Original Article

A shorter version of the revised Francis Psychological Type and Emotional Temperament Scales (FPTETS-R)

Andrew Village[1]

Leslie J. Francis[2]

[1]School of Humanities, York St John University, York, UK. ORCID: 0000-0002-2174-8822

[2]Centre for Educational Development, Appraisal and Research (CEDAR), University of

Warwick, Coventry, UK. ORCID: 0000-0003-2946-9980

Temperament Scales, factor structure, internal consistency reliability, psychological type

Corresponding author:
Andrew Village
School of Humanities
York St John University
Lord Mayor's Walk
YORK YO31 7EX
a.village@yorksj.ac.uk

Publication Ethics

Abstract

The Francis Psychological Type and Emotional Temperament Scales (FPTETS) operationalize the psychological-type model of personality alongside emotional temperament. The scales have been widely used in research as continuous variables that explain a wide range of religious beliefs and attitudes. The full instrument consists of five ten-item scales so a shorter version would be useful in longer surveys where completion time needs to be minimized. This study uses data from 700 Church of England clergy who completed the revised version of the FPTETS to reduce the ten-item scales to six-item scales. Ant colony optimization was found to be a better way of selecting the final items than reliability optimization alone because it balanced individual scale reliabilities with maintaining the factor structure of the overall instrument. The selected scales were validated using data from 1194 lay people from Church of England, and two samples of 884 clergy and 2765 lay people from the Episcopal Church (USA). The short scales are commended for use where the need is for continuous scale scores rather than producing psychological typologies.

Keywords: Ant colony optimization, Francis Psychological Type and Emotional Temperament Scales, factor structure, internal consistency reliability, psychological type

A shorter version of the revised Francis Psychological Type and Emotional Temperament Scales (FPTETS-R)

Introduction

The Francis Psychological Type and Emotional Temperament Scales (FPTETS) are an extension of the Francis Psychological Type Scales (FPTS) which operationalize the psychological type model of personality proposed by Carl Jung and modified by Katherine Briggs and Isabell Briggs Myers. The FPTETS added a fifth component, emotional temperament (Village & Francis, 2023a), which is closely correlated with the trait of neuroticism found in other models of personality such as Eysenck major three (Eysenck et al., 1985). Each scale consists of ten forced-choice responses that present characteristics of the paired preferences in each component of the model: orientation (extraversion versus introversion), the perceiving process (sensing versus intuition), the judging process (feeling versus thinking), attitude towards the outer world (judging versus perceiving), and emotional temperament (calmness versus volatility). The factor-structure of both the FPTS and FPTETS have been tested in several samples and shown to be generally robust (Village & Francis, 2022a, 2023a), but with a few items that loaded poorly on the expected scale. A revised version (FPTETS-R), with seven new items was shown to have a better factor structure than the original (Village & Francis, 2023b). This paper introduces a shorter version of FPTETS-R designed to be used in longer surveys and where the prime use of the scales is as continuous predictor variables in research.

Psychological type is a widely used model of personality, notably by those wanting to apply personality theory outside the narrow confines of academic research (Furnham, 2018). Perhaps because of this, it has been criticised by some who prefer trait models (McCrae & Costa, 1989). While some of the criticism is justified, trait models such as the Five-factor model also have their problems (see, for example Lloyd, 2015), and may be different ways of

conceptualising the same underlying processes that shape human personality (Lloyd, 2022).

The dimensions of the psychological type model have been shown to correlate well with the

major components of other personality models such as the three major dimensions model

(Village & Francis, 2022a) and the five factor model (McCrae & Costa, 1989). Research

using these two instruments has also shown the advantages of using scales both as trait-like

variables and in person-centred research which uses scores to create types. The disadvantage

of using just a 'variable' (trait) approach to personality has been well set out by Asendorpf

and others in regard to the five-factor model (Asendorpf, 2015; De Fruyt et al., 2002).

However, trait models such as the five-factor model are not easily converted to type models

because the creation of typologies depends on statistical methods that can yield different

results in different samples (De Fruyt et al., 2002; McCrae et al., 2006; Pittenger, 2004).

Psychological type has a stronger theoretical basis in person-centred approaches (Lloyd,

2008, 2022) and the scales scores in instruments such as the FPTS and FPTETS can also be

used in research as a manner similar to trait scores (Village, 2011).

Although the FPTS and FPTETS have been used mainly among religious samples

such as clergy or churchgoers, they are not measuring any specific religious beliefs or

characteristics. As with other psychological-type instruments such as the MBTI (Myers et al.,

1998) they could be used equally well on the general public, and have been used, for

example, on samples of students or teachers (Chaim, 2022; Francis & Lankshear, 2019).

They have been used largely in religious contexts partly because the underlying

conceptualization values all type preferences and carries no implicit or explicit pathological

or moral evaluation of dispositions, and partly because the scales lend themselves to profiling

different religious groups.

Studies where the main use of the scales is as continuous predictor variables often

include them in larger survey instruments, which may contain many items measuring other

variables. In these situations, where space is at a premium, shorter scales may improve instrument completion rates. This paper uses a fairly recent method, ant colony optimization (ACO) to create a shorter version of the FPETS-R for use in research questionnaires.

Methods of shortening psychological scales

The advantages of producing shorter versions of established psychometric scales are widely understood, and many personality scales have long and short versions (Eysenck et al., 1985; Jankowsky et al., 2020; Rammstedt et al., 2020). Shortening scales tends to reduce their internal consistency reliability as measured by indices such as Cronbach's alpha (Cronbach, 1951) or McDonald's omega (McDonald, 1999). A well-established method for selecting the best items for shorter uni-dimensional scales or subscales is to iteratively remove items that have the lowest factor loadings or lowest item-rest of scale correlations, thereby minimising the decrease in internal consistency reliability. Scales may be reduced by balancing length against reliability. Several authors (for example, Hayes & Coutts, 2020) have pointed out that these reliability indices do not indicate the factor structure of a scale, which might be influenced by more than one construct. Shortening scales using internal consistency reliability as the sole criterion can risk altering the dimensionality of scales (Schroeders et al., 2016).

Ant colony optimization (Dorigo & Stützle, 2019) has been applied to shortening personality scales by a number of authors (Jankowsky et al., 2020; Kilmen, 2021; Olaru & Jankowsky, 2022). The algorithms used in ACO mimic the way in which ants use pheromones to optimize their foraging routes around a nest. Applying ACO to shortening psychometric scales means selecting a set of criteria which identify the optimal characteristics required of the new scale. This could simply be internal consistency reliability but could also include measures of how closely the data fit the expected factor structure of the instrument. A frequent procedure is to use confirmatory factor analysis (CFA) in latent

models and optimize goodness of fit statistics (see Olaru et al., 2019 for further details and

worked examples). Optimization criteria can be combined to achieve a balance between

different measures of model fit and scale reliability, thereby offering a better chance of

achieving short scales that meet the requirements for internal consistency reliability and

which match the factor structure of the full scale.

Research question

This paper explores ways to shorten the FPTETS-R scales, which consist of ten items

measuring each of five components. The aim is to produce five short individual uni-

dimensional scales that have adequate internal consistency reliability, and which maximize

the discriminate validity of each scale. The traditional method of optimising the reliability of

individual scales will be compared to shortening scales using an ACO algorithm.

Method

Participants

The datasets were from the *Covid-19 and Church-21* survey, which surveyed readers of the

Church Times from January to July 2021 during the third national COVID-19 lockdown in

the UK (For details, see Village & Francis, 2022b). This was an online survey that included

the FPTETS-R alongside a range of other items exploring responses to the pandemic among

various Christian denominations. The subset of data used here was determined before the

analyses to be Anglican clergy and lay people who lived in England. Of 5,853 survey

responses, 2,292 identified as Anglicans living in England (790 clergy, 1502 lay), of which

1,892 (700 clergy, 1192) had no missing data and could be included in the final sample. A

modified version of the survey, containing the same items of the FPTETS-R, was also

distributed in the United States through the Episcopal Church (TEC) (Village et al., 2021a;

Village et al., 2021b). The subset of data used here was determined before the analyses to be

Episcopal clergy and lay people who lived in the USA. Of 5,229 survey responses, 4,858

identified as Episcopalians living in the US (1144 clergy, 3714 lay), of which 3,649 (884 clergy, 2765) had no missing data and could be included in the final sample. Details of the profiles of clergy and laity from each survey are shown in table 1.

- insert table 1 about here -

Instrument

The FPTETS assess preferences between the two orientations (extraversion and introversion), the two perceiving functions (sensing and intuition), the two judging functions (thinking and feeling), the two attitudes (judging and perceiving), and the two emotional temperaments (calm and volatile). (For details, see Village & Francis, 2023a, 2023c).

Overall procedure

The ten items in each component of the FPTETS-R are used to create two complementary scores in each component that sum to ten. Only one score in each of the five components is required to test reliability and factor structure: in this analysis the E, S, F, J, and V scales, which were the mirror image of responses to the I, N, T, P, and C scales respectively. Shortening the FPTETS-R consisted of two main stages: training and validation. The training stage used the sample of 700 Church of England clergy to identify the best five short scales of the instrument. The validation stage tested the reliability and factor structure of the individual scales and whole instrument on the remaining samples.

Short scales need to balance the time saved for completing the items with reliability of the new scales compared with the old. There are no universal guidelines for deciding the length of short scales and in this case we decided to produce five six-item scales, giving a total of 30 items. Trials of reducing items and testing alpha reliability suggested the latter declined in scales of less than six items, so six items per scale was chosen as the best compromise.

The ACO protocol

The ACO procedure used R- foundation software (R 4.2.1, Venables et al. (2022)) and the

instructions provided by Olaru et al. (2019). Their software (available at

https://osf.io/yx4km/) uses both internal reliability and CFA model fit statistics to decide

between scales. Reliability was measured using McDonald's omega and model fit by the

Comparative Fit Index (CFI), Root Mean Squared Error of Approximation (RMSEA), and the

Standard Root mean Square Residual (SRMR). McDonald's omega is a more general

measure of reliability that does not assume 'tau equivalence' (Hayes & Coutts, 2020). The

other measures are standard indices used to assess the goodness of fit of latent models

(Byrne, 2010). The cut-off values for these criteria were those widely used: CFI $\geq$ .95;

RMSEA $\leq$ .06; SRMR .06; $\omega \leq$ .70). These values were transformed using a logit

transformation as suggested by Olaru et al. (2019), which allows them to be combined

meaningfully. Scales were therefore selected using a balance of internal consistency

reliability and CFA.

   The CFA function used a weighted least squares estimator (WLSMV) rather than the

default maximum likelihood, because it is a better option for categorical scale variables. The

ACO function is a probabilistic procedure that selects starting items at random, so it does not

necessarily identify the best procedure. Olaru et al. (2019) recommend running the procedure

several times until the solution is replicated across different runs. We ran the program five

times, and the same solutions were returned in at least four out of five runs.

The training stage

The training stage employed three different ways to identify the optimal six-item scales:

1. *Reliability optimization*. The reliability procedure in SPSS 28 was used to

   determine Cronbach's alpha, corrected item-total correlations (CITC), and the

   alpha-if-item-removed for each of the five ten-item scales. For each scale, items

   with the lowest CITC were removed and the procedure was repeated until six

items remained. CITC scores tend to be closely correlated with factor loadings

and there was no additional benefit in using factor loading rather than CITC.

2. *ACO of individual scales*. The five component scales of the instrument were tested

   separately using the ACO protocol. This assessed the reliability of each

   component scale alongside the fit of a model that assumed the items were

   assessing a single latent construct.

3. *ACO of all five scales*. Testing scales individually did not test the discriminate

   validity of the different scales. Testing all scales at the same time took much

   longer but was more likely to select an instrument that maintained the overall

   factor structure of the original FPTETS-R. The procedure selected those items in

   each of the five scales that showed the best fit to a five-factor model alongside the

   best average internal consistency reliability of the competent scales.

   The scales identified for each of these procedures were then compared by

reliability analysis using both alpha and omega indices calculated in SPSS

(McDonald's omega was calculated using the macro provided by Hayes and Coutts

(2020)) and by CFA using Mplus (Muthén & Muthén, 1998-2017).

The validation stage

The items selected for the six-item scales were tested on the three remaining samples: Church

of England laity, Episcopal Church clergy, and Episcopal Church laity. In each case, scales

were tested individually using reliability and CFA, and the overall fit of the five-factor model

tested using CFA. The aim was to test whether the scales produced a satisfactory balance

between reliability and factor structure. Two other types of validation were applied to the

Church of England sample, using clergy and laity separately. The first was to test correlations

between the full and the shortened scales, the second was to test the short scales as predictors

of other variables collected in the survey. Previous studies have examined the predictive

power of psychological type variables on changes in positive and negative affect during the

pandemic (Village & Francis, 2022b). The two measures of affect were correlated against

both the full and the shortened scales and the results compared.

Results

Item endorsement for the FPTETS-R in the training sample

The items in the full instrument, and their endorsement in the training sample are shown in

Online Resource Table S1. Scores for Choice 1 responses were used in the scale shortening

procedure.

Selecting the best short scale

Details of reliabilities for the full 10-item scales and three versions of the 6-item scales for

the Church of England clergy sample are shown in Online Resource S2. Alpha reliabilities

for the 10-item scales were in the range .761 to .864, with omega reliabilities being similar, or

slightly higher, in each case (Online Resource S2a). The CFA model fits for individual scales

suggested that the extraversion and judging scales fitted a uni-dimensional scale model well

(extraversion: CFI = .995, RMSEA = .042, and SRMR = .040; judging: CFI = .984, RMSEA

= .052, and SRMR = .052;), with other scales showing slightly poorer fit (Online Resource

S2a). The overall model had a satisfactory RMSEA index (.051), but the CFI and TLI were

below the acceptable limits of .95 (CFI = .875, TFI = .869).

The three methods of shortening the scales selected slightly different six-items sets,

which had varying balances between scale reliability and CFA model fit (Online Resource

S2b-d). Reliability optimization of individual scales (Online Resource S2b) produced alpha

reliabilities in the range .75 to .86 and acceptable fits in all but the sensing scale. The overall

model fit was also good (CFI = .96; RMSEA = .04) suggesting it maintained the five-factor

model. Applying ACO to individual scales (Online Resource S2c) produced a slightly lower

range of alpha reliabilities (.72 to .81) but much better CFA model fits. This was as would be

expected since this procedure was optimising both reliability and the factor structure. When

the selected items were tested together for in five-factor latent model the fit was less good,

and none of the indices reached acceptable threshold limits. This method produced good

individual scales, but at the cost of poor factorial structure in the overall model. Selecting

scales using optimization of the full model (Online Resource S2d) produced reasonable alpha

reliabilities for each scale (.68 to .84) and good model fit for each individual scale. There was

an exception for feeling scale, where RMSEA = .065, which is slightly larger than the

minimum threshold of .06. The overall model fit was good according to all indices (CFI =

.997, TFI = .974, RMSEA = .031, and SRMR = .067), and the selected items seem to fit the

five-factor model slightly better than those selected by reliability optimization.

Overall, the best short scales seemed to be those selected by ACO of all scales:

although the alpha reliabilities were slightly lower than the reliability optimization versions,

the factorial structure of all five sets of items showed better discriminant validity between the

individual scales. The items from this model (Online Resource S2d) were used in subsequent

validation.

Validation of the selected short scales

Online Resource S3 shows how the selected short scales functioned in Church of England

laity sample (S3a), the Episcopal Church clergy sample (S3b) and the Episcopal Church laity

sample (S3c). The Church of England laity showed satisfactory alpha reliabilities (.66 to .81)

and good model fit for the individual scales. The fit for the five scales together passed the

threshold for RMSEA (.045) but was slightly short of the thresholds for the other indices. In

the Episcopal Church, the clergy sample showed slightly higher reliabilities and better CFA

model fits for both individual and overall scales compared with the lay sample. These results

suggest that the short scales generally worked well in these three samples, though it might be

that they perform slightly better among clergy than among lay churchgoers. Even so, they

met required reliabilities ($\alpha > .65$) in 14 or 15 tests, required CFI ($> .95$) in 17 of 18 tests, and

required RMSEA ($\leq .06$) in 16 of 18 tests. In all three samples the fit statistics for the short

scales were better than those for the original 10-item scales (see S3a-c).

Correlations of long and short scales

Correlations of summated scores for the ten-item and six-item scales in the training and

validation samples are shown in table 2. Pearson correlation coefficients ranged from .92 to

.96 and were similar for a given scale across all four samples. Correlations were highest for

the extraversion and judging scales and lowest for the sensing and volatility scales.

- insert table 2 about here –

Predictive validity of the short scales

This was tested on two measures of psychological affect change collected from the Church of

England sample (table 3). Two measures of affect (positive and negative) were correlated

against scale scores for clergy and laity separately. In 17 of the 20 tests the long and short

scales would have led to the same conclusions in terms of the approximate strength and

significance of the correlations. In two cases the short scales would have led to a Type II

error (false negative) and in one case to a Type I error (false positive). These were all cases

where the trends suggested marginal statistical significance.

- insert table 3 about here -

Properties of the short scales

 The item rest-of-scale correlations for the short scale items are shown in Online Resource S4

for each of the four samples. The average value for all items ranged from .50 to .54 across the

four samples and only three items ('Concerned about details', 'Present realities', and 'Seek

for peace') had average correlations of less than .35. The range suggested that the short items all loaded satisfactorily on their respective scale while not being so closely correlated that the construct range was too narrowed by the process of shortening.

Discussion

This paper tested three methods of shortening the revised version of the Francis Psychological Type and Emotional Temperament Scales (FPTETS-R) on a sample of 700 clergy from the Church of England. ACO gave the best compromise between individual scale reliabilities and the factor structure of the overall instrument. Validation suggested that the short scales retained good internal consistency reliability and factor structure across samples of clergy and lay people from two different countries, which represented two distinct groups of Anglicans from two different cultures. The short scales gave similar predictions of correlations with psychological wellbeing as did the full scales, with only two Type II errors and one Type I error across 20 correlations.

The benefits of using ACO over reliability optimization were marginal for individual items but more obvious in terms of maintaining (and clarifying) the factor structure of the five-factor instrument. When scales were selected individually reliability optimization gave short scales with better reliabilities but poorer model fit statistics than selecting scales individually using ACO. When these two sets of differently selected short scales were tested for overall model fit, the reliability-optimized scales showed slightly better fit statistics than the ACO selected scales. However, when scales were selected using ACO on the full model from the start, the final fit statistics were better than the full model selected by reliability optimization because the algorithm balanced reliability and model fit at each iteration. The ACO algorithm can test a huge range of combinations of items and select ones that show the

clearest separation of factors (minimizing cross loading of items). Although it takes more

computation and specialist software, it may produce a better overall outcome.

The short scales were developed to be used as continuous variables in research studies

that use psychological type to predict a range of attitudes, beliefs, and behaviours among

religious samples. The original 10-item scales have been widely used to assign preferences in

each of the four components of the type model. When this is done, the type profiles are

known to align with those that are calculated using the Myers-Briggs Type Inventory, MBTI

(Francis & Village, 2022). The short scales have not yet been shown to have concurrent

validity with the MBTI, so it may be unwise to use them where the aim is to assign

respondents to one of the 16 psychological types.

References

Asendorpf, J. B. (2015). Person-centered approaches to personality. In M. S. Mikulincer, Phillip R. Cooper, M. Lynne Larsen, Randy J. (Ed.), *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences.* (pp. 403-424). American Psychological Association. https://doi.org/10.1037/14343-018

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Chaim, W. (2022). The Polish adaptation of the Francis Psychological Type and Emotional Temperament Scales (FPTETS): an overview of recent research. *Mental Health, Religion & Culture*, *25*(9), 910-920. https://doi.org/10.1080/13674676.2022.2035338

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

De Fruyt, F., Mervielde, I., & Van Leeuwen, K. (2002). The consistency of personality type classification across samples and five-factor measures. *European Journal of Personality*, *16*(1_suppl), S57-S72. https://doi.org/10.1002/per.444

Dorigo, M., & Stützle, T. (2019). Ant colony optimization: Overview and recent advances. In M. Gendreau & J.-Y. Potvin (Eds.), *Handbook of metaheuristics* (pp. 311-351). Springer International Publishing. https://doi.org/10.1007/978-3-319-91086-4_10

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences*, *6*(1), 21-29. https://doi.org/10.1016/0191-8869(85)90026-1

Francis, L. J., & Lankshear, D. W. (2019). Work-related psychological health and psychological type: A study among primary school teachers in Wales. *Research Papers in Education*, *34*(4), 391-404. https://doi.org/10.1080/02671522.2018.1452959

Francis, L. J., & Village, A. (2022). The Francis Psychological Type Scales (FPTS): Factor structure, internal consistency reliability, and concurrent validity with the MBTI. *Mental Health, Religion & Culture*, 1-21. https://doi.org/10.1080/13674676.2022.2041584

Furnham, A. (2018). The great divide: Academic versus practitioner criteria for psychometric test choice. *Journal of Personality Assessment*, *100*(5), 498-506. https://doi.org/10.1080/00223891.2018.1488134

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross–cultural personality assessment using ant colony optimization. *European Journal of Personality*, *34*(3), 470-485. https://doi.org/10.1002/per.2260

Kilmen, S. (2021). Designing a shorter form of the big three perfectionism scale: An application of ant colony optimization. *Journal of Psychoeducational Assessment*, *40*(2), 190-201. https://doi.org/10.1177/07342829211055880

Lloyd, J. B. (2008). Myers-Briggs theory: How true? How necessary? *Journal of Psychological Type*, *68*(6), 43-50.

Lloyd, J. B. (2015). Unsubstantiated beliefs and values flaw the Five-Factor Model of Personality. *Journal of Beliefs & Values*, *36*(2), 156-164. https://doi.org/10.1080/13617672.2015.1033209

Lloyd, J. B. (2022). Seeking truth in personality science: Reconciling trait theory and psychological type. *Mental Health, Religion & Culture*, *25*(9), 817-828. https://doi.org/10.1080/13674676.2022.2158794

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the Five-Factor Model of personality. *Journal of Personality*, *57*(1), 17-40. https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

McCrae, R. R., Terracciano, A., Costa Jr., P. T., & Ozer, D. J. (2006). Person-factors in the California Adult Q-Set: closing the door on personality trait types? *European Journal of Personality*, *20*(1), 29-44. https://doi.org/https://doi.org/10.1002/per.553

McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press. https://doi.org/https://doi.org/10.4324/9781410601087

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus users guide. Eighth edition*. Muthén & Muthén.

Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator* (3rd ed.). Consulting Psychologists Press.

Olaru, G., & Jankowsky, K. (2022). The HEX-ACO-18: Developing an age-invariant HEXACO short scale using ant colony optimization. *Journal of Personality Assessment*, *104*(4), 435-446. https://doi.org/10.1080/00223891.2021.1934480

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant colony optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality*, *33*(3), 400-419. https://doi.org/10.1002/per.2195

Pittenger, D. J. (2004). The limitations of extracting typologies from trait measures of personality. *Personality and Individual Differences*, *37*(4), 779-787. https://doi.org/https://doi.org/10.1016/j.paid.2003.10.006

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*, *36*(1), 149-161. https://doi.org/10.1027/1015-5759/a000481

Schroeders, U., Wilhelm, O., & Olaru, G. (2016). Meta-heuristics in short scale construction: Ant colony optimization and genetic algorithm. *PloS one*, *11*(11), e0167110. https://doi.org/10.1371/journal.pone.0167110

Venables, W. N., Smith, D. M., & the R Core Team. (2022). *An introduction to R*. https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf

Village, A. (2011). Introduction to special section: Psychological type and Christian ministry. *Research in the Social Scientific Study of Religion*, *22*, 157-164. https://doi.org/https://doi.org/10.1163/ej.9789004207271.i-360.28

Village, A., & Francis, L. J. (2022a). Factorial structure and validity of the Francis Psychological Type and Emotional Temperament Scales (FPTETS). *Mental Health, Religion & Culture*, *25*(9), 897-909. https://doi.org/10.1080/13674676.2022.2026311

Village, A., & Francis, L. J. (2022b). Psychological wellbeing and sources of support for Church of England clergy and laity during the third national COVID-19 lockdown. *Rural Theology*, *20*(2), 72-90. https://doi.org/10.1080/14704994.2022.2120972

Village, A., & Francis, L. J. (2023a). Introducing the Francis Psychological Type and Emotional Temperament Scales (FPTETS): A study among church leaders and church members. *Religion, Brain & Behavior*, 1-21. https://doi.org/10.1080/2153599x.2022.2160800

Village, A., & Francis, L. J. (2023b). Revising the Francis Psychological Type and Emotional Temperament Scales (FPTETS). *Mental Health, Religion & Culture*, 1-10. https://doi.org/10.1080/13674676.2023.2232330

Village, A., & Francis, L. J. (2023c). Revising the Francis Psychological Type and Emotional Temperament Scales (FPTETS) [Manuscript submitted for publication]. School of Humanties, York St John University.

Village, A., Francis, L. J., & McKenzie, J. (2021a). *Covid-19 & Church-21: Paying attention to the voices of lay people in The Episcopal Church*. https://www.yorksj.ac.uk/media/content-assets/document-directory/documents/TEC-Covid-19--Church-21-Laity-report.pdf

Village, A., Francis, L. J., & McKenzie, J. (2021b). *The Episcopal Church in the USA during the coronavirus pandemic: Clergy and churches*. https://www.yorksj.ac.uk/media/content-assets/document-directory/documents/TEC-Covid-19--Church-21-Churches-and-Clergy-report.pdf

Open Science

We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact p values, effect sizes, and 95% confidence or credible intervals.

Open Data: The datasets generated during and/or analysed during the current study are available on request from the corresponding author.

Open Materials: The information needed to reproduce all of the reported methodology is available on request from the corresponding author.

Preregistration of Studies and Analysis Plans: This study was not preregistered.

Open Analytic Code: The scripts, code, and outputs needed to reproduce all of the reported results are available on request from the corresponding author.

The online supplementary tables (S1 to S4) are available at:
https://osf.io/rasdf/?view_only=1adad253ff9042f5aceb6378bafc4597

Appendix: suggested presentation of the FPTETS-RS

Note column one is not part of the presented scale, but indicated which component the item related to: O Orientation; P Perceiving process; J Judging process; A Attitude to the outer world; M emotional temperament.

> *Please tick (✔) ONE box next to that characteristic which is **closer** to the real you, even if you feel both characteristics apply to you.*
> *Tick the characteristic that reflects the real you, even if other people see you differently.*

PLEASE COMPLETE EVERY QUESTION

P	**Do you tend to be more** .... interested in facts	☐	or	☐	interested in theories
J	**Are you concerned more** .... with empathy	☐	or	☐	with objectivity
M	**Do things** ....rarely worry you	☐	or	☐	often worry you
O	**Are you more** .... private	☐	or	☐	sociable
P	**Are you more** .... inspirational	☐	or	☐	practical
J	**Are you more** .... analytic	☐	or	☐	sympathetic
A	**Are you more** .... structured	☐	or	☐	open-ended
O	**Do you prefer** ...... having many friends	☐	or	☐	a few deep friendships
J	**Do you prefer** .... feeling	☐	or	☐	thinking
M	**Do you mostly** .... feel secure	☐	or	☐	feel insecure
O	**Are you** .... energised by others	☐	or	☐	drained by too many people
O	**Are you** ... happier working alone	☐	or	☐	happier working in groups
P	**Do you tend to be more** ... concerned for meaning	☐	or	☐	concerned about details
J	**Do you tend to be more** .... logical	☐	or	☐	humane
A	**Do you tend to be more** .... orderly	☐	or	☐	easy going
M	**Do you tend to** ... feel guilty about things	☐	or	☐	feel guilt-free
O	**Are you more** .... talkative	☐	or	☐	reserved
P	**Are you more** .... sensible	☐	or	☐	imaginative
A	**Are you more** .... spontaneous	☐	or	☐	organised
M	**Are you generally** .... at ease	☐	or	☐	anxious about things
O	**Are you mostly** .... an introvert	☐	or	☐	an extravert
P	**Do you mostly focus on** .... present realities	☐	or	☐	future possibilities
A	**Are you mostly** .... leisurely	☐	or	☐	punctual

M	**Do you tend to** ……….………… stay calm	☐	or	☐	panic easily
J	**Do you** ….………... seek for truth	☐	or	☐	seek for peace
A	**Do you** . dislike detailed planning	☐	or	☐	like detailed planning
P	**Are you** ……….……. up in the air	☐	or	☐	down to earth
J	**Are you led mostly by** …………..….. your heart	☐	or	☐	your head
M	**Are you mostly** ..… unbothered by things	☐	or	☐	easily bothered by things
A	**Are you** ………..…… systematic	☐	or	☐	casual

Table 1 Profile of samples

		Church of England		Episcopal Church	
		Clergy	Laity	Clergy	Laity
	$N =$	700	1194	884	2765
		%	%	%	%
Sex	Female	44	62	53	69
	Male	56	38	47	31
Age	20s	< 1	2	< 1	1
	30s	5	4	4	3
	40s	13	8	9	4
	50s	25	17	16	11
	60s	38	32	31	30
	70s	17	31	30	39
	80s+	2	6	10	13

Table 2 Correlations of full (10-item) and short (6-item scales) in the training and validation samples

	Church of England		The Episcopal Church	
	Clergy	Laity	Clergy	Laity
$N =$	700	1194	884	2765
Extraversion	.962	.945	.959	.947
Sensing	.923	.916	.920	.916
Feeling	.940	.940	.940	.942
Judging	.957	.953	.956	.956
Volatility	.938	.929	.922	.922

Table 3 Predictive validity of full (10-item) and short (6-item scales) in the Church of England samples

		Ten-item scale		Six-item scale	
		r	*p*	*r*	*p*
a) Positive affect					
Laity	*Extraversion*	*.063*	*.029*	*.055*	*.059*
	Sensing	-.126	.001	-.135	.001
	Feeling	*.058*	*.046*	*.041*	*.164*
	Judging	-.052	.074	-.050	.085
	Volatility	-.300	.001	-.295	.001
Clergy	Extraversion	.150	.001	.131	.001
	Sensing	-.053	.166	-.069	.067
	Feeling	.042	.270	.029	.449
	Judging	-.051	.176	-.059	.122
	Volatility	-.259	.001	-.249	.001
b) Negative affect		*r*	*p*	*r*	*p*
Laity	Extraversion	-.050	.087	-.052	.075
	Sensing	.065	.025	.072	.013
	Feeling	-.025	.388	-.004	.886
	Judging	.047	.108	.048	.100
	Volatility	.373	.001	.354	.001
Clergy	Extraversion	-.015	.696	.004	.917
	Sensing	*-.062*	*.102*	*-.087*	*.021*
	Feeling	.033	.379	.057	.132
	Judging	-.027	.483	-.030	.433
	Volatility	.356	.001	.349	.001

Note. Rows in italic indicate cases where long and short scales indicated a difference in significance at the 5% level.