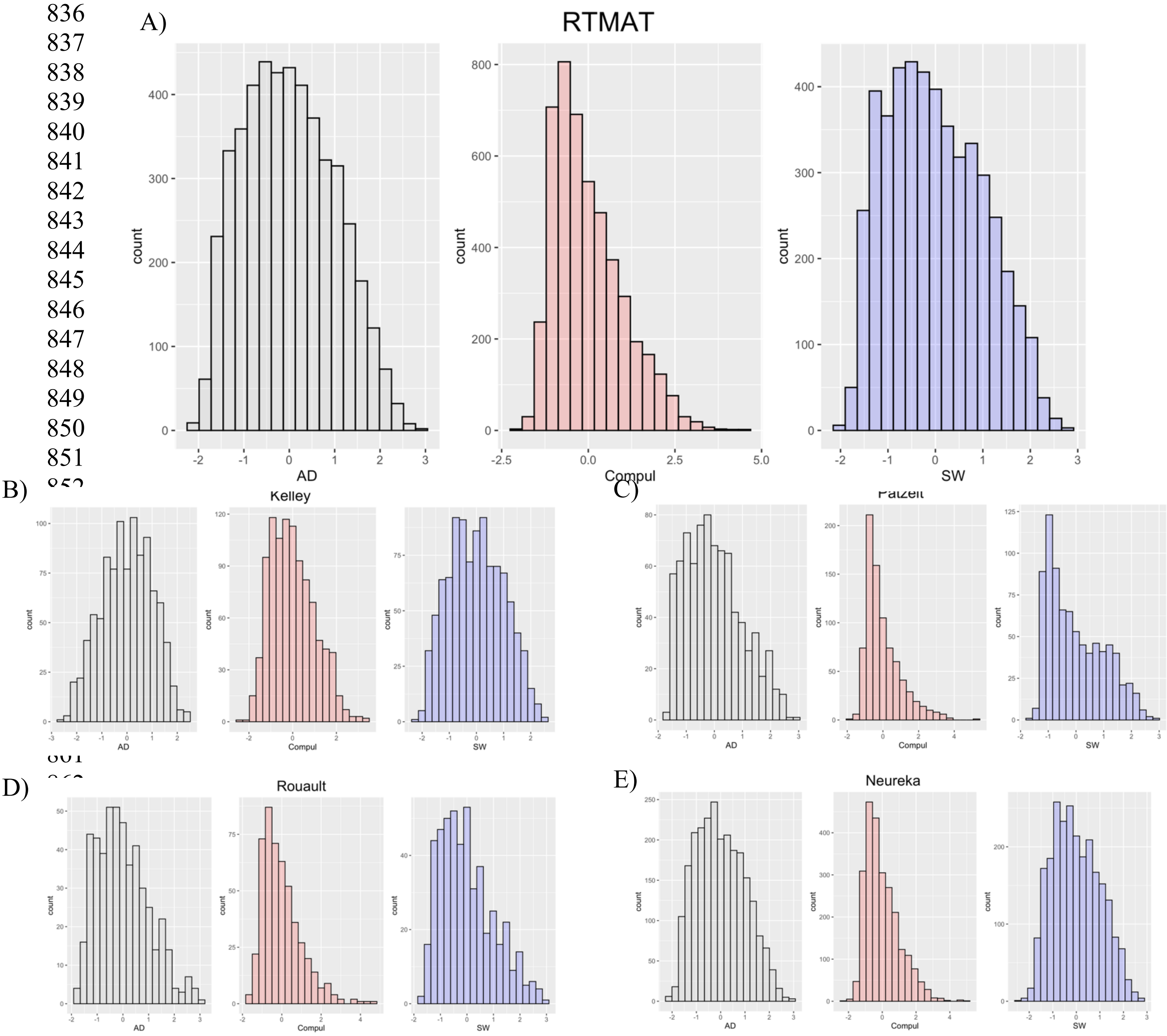

Figure S5. Distributions of predicted factor scores for A) the pooled dataset B) Kelley C) Patzelt D) Rouault E) Neureka.

Assessing different factor solutions

The 23-factor model could not be inverted, suggested the model may be unidentifiable, therefore we opted to fit up to an 8-factor model, similar to Gillan et al. (2016). As can be seen in Table S6, across the fit metrics, there was little benefit to adding more factors. s

Factor Solutions	Chi Sq	CFI	RMSEA	AIC	BIC
1	6031.55	0.84	0.1	6123.55	4242.37
2	9230.36	0.84	0.07	9388.36	3777.25
3	11782.35	0.82	0.06	11986.35	3046.48
4	16279.69	0.82	0.06	16549.69	1149.44
5	19224.53	0.83	0.05	19546.53	-1693.34
8	29010.97	0.81	0.04	29500.97	-14030.5

Table S6. Fit metrics from the CFA models, assessing multiple factor solutions to the data. Although the 3 factor model was not the best-fitting model across fit metrics in absolute terms, there was little improvement in relative terms by adding more factors.