Is the construct of statistics anxiety distinct from maths anxiety?

Pre-registration of phase one.

Many non-STEM students of introductory statistics report experiencing high levels of anxiety about the subject, which might have a detrimental effect upon their grades (for reviews see Chew & Dillon, 2014; Onwuegbuzie & Wilson, 2003). Statistics anxiety has most recently been defined as “a negative state of emotional arousal experienced by individuals as a result of encountering statistics in any form and at any level [...] and is related to but distinct from mathematics anxiety” (Chew & Dillon, 2014, p. 199). Few empirical studies have tested whether statistics anxiety and mathematics anxiety are separate constructs, but the ones that have broadly agree that whilst strongly related to maths anxiety, statistics anxiety contains a unique component (e.g. Birenbaum & Eylath, 1994; Paechter et al., 2017; Zeidner, 1991). However, these studies have methodological flaws such as using unvalidated measures (Birenbaum & Eylath, 1994; Zeidner, 1991) and comparing statistics anxiety with maths anxiety measured at a different time (Paechter et al., 2017). The present research is a multi-faceted exploration of the constructs of maths and statistics anxiety. The investigation will be broken down into two phases, each of which examines the question in different ways (see Figure 1 and descriptions below). This pre-registration outlines the first phase.

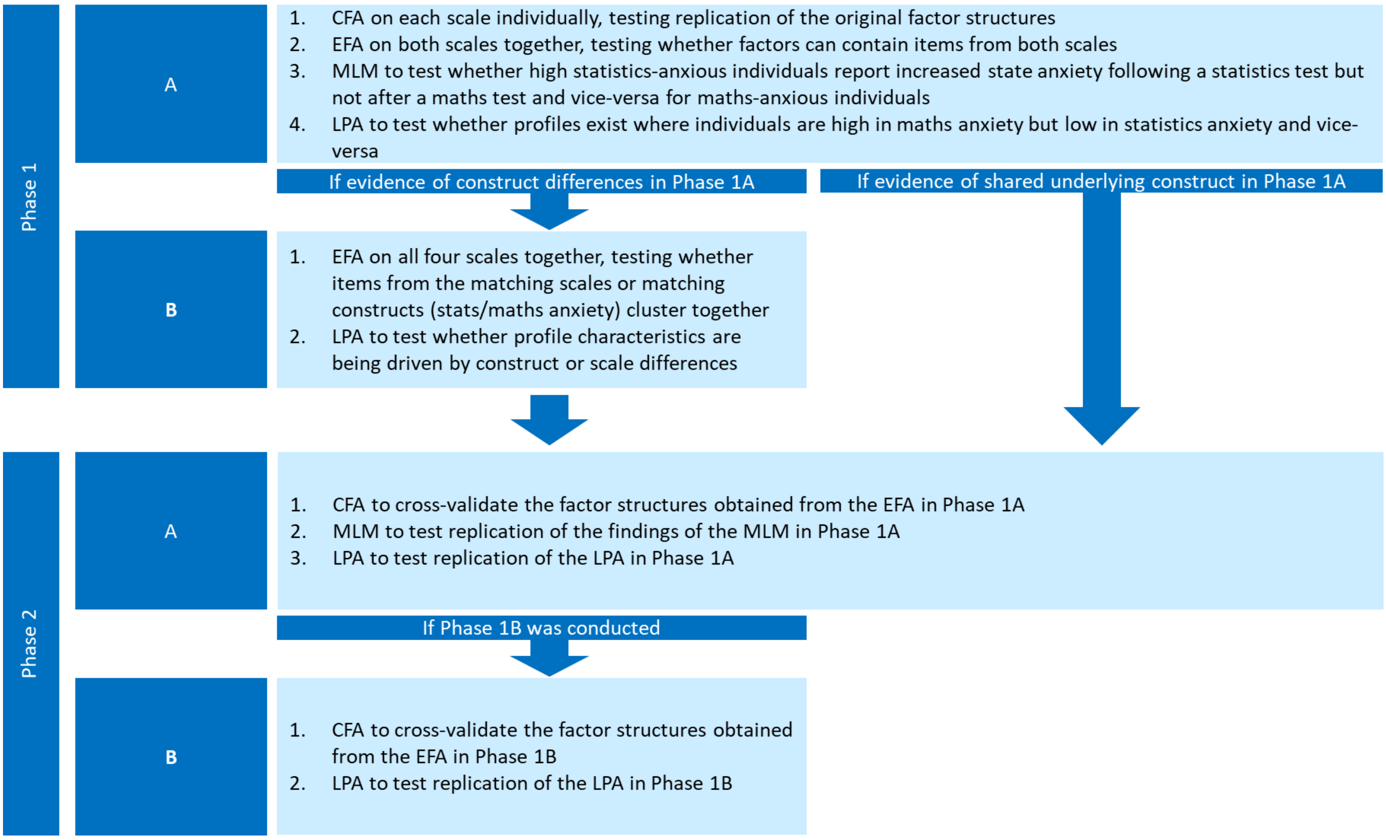

Figure 1: An overview of the analyses to be conducted in each phase of this study.

Phase 1A

Phase 1A will involve a series of analyses of the same dataset. First, a confirmatory factor analysis (CFA) will examine whether the factor structures of the Statistics Anxiety Rating Scale (STARS; Cruise et al., 1985) and Revised Maths Anxiety Rating Scale (R-MARS; Baloğlu & Zelhart, 2007) are replicated in our sample. Second, an exploratory factor analysis (EFA) will be conducted upon all items from the STARS and R-MARS to identify whether and which items from the two scales load on to the same factors. If items from the STARS and R-MARS load onto common factors it suggests that they are indicators of a construct common to both measures.

Next, we will test whether the two constructs show specificity in their predictive validity. That is, if the two constructs are meaningfully different, scores on statistics anxiety should predict state anxiety (anxiety symptoms at a specific point in time) following a statistics test but not necessarily following a maths test, whereas maths anxiety should predict the opposite (i.e. two single dissociations)

In the final stage of phase 1A, a latent profile analysis (LPA) will examine whether there are individuals that report high levels of statistics anxiety but low maths anxiety and vice-versa. Such profiles would be indicative of differences between the two constructs.

If there is evidence from the phase 1A analyses that statistics anxiety and maths anxiety could be unique constructs (or the evidence is inconsistent), analysis will continue to phase 1B. If the phase 1A results indicate that the two constructs are the same, then analysis will move directly to phase 2.

Phase 1B

Phase 1B tests the possibility that any differences between the constructs of maths and statistics anxiety indicated in phase 1A are a function of the measurement scales used, rather than reflecting a true difference in the underlying constructs. The STARS and R-MARS consist of different combinations of subscales each measuring different triggers of subject-specific anxiety. The STARS has three subscales that pertain to test and class anxiety, interpretation anxiety, and fear of asking for help, whereas the R-MARS has three subscales relating to maths test anxiety, numerical task anxiety, and mathematics course anxiety. It is plausible that some of the variance that each scale captures is not specific to the construct being measured (e.g. items relating to asking for help tap a general fear of asking for help, rather than an anxiety specific to statistical tasks). In the absence of comparable items across scales, apparent differences between maths and statistics anxiety as measured by the R-MARS and STARS might reflect the non-shared variance derived from their respective items tapping different aspects of general anxiety. For example, the R-MARS has no items relating to asking for help, therefore, the STARS's items tap an aspect of general anxiety (asking for help) not tapped by R-MARS.

To rule this possibility out, the EFA from phase 1A will be repeated but, as well as the original STARS and R-MARS scales, these analyses will also include versions of each scale that have been modified to reflect the other construct (i.e. there will be a version of the STARS that asks about maths instead of statistics - thus creating a fear of asking for help about maths scale, for example - and vice-versa). Across the four scales, if statistics and maths items cluster onto different factors to each other it suggests that the differences between maths and statistics anxiety from phase 1A reflect differences in the underlying constructs. However, if statistics and maths items cluster onto the same factors to each other it suggests that the differences between maths and statistics anxiety from phase 1A reflect the differences in the scales themselves rather than the underlying constructs.

The LPA from phase 1A will also be repeated with the inclusion of the modified scales. If the two maths perform similarly to one another (i.e. have similar scores within a given estimated profile) and the two statistics scales also perform similarly then this will be indicative that the type of anxiety (maths or statistics) is what is driving the characteristics of the profiles and, therefore, supportive of distinctiveness between the constructs of statistics anxiety and maths anxiety. If the two R-MARS or the two STARS scales perform similarly to one another in a given profile, then this will be indicative that the scales themselves are what are driving the profile differences and, therefore, suggestive that statistics anxiety and maths anxiety are not meaningfully different.

Regardless of the results of phase 1B, the analysis will then move on to phase 2.

Phase 2 (A & B)

Phase 2 will cross-validate the factor structures obtained in the phase 1A EFA (and phase 1B, if conducted) and test replication of the double dissociation and LPA with different samples. We are still organising data collection for these samples so a full pre-registration for phase 2 will be posted at a later date.

Hypotheses

Due to insufficient empirical evidence or relevant theory, no directional predictions are being made.

Method

Participants

Phase 1A and 1B participants are undergraduate psychology students from a UK university that have begun or completed at least one introductory statistics course on their degree programmes at the time of data collection. The number of eligible students is approximately 1500.

Data Collection. As this pre-registration was being drafted, the University and Colleges Union (UCU) strikes were announced and took place on 14 days between the 20th February 2020 and 13th March 2020. From the 16th March 2020, the university involved in this study switched to remote learning due to COVID-19. These circumstances meant that the data collection plans were revised, and this pre-registration was not finalised before data collection had to begin (to coincide with teaching schedules).[1] Data collection has now ended but data has not been analysed by the researchers.

The study was advertised to all undergraduate psychology students (except second years) via the participant pool system in exchange for course credits throughout the middle of

[1] Further delays to the pre-registration were due to the decision to include a detailed analysis plan.

the Spring term (March/April 2020). Due to COVID 19, the deadline for students to complete studies for course credits was extended and, accordingly, data collection was extended until May 22nd. A total of 295 responses were recorded via the participant pool.

Second-year students were invited to take part by teaching staff in-class and via e-mail in the middle of the Spring term (March 2020). Due to the UCU strikes in the UK at that time, data could not be collected in class as planned and students were instead invited to complete the study remotely. The second-year students used some of the data for their course assessments and did not receive further course credits for taking part in the study itself. Second year data collection has also now closed and 293 responses were recorded, bringing the total responses to 588.

Sample Size. The exact sample size required for informative analyses could not be determined a-priori for the present study because the information that would inform such a calculation (e.g. number of factors, number of latent profiles) are unknowable in advance for the exploratory factor analysis (EFA), by extension the confirmatory factor analysis (CFA), and the latent profile analysis (LPA). However, simulation studies suggest that the minimum recommended sample sizes for both EFA and LPA under the known conditions of the current study (e.g. a large number of items) are approximately 300, but as sample sizes increase these analyses tend to perform better on an increasing number of indicators (MacCallum et al., 1994; Tein et al., 2013). As such, the intention was to recruit a minimum of 300 participants (after exclusions as outlined below) with no upper limit. If the final sample size is lower than 300, the analysis plan will be revised accordingly before proceeding.

The decision to use multi-level modelling to analyse the single dissociations was made after data collection had ceased so was not included in sample size planning. However, the focus in these models is estimation of the effects (i.e. interpreting parameter estimates and their confidence intervals), and *p*-values will only be interpreted within this context as secondary information.

Exclusions. Cases will be removed according to the following criteria:

1. Duplicates: Cases will be deemed duplicates if they have the same self-generated ID code. The case with the most complete data will be retained, or, if both cases contain the same amount of data, the case with the earliest start date (i.e. the participant's first attempt) will be retained. Cases where the ID code is missing will not be treated as duplicates and will all be retained.
2. Speedy-Responders: To try to reduce careless responses, cases where the participant has responded too quickly to have plausibly been paying attention will be removed. To gauge what a reasonable survey completion time would be, three trial runs of the survey were undertaken which took 507, 462, and 486 seconds (485 on average), excluding the 300 seconds taken to answer the MCQ questions each time (because the time taken by participants to complete the MCQs will also be omitted from their total time so that we do not remove participants that answered the MCQs quickly which, for example, may be a strategy used disproportionately by high maths/statistics anxious participants). Not every participant will have completed 100% of the survey so instead of using the total plausible completion time, the number of seconds it would take to complete 1% progress will be used. By dividing the average plausible completion time by 100, we get a rate of 4.85 seconds per 1%. Participants that completed the survey at a rate faster than 4.85 per 1% (excluding their time taken on the MCQs) will be deemed careless responders and will be removed.

3. Missing data: No exclusions will be made based upon missing data. Instead, missing data will be handled via FIML for EFA, CFA, and MLM and an iterative imputation method based on a random forest for LPA.
4. Outliers: Outliers will not be removed because extreme scores are plausibly genuine and because we anticipate large enough samples for outliers to not bias estimates to a problematic extent.

Ethics. This research has been approved (ER/JLT26/5) by the Sciences & Technology Cross-Schools Research Ethics Committee (C-REC) at the University of Sussex in adherence to the British Psychological Society's Code of Human Research Ethics (2018).

Procedure

Participants were provided with a link to the online survey where they were first asked to read the information sheet carefully before giving their consent to take part. Participants then completed all four measures of statistics anxiety and mathematics anxiety, randomised at the question and item level. Next, they completed the trait anxiety scale followed by the state anxiety scale, both randomised at the item level. Participants were then randomly allocated to take either a statistics or maths version of a five-question multiple-choice question (MCQ) test. Multiple choice options for each question were randomised but questions kept in order. Participants were instructed that they were not allowed to use a calculator and had five minutes to complete the test. If they did not complete the test within that time, the survey moved on automatically. A countdown timer was visible to participants throughout the test. Following the MCQs, participants completed a post-test measure of state anxiety randomised at the item level and, finally[2], some questions about their prior maths and

[2] Students that have taken part as part of their second-year statistics module also completed an additional questionnaire about statistics self-efficacy after the post-test state anxiety measures which were randomised at the item level. These items will not be analysed in the present study.

statistics education and demographics. Following completion, participants were debriefed and thanked. The survey was estimated to take approximately 15 minutes to complete.

Materials

Participants completed the study remotely on PCs or mobile devices. The study was hosted on Qualtrics online survey software.

Measures.

Reliability. The reliability of all measures will be estimated using McDonald's ω (coefficient omega) because omega is a less-biased estimate than Cronbach's alpha (Dunn et al., 2014). Ninety-five percent confidence intervals (CIs) will also be reported. Acceptable omega scores are comparable to the widely adopted cut-off for Cronbach's alpha of .70 for psychological research (Dunn et al., 2014; Nunnally & Bernstein, 1994) and, therefore, any items with a lower 95% CI boundary of < .70 will be considered unreliable. Unreliable items will be reported but not removed because the aim of the present research is to evaluate the original scales in their commonly used form.

Statistics Anxiety. Statistics anxiety was measured with the Statistics Anxiety Rating Scale (STARS; Cruise et al., 1985). The three anxiety subscales (Hanna et al., 2008; Papousek et al., 2012) of the STARS were used; test and class anxiety (8 items), interpretation anxiety (11 items), and fear of asking for help (4 items). Each item describes a situation involving statistics such as "Doing an examination in a statistics course" (test and class anxiety), "Interpreting the meaning of a table in a journal article" (interpretation anxiety), or "Going to ask my statistics teacher for individual help with material I am having difficulty understanding" (fear of asking for help). Participants were asked to indicate how much anxiety they feel in those situations on a Likert scale ranging from 1 = "no anxiety" to 5 = "very much anxiety".

Several items used outdated language and were modified to reflect modern equivalents (e.g. “Asking one of my teachers for help in understanding a printout” was changed to “Asking one of my teachers for help in understanding statistical output”).

Maths Anxiety. Maths anxiety was measured with the Revised Maths Anxiety Rating Scale (R-MARS; Baloğlu & Zelhart, 2007). There are three subscales in the R-MARS which measure mathematics test anxiety (15 items), numerical task anxiety (5 items), and mathematics course anxiety (5 items). Each item describes a situation involving maths such as “taking an exam in a math course” (mathematics test anxiety), “being given a set of division problems to solve” (numerical task anxiety), or “listening to another student explain a math formula” (mathematics course anxiety). Participants were asked to indicate how much anxiety they feel in those situations on a Likert scale ranging from 1 = “no anxiety” to 5 = “very much anxiety”.

Several items were modified to reflect UK equivalents of US terms (e.g. “Taking the mathematics section of a college entrance exam” was changed to “Taking the mathematics section of a university entrance exam”).

Modified STARS and R-MARS. Versions of the STARS and R-MARS have been created by the researchers in which the STARS items were revised to reflect maths-related situations (e.g. “Doing the coursework for a statistics course” has been changed to “Doing the coursework for a maths course”) and the R-MARS statements were revised to reflect statistics-related situations (e.g. “Walking into a maths class” has been changed to “Walking into a statistics class”).

Three items in the STARS were not easily distinguishable as being about either maths or statistics so equivalent items were not created (“Arranging to have a body of data put into the computer”, “Reading an advertisement for a car which includes figures on miles per gallon, depreciation, etc.", and "Trying to understand the odds in a lottery"). Additionally, one item on the R-MARS was deemed untranslatable to a statistics context so, again, an equivalent was not created ("Reading a cash register receipt after your purchase").

Trait and State Anxiety. Trait and state anxiety were measured using the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA; Ree et al., 2008). The STICSA has been developed and evidenced to differentiate anxiety from depression more effectively than other popular anxiety measures (e.g. the State-Trait Anxiety Inventory (STAI) Spielberger, 1983; Grös et al., 2007).

There are two sets of dimensions in the STICSA: Trait or state anxiety, each further broken down into cognitive or somatic symptoms. The combinations of these dimensions result in four subscales that measure trait-somatic symptoms (11 items), trait-cognitive symptoms (10 items), state-somatic symptoms (11 items), and state-cognitive symptoms (10 items). The trait and state scales contain identical items, but the questions are modified; Trait anxiety is ascertained by asking participants to rate how often a statement is true in general and state anxiety is measured by asking participants to rate how often a statement is true at the moment of assessment. Within each of the trait and state scales, cognitive symptoms are identified with statements such as "I cannot concentrate without irrelevant thoughts intruding" and somatic symptoms are identified with statements such as "My heart beats fast". Participants were asked to indicate the extent to which each item is true of them on a Likert scale ranging from 1 = "not at all" to 4 = "very much so". The scales will be used at the state and trait levels and not further broken down into the cognitive and somatic subscales.

Some statements in the state anxiety scale were modified to be in the present tense instead of past tense (e.g. "My heart beats fast" was changed to "My heart is beating fast").

Prior Maths and Statistics Education. Participants were asked to indicate their highest level of both maths and statistics education (including modules on their current degree courses) and to provide the final grade (the highest if more than one) awarded at each level.

Demographics*[3]*. Participants were also asked to indicate their age (in years), gender identity, ethnicity, and whether they have been diagnosed with a specific learning disability (SpLD), such as dyslexia or dyscalculia.

A copy of all measures and any modifications are available in the supplementary materials.

Multiple-Choice Questions. The maths and statistics multiple-choice tests have been designed by the two lead researchers. The purpose of the MCQs is to elicit anxiety and not to measure statistics or maths knowledge. Each test consists of five questions, matched across conditions so that each has a question about the following concepts, presented in this order: mean, standard deviation, confidence intervals, raw beta coefficient, and standard error. The questions in each condition also uses the same numerical information (e.g. a mean of 61 was the focus of the mean question in both conditions). In the maths condition, the questions ask students to perform purely numerical operations (e.g. "What is the mean of the following set of numbers?") and in the statistics condition, the questions provide participants with a research context and asks them to make inferences from numerical information (e.g. "A researcher asked people how likely they would be to purchase an environmentally friendly alternative to their favourite product, even if it was more expensive. Possible scores ranged from 1 to 100 and the mean rating was 61. Which of the following statements is correct?").

[3] The raw data from demographic questions will not be made available in accordance with the conditions of our ethics approval and in line with the General Data Protection Regulations (GDPR) of the Data Protection Act 2018.

Data Analysis

All data analysis will be conducted using RStudio (RStudio Team, 2020) for R (R Core Team, 2020). Anonymised data and R code will be made available via GitHub.

Models will be interpreted using fit statistics and parameter estimation as appropriate (detailed below), whilst *p*-values (alpha = .05) will be used only as secondary evidence.

If, during analysis, unexpected problems emerge that mean changes need to be made to the analysis plan (e.g. violated assumptions, models not converging), analysis will be paused whilst a revised analysis plan is drawn up and pre-registered.

Confirmatory Factor Analysis

CFA will be conducted using the lavaan package for R (Rosseel, 2012). The factor structure of the STARS anxiety subscales (Hanna et al., 2008) and RMARS scales (Baloğlu & Zelhart, 2007) will be imposed. The following steps will be taken to assess model fit:

1) The measurement model will be specified. By default, lavaan constrains each latent factor's scale to that of its first observed variable but, because this means no coefficient will be obtained for these indicators, the latent factors will instead be standardised (by giving them a mean of zero and a variance of one). The results will be no different except the covariances between the indicators become correlations. The size (but not statistical significance) of the factor loadings will be interpreted with loadings > 0.4 considered high.
2) The CFA model will be fitted using the MLR estimator which uses robust (Huber-White) standard errors and a scaled test statistic that is (asymptotically) equal to the Yuan-Bentler test statistic, and FIML to handle missing data.
3) The following fit statistics will be interpreted:

a. *Chi-squared.* A non-significant chi-squared value ($p > .05$) would indicate that the difference between the observed covariance matrix and the estimated population covariance matrix implied by the model is likely to be due to sampling error and, therefore, is a good fitting model. However, the chi-squared test is highly powered and sample sizes above 400 will almost always produce statistically significant chi-squared values (Kenny, 2020). As such, the chi-squared test will be reported but not used to assess the fit of the model.

b. *Comparative fit index (CFI) and Tucker-Lewis index (TLI).* The CFI and TLI compare the specified model to the baseline (saturated) model. CFI and TLI indices > .95 will be considered indicative of good model fit, with values > .90 considered acceptable. However, CFI and TLI both depend on the average correlations in the data and a low mean will result in lower fit values, which is possible even with a good fitting model (Kenny, 2020). This problem will be checked for via the null root mean square error of approximation (null RMSEA), whereby a value > 0.158 will indicate that CFI and TLI indices are informative and can be used to assess the fit of the model.

c. *RMSEA.* An RMSEA value < 0.5 will be considered indicative of good model fit. The 90% confidence interval (CI) will also be interpreted and a lower bound near zero and an upper bound no larger than .08 will be preferred (Kenny et al., 2015).

d. *Standardised root mean square residual (SRMR).* An SRMR value less than .08 will be considered indicative of good fit, the closer to zero the better.

e. *Standardised residuals*. Standardised residuals > 2 will indicate that the difference between the expected and observed covariances are statistically significantly different from zero and, therefore, the model is not capturing the relationship between the variables adequately. These residuals drive the RMSEA value so, even if some residuals > 2, providing the RMSEA value < .05, we can be confident that the overall fit is fine, and just notice where the relationships between variables might not be well-captured.

f. *Modification indices*. Modification indices above 3.86 are statistically significant and indicate that model fit would improve if the corresponding parameter was specified in the model. Modification indices will be reported, along with the expected parameter change (EPC; because the modification index can be significant in large samples even when the EPC is negligible; Kline, 2015), however, no actual modifications will be made.

Exploratory Factor Analysis

EFA will be conducted using the psych package for R (Revelle, 2019). The following steps will be taken, using the STARS and R-MARS scales for phase 1A and with the addition of their modified versions for phase 1B:

1) Descriptive statistics (means and standard deviations) will be obtained for all items.
2) A covariance matrix will be calculated using full information maximum likelihood (FIML) to handle missing data.
3) Correlations will be calculated from the covariance matrix (to avoid dropping cases via pairwise deletion). Variables will be screened for low correlations ($r < 0.3$) with most other variables and for very high correlations ($r > 0.9$). Ordinarily

in EFA, variables with such low/high correlations may be removed but, because this project is an exploration of the factor structure of the scales in their original form, we don't want to exclude any problematic variables, just to notice them. An exception may have to be made if there are computational complications owing, for example, to extreme multicollinearity, but this is unlikely.

4) Multicollinearity will also be checked via the determinant which should be > 0.00001.
5) The KMO test will be conducted to ascertain the sampling adequacy of each item. Scores > 0.5 will indicate that there is likely to be enough data to calculate distinct and reliable factors.
6) The factors will then be extracted via parallel analysis using principal axis factor analysis, squared multiple correlations to estimate communalities, and 100 iterations. Additionally, also using principal axis factor analysis, the Very Simple Structure (VSS) and Minimum Average Partial (MAP) criteria will be obtained, along with these fit indices: Root Mean Square Error of Approximation (RMSEA), Bayesian Information Criterion (BIC), and Sample Size Adjusted BIC (SABIC). The number of factors to retain will be decided based upon converging evidence from the parallel analysis, scree plot, VSS, MAP, RMSEA, BIC, and SABIC.
7) Having derived the number of factors, the factor solution will then be rotated using oblique oblimin rotation. Again, the principle axis factoring option will be set and squared multiple correlations used to estimate communalities. Bootstrapped 95% confidence intervals for the factor loadings & inter-factor correlations will also be obtained using 1000 iterations.

8) The standardised factor loadings from the rotated pattern matrix will be interpreted to determine which factors each item loads mostly strongly on to. Items will be considered to load highly onto a factor where the loading is > 0.4.

Multilevel Model

Multilevel modelling (MLM) will be conducted using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages for R to test for evidence of two dissociations. The first model (Equation 1) is a three-way cross-level interaction between time (pre, post; level two), statistics anxiety (level one) and MCQ type (stats, maths; level one). The second model (Equation 2) is a replica of the first, except math anxiety is substituted for statistics anxiety. In both models, trait anxiety will be adjusted for by adding it as an additional predictor. Models will be fitted using full maximum likelihood estimation. The models take the following forms:

MLM 1 (statistics anxiety)

Level 1:

$$StateAnx_{it} = \beta_{0i} + \beta_{1i}Time_{it} + e_{it}$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(TraitAnx_i) + \gamma_{02}(MCQType_i) + \gamma_{03}(StatsAnx_i) + \gamma_{04}(MCQType_i \times StatsAnx_i) + r_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(TraitAnx_i) + \gamma_{12}(MCQType_i) + \gamma_{13}(StatsAnx_i) + \gamma_{14}(MCQType_i \times StatsAnx_i) + r_{1i}$$

Equation 1.1

Combined Model:

$$StateAnx_{it} = \gamma_{00} + \gamma_{01}(TraitAnx_i) + \gamma_{02}(MCQType_i) + \gamma_{03}(StatsAnx_i) + \gamma_{04}(MCQType_i \times StatsAnx_i) + (\gamma_{10} + \gamma_{11}(TraitAnx_i) + \gamma_{12}(MCQType_i) + \gamma_{13}(StatsAnx_i) + \gamma_{14}(MCQType_i \times StatsAnx_i))(Time_{it}) + r_{0i} + r_{1i}(Time_{it}) + e_{it}$$

Equation 1.2

MLM 2 (maths anxiety)

Level 1:

$$StateAnx_{it} = \beta_{0i} + \beta_{1i}Time_{it} + e_{it}$$

Level 2:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(TraitAnx_i) + \gamma_{02}(MCQType_i) + \gamma_{03}(MathsAnx_i) + \gamma_{04}(MCQType_i \times MathsAnx_i) + r_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}(TraitAnx_i) + \gamma_{12}(MCQType_i) + \gamma_{13}(MathsAnx_i) + \gamma_{14}(MCQType_i \times MathsAnx_i) + r_{1i}$$

Equation 2.1

Combined Model:

$$StateAnx_{it} = \gamma_{00} + \gamma_{01}(TraitAnx_i) + \gamma_{02}(MCQType_i) + \gamma_{03}(MathsAnx_i) + \gamma_{04}(MCQType_i \times MathsAnx_i) + (\gamma_{10} + \gamma_{11}(TraitAnx_i) + \gamma_{12}(MCQType_i) + \gamma_{13}(MathsAnx_i) + \gamma_{14}(MCQType_i \times MathsAnx_i))(Time_{it}) + r_{0i} + r_{1i}(Time_{it}) + e_{it}$$

Equation 2.2

The following analytical steps will be taken for each model:

1) Composite scores (means) for the relevant measures (STARS, RMARS, STICSA trait, STICSA pre-state, STICSA post-state) will be calculated. Any missing values at the item level will be imputed using single imputation with predictive mean matching.
2) The STARS and RMARS are on different scales so will be standardised (given a mean of zero and standard deviation of one) prior to converting the data to long format.
3) The remaining non-standardised continuous level one predictor variable (trait anxiety) will be grand mean centred.
4) In converting to long format, the pre- and post- state anxiety scores will be transformed to a single continuous variable (state_anx), along with a factor variable that indicates whether the score was pre or post (time: pre, post).

5) An unconditional means model will be fitted to the data which will estimate the grand mean of state anxiety across all individuals and time points (pre/post MCQ test). The amount of within- and between-person variance will be reported and interpreted.
6) Next, an unconditional growth model will be fitted to the data which will partition and quantify variance across people and time. The fixed effects will estimate the starting point and slope of the average change trajectory, thus indicating the number of units change in state anxiety from pre- to post-MCQ test. Then, the interaction model will be fitted (Equations 1 or 2). We are only interested in the interaction so all predictors will be entered into the model simultaneously.
    a. The parameter estimate for the main effect of trait anxiety will be interpreted to establish whether and how adjusting for trait anxiety effects the model.
    b. The parameter estimates for the interaction will indicate whether the difference in the change in state anxiety scores from pre to post between conditions of the MCQ tests is a function of statistics anxiety in model one or maths anxiety in model two. A single-dissociation will be indicated if the confidence intervals for the parameter estimates of the interaction for each condition of the MCQ test do not overlap.
    c. A double dissociation will be suggested (albeit not directly tested) if the interactions in each model are in opposite directions (i.e. if the Equation 1 model indicates that students high in statistics anxiety have higher state anxiety scores following a statistics MCQ than the maths MCQ, and the Equation 2 model indicates that students high in maths anxiety have higher

state anxiety scores following a maths MCQ than the statistics MCQ). This pattern of results will suggest that the two constructs are distinct.

7) Finally, standard diagnostic plots will be generated to examine the validity of the assumptions of linearity, homoscedasticity, and normality at the level of the residuals. Additionally, influence diagnostics will be checked via the HLMdiag (Loy & Hofmann, 2014) package.

Latent Profile Analysis

Latent profile analysis will be conducted using the tidyLPA package for R (Rosenberg et al., 2018). The following steps will be taken, using the STARS and R-MARS scales for phase one and their modified versions for phase two:

1) The STARS and RMARS composites obtained for the MLM analysis will be used. If the modified versions are to be analysed, composites will be created, also using their means.
2) Descriptive statistics (means, standard deviations, and correlations) for each composite variable will be obtained.
3) All composite variables will be mean centred.
4) A series of latent profile models will be fitted iteratively with k = 1:6 solutions (where k is the number of profiles). Each will be compared to the k – 1 model to ascertain the best fitting model and, thus, the number of profiles to retain. Note that testing solutions up to k = 6 is usually sufficient (Ferguson et al., 2019) but k will be increased if suggested by tidyLPA.
5) Following Masyn (2013), the retained model will also be iteratively fitted and compared with different covariance structures:
    a. Variances and covariances will be restricted to be the same across profiles.

    b. Variances will be allowed to vary across profiles whilst covariances remain restricted.
    c. Covariances will be allowed to vary across profiles whilst variances remain restricted.
    d. Variances and covariances will be allowed to vary across profiles, and the error variances of the indicator variables are allowed to covary.

6) Following Ferguson et al., (2019), the model to retain (number of profiles and covariance structure) will be determined by evaluation of the following:
    a. Bayesian information criterion (BIC) – a lower BIC indicates the preferred model.
    b. Akaike's information criterion (AIC) – a lower AIC indicates the preferred model.
    c. Sample-adjusted BIC (SABIC) – a lower SABIC indicates the preferred model. SABIC is the most accurate of the three information criteria indexes so more weight will be given to SABIC compared to BIC and AIC when evaluating model fit.
    d. Entropy – a measure of classification uncertainty that ranges from zero to one with higher values indicating better fit and values > .80 indicating minimal uncertainty. Entropy measures perform inconsistently in simulations so will be given low weighting in model fit evaluations.
    e. Posterior probabilities – the probability of an individual being assigned a specific profile given their scores on the indicator variables. The lowest posterior probability in the retained profile should be high (> .70), reflecting greater classification certainty.

f. Bootstrap likelihood ratio test (BLRT) – BLRT compares the model to the $k - 1$ model. A statistically significant BLRT indicates that the more parsimonious model ($k - 1$) is the better fit.

g. Small proportion profiles – where there is < 5% of the sample in a single profile, the profile may be spurious. Lack of support for a small proportion profile will be indicated where the profile is not present in the $k + 1$ model.

h. Theoretical support – the retained profiles should be distinct (i.e. not too similar) and interpretable (i.e. make sense theoretically).

7) The mean values of the variables within each class will be compared and, along with the posterior probabilities of individuals and profile plots, will be interpreted to determine the characteristics of each profile (e.g. high statistics anxiety but low maths anxiety). If there are no profiles in the retained solution whereby individuals score differently on the statistics and maths anxiety scales, this will suggest that the underlying constructs are the same.

8) Finally, the assumption of normally distributed indicator variables within each class will be checked via plots.

References

Baloğlu, M., & Zelhart, P. F. (2007). Psychometric Properties of the Revised Mathematics Anxiety Rating Scale. *The Psychological Record*, *57*(4), 593–611. https://doi.org/10.1007/BF03395597

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Birenbaum, M., & Eylath, S. (1994). Who is afraid of statistics? Correlates of statistics anxiety among students of educational sciences. *Educational Research*, *36*(1), 93–98. https://doi.org/10.1080/0013188940360110

Chew, P. K. H., & Dillon, D. B. (2014). Statistics Anxiety Update: Refining the Construct and Recommendations for a New Research Agenda. *Perspectives on Psychological Science*, *9*(2), 196–208. https://doi.org/10.1177/1745691613518077

Cruise, R., J., Cash, R. W., & Bolton, D., L. (1985). Development and Validation of an Instrument to Measure Statistical Anxiety. *American Statistical Association Proceedings of the Section on Statistical Education*, *4*, 92–97.

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*(3), 399–412. https://doi.org/10.1111/bjop.12046

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2019). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development*, 0165025419881721. https://doi.org/10.1177/0165025419881721

Grös, D. F., Antony, M. M., Simms, L. J., & McCabe, R. E. (2007). Psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA): Comparison to the State-Trait Anxiety Inventory (STAI). *Psychological Assessment*, *19*(4), 369–381. https://doi.org/10.1037/1040-3590.19.4.369

Hanna, D., Shevlin, M., & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Differences*, *45*(1), 68–74. https://doi.org/10.1016/j.paid.2008.02.021

Kenny, D. A. (2020). *SEM: Fit*. http://davidakenny.net/cm/fit.htm

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Loy, A., & Hofmann, H. (2014). HLMdiag: A Suite of Diagnostics for Hierarchical Linear Models in R. *Journal of Statistical Software*, *56*(5), 1–28.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1994). Sample Size in Factor Analysis. *Psychological Methods, 4*(1), 84–99.

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods*. Oxford University Press.

Nunnally, J., & Bernstein, I. (1994). *Psychological methods*. McGraw-Hill.

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments–a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195–209. https://doi.org/10.1080/1356251032000052447

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics Anxiety and Statistics Anxiety. Shared but Also Unshared Components and Antagonistic Contributions to Performance in Statistics. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01196

Papousek, I., Ruggeri, K., Macher, D., Paechter, M., Heene, M., Weiss, ElisabethM., Schulter, G., & Freudenthaler, H. H. (2012). Psychometric Evaluation and

Experimental Validation of the Statistics Anxiety Rating Scale. *Journal of Personality Assessment*, *94*(1), 82–91. https://doi.org/10.1080/00223891.2011.627959

R Core Team. (2020). *R: A Language and Environment for Statistical Computing* (4.01) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ree, M. J., French, D., MacLeod, C., & Locke, V. (2008). Distinguishing Cognitive and Somatic Dimensions of State and Trait Anxiety: Development and Validation of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA). *Behavioural and Cognitive Psychotherapy*, *36*(03). https://doi.org/10.1017/S1352465808004232

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Lissa, C. J. V., & Schmidt, J. A. (2018). TidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Rosseel, Y. (2012). *lavaan: An R Package for Structural Equation Modeling*. http://www.jstatsoft.org/v48/i02/

RStudio Team. (2020). *RStudio: Integrated Development for R.* (1.3.959) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Spielberger, C. D. (1983). *Manual for the State–Trait Anxiety Inventory (Form Y)*. Mind Garden.

Tein, J.-Y., Coxe, S., & Cham, H. (2013). Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling : A Multidisciplinary Journal*, *20*(4), 640–657. https://doi.org/10.1080/10705511.2013.824781

Zeidner, M. (1991). Statistics and Mathematics Anxiety in Social Science Students: Some Interesting Parallels. *British Journal of Educational Psychology*, *61*(3), 319–328. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x