The Perth Alexithymia Questionnaire: A Meta Analysis

and

Psychometric Properties of the Perth alexithymia Questionnaire in a Forensic Population

by

BENJAMIN ROBERT JACKSON

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF CLINICAL PSYCHOLOGY

Centre for Applied Psychology

School of Psychology

The University of Birmingham

August 2023

Thesis Overview

This volume presents a meta-analysis, an original empirical study, and a press release for each, submitted as the clinical research component of the degree of Doctor in Clinical Psychology. The topic of this volume is namely alexithymia and how it is measured using the Perth Alexithymia Questionnaire. Alexithymia is a psychological construct defined by a person's inability to identify and describe their own internal emotional states. A meta-analysis explores the psychometric properties of the Perth Alexithymia Questionnaire according to the current existing literature, including internal consistency and test-retest reliability. An original empirical study describes the psychometric properties of the PAQ in a forensic population, and explores the relationship between alexithymia and personal, social, and psychological characteristics of a forensic population.

Table of Contents

Table of Tables

Table of Figures

Chapter 1: The Perth Alexithymia Questionnaire: A Meta Analysis

Abstract

The Perth Alexithymia Questionnaire (PAQ) measures the construct alexithymia; a person's inability to identify and describe their own emotional states. This meta-analysis examined the psychometric properties of the PAQ according to the current existing literature, including internal consistency and test-retest reliability. Fourteen samples across eight separate studies were included in this review (n=5161). Results are presented for both PAQ subscale and composite scores. This study found the PAQ to have excellent internal reliability, both across its subscale and composite scores. The total scale internal consistency alpha coefficient was very high ($\alpha = 0.947$), while all eleven weighted average alpha coefficients were above the generally accepted 0.7 for use in clinical research. Future research on the psychometric properties of the PAQ should focus on test-retest reliability as there is currently very little data available on the application of the PAQ over time.

Introduction

Over the past two decades, an abundance of empirical research has focussed on the trait alexithymia. Alexithymia literally means "a lack of words for mood" and was first introduced by American psychoanalysts in the 1970's (e.g. Sifneos, 1973). This was characterised by a person's inability to describe their feelings or differentiate between them, coupled with a distinct focus on external rather than internal experiences (Nemiah, 1984). Emotion processing deficits have been a recent focus of empirical research due to alexithymia becoming an important transdiagnostic risk factor for a number of clinical pathologies, while it can also reduce the effectiveness of some therapeutic intervention strategies (Taylor et al., 1999). Areas of research include depression (Honkalampi et al., 2001), anxiety (Zeitlin & McNally, 1993), personality disorders (Berenbaum, 1996), eating disorders (Taylor et al., 1996), substance abuse (Thorberg et al., 2009) and offending behaviours (Garofalo et al., 2017; Gillespie et al., 2018).

Recently, alexithymia has been more explicitly defined as a construct comprised of three components: difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and externally orientated thinking (EOT) (Preece et al., 2017). Energised by the immediate clinical relevance of the construct, a number of measures were developed in an attempt to measure the alexithymia construct, including the Toronto Alexithymia Scale (TAS-20) (Bagby et al., 1994), the Bermond-Vorst Alexithymia Questionnaire (BVAQ) (Vorst & Bermond, 2001) and the more recent Perth Alexithymia Questionnaire (PAQ) (Preece, Becerra, Robinson, Dandy, et al., 2018).

The TAS-20 (Bagby et al., 1994) is a 20-item self-report measure of alexithymia. Items are designed to measure three subscales (DIF, DDF, and EOT) as well as a total score as an overall marker of alexithymia. Although the TAS-20 is the most commonly used measure of alexithymia, it has come under scrutiny for the poor internal reliability of its EOT subscale; while Cronbach's Alpha scores of 0.7 or higher are widely considered acceptable for internal reliability, the reliability coefficients for the EOT subscale of the TAS-20 have often been found to be below 0.6 (Preece, Becerra, Robinson, & Dandy, 2018). The BVAQ (Vorst & Bermond, 2001) is a 40-item self-report measure of alexithymia. Items correspond to the standard three subscales of alexithymia (DIF, DDF, EOT) as well as two further subscales: difficulty fantasising (DFAN) and difficulty emotionalising (DEMO). Previously, DFAN and

DEMO had been purported to be further components of the alexithymia construct, but there was little statistical support for either (Preece et al., 2017). DFAN and DEMO were found to be uncorrelated or negatively correlated with the core subscales (DIF, DDF, EOT) (Preece et al., 2017 to Vorst & Bermond, 2001), so the total score for alexithymia in the BVAQ is scored without them. Similar to the TAS-20, while the DIF and DDF subscales of the BVAQ meet minimum reliability requirements for use in research (Cronbach's Alpha >0.7) (Vorst & Bermond, 2001), EOT has been found to be a less reliable subscale (Bermond et al., 2007). It also worth noting that no known studies have reported a reliability of 0.9 for the total alexithymia composite score, which is the desired value for clinical decision making (Preece, Becerra, Robinson, Dandy, et al., 2018).

In 2018, Preece and colleagues developed a new tool to measure alexithymia, with the aim of meeting their own strict evaluation criteria (Preece, Becerra, Robinson, Dandy, et al., 2018): to allow separate DIF, DDF, and EOT subscales to be derived that accounts for emotional valence when assessing functioning at the appraisal stage of emotion valuation (i.e., includes items on both positive and negative emotions in DIF and DDF subscales). The reason for this is that there is evidence to suggest that those with emotional processing deficits process negative and positive emotions differently (van der Velde et al., 2013), which has never been accounted for in previous tools to measure alexithymia.

The Perth Alexithymia Questionnaire (PAQ) (Preece, Becerra, Robinson, Dandy, et al., 2018) is a 24-item self-report measure of alexithymia. All items comprise a statement designed to assess either the DIF, DDF, or EOT component of alexithymia. Respondents answer each item on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating higher levels of alexithymia. This differs from previous alexithymia measures which used a 5-point Likert scale and there is some evidence to suggest that 7-point scales perform better when measuring continuous constructs (e.g. Preston & Colman, 2000). Items corresponding to the DIF and DDF components of alexithymia include negatively and positively valenced items (e.g. "when I'm feeling bad…" or "when I'm feeling good…"). The 24-item measure includes an equal number of items that attempt to measure each component of alexithymia (eight items for each of the DIF, DDF, and EOT subscales) including four positively and negatively valenced items in each of the DIF and DDF subscales. This allows the PAQ to have five separate subscales; positively-valenced difficulty identifying feelings (P-DIF), negatively-valenced difficulty identifying feelings (N-DIF), positively-valenced

difficulty describing feelings (P-DDF), negatively-valenced difficulty describing feelings (N-DDF), and externally oriented thinking (EOT). In addition to these, the PAQ was designed for the subscales to be combined into theoretically meaningful composite scores. As well as a total score for alexithymia (ALEXI, 24 items), the PAQ can sum across negatively and positively-valenced items to achieve an overall DIF score (G-DIF, 8 items), overall DDF score (G-DDF, 8 items), combine the positively valenced subscales to create a *positive-difficulty appraising feelings composite* (P-DAF, 8 items), combine the negatively valenced subscales to create a *negative-difficulty appraising feelings composite* (N-DAF, 8 items), while the P-DAF and N-DAF composites can be combined to create a *general-difficulty appraising feelings* composite (G-DAF, 16 items).

The purpose of this study was to provide a meta-analytic review of the reliability of the PAQ based on the current literature that has examined the psychometric properties of the PAQ. The present study analyses reliability coefficients that pertain to both internal consistency and test–retest reliability, and thus comment on the diagnostic utility of the PAQ.

Methods

Identifying primary studies

Search of electronic databases

An electronic search was conducted in November 2022 using Ovid (MEDLINE, EMBASE, PsycINFO). This was updated in January 2023. There was no date limit for identified articles, however only those in English language were considered due to the inability to accurately translate foreign language papers. Search terms (and their derivatives) focused on the variables of interest; "psychometric properties" of the "Perth Alexithymia Questionnaire". See Table 1. The reference list of included articles and key papers within the field were examined for further relevant publications.

Table 1. Search terms

Construct	Free Text Search Terms	Method of Search	Limits
Perth Alexithymia Questionnaire	"Perth Alexithymia Questionnaire" "PAQ"	Free search terms All search terms combined with *OR*	Peer reviewed articles

Construct	Free Text Search Terms	Method of Search	Limits
Psychometric properties	"Reliability" "Test Reliability" "Test-Retest Reliability" "Interrater Reliability" "Statistical Reliability" "Confirmatory Factor Analysis" "Factor Structure" "Psychometric*" "Internal Consistency" "Alpha" or "Cronbach*"	The two constructs were combined with *AND*	

Inclusion criteria

The main inclusion criteria for this review was that the article must report the psychometric properties of the PAQ in a novel sample, while reporting the Cronbach's Alpha value for the five PAQ subscales (P-DIF, N-DIF, P-DDF, N-DDF, EOT) and the total alexithymia score (ALEXI) at a minimum (some articles may report alpha values for remaining composite scores). As the PAQ is a relatively newly developed questionnaire, with much of the alexithymia literature having used other scales (such as the TAS-20), the inclusion criteria for this review were kept broad in order to maximise the chance of identifying relevant studies. No restrictions were placed on disorder, setting, country of origin, type of methodology, purpose of study, or timeframe. Full inclusion/exclusion criteria are detailed in Table 2, with Figure 1 depicting the flow diagram of how the inclusion was applied to the systematic search.

Table 2. Inclusion criteria

Inclusion criteria	Justification
Psychometric properties	
The study must explore the psychometric properties of the 24-ITEM PAQ and, at a minimum, reported the reliability of the total and subscale scores.	The reliability of the PAQ is the focus of this review.
Administration	
The scale can be delivered in any capacity as described by the original author, including face-face or online	The PAQ was developed prior to covid-19, while many reliability studies took place during the pandemic, meaning administration was often online. This is considered adequate by the authors and warrant inclusion in this review.
The entire scale must have been delivered and completed in one sitting by the participant.	This is how the scale was designed to be delivered and variations of this method would influence the reliability of the measure which may affect the outcome of the review.

Inclusion criteria	Justification
Participant focus and characteristics	
No restrictions on participant language, gender, education, or demographics. Age restricted to an 18+ adult population.	This review aimed to explore the reliability of the PAQ across all available population groups where the questionnaire has been delivered. However, restriction placed on age due to the potential developmental differences for processing emotions in younger people.
Outcome data and study design	
No restriction on study size or design, providing presented results are not reproduced or manufactured from previous published work or participant pool.	Unique data from studies without use of overlapping data will be used to ensure that the overall reliability is calculated without overwhelming influence from individual study data.
The studies are required to report at minimum the total 'ALEXI' alexithymia reliability value, and reliability values for each of the five subscales.	This will ensure that, at a minimum, the reliability of the Total PAQ score, as well as its subscales, will be analysed in this review.
Type of article	
The following article types were excluded: meta-analysis/theoretical papers/ reviews/commentaries/ clinical guidance/qualitative papers	These articles do not provide the reliability data needed for this meta-analysis.

The application of these inclusion criteria to the search results is provided in figure 1.

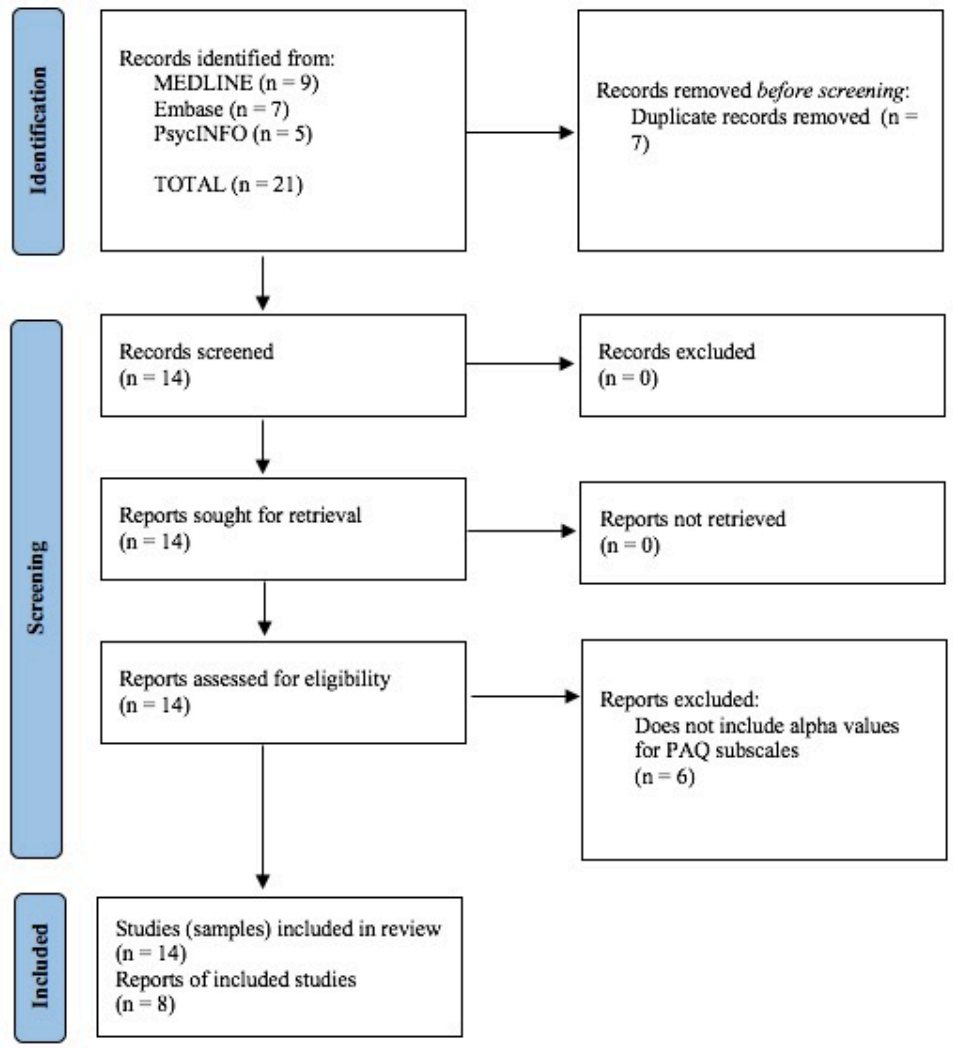


Figure 1. Flow diagram of the application of the inclusion criteria to the results of this systematic search

Data Extraction

The data was extracted by the corresponding author. The following data were extracted from included studies: authors; year of publication; measurement properties evaluated: internal reliability, and/or test-retest reliability, number of items, sample size; study recruitment strategy; and participant demographics.

It is expected that internal reliability of the PAS will be reported in the form of a Cronbach's alpha values for internal reliability Pearson's correlation coefficient for test-retest reliability. If internal reliability and/or test-retest reliability were reported in another form (i.e., non-parametric correlations or averaged inter-item correlations) these were transformed into the expected form prior to analysis.

Study Design Hierarchy

Numerical quality weightings were used for each study to evaluate the study design (Table 3.) and to account for methodological flaws and biases to the data (Table 4).

Table 3. Study design hierarchy

	Score	Description
Psychometric study	10	The study that was designed to assess the psychometric properties of the PAQ in greater than 50 participants
Other Study	0	Reported psychometric properties in a unique sample but the study was designed to address another question OR study that was designed to assess the psychometric properties of the PAQ but in less than 50 participants.

An "Overall Quality Index" percentage was calculated for each study based on study design (psychometric study vs other) and risk of bias scores (see below). This value was then expressed as a percentage of the maximum possible score (i.e., a psychometric study with greater than 50 participants and no identified risk bias). Studies that fell below 75% quality were to be removed from this review.

Risk of bias assessment

The risk of bias quality framework to asses any risk of bias within the extracted literature, were adapted from existing risk of bias frameworks including The Cochrane Collaboration Risk of Bias Tool (Higgins et al., 2011) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013). As reported by Higgins, when assessing bias, studies should operationalise general quality criteria to be specific and suitable for the particular biases expected from the literature under review. As such, risk of bias was assessed through seven domains applicable to this review: selection bias; performance bias; measurement fidelity; detection bias; statistical bias; reporting bias; generalisability (see Table 4). Domains were rated as either low, unclear or high risk using the criteria described in Table 5).

Table 4. Risk of bias quality framework

Domain	Details	Risk of Bias
Selection bias	Selection bias occurs when there is a systematic difference between the characteristics of those selected for the study and those who are not. Have the selection method and characteristics of participants been described adequately?	High Risk – No method of how participants were selected, or characteristics of participants are described. Unclear Risk – The characteristics of the study population are not clearly or fully reported. This includes age range, education years, socioeconomic status, ethnicity, where participants were recruited from (how).

Domain	Details	Risk of Bias
		Low Risk - The characteristics of the study population are clearly described and without evidence of bias.
Performance bias	Performance bias may occur through participants completing measures differently due to social desirability, as well as other factors such as incentives to participate. Were these adequately controlled for?	High Risk- Clear evidence of performance bias Low Risk- No obvious internal or external incentives to complete the PAQ in a particular way.
Measurement Fidelity	Was the delivery of the test sufficiently well described that it could be replicated? Were procedures in place to assess the fidelity of the administered test? Was the delivery of the test completed in an acceptable way as per the recommendations of the test's author's?	High Risk – No mention of processes used to ensure fidelity. No description of application of test. Unclear Risk – Unclear if study protocol was followed. Training of those delivering the intervention not reported. This included where the procedure wasn't reported - not clear whether the test was administered 1-2-1 or in group, at home or in a different setting. Low Risk - Test delivery and completion described and adequate adherence to the test author's recommendations demonstrated. Valid test application conducted by someone with suitable experience.
Detection bias	Was the PAQ delivered in its original or agreed format? Was the scoring of the test completed as per the author's recommendations (matrix)?	High Risk – Major alterations to the test, including wording and/or scoring matrix. Combined with or amalgamated with a different test Unclear Risk – Minor changes made to the wording of questions to changes made to the scoring matrix (i.e. changed from 5-point to 3-point scale) to changes to questions due to translation. Low Risk - Test administered in its original or agreed format and scored following the recommended matrix.
Statistical bias	Bias resulting from the (inappropriate) statistical treatment of the data.	High Risk – Analysis does not produce a Cronbach's Alpha value. Unclear Risk – A variation or alternative value is provided in place of a Cronbach's Alpha value. Low Risk - Analysis as expected to produce a Cronbach alpha
Reporting bias	Is there evidence of selective outcome reporting? Are there measures that have not been reported in the results that have been mentioned in the method section?	High Risk- Not reported Cronbach's Alpha value for the PAQ. Unclear Risk- Not all descriptive statistics are presented. Values not presented for the total and all sub-scales. Low Risk- Reported Cronbach's Alpha values for all sub-scales and Total test.
Generalisability	Can the results be applied to other populations groups or settings based on the sample used?	High risk – Small sample with or without idiosyncratic features (<20 per group). Unclear risk - Sufficient sample for generalisation but with some idiosyncratic feature (> 20 per group). Sample

Domain	Details	Risk of Bias
		taken from only one population group (i.e. students) with attempts to generalise to entire population. Low risk- Sufficient sample for generalisation and representative of target population (>20 per group)

The application of the risk of bias criteria to the included studies is shown in Table 5, which also reports an overall quality index. This index was calculated by first assigning a numerical weighting according to the methodological rigour of the study's overall design (Table 3). A total risk of bias score was then calculated by summing the seven risks of bias domains (low risk = 2 points, unclear risk 1 point, high risk = 0 points) such that the total risk of bias score could vary between 0 and 14 points. The study design score and the total risk of bias score were then summed and the overall quality index for each study was expressed as a percentage of the theoretical maximum score (i.e., the highest quality design without risk of bias).

Table 5. Risk of bias ratings

Study	N	Selection.Bias	Performance.Bias	Treatment.Fidelity	Detection.Bias	Statistical.Bias	Reporting.Bias	Generalisability	Overall Quality Index
Preece et al. (study 1) 2018	231	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	95.83%
Preece et al. (study 2) 2018	748	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	95.83%
Preece et al. 2020a	675	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	91.67%
Preece et al. (sample 1) 2020b	148	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	91.67%
Preece et al. (sample 2) 2020b	103	Low risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	91.67%
Greene et al. (No NSSI) 2020	373	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	87.50%
Greene et al. (NSSI) 2020	267	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	87.50%
Greene et al. (low risk) 2020	370	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	87.50%
Greene et al. (high risk) 2020	270	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	87.50%
Asl et al. 2020	254	Unclear risk	Low risk	High risk	Low risk	Low risk	Low risk	Unclear risk	83.33%
Becerra et al. 2021	370	Unclear risk	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Unclear risk	83.33%
Lashkari et al. 2021	429	Unclear risk	Low risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	87.50%
Chan et al. (Singaporean) 2022	434	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	95.83%
Chan et al. (Australian) 2022	489	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Unclear risk	95.83%

Selection Bias

Overall, selection bias represented the second largest cause of risk, with seven studies rated as demonstrating an unclear risk of bias. The main reason for this was studies not adequately reporting the characteristics of their study population, for example not reporting years of education. Without key information such as years of education and socioeconomic status, there is the possibility that a selection bias could have influenced the reliability of the measurement of the PAQ scale for the given study.

Performance bias

Performance bias only included the option to be rated as high or low risk, based on whether or not there was any evidence of an incentive for participants to complete the PAQ, or any other variable that might result in social desirability effects. No studies indicated any such risk, with all studies utilising voluntary participation without the use of incentives.

Measurement fidelity

Measurement fidelity represented the largest potential cause of risk in measuring the PAQ, with nine studies rated as unclear risk, and one study as high risk. The main cause of this was inadequate explanation of the administration procedure. One study (Asl et al., 2020) included no description of the application of the test, thus prompting a high-risk rating, while only two papers (Preece et al., 2018; Chan et al., 2022) reported whether the questionnaire was administered online or face-to-face.

Detection bias

All studies, with the exception of one (Becerra et al., 2021), were rated as having low detection bias as there was no evidence that the PAQ was delivered in a way that differed from its original intention. Becerra and colleagues (2021) translated the original PAQ into Spanish and reported needing to make ‘minor modifications’ as a result of translational difficulties. Without further clarification, this study was rated as having unclear risk of detection bias.

Statistical and reporting bias

All studies were rated as demonstrating a low risk of both statistical and reporting bias, which was due to the primary studies all reporting a Cronbach’s Alpha value for the reliability of the

scale and its subtests, with no variations on reporting method (i.e., no different statistical measure of reliability other than Cronbach's alpha).

Generalisability

All 14 studies were rated as demonstrating an unclear risk of bias for generalisability. The main reasons for this were that the majority of studies only included university staff or students as participants, or selected a distinct group of participants due to the intended nature of the study (i.e. assessing the psychometric properties of the PAQ in a given population).

Summary

Overall, the level of bias demonstrated by the studies included in this review was low. There was only one study which demonstrated a high risk of bias in any domain (Asl, et al., 2021). Notably there was a particularly high number of studies demonstrating unclear risk of measurement fidelity and there was a lack of heterogeneity amongst participants in sample populations leading to unclear risk of bias in generalisability. However, results from this meta-analysis should be an accurate representation of the current research literature and the Cronbach's alpha values achieved by the total, subscale and composites of the PAQ.

Results

Internal Consistency of the PAQ subscales

The 24-item PAQ has equal numbers of items (eight) pertaining to each component of the alexithymia construct (DIF, DDF, EOT). DIF and DDF are further grouped into positive and negatively valenced items, resulting in five subscales (P-DIF, N-DIF, P-DDF, N-DDF, EOT). The internal consistency of the five subscales is presented below.

Selection of the overall meta-analytic model

Selecting an appropriate model for meta-analysis is essential in ensuring that the following statistics are estimated correctly. Under the fixed-effect model it is assumed that there is one true effect size that underlies all the studies in the analysis, and that all differences in observed effects are entirely due to sampling error, while under the random-effects model the study level

effect sizes may differ due to the idiosyncratic constitution of the individual studies (e.g., variation in participant characteristics, measures, study design and other study level factors).

The distribution of primary study effects is shown in Figure 2, for the fixed effects model and the random effects model with between studies variance (tau^2) was calculated using the DerSimonian-Laird estimator.

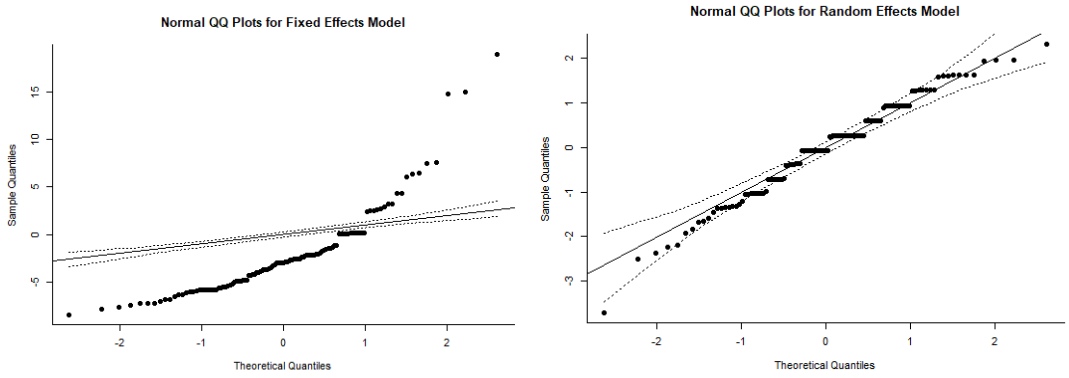


Figure 2. QQ plot of the distribution of alpha coefficients within the included studies

As can be seen from Figure 2, there is clear evidence of non-normality in the distribution of *alpha coefficients* when using the fixed effects model, however, this non normality is largely absent when using the random effects model in which between studies variation was estimated using the DerSimonian-Laird method. Therefore, this suggests the use of the random effects model, and the DerSimonian-Laird estimate of between studies variation, as the most appropriate method for the calculation of the weighted average internal consistency coefficient.

The omnibus test of the PAQ subscales

The alpha coefficients of the PAQ subscales reported in each of the studies are shown in Table 6. There were 14 studies reporting a total of 5161 participants for each of the subscales. Participants were selected from a range of populations as the intention of some studies was to test the reliability of the PAQ in a specific population (e.g. a Spanish population). Study participants were predominantly female, with almost all samples reporting higher percentages of females recruited. Ages ranged from 18 to 85, though the mean age for most studies was less than 30 years of age, while the highest level of qualification for the majority of participants was high school. Study samples came from a range of geographical locations including Spain, Australia, Singapore and the United States.

Table 6: Study level alpha coefficients for the subscales of the PAQ

	Alpha Coefficient	95%-CI	Percent weighting in REM
EOT			
Preece et al. (study 1)	0.9	0.8804 to 0.9196	1.5
Preece et al. (study 2)	0.9	0.8892 to 0.9108	1.5
Preece et al.	0.88	0.8663 to 0.8937	1.5
Preece et al. (sample 1)	0.89	0.8630 to 0.9170	1.4
Preece et al. (sample 2)	0.89	0.8576 to 0.9224	1.3
Greene et al. (No NSSI)	0.93	0.9192 to 0.9408	1.5
Greene et al. (NSSI)	0.91	0.8936 to 0.9264	1.5
Greene et al. (low risk)	0.91	0.8961 to 0.9239	1.5
Greene et al. (high risk)	0.93	0.9173 to 0.9427	1.5
Asl et al.	0.85	0.8220 to 0.8780	1.4
Becerra et al.	0.92	0.9076 to 0.9324	1.5
Lashkari et al.	0.85	0.8285 to 0.8715	1.4
Chan et al. (Singaporean)	0.88	0.8629 to 0.8971	1.5
Chan et al. (Australian)	0.9	0.8866 to 0.9134	1.5
N-DDF			
Preece et al. (study 1)	0.9	0.8788 to 0.9212	1.4
Preece et al. (study 2)	0.91	0.8995 to 0.9205	1.5
Preece et al.	0.86	0.8427 to 0.8773	1.5
Preece et al. (sample 1)	0.91	0.8862 to 0.9338	1.4
Preece et al. (sample 2)	0.85	0.8022 to 0.8978	1.1
Greene et al. (No NSSI)	0.91	0.8950 to 0.9250	1.5
Greene et al. (NSSI)	0.91	0.8923 to 0.9277	1.5
Greene et al. (low risk)	0.91	0.8950 to 0.9250	1.5
Greene et al. (high risk)	0.93	0.9163 to 0.9437	1.5
Asl et al.	0.84	0.8077 to 0.8723	1.3
Becerra et al.	0.86	0.8366 to 0.8834	1.4
Lashkari et al.	0.87	0.8499 to 0.8901	1.4
Chan et al. (Singaporean)	0.87	0.8500 to 0.8900	1.4
Chan et al. (Australian)	0.9	0.8855 to 0.9145	1.5
N-DIF			
Preece et al. (study 1)	0.87	0.8425 to 0.8975	1.4
Preece et al. (study 2)	0.89	0.8771 to 0.9029	1.5
Preece et al.	0.87	0.8540 to 0.8860	1.5
Preece et al. (sample 1)	0.89	0.8609 to 0.9191	1.3
Preece et al. (sample 2)	0.91	0.8813 to 0.9387	1.4
Greene et al. (No NSSI)	0.91	0.8950 to 0.9250	1.5
Greene et al. (NSSI)	0.91	0.8923 to 0.9277	1.5
Greene et al. (low risk)	0.91	0.8950 to 0.9250	1.5
Greene et al. (high risk)	0.91	0.8924 to 0.9276	1.5
Asl et al.	0.86	0.8318 to 0.8882	1.4
Becerra et al.	0.86	0.8366 to 0.8834	1.4
Lashkari et al.	0.83	0.8037 to 0.8563	1.4
Chan et al. (Singaporean)	0.88	0.8615 to 0.8985	1.5
Chan et al. (Australian)	0.88	0.8626 to 0.8974	1.5
P-DDF			
Preece et al. (study 1)	0.91	0.8910 to 0.9290	1.5
Preece et al. (study 2)	0.9	0.8883 to 0.9117	1.5
Preece et al.	0.85	0.8315 to 0.8685	1.5
Preece et al. (sample 1)	0.86	0.8229 to 0.8971	1.2
Preece et al. (sample 2)	0.9	0.8682 to 0.9318	1.3
Greene et al. (No NSSI)	0.88	0.8601 to 0.8999	1.4
Greene et al. (NSSI)	0.9	0.8803 to 0.9197	1.5
Greene et al. (low risk)	0.89	0.8716 to 0.9084	1.5
Greene et al. (high risk)	0.9	0.8804 to 0.9196	1.5
Asl et al.	0.76	0.7116 to 0.8084	1.1
Becerra et al.	0.84	0.8133 to 0.8667	1.4
Lashkari et al.	0.87	0.8499 to 0.8901	1.4
Chan et al. (Singaporean)	0.86	0.8384 to 0.8816	1.4
Chan et al. (Australian)	0.97	0.9656 to 0.9744	1.6
P-DIF			
Preece et al. (study 1)	0.88	0.8546 to 0.9054	1.4
Preece et al. (study 2)	0.89	0.8771 to 0.9029	1.5
Preece et al.	0.87	0.8540 to 0.8860	1.5
Preece et al. (sample 1)	0.85	0.8103 to 0.8897	1.2
Preece et al. (sample 2)	0.93	0.9077 to 0.9523	1.4
Greene et al. (No NSSI)	0.88	0.8601 to 0.8999	1.4
Greene et al. (NSSI)	0.92	0.9043 to 0.9357	1.5
Greene et al. (low risk)	0.9	0.8833 to 0.9167	1.5
Greene et al. (high risk)	0.91	0.8924 to 0.9276	1.5
Asl et al.	0.82	0.7837 to 0.8563	1.3
Becerra et al.	0.83	0.8016 to 0.8584	1.4
Lashkari et al.	0.82	0.7921 to 0.8479	1.4
Chan et al. (Singaporean)	0.86	0.8384 to 0.8816	1.4
Chan et al. (Australian)	0.87	0.8511 to 0.8889	1.5

As can be seen from Table 6, all reported alpha coefficients for the subscales of the PAQ were above the minimum acceptable value of 0.70 for use in clinical research. The lowest reported

alpha coefficient was 0.76 for P-DDF (Asl et al., 2020). A random effects models was calculated using the generic inverse variance method. The weighted average alpha coefficients and associated measures of heterogeneity for the subscales of the PAQ are shown in Table 7 and the study level effects are reported in Figure 4.

Table 7: Weighted average alpha coefficients and associated measures of heterogeneity for the subscales of the PAQ

	Number of studies	**Weighted Average Alpha**	**95%-CI**	**tau^2**	**tau**	**Q**	**I^2**
G-EOT	14	0.8974	0.8854 to 0.9094	0.0004	0.0209	106.69	87.80%
N-DDF	14	0.8907	0.8777 to 0.9037	0.0005	0.0223	93.57	86.10%
N-DIF	14	0.8855	0.8738 to 0.8972	0.0004	0.0196	66.43	80.40%
P-DDF	14	0.8793	0.8485 to 0.9101	0.0033	0.0574	651.2	98.00%
P-DIF	14	0.8759	0.8601 to 0.8917	0.0008	0.0277	108.09	88.00%

As can be seen from Table 7, each of the weighted average alpha coefficients exceeded the minimum acceptable value of 0.70. However, a high level of heterogeneity between the included studies was also observed for each of the subscales. This suggests that the estimates of internal consistency in the included studies may be biased by the presence of uncontrolled or confounding factors. Therefore, the focus of the subsequent analyses will be upon the identification of the sources of heterogeneity. The studies contributing to each of the PAQ subscales is shown in Figure 3.

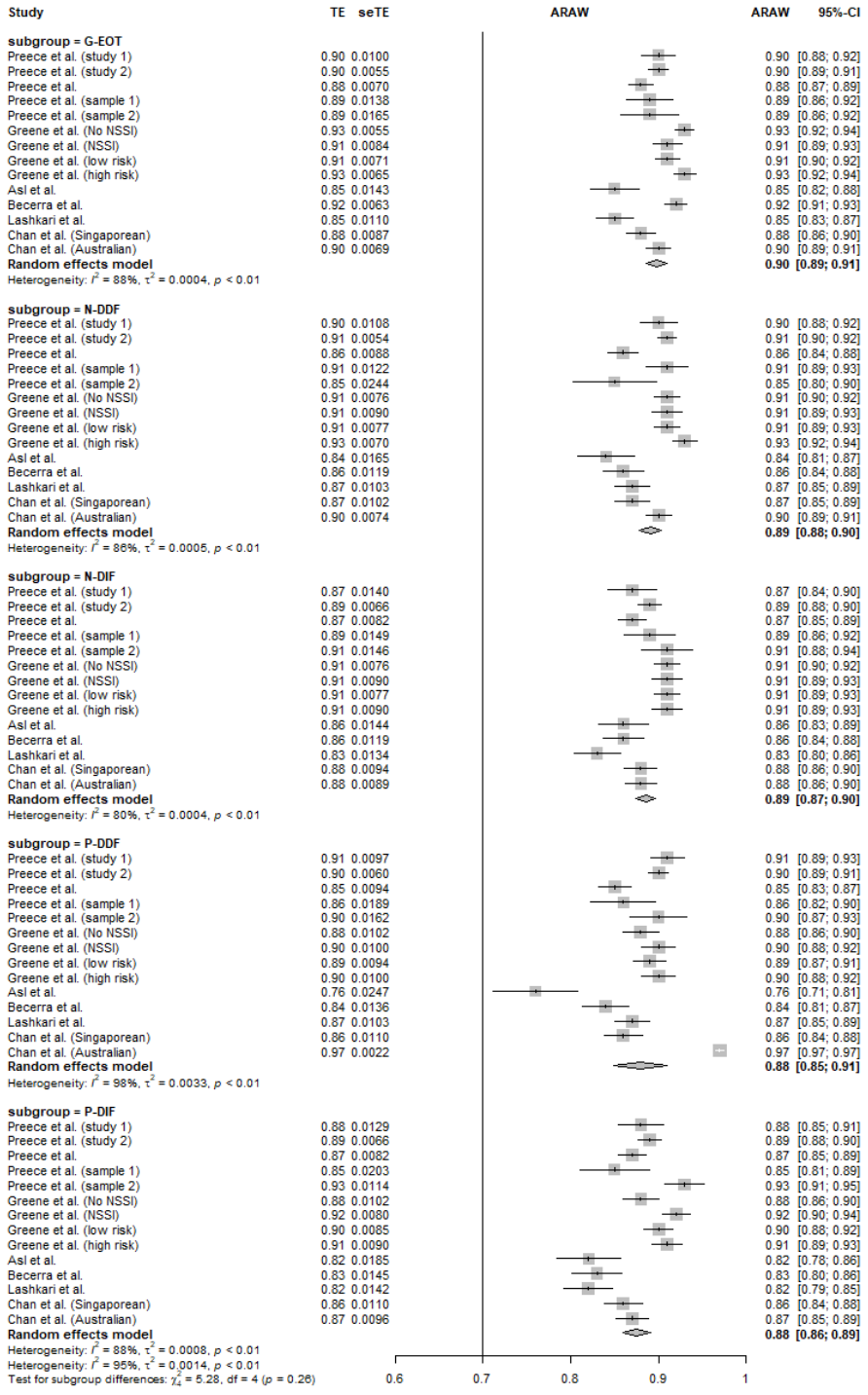
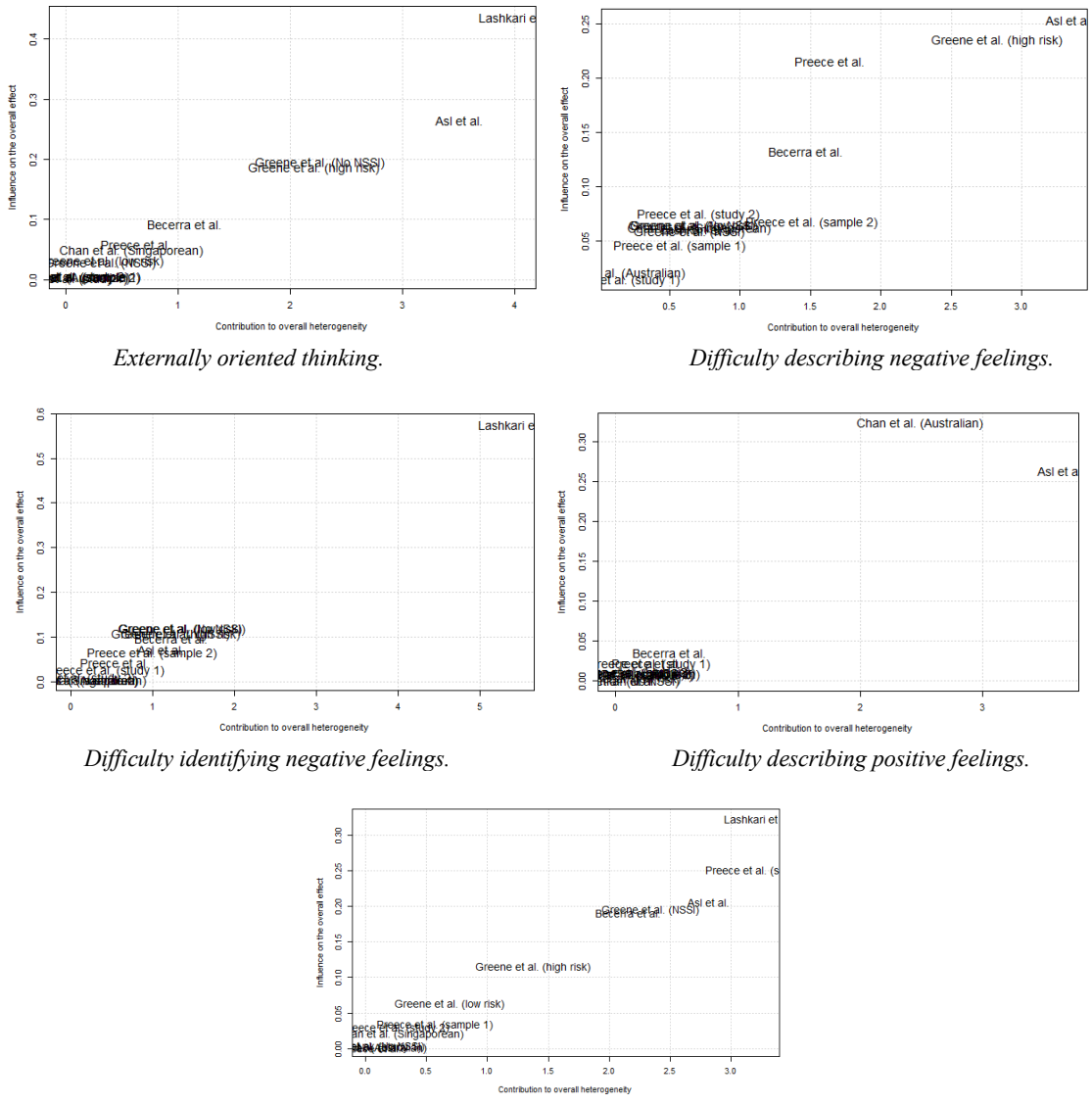


Figure 3. Forest plot of alpha coefficients for the subscales of the PAQ

The impact of influential studies

Some study level effects may be quite discrepant from the meta analytics synthesis, and due to this discrepancy may be disproportionately influential on the weighted average effect size. The

impact of disproportionately influential studies was assessed using a "leave-one-out" analysis, in which the random effects model was calculated with each of the included studies removed in turn, while the change in weighted average effect size (i.e., influence) and the change in heterogeneity (i.e., discrepancy) was recorded for each of the five subscales. The result of this "leave-one-out" analysis is presented in the Baujat plots (Baujat et al., 2002) in Figure 4.



Externally oriented thinking.

Difficulty describing negative feelings.

Difficulty identifying negative feelings.

Difficulty describing positive feelings.

Difficulty identifying positive feelings.

Figure 4. Baujat diagnostic plot of sources of heterogeneity. The vertical axis reports the influence of the study on the overall effect and the horizontal axis reports the discrepancy of the study with the rest of the literature.

For the *externally oriented thinking* subscale (G-EOT), the Baujat plot suggested two studies (Asl et al., 2021; Lashkari et al., 2021) were influential on the overall meta-analytic result and were discrepant from the main body of literature. When Lashkari and colleagues (2021) paper was omitted from the analysis the average alpha coefficient changed to 0.8924 (95% CI

0.8791;0.9057) and when Asl and colleagues (2021) paper was omitted the average alpha coefficient change to 0.8939 (95% CI 0.8813;0.9065). These papers were re-examined with a view to removal from the meta-analysis if substantial risk of bias was identified. However, no such bias could be identified and the overall conclusions of the meta-analysis remained the same with or without the inclusion of these papers.

For the *difficulty describing negative feelings* subscale (N-DDF), three studies (Asl et al., 2021 to Lashkari et al., 2021; Preece et al., 2020a) were shown to be influential on the overall meta-analytic result. When Asl and colleagues (2021) paper was omitted from the analysis the average alpha coefficient changed to 0.9005 (95%CI 0.8888 to 0.9121); when Lashkari and colleagues (2021) paper was omitted the average alpha coefficient change to 0.9012 (95%CI 0.8901 to 0.9123) and when Preece and colleagues (2020a) paper was omitted the average alpha coefficient changed to 0.8989 (95%CI 0.8867 to 0.9112). However, again, the overall conclusions of the meta-analysis remained the same with or without the inclusion of these papers.

For the *difficulty identifying negative feelings* subscale (N-DIF), the Baujat plot suggested one study (Lashkari et al., 2021) was influential on the overall meta-analytic result and was discrepant from the main body of literature. When Lashkari and colleagues (2021) paper was omitted from the analysis the average alpha coefficient changed to 0.8895 (95%CI 0.8792 to 0.8999). This paper was re-examined with a view to removal from the meta-analysis if substantial risk of bias was identified, however no such bias could be identified and the overall conclusions of the meta-analysis remained the same with or without the inclusion of this paper.

For the *difficulty describing positive feelings* subscale (P-DDF), the Baujat plot suggested two studies (Chan et al., 2022 Australian sample; Asl et al., 2021) were influential on the overall meta-analytic result. When Chan and colleagues (2022) Australian sample was omitted from the analysis the average alpha coefficient changed to 0.875 (95%CI 0.8600 to 0.8899) and when Asl and colleagues (2021) paper was omitted the average alpha coefficient change to 0.8874 (95%CI 0.8566 to 0.9181). These papers were re-examined with a view to removal from the meta-analysis if substantial risk of bias was identified, however no such bias could be identified and the overall conclusions of the meta-analysis remained the same with or without the inclusion of these papers.

Finally, for the *difficulty identifying positive feelings* subscale (P-DIF), the Baujat plot suggested one study (Lashkari et al.,2021) was influential on the overall meta-analytic result and was discrepant from the main body of literature. When Lashkari and colleagues (2021) paper was omitted from the analysis the average alpha coefficient changed to 0.8803 (95%CI 0.8653 to 0.8952). Once again, when removed from the analysis the overall conclusions of the meta-analysis remained the same with or without the inclusion of this paper.

The effect of risk of bias in the primary studies

In order to assess the impact of study level risk of bias upon heterogeneity, a series of subgroup analysis were conducted on the reported alpha coefficients for the risk of bias ratings of "low risk" and "any risk" (i.e., unclear risk and high risk of bias combined) for each of the seven types of risk of bias.

Table 8: The effect of risk of bias in the included studies

	Low Risk			Any Risk				
	EFFECT	**95% CI**	**k**	**EFFECT**	**95% CI**	**k**	**X^2**	**P**
Selection bias	0.8871	0.8723 to 0.9019	35	0.8845	0.8745 to 0.8946	35	0.08	0.7792
Performance bias	0.8853	0.8761 to 0.8945	70					
Measurement fidelity	0.8914	0.8719 to 0.9109	20	0.8837	0.8753 to 0.8921	50	0.50	0.4791
Detection bias	0.8870	0.8775 to 0.8964	65	0.8629	0.8241 to 0.9018	5	1.39	0.2383
Statistical bias	0.8853	0.8761 to 0.8945	70					
Reporting bias	0.8853	0.8761 to 0.8945	70					
Generalisability				0.8853	0.8761 to 0.8945	70		

None of the areas of risk of bias showed significantly different estimates of internal reliability for those studies at low or any risk of bias. Accordingly, adjustment for the risk of bias in the primary studies would not result in substantive changes to the conclusions of this review.

The impact of publication and small study biases

Publication bias is caused by the tendency for statistically significant results to be published and the reticence to publish papers with non-significant results. Small study bias is the tendency for studies with smaller sample sizes to show greater variability in their measurement of internal consistency. These biases can be identified in a funnel plot, which plots the magnitude of the study's alpha coefficient (i.e., the importance of the study in the synthesis) estimate the studies deviation from the meta-analytic average (i.e., the discrepancy of the study within the literature). If there is an absence of publication bias, the effects from the studies with small sample sizes which show greater variability will scatter more widely at the bottom of the plot compared to studies with larger samples at the top which will lie closer to the overall meta-analytic effect, creating a symmetrical funnel shape. If there is an absence of studies in the area

of the plot associated with small sample sizes and non-significant results, then it is likely there is some publication bias leading to an overestimation of the true effect. The funnel plot of study level alpha coefficients is presented in Figure 5.

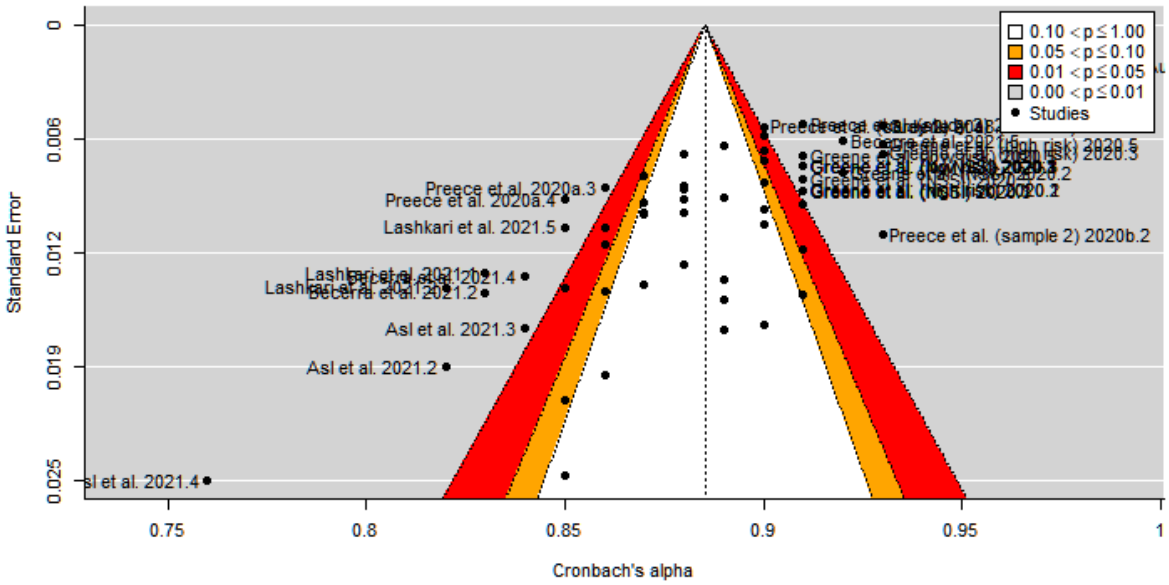


Figure 5. Contour enhanced funnel plot of the EFFECT. The 95% confidence interval of the expected distribution of EFFECT is shown as an inverted "funnel".

As can be seen from Figure 5, there is clear evidence of the previously reported heterogeneity within these studies, with more studies than expected reporting both higher and lower alpha coefficients than the meta-analytic weighted average. However, as this data set contains small studies reporting results That are significantly below the weighted average alpha, there is minimal evidence of publication bias in the distribution of alpha coefficients. Therefore, no simulation of an adjustment for publication bias and small study effects was undertaken.

Orwin (1983) describes the calculation of a failsafe number which calculates the number of with non-significant results which would need to be included in the meta-analysis for the overall effect to be reduce to a minimally interpretable value. This procedure suggests that 65 studies with an average effect size of alpha=0.5 would be required to reduce the observed effect to alpha=0.7, suggesting that the observed weighted average alpha coefficient are robust to studies missing due to publication bias.

Internal Consistency of the PAQ composite scores

The five subscales of the PAQ (results presented above) are, furthermore, designed to be combined into a number of theoretically meaningful composite scores (Preece, Becerra, Robinson, Dandy, et al., 2018). N-DIF and P-DIF are combined to become G-DIF. N-DDF and P-DDF are combined to become G-DDF. N-DIF and N-DDF are combined to become N-DAF. P-DIF and P-DDF are combined to become P-DAF. N-DAF and P-DAF (effectively N-DIF, N-DAF, P-DIF and P-DAF) are combined to become G-DAF. Lastly all five subscales can be combined into a total score alexithymia composite (ALEXI). The internal consistency of the composite scales is presented below.

Selection of the overall meta-analytic model

The distribution of primary study alpha coefficient for the composite subscales is shown in Figure 6, for the fixed effects model and the random effects model where between studies variance (tau^2) was calculated using the DerSimonian-Laird estimator.

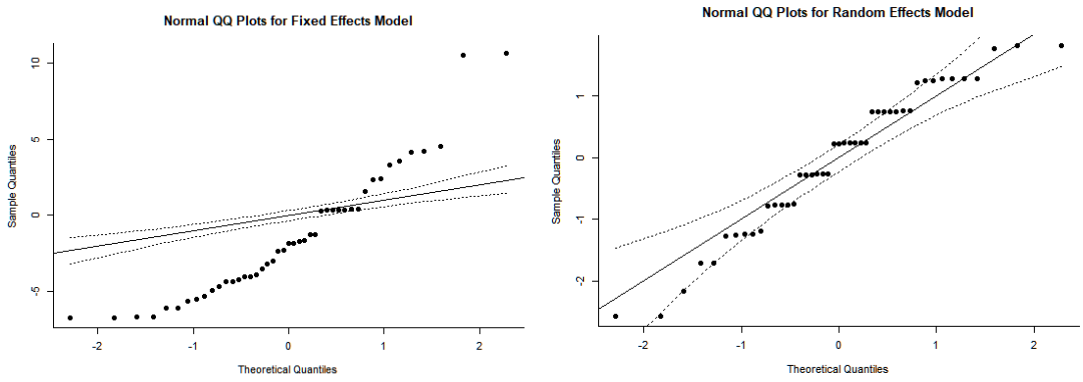


Figure 6. QQ plot of the distribution of alpha coefficients within the included studies.

As can be seen from Figure 6, there is clear evidence of non-normality in the distribution of *alpha coefficients* when using the fixed effects model, however, this non normality is largely absent when using the random effects model in which between studies variation was estimated using the DerSimonian-Laird method. Therefore, this indicates that the use of the use of the random effects model and the DerSimonian-Laird estimate is an appropriate method for the calculation of the weighted average internal consistency coefficient.

The omnibus test of the PAQ composite scores

The alpha coefficients of the composite scores of the PAQ reported in each of the included studies are shown in Table 9. Not all studies reported results for PAQ composites, with 10 studies reporting a total of 3881 participants for general alexithymia (ALEXI) and 7 studies reporting a total of 3376 participants for the rest of the composite scores.

Table 9: Study level alpha coefficients for the subscales of the PAQ

Study	Effect	Std.Er	95% Confidence interval	Weight(random)
ALEXI				
Preece et al. (study 1) 2018	0.95	0.004773	0.9406 to 0.9594	2556
Preece et al. (study 2) 2018	0.96	0.002116	0.9559 to 0.9641	2682
Preece et al. 2020a	0.95	0.002784	0.9445 0.9555	2658
Preece et al. (sample 1) 2020b	0.96	0.004782	0.9506 to 0.9694	2555
Preece et al. (sample 2) 2020b	0.95	0.007187	0.9359 to 0.9641	2380
Asl et al. 2021	0.91	0.00819	0.8939 to 0.9261	2296
Becerra et al. 2021	0.94	0.004518	0.9311 to 0.9489	2572
Lashkari et al. 2021	0.94	0.004195	0.9318 to 0.9482	2590
Chan et al. (Singaporean) 2022	0.95	0.003475	0.9432 to 0.9568	2628
Chan et al. (Australian) 2022	0.95	0.003273	0.9436 to 0.9564	2637
G-DAF				
Preece et al. (study 1) 2018	0.95	0.004826	0.9405 to 0.9595	2553
Preece et al. (study 2) 2018	0.96	0.002139	0.9558 to 0.9642	2681
Preece et al. 2020a	0.95	0.002815	0.9445 to 0.9555	2657
Becerra et al. 2021	0.93	0.00533	0.9196 to 0.9404	2520
Lashkari et al. 2021	0.94	0.004241	0.9317 to 0.9483	2588
Chan et al. (Singaporean) 2022	0.94	0.004216	0.9317 to 0.9483	2589
Chan et al. (Australian) 2022	0.94	0.003971	0.9322 to 0.9478	2603
G-DDF				
Preece et al. (study 1) 2018	0.92	0.007993	0.9043 to 0.9357	2313
Preece et al. (study 2) 2018	0.93	0.003875	0.9224 to 0.9376	2608
Preece et al. 2020a	0.9	0.005828	0.8886 to 0.9114	2485
Becerra et al. 2021	0.87	0.010245	0.8499 to 0.8901	2112
Lashkari et al. 2021	0.9	0.007316	0.8857 to 0.9143	2370
Chan et al. (Singaporean) 2022	0.89	0.008001	0.8743 to 0.9057	2312
Chan et al. (Australian) 2022	0.91	0.006166	0.8979 to 0.9221	2460
G-DIF				
Preece et al. (study 1) 2018	0.9	0.009991	0.8804 to 0.9196	2136
Preece et al. (study 2) 2018	0.92	0.004428	0.9113 to 0.9287	2577
Preece et al. 2020a	0.91	0.005245	0.8997 to 0.9203	2526
Becerra et al. 2021	0.87	0.010245	0.8499 to 0.8901	2112
Lashkari et al. 2021	0.88	0.00878	0.8628 to 0.8972	2245
Chan et al. (Singaporean) 2022	0.89	0.008001	0.8743 to 0.9057	2312
Chan et al. (Australian) 2022	0.9	0.006851	0.8866 to 0.9134	2407
N-DAF				
Preece et al. (study 1) 2018	0.93	0.006993	0.9163 to 0.9437	2396
Preece et al. (study 2) 2018	0.94	0.003321	0.9335 to 0.9465	2635
Preece et al. 2020a	0.92	0.004662	0.9109 to 0.9291	2563
Becerra et al. 2021	0.92	0.006305	0.9076 to 0.9324	2450
Lashkari et al. 2021	0.91	0.006585	0.8971 to 0.9229	2428

Study	Effect	Std.Er	95% Confidence interval	Weight(random)
Chan et al. (Singaporean) 2022	0.93	0.005092	0.9200 to 0.9400	2536
Chan et al. (Australian) 2022	0.93	0.004796	0.9206 to 0.9394	2555
P-DAF				
Preece et al. (study 1) 2018	0.93	0.006993	0.9163 to 0.9437	2396
Preece et al. (study 2) 2018	0.94	0.003321	0.9335 to 0.9465	2635
Preece et al. 2020a	0.92	0.004662	0.9109 to 0.9291	2563
Becerra et al. 2021	0.9	0.007881	0.8846 to 0.9154	2323
Lashkari et al. 2021	0.91	0.006585	0.8971 to 0.9229	2428
Chan et al. (Singaporean) 2022	0.92	0.005819	0.9086 to 0.9314	2486
Chan et al. (Australian) 2022	0.93	0.004796	0.9206 to 0.9394	2555

As can be seen from Table 9, all reported alpha coefficients for the composites of the PAQ were above the minimum acceptable value of 0.70 for use in clinical research. A random effects models was calculated using the generic inverse variance method. The weighted average alpha coefficients and associated measures of heterogeneity for the composite scores of the PAQ are shown in Table 10 and the study level effects are reported in Figure 8.

Table 10: Weighted average alpha coefficients and associated measures of heterogeneity for the composite scales of the PAQ

	Number of studies	Weighted Average Alpha	95%-CI	tau^2	tau	Q	I^2
ALEXI	10	0.9474	0.9412 to 0.9536	<0.0001	0.0089	59.78	84.90%
G-DAF	7	0.9448	0.9367 to 0.9528	0.0001	0.01	53.34	88.80%
G-DDF	7	0.9038	0.8893 to 0.9184	0.0003	0.0183	53.97	88.90%
G-DIF	7	0.8971	0.8842 to 0.9099	0.0002	0.0156	36.39	83.50%
N-DAF	7	0.9263	0.9185 to 0.9341	<0.0001	0.009	25.16	76.20%
P-DAF	7	0.920	0.910 to 0.930	0.0003	0.01		84.00%

As can be seen from Table 10, each of the weighted average alpha coefficients for the PAQ composites exceeded the minimum acceptable value of 0.7. However, a high level of heterogeneity between the included studies was also observed for each of the composite scores. This suggests that the estimates of internal consistency in the included studies may be biased by the presence of uncontrolled or confounding factors. Therefore, the focus of the subsequent analyses will be upon the identification of the sources of heterogeneity. The weighted average alpha coefficients along with the study level effects are shown in Figure 7.

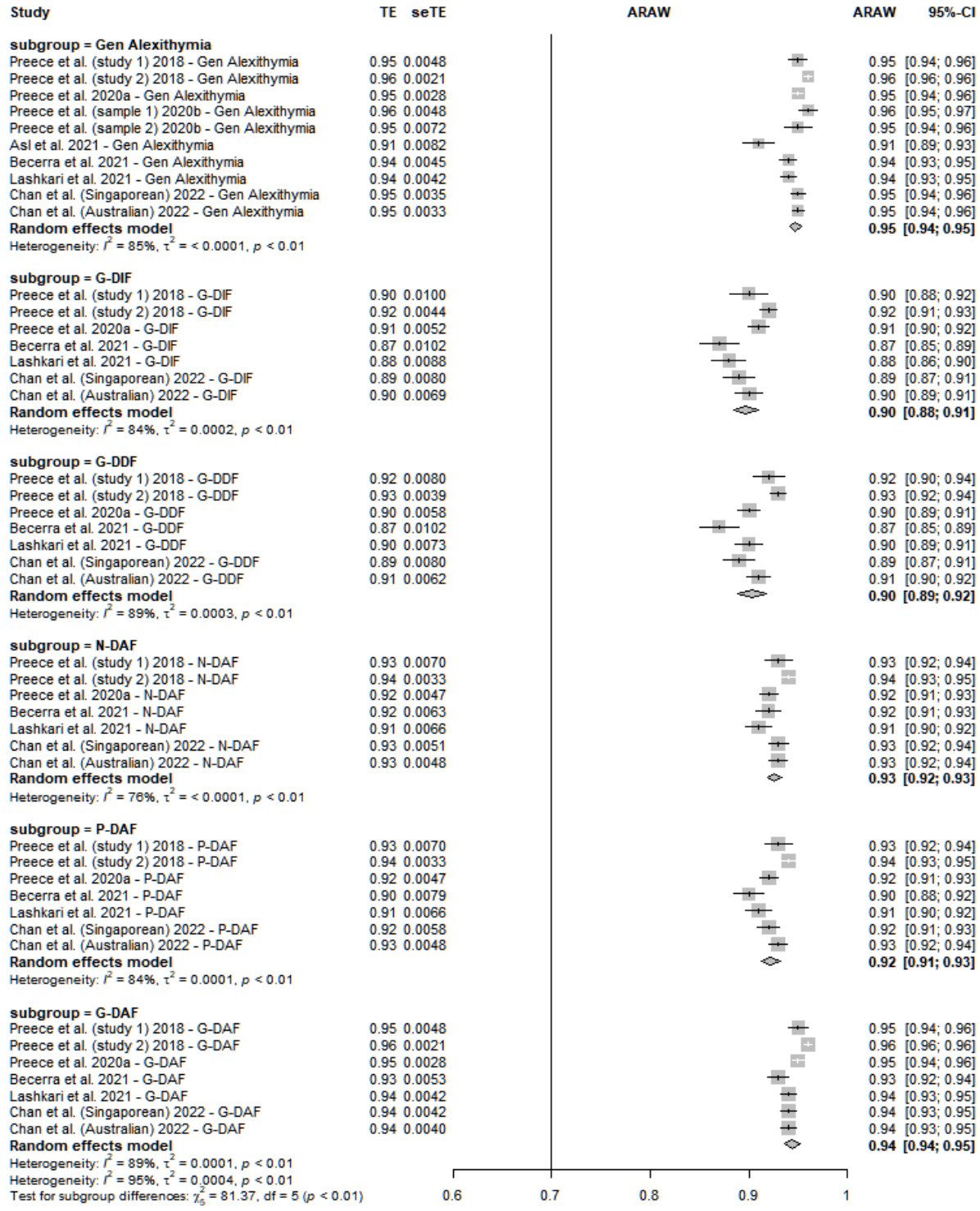


Figure 7. Forest plot of alpha coefficients of the composite scores of the PAQ

The impact of influential studies

The impact of disproportionately influential studies was assessed using a "leave-one-out" analysis and the results are presented on the Baujat plots (Baujat et al., 2002) in Figure 8 for each composite.

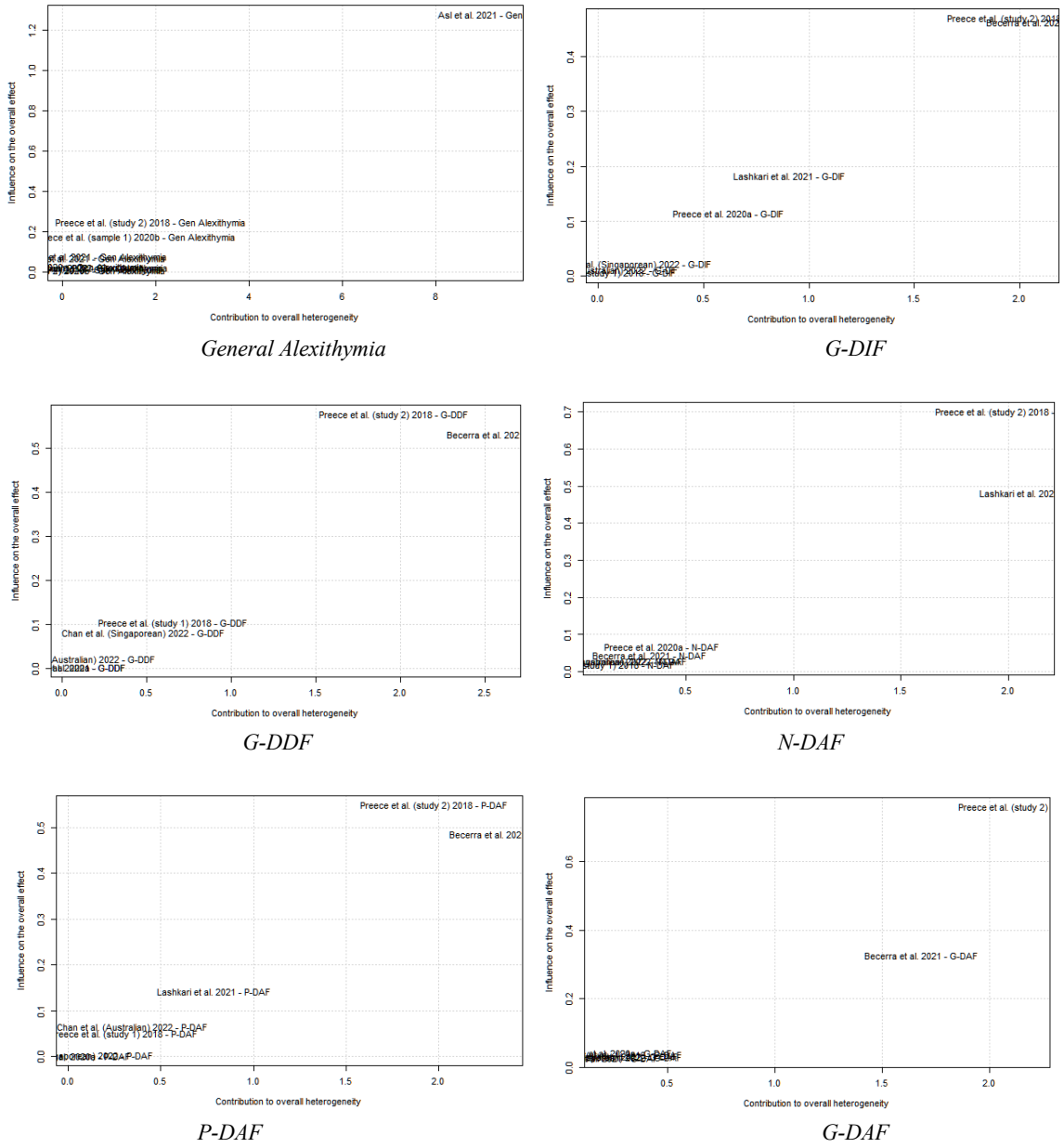


Figure 8. Baujat diagnostic plot of sources of heterogeneity. The vertical axis reports the influence of the study on the overall effect and the horizontal axis reports the discrepancy of the study with the rest of the literature.

As can be seen from Figure 8 there were a number of influential studies for each composite scale score. As above, these studies were omitted from the analysis and the alpha coefficient change was recorded. However, with each omission the overall conclusions of the meta-analysis remained the same with or without the inclusion of this paper, so no studies were considered influential in any composite analysis.

The effect of risk of bias in the primary studies

In order to assess the impact of study level risk of bias upon heterogeneity, a series of subgroup analysis were conducted on the reported alpha coefficients for the risk of bias ratings of "low risk" and "any risk" (i.e., unclear risk and high risk of bias combined) for each of the seven types of risk of bias.

Table 11: The effect of risk of bias in the included studies

	Low Risk			Any Risk				
	EFFECT	**95% CI**	**k**	**EFFECT**	**95% CI**	**k**	**X^2**	**P**
Selection bias	0.9312	0.9249 to 0.9375	32	0.9105	0.8983 to 0.9228	13	8.65	0.0033
Performance bias	0.9254	0.9195 to 0.9313	45					
Measurement fidelity	0.9307	0.9235 to 0.9379	24	0.9192	0.9096 to 0.9288	21	3.52	0.0608
Detection bias	0.9281	0.9221 to 0.9341	39	0.9063	0.8851 to 0.9275	6	3.75	0.0527
Statistical bias	0.9254	0.9195 to 0.9313	45					
Reporting bias	0.9254	0.9195 to 0.9313	45					
Generalisability				0.9254	0.9195 to 0.9313	45		

For composite scores, selection bias evidenced statistically significant difference estimates of alpha coefficients, with lower levels of bias being associated with higher estimates of internal consistency. Accordingly, adjustment for the risk of bias in the primary studies would not result in substantive changes to the conclusions of this review.

The impact of publication and small study biases

The funnel plot of study level alpha coefficients for the PAQ composites is presented in Figure 9.

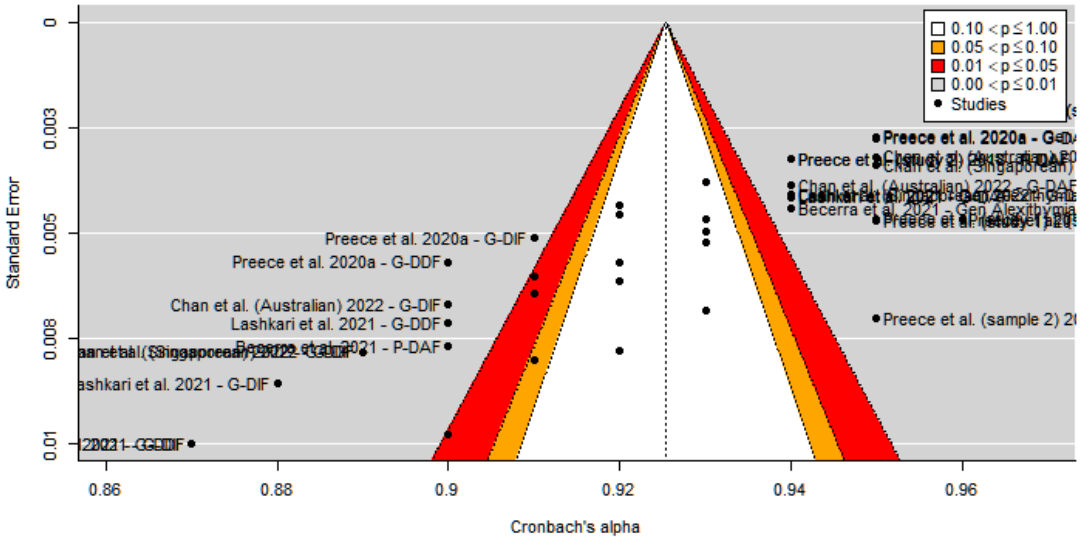


Figure 9. Contour enhanced funnel plot of the EFFECT. The 95% confidence interval of the expected distribution of EFFECT is shown as an inverted "funnel".

As can be seen from Figure 9, there is, again, clear evidence of the previously identified heterogeneity within these data, with more studies than expected reporting both higher and lower alpha coefficients than the meta-analytic weighted average. However, as this data set contains small studies reporting results significantly below the meta-analytic average, there is minimal evidence of publication bias in the distribution of alpha coefficients. Again, no simulation of and adjustment for publication bias and small study effects was undertaken.

Orwin (1983) describes the calculation of a failsafe number which calculates the number of with non-significant results which would need to be included in the meta-analysis for the overall effect to be reduce to a minimally interpretable value. For the composites, this procedure suggests that 51 studies with an average effect size of alpha=0.5 would be required to reduce the observed effect to alpha=0.7, suggesting that the observed weighted average alpha coefficients are robust to studies missing due to publication bias.

Test-Retest Reliability

Only two studies reported test-retest reliability coefficients for the Perth Alexithymia Scale (Asl et al., 2021; Lashkari et al., 2021). Given the relative paucity of studies reporting test-retest reliability, the weighted average coefficient was calculated using the fixed effects model. The fixed effects model weighted average test-retest reliability is reported in Table 12.

Table 12: Fixed effects model weighted average test-retest reliability and measures of heterogeneity

	Asl et al. 2020	Lashkar et al. 2021	Weighted Average Coefficient	95%-CI	Heterogeneity Q	Heterogeneity I^2
Subscales						
P-DIF	0.87	0.71	0.75	0.7057 to 0.7875	10.36	90.30%
N-DIF	0.93	0.72	0.82	0.7870 to 0.8526	39.31	97.50%
P-DDF	0.84	0.69	0.72	0.6731 to 0.7621	6.57	84.80%
N-DDF	0.84	0.71	0.73	0.6894 to 0.7744	5.03	80.10%
G-EOT	0.88	0.74	0.77	0.7337 to 0.8087	9.28	89.20%
Composites						
ALEXI	0.86	0.85	0.85	0.8257 to 0.8758	0.04	0.00%
N-DAF		0.75	0.75	0.7089 to 0.7911	--	--
P-DAF		0.76	0.76	0.7203 to 0.7997	--	--
G-DAF		0.81	0.81	0.7777 to 0.8423	--	--
G-DDF		0.77	0.77	0.7317 to 0.8083	--	--
G-DIF		0.76	0.76	0.7203 to 0.7997	--	--

Asl and colleagues (2020) reported test-retest data only for the PAQ subscales and the general alexithymia total score, with all demonstrating high test-retest reliability (all alpha coefficients above 0.8). Lashkari and colleagues (2021) provided test-retest data for all 11 subscales and composites. One of the 11 alpha coefficients (P-DDF) fell below the 0.70 minimum standard for use in clinical research.

Discussion

The purpose of this review was to synthesise the literature that has examined the psychometric properties of the Perth Alexithymia Questionnaire (PAQ) (Preece, Becerra, Robinson, Dandy, et al., 2018), namely its internal reliability. The results above demonstrate the reliability of the measures subscales and composite scores including a total score for alexithymia.

The PAQ shows good to excellent internal reliability, both across its subscale and composite scores. The lowest subscale weighted average coefficient was P-DIF (0.876), while the highest was G-EOT (0.897) (see Table 7 for subscale scores). The lowest composite weighted average coefficient was G-DIF (0.897), while the highest was ALEXI (0.947) (see Table 10 for composite scores). All eleven weighted average alpha coefficients were above the generally accepted 0.7 standard score for use in clinical research, while some composite scores reached a level of 0.9 that was aimed for by the original authors (Preece, Becerra, Robinson, Dandy, et al., 2018). It has been argued that in order for a measure to be used in clinical decision-making

it should achieve an internal reliability score of at least 0.9 (Groth-Marnat, 2009). As such, the weighted alpha coefficient for the total scale composite score (ALEXI) demonstrated in this review (0.947) could be one of significance for Preece and colleagues. While the 20-item Toronto Alexithymia Scale (TAS-20) (Bagby et al., 1994) has been the most commonly adopted measure of alexithymia, it has been criticised for its poor internal reliability of the externally oriented thinking (EOT) subscale (see Preece, Becerra, Robinson, & Dandy, 2018). This review demonstrates the PAQ's superiority in internal reliability generally, and a more acceptable weighted-average alpha coefficient of 0.897 for its G-EOT subscale. Although we did not conduct a confirmatory factor analysis (CFA) as part of this review, the PAQ's five-factor model that accounts for valence-specific factors (N-DIF, P-DIF, N-DDF, P-DDF, G-EOT) has been shown to be a superior fit than three-factor models of the PAQ (G-DIF,G-DDF, G-EOT) and the TAS-20 (DIF, DDG, EOT) (Preece et al., 2020). As such, the PAQ might be a more favourable measure for diagnostic utility and use in clinical decision-making.

Robust associations continue to be observed between alexithymia and variety of psychopathologies, including depression (Honkalampi et al., 2001), anxiety (Zeitlin & McNally, 1993), personality disorders (Berenbaum, 1996), eating disorders (Taylor et al., 1996), substance abuse (Honkalampi et al., 2022; Thorberg et al., 2009) offending behaviours (Garofalo et al., 2017; Gillespie et al., 2018), sleep problems (Alimoradi et al., 2022) and more. Alexithymia is widely regarded as an important transdiagnostic factor in various clinical populations, and it is becoming more frequently assessed in clinical settings (Preece et al., 2022). While historically this has been largely done using the TAS-20 (Bagby et al., 1994), this review demonstrates the potentially superior utility of the PAQ given that the TAS-20 has been criticised for poor internal reliability of the EOT subscale (Preece, Becerra, Robinson, & Dandy, 2018).

The measurement of alexithymia also has clinical relevance in relation to therapeutic intervention. The presence of alexithymia has been shown to be a barrier to therapeutic treatment due to the recipients' inability to communicate emotional states (Ogrodniczuk et al., 2011). One explanation for this is that the lack of emotion portrayed resulted in a poor therapeutic alliance with the therapy professional. This may provide further justification for alexithymia to be accurately measured prior to any intended therapeutic intervention delivery. However, there is some debate as to whether alexithymia itself is modifiable or a more stable personality trait (Cameron et al., 2014). There is some evidence to suggest that when

alexithymia has been appropriately identified, interventions that directly targeted alexithymic symptoms found significant reductions in alexithymia scores following treatment (Cameron et al., 2014). A recent meta-analysis combining study endpoint data showed a statistically significant effect of mindfulness-based treatment on alexithymia compared with the control group (p=0.010) (Norman et al., 2019). Recent guidelines have been developed to improve the therapeutic alliance between therapist and client with alexithymia (Nunes da Silva, 2021), which includes an initial assessment of alexithymia, a collaborative case conceptualisation that characterises alexithymia as a psychopathology symptom rather than a disorder, and an intervention that accounts for difficulties in communication (i.e. ensuring emotional stimuli is conveyed through body language and facial expressions, rather than spoken language. Ogrodniczuk and colleagues (2005) observed that patients with alexithymia may have poorer therapeutic outcomes as their therapists experience them as having less compatibility with the therapist or may be perceived as less significant members in group therapy. Recognising this as a potential symptom of alexithymia and including this interpersonal cycle in case conceptualisation could help to improve therapeutic outcomes. More research, however, is needed to assess whether reductions in alexithymic traits or an increase in affect awareness lead to better outcomes in psychotherapeutic treatment.

This is the first meta-analysis of the internal consistency of the Perth Alexithymia Questionnaire, and there are a number of limitations to note. Firstly, the number of studies explicitly investigating the psychometric properties of the PAQ is limited, most likely as the PAQ is a very recently developed measure. This study included a relatively modest sample size across 14 samples from eight empirical papers all of which were designed to assess the psychometric properties of the Perth Alexithymia Questionnaire. Is this questionnaire becomes more widely used it is expected that studies will report psychometric properties from samples that were collected in the service of other research needs. Therefore, future versions of this review should conduct sensitivity analyses between effect sizes reported by studies at different levels of the design hierarchy.

Notably, the results of this review also demonstrate the need for caution when interpreting reliability coefficients. There is a distinct lack of research data describing the test-retest reliability of the PAQ and its subscale. This test was also omitted by the original authors (Preece, Becerra, Robinson, Dandy, et al., 2018). Test-retest reliability is a fundamental psychometric property and forms the basis for the calculation of confidence intervals. Without

accurate test-retest reliability data, it would be difficult to calculate ranges for true values and to compare performance on its composite subscales. Accordingly, future studies assessing the psychometric properties of the PAQ should seek to evaluate this important psychometric characteristic. Moreover, for the composite indices, studies with a low risk of selection bias were shown to report greater coefficients than those studies at some risk of selection bias. Further, substantial heterogeneity was observed for all of the PAQ subscales and composites indices, the source for which could not be identified within this review. Accordingly, researchers and clinicians should be mindful that there may be, as yet, unreported factors that affect the internal reliability of the PAQ. Future research should seek to identify these factors.

References

Alimoradi, Z., Majd, N. R., Broström, A., Tsang, H. W. H., Singh, P., Ohayon, M. M., Lin, C.-Y., & Pakpour, A. H. (2022). Is alexithymia associated with sleep problems? A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *133*, 104513. https://doi.org/https://doi.org/10.1016/j.neubiorev.2021.12.036

Asl, E.M., Mahaki, B., & Khanjani, S. (2020). The assessment of alexithymia across positive and negative emotions: The psychometric properties of the Iranian version of the perth alexithymia questionnaire. *Iranian Journal of Psychiatry and Behavioral Sciences*, *14*(4), e102317. https://doi.org/https://dx.doi.org/10.5812/ijpbs.102317

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/https://doi.org/10.1016/0022-3999(94)90005-1

Baujat, B., Mahé, C., Pignon, J.-P., & Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in Medicine*, *21*(18), 2641–2652. https://doi.org/10.1002/sim.1221

Becerra, R., Baeza, C. G., Fernandez, A. M., & Preece, D. A. (2021). Assessing Alexithymia: Psychometric Properties of the Perth Alexithymia Questionnaire in a Spanish-Speaking Sample. *Front. Psychol., Sec. Psychological Therapy and Psychosomatics, 12.* https://doi.org/10.3389/fpsyt.2021.710398

Berenbaum, H. (1996). Childhood abuse, alexithymia and personality disorder. *Journal of Psychosomatic Research*, *41*(6), 585–595. https://doi.org/https://doi.org/10.1016/S0022-3999(96)00225-5

Bermond, B., Clayton, K., Liberova, A., Luminet, O., Maruszewski, T., Ricci Bitti, P. E., Rimé, B., Vorst, H. H., Wagner, H., & Wicherts, J. (2007). A cognitive and an affective dimension of alexithymia in six languages and seven populations. *Cognition and Emotion*, *21*(5), 1125–1136. https://doi.org/10.1080/02699930601056989

Borenstein, M., Hedges, L. V, Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Cameron, K., Ogrodniczuk, J., & Hadjipavlou, G. (2014). Changes in Alexithymia Following Psychological Intervention: A Review. *Harvard Review of Psychiatry*, *22*(3). https://journals.lww.com/hrpjournal/Fulltext/2014/05000/Changes_in_Alexithymia_Following_Psychological.3.aspx

Chan, J., Becerra, R., Weinborn M., & Preece, D.; (2022). Assessing Alexithymia across Asian and Western Cultures: Psychometric Properties of the Perth Alexithymia Questionnaire and Toronto Alexithymia Scale-20 in Singaporean and Australian Samples. *Journal of Personality Assessment.* DOI: 10.1080/00223891.2022.2095641

Garofalo, C., Velotti, P., & Zavattini, G. C. (2017). Emotion Regulation and Aggression: The Incremental Contribution of Alexithymia, Impulsivity, and Emotion Dysregulation Facets. *Psychology of Violence*, *8*, 470–483.

Gillespie, S. M., Garofalo, C., & Velotti, P. (2018). Emotion regulation, mindfulness, and alexithymia: Specific or general impairments in sexual, violent, and homicide offenders? *Journal of Criminal Justice*, *58*, 56–66. https://doi.org/https://doi.org/10.1016/j.jcrimjus.2018.07.006

Groth-Marnat, G. (2009). Handbook of psychological assessment, 5th ed. In *Handbook of psychological assessment, 5th ed.* John Wiley & Sons Inc.

Greene, D., Hasking, P., Boyes, M., & Preece, D. (2020). Measurement Invariance of Two Measures of Alexithymia in Students Who Do and Who Do Not Engage in Non-suicidal Self-Injury and Risky Drinking. *Journal of Psychopathology and Behavioral Assessment (2020) 42:808–825.* https://doi.org/10.1007/s10862-020-09806-7

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration{\textquoteright}s tool for assessing risk of bias in randomised trials. *BMJ*, *343*. https://doi.org/10.1136/bmj.d5928

Honkalampi, K., Hintikka, J., Laukkanen, E., & Viinamäki, J. L. H. (2001). Alexithymia and Depression: A Prospective Study of Patients With Major Depressive Disorder. *Psychosomatics*, *42*(3), 229–234. https://doi.org/https://doi.org/10.1176/appi.psy.42.3.229

Honkalampi, K., Jokela, M., Lehto, S. M., Kivimäki, M., & Virtanen, M. (2022). Association between alexithymia and substance use: A systematic review and meta-analysis. *Scandinavian Journal of Psychology*, *63*(5), 427–438. https://doi.org/https://doi.org/10.1111/sjop.12821

Kim, S. Y., Park, J. E., Lee, Y. J., Seo, H.-J., Sheen, S.-S., Hahn, S., Jang, B.-H., & Son, H.-J. (2013). Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *Journal of Clinical Epidemiology*, *66*(4), 408–414. https://doi.org/10.1016/j.jclinepi.2012.09.016

Lashkari, A., Dehghani, M., Sadeghi-Firoozabadi, V., Heidari, M., & Khatibi, A. (2001). Further Support for the Psychometric Properties of the Farsi Version of Perth

Alexithymia Questionnaire. *Front. Psychol., Sec. Quantitative Psychology and Measurement.* https://doi.org/10.3389/fpsyg.2021.657660

Nemiah, J. C. (1984). The psychodynamic view of anxiety. *Diagnosis and Treatment of Anxiety Disorders*, 117–137.

Norman, H., Marzano, L., Coulson, M., & Oskis, A. (2019). Effects of mindfulness-based interventions on alexithymia: a systematic review. *Evidence Based Mental Health*, *22*(1), 36 LP – 43. https://doi.org/10.1136/ebmental-2018-300029

Nunes da Silva, A. (2021). Developing Emotional Skills and the Therapeutic Alliance in Clients with Alexithymia: Intervention Guidelines. *Psychopathology*, *54*(6), 282–290. https://doi.org/10.1159/000519786

Ogrodniczuk, J. S., Piper, W. E., & Joyce, A. S. (2005). The negative effect of alexithymia on the outcome of group therapy for complicated grief: what role might the therapist play? *Compr Psychiatry, 46,* 206-213. https://doi.org/10.1016/j.comppsych.2004.08.005

Ogrodniczuk, J. S., Piper, W. E., & Joyce, A. S. (2011). Effect of alexithymia on the process and outcome of psychotherapy: a programmatic review. *Psychiatry Research*, *190*(1), 43–48. https://doi.org/10.1016/j.psychres.2010.04.026

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157–159. https://doi.org/10.2307/1164923

Preece, D. A., Becerra, R., Allan, A., Robinson, K., Chen, W., Hasking, P., & Gross D. A.; (2020a). Assessing alexithymia: Psychometric properties of the Perth Alexithymia Questionnaire and 20-item Toronto Alexithymia Scale in United States adults. *Personality and Individual Differences*, *166*. https://doi.org/https://dx.doi.org/10.1016/j.paid.2020.110138

Preece, D. A., Becerra, R., Boyes, M., Northcutt, C., McGillivray, L., Hasking, P. A.; (2020b). Do self-report measures of alexithymia measure alexithymia or general psychological distress? A factor analytic examination across five samples. *Personality and Individual Differences*, *155*. https://doi.org/10.1016/j.paid.2019.109721

Preece, D. A., Mehta, A., Becerra, R., Chen, W., Allan, A., Robinson, K., Boyes, M., Hasking, P., & Gross, D. A.; (2022). Why is alexithymia a risk factor for affective disorder symptoms? The role of emotion regulation. *Journal of Affective Disorders*, *296*, 337–341. https://doi.org/https://dx.doi.org/10.1016/j.jad.2021.09.085

Preece, D., Becerra, R., Allan, A., Robinson, K., & Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences*,

119, 341–352. https://doi.org/10.1016/j.paid.2017.08.003

Preece, D., Becerra, R., Robinson, K., & Dandy, J. (2018). Assessing Alexithymia: Psychometric Properties and Factorial Invariance of the 20-Item Toronto Alexithymia Scale in Nonclinical and Psychiatric Samples. *Journal of Psychopathology and Behavioral Assessment*, *40*(2), 276–287. https://doi.org/10.1007/s10862-017-9634-6

Preece, D., Becerra, R., Robinson, K., Dandy, J., & Allan, David; (2018). The psychometric assessment of alexithymia: Development and validation of the Perth Alexithymia Questionnaire. *Personality and Individual Differences*, *132*, 32–44. https://doi.org/https://dx.doi.org/10.1016/j.paid.2018.05.011

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/https://doi.org/10.1016/S0001-6918(99)00050-5

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*, 255–262. https://doi.org/10.1159/000286529

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1999). *Disorders of affect regulation: Alexithymia in medical and psychiatric illness*. Cambridge University Press.

Taylor, G. J., Parker, J. D. A., Bagby, R. M., & Bourke, M. P. (1996). Relationships between alexithymia and psychological characteristics associated with eating disorders. *Journal of Psychosomatic Research*, *41*(6), 561–568. https://doi.org/https://doi.org/10.1016/S0022-3999(96)00224-3

Thorberg, F. A., Young, R. M., Sullivan, K. A., & Lyvers, M. (2009). Alexithymia and alcohol use disorders: A critical review. *Addictive Behaviors*, *34*(3), 237–245. https://doi.org/https://doi.org/10.1016/j.addbeh.2008.10.016

van der Velde, J., Servaas, M. N., Goerlich, K. S., Bruggeman, R., Horton, P., Costafreda, S. G., & Aleman, A. (2013). Neural correlates of alexithymia: A meta-analysis of emotion processing studies. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1774–1785. https://doi.org/https://doi.org/10.1016/j.neubiorev.2013.07.008

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond–Vorst Alexithymia Questionnaire. *Personality and Individual Differences*, *30*(3), 413–434. https://doi.org/https://doi.org/10.1016/S0191-8869(00)00033-7

Zeitlin, S. B., & McNally, R. J. (1993). Alexithymia and anxiety sensitivity in panic disorder and obsessive-compulsive disorder. *The American Journal of Psychiatry*, *150*(4), 658–

660. https://doi.org/10.1176/ajp.150.4.658

Chapter 2: Psychometric properties of the Perth Alexithymia Questionnaire in a prison population

Abstract

Alexithymia, the inability to identify and describe internal emotional states, is an important transdiagnostic risk factor for a number of psychopathologies, notably including offending behaviour. Furthermore, prevalence of alexithymia is estimated to be higher than that of the normal population. The Perth Alexithymia Questionnaire (PAQ) has been recently developed to measure the construct of alexithymia. This study sought to assess the psychometric properties of the PAQ within a forensic population. Data was collected from a UK prison. This study found the PAQ to have high convergent validity with the TAS-20, an established measure of alexithymia, but even higher internal consistency. This study found no significant relationship between PAQ scores and offence type, maternal and paternal separation, or substance misuse. This study found significant relationships between alexithymia scores on the PAQ and current self-esteem and blame attribution, and higher rates of alexithymia in prison compared to the normal population.

Introduction

Alexithymia is described as the inability to identify and describe internal emotional states, and was first coined by American Psychiatry in the early 1970's (e.g. Sifneos, 1973). It has been more recently defined as a construct of three components: difficulty identifying feelings (DIF), difficulty describing feelings (DDF), and externally orientated thinking (EOT) (Preece et al., 2017). Since its emergence into the literature, alexithymia has been considered an important transdiagnostic risk factor for a number of psychological difficulties including depression (Honkalampi et al., 2001), anxiety (Zeitlin & McNally, 1993), personality disorders (Berenbaum, 1996), eating disorders (Taylor et al., 1996), substance abuse (Thorberg et al., 2009), low self-esteem (Yelsma, 1995) and offending behaviours (Garofalo et al., 2017; Gillespie et al., 2018).

There are a number of well-established correlates of violence and offending behaviours, most notably impulsivity and anger (Krakowski & Czobor, 2014; Norlander & Eckhardt, 2005; Seroczynski et al., 1999). Specific relationships between alexithymia and violent behaviours have also been found in people with substance misuse (Evren et al., 2015), those with antisocial personality disorder (Ates et al., 2009) and post-traumatic stress disorder (Teten et al., 2008).

While the relationships between alexithymia and other types of behavioural offending (e.g. sexual offending, theft) are less prominent in the literature (e.g. Gillespie et al., 2018), rates of alexithymia amongst the prison population are estimated to range between 31 and 47% (Chen et al., 2017; Parker et al., 2005; Zimmermann, 2006); three times higher than the estimated 10 to 19% prevalence of alexithymia in the general population (Franz et al., 2008; Mason et al., 2005; Mattila et al., 2006; Parker et al., 1989).

Higher rates of alexithymia within the prison population have been explained by the stress of incarceration. Imprisonment can be experienced as a highly stressful situation as prisoners are forcibly confined into a limited space while social interaction is restricted. Chen and colleagues (2017) argue that alexithymia is more likely to be more prevalent in prisons due to its stressful nature. An alternative explanation is that those who commit crimes that result in imprisonment already experience higher levels of alexithymia, rather than it being a direct consequence of the imprisonment itself. As a consequence of not having the ability to discuss emotions, it has been found that those with alexithymia in prison can experience either emotional overload or an absence of emotions, both of which can lead to harm to self and harm to others (Hemming et al., 2020). Emotional regulation difficulties are often cited as a correlate with offending behaviours; for example, the general theory of crime (Gottfredson & Hirschi, 1990) posits that low social control (emotional regulation) is the main cause for antisocial behaviour, which much subsequent research seems to confirm (Hudson et al., 1999; Polaschek & Ward, 2002; Roberton et al., 2014; Roberton et al., 2015). More recent research has led to the hypothesis that alexithymia has impairing effects on emotional regulation (Preece et al., 2022), as decisions about affect regulation are based on how effectively an emotion is appraised (Gross, 2015; Preece et al., 2017). Emotional regulation, therefore, can be seen to act as the mediator between alexithymia and affective disorder symptoms, which may lead to offending behaviours.

Reflecting its clinical relevance, a number of measures have been developed in order to measure the construct alexithymia; most prominently the Toronto Alexithymia Scale (TAS-20)

(Bagby et al., 1994) and the more recent Perth Alexithymia Questionnaire (PAQ) (Preece, Becerra, Robinson, Dandy, et al., 2018). The TAS-20 (Bagby et al., 1994) is a 20-item self-report measure of alexithymia. Items are designed to measure three subscales and provide scores for each (DIF, DDF, and EOT) as well a total score as an overall marker of alexithymia. Although the TAS-20 is the most commonly used measure of alexithymia, it has come under scrutiny for its poor internal reliability of the EOT subscale; while Cronbach's Alpha scores of 0.7 or higher are widely considered acceptable for internal reliability, the reliability coefficients for the EOT subscale of the TAS-20 have often been found to be below 0.6 (Preece, Becerra, Robinson, & Dandy, 2018). It is also worth noting that no known studies have reported a reliability of 0.9 for the total alexithymia composite score, which is the desired value for clinical decision-making (Preece, Becerra, Robinson, Dandy, et al., 2018).

In 2018, Preece and colleagues developed a tool to measure alexithymia, based on their own *attention-appraisal model of alexithymia* (Preece et al., 2017). The attention-appraisal model of alexithymia is a valuation model that posits emotional responses occur after a particular stimulus has been attended to and appraised. Alexithymia manifests within this model, with EOT conceptualised as a difficulty at the *attention* stage, and DIF and DDF conceptualised as a difficulty within the *appraisal* stage. Their new measure was developed with the aim of meeting the following evaluation criteria (Preece, Becerra, Robinson, Dandy, et al., 2018): to allow separate DIF, DDF, and EOT subscales to be derived; to account for emotional valence when assessing functioning at the appraisal stage of emotion valuation (i.e., includes items on both positive and negative emotions in DIF and DDF subscales); and that subscale and composite scores have adequate validity and reliability when tested statistically.

The Perth Alexithymia Questionnaire (PAQ) (Preece, Becerra, Robinson, Dandy, et al., 2018) is a 24-item self-report measure of alexithymia. All items comprise a statement designed to assess the DIF, DDF, or EOT component of alexithymia. Respondents answer each item on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating higher levels of alexithymia. This differs from previous alexithymia measures which used a 5-point Likert scales and there is some evidence to suggest that 7-point scales perform better when measuring continuous constructs (e.g. Preston & Colman, 2000). Items corresponding to the DIF and DDF components of alexithymia include negatively and positively valenced items (e.g. "when I'm feeling bad…" or "when I'm feeling good…") as there is evidence to suggest that those with emotional processing deficits process negative and

positive emotions differently (van der Velde et al., 2013). The 24-item measure includes an equal number of items that attempt to measure each component of alexithymia (eight items for each of the DIF, DDF, and EOT subscales) including four positively and negatively valenced items in each of the DIF and DDF subscales. This allows the PAQ to sum five separate subscales; P-DIF, N-DIF, P-DDF, N-DDF, EOT. In addition to these, the PAQ was designed for the subscales to be combined into theoretically meaningful composite scores. As well as a total score for alexithymia (ALEXI, 24 items), the PAQ can sum an overall DIF score (G-DIF, 8 items), overall DDF score (G-DDF, 8 items), combine the positively valenced subscales to create a *positive-difficulty appraising feelings* composite (P-DAF, 8 items), combine the negatively valenced subscales to create a *negative-difficulty appraising feelings* composite (N-DAF, 8 items), while the P-DAF and N-DAF composites can be combined to create a *general-difficulty appraising feelings* composite (G-DAF, 16 items).

Alexithymia is believed to be normally distributed within the general population, with interest in its academic and clinical utility now extending to forensic settings. Despite growing interest, very few studies have examined the reliability of its measurement in a forensic population. A reason for this could be the difficulty in drawing inferences from test performance within forensic populations from tests that are normed outside a forensic population (i.e., normed on a sample deemed representative of the general population).

This study aims to:

1. describe the psychometric properties of the PAQ in a forensic population, and;
2. explore the relationship between alexithymia and personal, social, and psychological characteristics of an incarcerated population.

Methods

Participants & Procedure

Data for this study was collected from an opportunistic sample of male inmates from a prison in the South of East of England, UK. A range of data is routinely collected from all inmates as standard on admission to the prison, which includes measures on alexithymia, criminal behaviour history, family and social history (i.e. maternal and paternal separation), substance misuse, as well as general demographic and personal characteristics. The author requested

access to this existing set of data (recorded in SPSS) specifically for the purposes of this research, which was granted.

Measures

Perth Alexithymia Questionnaire

The Perth Alexithymia Questionnaire (PAQ; Preece, et al., 2018) is a 24-item self-report measure of alexithymia. All items comprise a statement designed to assess the DIF, DDF, or EOT component of alexithymia, with respondents answering on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating higher levels of alexithymia. Items corresponding to the DIF and DDF components of alexithymia include negatively and positively valenced items (i.e. statements eluding to both positive and negative emotions) allowing for separate positive and negative subscales to be derived. The PAQ includes no negatively-keyed items.

Toronto Alexithymia Scale

The TAS-20 (Bagby et al., 1994) is a 20-item self-report measure of alexithymia. Items are designed to measure three subscales and provide scores for each (DIF, DDF, and EOT) as well a total score as an overall marker of alexithymia, with higher scores indicating higher levels of alexithymia. Each item consists of a statement that participants can agree to on a 5-point Likert scale ranging from 1 ‘strongly disagree’ to 5 ‘strongly agree’. The TAS-20 does not include valence-specific subscales (i.e. separate subscales for both positive and negative emotions), and includes five negatively keyed (thus reverse-scored) items.

Reverse scoring of TAS-20

Reverse-scored items are items that indicate the opposite to the other items in the scale. In the TAS-20, most items are worded so that agreement with the item indicates difficulties (i.e., high alexithymia); however, five items are worded in the opposite way whereby agreement with the item indicates a lack of difficulties. These latter five items must therefore be reverse scored prior to calculating scale scores. In factor analysis, reverse-scored items can often cluster together due to the similarity in their response format (rather than the item content), thus creating a “method” factor within the factor model (see van Sonderen et al., 2013). In this way, reverse-scored items can be disruptive to the factor structure and internal consistency of a measure (Meganck et al., 2008).

Data Analysis

Descriptive statistics and internal consistency

Distribution of the PAQ was examined using the one-sample Kolmogorov-Smirnov test to assess the distribution of scores from normality.

Cronbach's alpha coefficients were calculated for the TAS-20 subscales and PAQ subscales and composite scales, with a value of 0.7 or greater widely considered as an indication for acceptable levels of internal consistency.

The prevalence of alexithymia was calculated using the mean value and standard deviation for each PAQ subtest and composite score for the normal population taken from the normative data provided by Preece and colleagues (2018). These data were compared with the mean value of scores obtained from the prison population in this study in order to ascertain whether scores on the PAQ differ in a prison population vs the general population.

Reliable change was calculated from standard deviations for each of the PAQ composites and subscales.

Convergent validity

Convergent validity refers to how closely a test relates to other tests that measure the same construct. Pearson correlations were calculated to assess associations between the TAS-20 subscales and PAQ subscales and composite scales.

Correlations with other variables

The relationship between alexithymia scores on the PAQ and a number of variables collected on admission to prison were examined; including maternal and paternal separation, and substance misuse. Associations were also explored between alexithymia and self-esteem, comparing scores on the PAQ and the Culture-free Self-esteem Inventory (CFSEI; (Battle, 1981) as associations between alexithymia and self-esteem have been established previously in the general population (e.g. Yelsma (1995), but not in a forensic setting. Associations were also explored between alexithymia and blame, comparing scores on the PAQ and the Gudjonsson Blame Attribution Inventory-Revised (GBAI-R; Gudjonsson & Singh, 1989). The GBAI-R is conducted routinely on admission to prison. As blame is essentially an emotional

regulation strategy, and alexithymia is underpinned by difficulties with emotional regulation, correlations were used to explore this relationship.

Factorial validity

The factor structure of the PAQ was assessed using a series of confirmatory factor analyses, using maximum likelihood estimation with robust standard errors. For the factor structure of the PAQ suggested by Preece and colleagues (2018) , see Figure 10. This includes a single-factor model (all 24 items), a two-factor model (G-EOT and G-DIF/G-DDF [G-DAF] combined), a non-valenced three-factor model (G-EOT, G-DIF, and G-DDF), a valenced three-factor model (G-EOT, N-DAF, and P-DAF), and a valenced five-factor model (G-EOT, N-DIF, P-DIF, N-DDF, and P-DDF). The factor structure of the PAQ is presented in Figure 10.

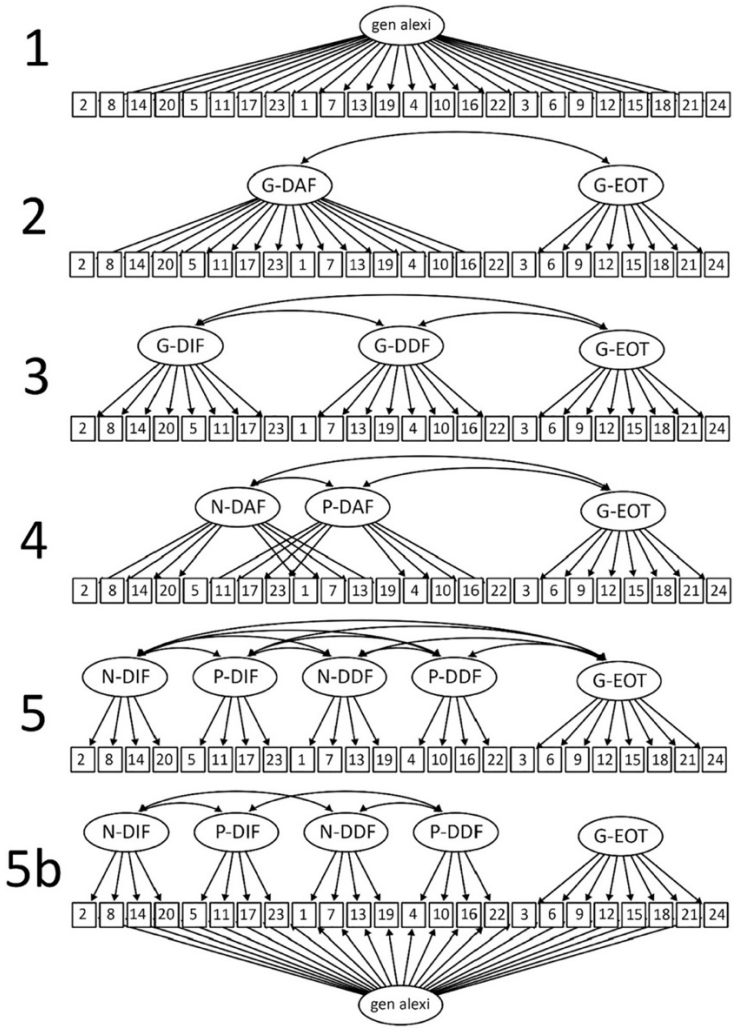


Figure 10. Factor structure of the PAQ

In order to evaluate the goodness-of-fit of these models the Satorra-Bentler scaled chi-squared test ($SB\chi^2$; (Satorra, 1988) was calculated. This procedure was selected as it has been demonstrated to be robust to non-normality, small sample sizes and high model complexity (Brown, 2015). In addition to the Satorra-Bentler scaled chi-squared test, it is recommended that supplementary fit indices are used to quantify the fit of a model (Hu & Bentler, 1999). In line with the guidelines of Kline and colleagues (2011), the comparative fit index (CFI), standardised root mean square residual (SRMR) and root mean square error of approximation (RMSEA) indices were calculated. Hu and Bentler (1999) provide approximate cut-off values for the good fit between a proposed model and the observed data: CFI values close to ≥ .95, SRMR values close to ≤ .08 and RMSEA values close to ≤ .06.

The corrected Akaike Information Criterion (AIC_c), which includes a correction for smaller sample sizes (Burnham & Anderson, 2004), was used to compare the fit of each model. The AICc includes a correction for smaller sample sizes and penalises model complexity. A smaller AICc value indicates a better model fit controlling for model complexity (Brown, 2015). The difference in AICc values between two models (ΔAICc) provides a measure of the relative quality of the two statistical models in terms of minimising information loss, with a difference of greater than ten points (ΔAICc > 10) indicating no support for the model with the higher AIC value (Fabozzi et al., 2014). A formal test of model adequacy is provided by calculating the "relative likelihood" (i.e., $\exp((AIC_{min} - AIC_i)/2)$). Relative likelihood values less than 0.05 do not support for the model with the higher AIC value.

Analysis strategy

Estimated relationships between the PAQ and other collected data were obtained by bootstrapping the normative sample 10,000 times and then averaging the results of the individual resampled estimates. Bootstrapping was employed for all analyses as recommended for deriving normative data from small sample sizes.

Bootstrapping

Traditional approaches to the norming of psychometric measures have relied upon obtaining large data sets in which average performance and variation is estimated either directly from the variability of data within relevant demographic and person specific factors (e.g., as derived from classical testing theory). Unfortunately, when even quite large samples are divided into

smaller groups on the basis of demographic and person specific factors then the sample sizes that inferences are actually based upon, can quickly become quite small and unrepresentative.

An alternative would be to estimate normative performance by fitting a detailed statistical model of test performance and respondent characteristics (e.g., as derived from item response theory). Unfortunately, these statistical models are based on explicit distributional assumptions of both test performance and the distribution of respondent characteristics in the general population. The history of psychometrics is replete with examples of where such assumptions have later been discovered to be incorrect. Even if the pertinent test performance and respondent characteristics are identified and their distributions in the general population are known, it still remains unclear whether such characteristics and distributions can be appropriately extended for specialist populations.

The bootstrap method, which is similar to the Monte Carlo method used by Crawford and colleagues (2007), involves making an estimate (such as sample mean) of a population parameter (such as population mean) from sample data. This estimate is obtained by repeatedly resampling, with replacement, a large number of times from the existing sample data. An estimate for the population parameter is calculated from each of the bootstrap samples. The final bootstrap estimate for the population parameter is, typically, the mean of the large number of resampling estimates and the sampling distribution of the population parameter (e.g. the standard error of measurement) is obtained from the empirical distribution of resampled parameter estimates. Therefore, the bootstrap method can be seen as a form of Monte Carlo simulation of the population parameter and the variability of this estimated population parameter within the bootstrap resamples provides a method of empirically assessing the variability of the parameter within the specific sample. Unlike traditional statistical methods, the statistical power of a bootstrap estimate is provided by the number of bootstrap samples are obtained from the original sample come out with larger numbers of bootstrap samples indicating greater statistical power. Also, the bootstrap method makes no specific claims to external validity and generalizability, the bootstrap parameter estimate is claimed to be valid for the sample from which it is derived, the generalizability of the bootstrap parameter estimate is given by the "similarity" of the participants in the sample to those persons outside of the sample for whom an inference is to be extended. Accordingly, the generalizability of the bootstrap parameter estimate is provided by the psychological and other pertinent factors within the sample from which the estimate was derived. Finally, as the bootstrap parameter

estimate and its distribution are empirically estimated from many repeated bootstrap samples, the bootstrap parameter estimate remains valid in non-normal distributions and in distributions that show unusual characteristics (e.g., bimodality).

Therefore, bootstrap methods are ideally suited to situations where normative data is to be derived from relatively small sample sizes and the distribution of the test parameters may deviate from normality.

Results

Demographic Characteristics

Demographic characteristics are presented in Table 13. This study included data from 128 male inmates. Ages ranged from 21 to 73 years, with a mean age of 38.8 years. The majority of participants were white (n= 88, 75%), a UK national (n=111, 95.7%), and single (n=100, 90.1%). The majority of participants were convicted of murder or manslaughter (n=54, 51.9%) vs other types of crime, while a mandatory sentence type was most common (n=43, 40.6%). 38.5% of participants reported having a problem with alcohol, with 92.9% of participants reporting having used illicit drugs.

Table 13. Demographics

		Count	Column N %	95.0% Lower CL for Column Total N %	95.0% Upper CL for Column Total N %
age in 10 year bands	21-30 yrs	39	33.3%	23.0%	38.8%
	31-40 yrs	35	29.9%	20.2%	35.5%
	41-50 yrs	21	17.9%	10.8%	23.5%
	51+ yrs	22	18.8%	11.4%	24.4%
	Total	117	100.0%	.	.
Ethnicity Grouped	White	88	75.2%	60.4%	76.3%
	Asian	6	5.1%	2.0%	9.4%
	Black	13	11.1%	5.8%	16.3%
	Mixed	9	7.7%	3.5%	12.4%
	Not stated	1	0.9%	0.1%	3.6%
	Total	117	100.0%	.	.
Marital Status from LIDS	Single	100	90.1%	70.4%	84.6%
	Married	7	6.3%	2.5%	10.4%
	Civil Union	2	1.8%	0.3%	4.9%
	Divorced	2	1.8%	0.3%	4.9%
	Total	111	100.0%	.	.
crime classification	Murder / manslaughter	54	51.9%	33.9%	50.8%

		Count	Column N %	95.0% Lower CL for Column Total N %	95.0% Upper CL for Column Total N %
	Violence to person	15	14.4%	7.0%	18.1%
	sexual offences	18	17.3%	8.9%	20.9%
	robbery	15	14.4%	7.0%	18.1%
	firearms	2	1.9%	0.3%	4.9%
	Total	104	100.0%	.	.
Sentence type	Discretionary	8	7.5%	3.0%	11.4%
	Mandatory	43	40.6%	25.8%	42.1%
	IPP	13	12.3%	5.8%	16.3%
	Determinate (pre Apr05)	1	0.9%	0.1%	3.6%
	Determinate (post Apr05)	17	16.0%	8.2%	20.0%
	EDS LASPO	21	19.8%	10.8%	23.5%
	EDS	3	2.8%	0.7%	6.1%
	Total	106	100.0%	.	.
	No. of guilty adjudications	128	100.0%	.	.
Drink problem outside	No	16	61.5%	42.4%	78.2%
	Yes	10	38.5%	21.8%	57.6%
	Total	26	100.0%	.	.
Treatment for alcohol abuse	No	22	84.6%	67.5%	94.6%
	Yes	4	15.4%	5.4%	32.5%
	Total	26	100.0%	.	.
Ever used illicit drugs	No	2	7.1%	1.5%	21.0%
	Yes	26	92.9%	79.0%	98.5%
	Total	28	100.0%	.	.
Ever been registered addict	No	22	81.5%	64.1%	92.6%
	Yes	5	18.5%	7.4%	35.9%
	Total	27	100.0%	.	.
Has he ever had drug treatment inside or outside prison	no	12	50.0%	31.0%	69.0%
	inside	7	29.2%	14.1%	48.9%
	outside and inside	5	20.8%	8.4%	39.8%
	Total	24	100.0%	.	.

Descriptive Statistics & Internal Consistency

Distribution of the PAQ

Table 14 presents the distribution of scores for the subscales and composites of the PAQ. The mean score for the total PAQ (24 items, lowest possible score = 24, highest possible score = 168), was 87.6 (SD = 34.466), with a median of 89. Scores ranged from 24 to 168. Scores of 27.8 sit around the 5th percentile, 62 the 25th percentile, 111 the 75th percentile, while scores of 149.2 are in the 95th percentile. See Table 14 for the relative distribution of all subscale and composite scores for the PAQ.

Table 14. Distribution of PAQ scores

		ALEXI (total)	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
Mean		87.60	14.65	12.79	16.41	13.57	30.19	31.06	26.35	57.41	27.43	29.98
Std. Error of Mean		3.058	0.569	0.578	0.595	0.593	1.117	1.115	1.143	2.152	1.091	1.118
Std. Deviation		34.466	6.415	6.517	6.705	6.686	12.586	12.565	12.876	24.251	12.298	12.600
Median		89.00	15.00	12.00	17.00	14.00	31.00	32.00	26.00	60.00	27.00	31.00
Minimum		24	4	4	4	4	8	8	8	16	8	8
Maximum		168	28	28	28	28	56	56	56	112	56	56
	2	24.00	4.00	4.00	4.00	4.00	8.00	8.00	8.00	16.00	8.00	8.00
	5	27.80	4.00	4.00	4.00	4.00	8.00	8.40	8.00	16.40	8.00	8.40
	16	45.48	7.48	4.00	9.00	6.00	15.48	17.48	10.48	30.00	14.00	16.48
	25	62.00	10.00	8.00	12.00	8.00	20.00	22.00	16.00	38.00	17.00	20.00
	50	89.00	15.00	12.00	17.00	14.00	31.00	32.00	26.00	60.00	27.00	31.00
	75	111.00	20.00	17.00	21.00	19.00	39.00	40.00	36.00	75.00	36.00	39.00
	90	134.40	23.00	22.00	26.00	23.00	46.40	48.00	44.00	89.20	44.00	48.00
Percentiles	95	149.20	24.60	24.60	27.00	25.00	54.00	51.20	48.60	98.60	49.00	51.00

Prevalence of alexithymia in prison and normal populations

Table 15 presents the mean value and standard deviation for each PAQ subtest and composite score for the normal population (provided by Preece and colleagues (2018) study 2) and participants from this study. Significant differences between scores were seen in the P-DIF and N-DIF subscales, and G-DAF and G-DIF composite scales.

Table 15. Mean scores on PAQ vs normal population taken from Preece et al. (2018).

	Normal Population			Prison Population				
	Mean	SD	N	Mean	SD	N	t	P
Subscales								
P-DIF	11.3	5.76	748	12.79	6.517	128	2.651	0.008
N-DIF	13.38	6.41	748	14.65	6.415	128	2.072	0.039
P-DDF	12.97	6.12	748	13.57	6.686	128	1.01	0.313
N-DDF	15.35	6.89	748	16.41	6.705	128	1.613	0.107
G-EOT	28.97	11.19	748	30.19	12.586	128	1.118	0.264
Composites								
ALEXI	81.97	30.91	748	87.60	34.466	128	1.872	0.062
N-DAF	28.73	12.71	748	31.06	12.565	128	1.919	0.055
P-DAF	24.27	11.39	748	26.35	12.876	128	1.872	0.062

	Normal Population			Prison Population				
	Mean	SD	N	Mean	SD	N	t	P
G-DAF	52.99	22.58	748	57.41	24.251	128	2.024	0.043
G-DDF	28.32	12.16	748	29.98	12.600	128	1.42	0.156
G-DIF	23.68	11.28	748	27.43	12.298	128	3.428	0.001

Internal reliability of the PAQ and TAS-20

The Cronbach's alpha coefficient for the PAQ (total scale) was 0.961, indicating very high internal reliability of the measure as a whole. For comparison, the internal reliability for the TAS-20 total scale was 0.879. The PAQ subscale and composites all demonstrate acceptable internal reliability (see Table 16) with the lowest alpha coefficient being 0.827 for the N-DDF (*negative-difficulty describing feelings*) subscale. Similar to other studies, this study found the TAS-20 to have acceptable internal reliability alpha coefficients across its subscales with the exception of EOT (*externally oriented thinking*) which had an alpha coefficient of 0.560.

Table 16. Internal consistency of PAQ and TAS-20 subscales

PAQ	Cronbach's alpha	items
Subscales		
P-DIF	0.868	4
N-DIF	0.834	4
P-DDF	0.876	4
N-DDF	0.827	4
G-EOT	0.917	8
Composites		
ALEXI (total)	0.961	24
N-DAF	0.908	8
P-DAF	0.935	8
G-DAF	0.953	16
G-DDF	0.910	8
G-DIF	0.914	8
TAS-20	**Cronbach's alpha**	**items**
Subscales		
Total scale	0.879	20
DIF	0.889	7
DDF	0.804	5

PAQ	Cronbach's alpha	items
EOT	0.560	8

Reliable Change

The test-retest reliability coefficients reported in the meta-analysis of Jackson (2023, this volume) were used to assess the reliable change on the PAQ subscales and composite scores. Table 17 presents reliable change estimates in the normal population (as presented by Preece and colleagues (2018) study 2) and in the prison population from the current study. Reliable change estimates are presented at 66%, 95% and 99% confidence intervals. For example, for a total alexithymia score (whole scale) change of 38 points within this prison population, confidence of reliable change would be at 99% (vs a change of 34 points in the normal population). A change of 19 points within this prison population would estimate 66% confidence of reliable change (vs a change of 17 points in the normal population).

Table 17. Fixed effects model weighted average test-retest reliability and measures of heterogeneity (reported from Jackson (2023, this volume)

	Test/retest Reliabilities for the PAQ				Reliable Change in Normal Population*			Reliable Change in Prison Population**		
	Asl et al. 2020	Lashkar et al. 2021	Weighted Average Coefficient	95%-CI	66%	95%	99%	66%	95%	99%
Subscales										
P-DIF	0.87	0.71	0.75	0.7057 to 0.7875	4.07	7.98	8.15	4.61	9.04	9.22
N-DIF	0.93	0.72	0.82	0.7870 to 0.8526	3.85	7.54	7.69	3.85	7.54	7.70
P-DDF	0.84	0.69	0.72	0.6731 to 0.7621	4.58	8.98	9.16	5.01	9.81	10.01
N-DDF	0.84	0.71	0.73	0.6894 to 0.7744	5.06	9.92	10.13	4.93	9.66	9.86
G-EOT	0.88	0.74	0.77	0.7337 to 0.8087	7.59	14.88	15.18	8.54	16.74	17.08
Composites										
ALEXI	0.86	0.85	0.85	0.8257 to 0.8758	16.93	33.18	33.86	18.88	37.00	37.76
N-DAF		0.75	0.75	0.7089 to 0.7911	8.99	17.62	17.97	8.89	17.42	17.78
P-DAF		0.76	0.76	0.7203 to 0.7997	7.89	15.47	15.78	8.92	17.49	17.85
G-DAF		0.81	0.81	0.7777 to 0.8423	13.92	27.28	27.84	14.95	29.30	29.90
G-DDF		0.77	0.77	0.7317 to 0.8083	8.25	16.16	16.49	8.55	16.75	17.09
G-DIF		0.76	0.76	0.7203 to 0.7997	7.82	15.32	15.63	8.52	16.70	17.04

** Standard deviations for the normal population reported by Preece et al. (2018) ** Standard deviations for the prison population were calculated from the data reported in this paper.*

Convergent Validity

Convergent validity of the PAQ with the TAS-20

Pearson correlations between the TAS-20 and PAQ subscales are presented in Table 18. All PAQ subscales and composite scales highly correlate with the TAS-20 subscales to a significance of <0.01, with correlations consistently lower with the TAS EOT (externally oriented thinking) subscale.

Table 18. Correlations between TAS-20 and PAQ scores

	PAQ ALEXI	PAQ N-DIF	PAQ P-DIF	PAQ N-DDF	PAQ P-DDF	PAQ G-EOT	PAQ N-DAF	PAQ P-DAF	PAQ G-DAF	PAQ G-DIF	PAQ G-DDF
TAS ALEXI	.721**	.722**	.661**	.711**	.643**	.521**	.752**	.671**	.749**	.732**	.724**
TAS DIF	.589**	.695**	.588**	.601**	.534**	.330**	.680**	.577**	.662**	.679**	.607**
TAS DDF	.768**	.736**	.674**	.772**	.686**	.579**	.792**	.700**	.786**	.746**	.780**
TAS EOT	.451**	.329**	.372**	.401**	.389**	.442**	.384**	.392**	.409**	.371**	.423**

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations with other personal variables

Developmental characteristics and experiences

PAQ association with maternal and paternal separation

35.3% of participants report being separated from their birth mother by the age of 15, while 53.1% report being separated from their father by the age of 15. Table 19 presents Pearson correlations between maternal and paternal separation during childhood and PAQ subscale and composite scores. No significant relationship was found between maternal or paternal separation and alexithymia scores.

Table 19. Correlations between PAQ scores and maternal and paternal separation

	ALEXI (total)	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
Separated from mother/stepmother Pearson Correlation	-0.178	-0.056	-0.158	-0.151	-0.233	-0.196	-0.107	-0.2	-0.161	-0.111	-0.204

Sig. (2-tailed)		0.314	0.752	0.373	0.395	0.184	0.268	0.549	0.256	0.363	0.53	0.247
N		34	34	34	34	34	34	34	34	34	34	34
Bootstrapc	Bias	0.004	0.005	0.006	0.005	0.005	0.004	0.005	0.005	0.005	0.006	0.005
	Std. Error	0.121	0.143	0.133	0.128	0.115	0.126	0.132	0.122	0.121	0.13	0.116
	Lower 95% CI	-0.434	-0.318	-0.419	-0.398	-0.458	-0.457	-0.358	-0.44	-0.409	-0.367	-0.436
	Upper 95% CI	0.048	0.231	0.094	0.104	-0.021	0.059	0.167	0.029	0.071	0.144	0.017
Separation from father/stepfather 1yr+												
Pearson Correlation		0.012	0.041	-0.048	0.007	-0.032	0.049	0.024	-0.04	-0.007	-0.001	-0.013
Sig. (2-tailed)		0.946	0.818	0.79	0.969	0.857	0.784	0.892	0.822	0.968	0.995	0.943
N		34	34	34	34	34	34	34	34	34	34	34
Bootstrapc	Bias	0.013	0.009	0.015	0.008	0.014	0.011	0.01	0.015	0.013	0.013	0.012
	Std. Error	0.176	0.18	0.172	0.178	0.178	0.173	0.179	0.175	0.176	0.175	0.177
	Lower 95% CI	-0.315	-0.285	-0.356	-0.318	-0.36	-0.286	-0.287	-0.361	-0.332	-0.318	-0.345
	Upper 95% CI	0.387	0.417	0.303	0.379	0.332	0.408	0.403	0.313	0.367	0.362	0.362

Offence type

PAQ association with offence type

Offence types were classified into murder and manslaughter, violence to person, sexual offence, robbery, and firearms offence (for descriptive statistics see Table 13). Table 20 presents Pearson correlations between offence type and PAQ scores. There was no significant relationship between offence type and alexithymia as scored on the PAQ.

Table 20. Correlations between PAQ scores and offence type

		ALEXI (total)	**N-DIF**	**P-DIF**	**N-DDF**	**P-DDF**	**G-EOT**	**N-DAF**	**P-DAF**	**G-DAF**	**G-DIF**	**G-DDF**
Offence type		-0.062	-0.096	-0.069	-0.062	-0.024	-0.036	-0.083	-0.047	-0.068	-0.088	-0.046
Sig. (2-tailed)		0.534	0.33	0.489	0.532	0.806	0.718	0.405	0.634	0.49	0.376	0.64
N		104	104	104	104	104	104	104	104	104	104	104
Bootstrap[c]	Bias	0.002	0.003	0.001	0.003	0.002	0.001	0.003	0.001	0.002	0.002	0.002
	Std. Error	0.097	0.094	0.093	0.096	0.095	0.094	0.096	0.095	0.096	0.094	0.097
95% Confidence Interval	Lower	-0.255	-0.277	-0.253	-0.252	-0.213	-0.217	-0.275	-0.228	-0.26	-0.272	-0.247
	Upper	0.124	0.105	0.114	0.126	0.163	0.156	0.108	0.138	0.124	0.1	0.141

Substance abuse

PAQ association with substance misuse

Substance abuse here combines both reported alcohol and drug use (see Table 13). Table 21 shows Pearson correlations between any substance misuse and PAQ scores, with no significant relationship.

Table 21. Correlation between PAQ and any substance misuse

			N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
Any substance abuse	Pearson Correlation		-0.035	0.234	-0.075	0.181	0.015	-0.057	0.210	0.077	0.099	0.054
	Sig. (2-tailed)		0.861	0.231	0.703	0.356	0.940	0.774	0.284	0.698	0.617	0.786
	N		28	28	28	28	28	28	28	28	28	28
	Bootstrap	Bias	-.017	.004	-.017	.003	-.012	-.018	.003	-.010	-.009	-.010
		Std. Error	.237	.162	.257	.189	.207	.253	.178	.190	.176	.199
		Lower 95% CI	-.471	-.117	-.536	-.213	-.353	-.525	-.168	-.266	-.215	-.309
		Upper 95% CI	.428	.523	.448	.493	.427	.447	.509	.450	.452	.441

Self Esteem

PAQ association with self-esteem

The Culture-free Self-esteem Inventory (CFSEI; (Battle, 1981) measures four different dimensions, including general self-esteem, personal self-esteem, social self-esteem, and lie scales (to indicate defensiveness). 40-items are summed for a total self-esteem score, with higher scores indicating more positive self-esteem. Table 23 presents the relationship between PAQ subscale and composite scores and the CFSEI subscale and total scores. There are many significant relationships between subscale scores (see Table 22), with all significant relationships being a negative correlation. In effect, this means that those who scored higher on the self-esteem measure, scored lower on the PAQ for alexithymia. For example, total self-esteem score had a significant negative correlation with total alexithymia score ($r = -.39$; $p = 0.02$). Total self-esteem score had higher negative correlations with the negatively-valenced items on the PAQ than those positively-valenced; for example N-DIF ($r = -.55$; $p < 0.01$), N-DDF ($r = -.49$; $p < 0.01$), N-DAF ($r = -.55$; $p < 0.01$).

Table 22. Correlations between PAQ and CFSEI

				ALEXI (total)	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
CFSEI Social se	Pearson Correlation			-0.30	-.44**	-0.29	-.35*	-.38*	0.03	-.41*	-.35*	-.41*	-.39*	-.40*
	Sig. (2-tailed)			0.09	0.01	0.10	0.04	0.03	0.84	0.02	0.05	0.02	0.02	0.02
	N			34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00
	Bootstrap[c]	Bias		-0.01	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00
		Std. Error		0.19	0.17	0.19	0.18	0.18	0.21	0.18	0.18	0.18	0.18	0.18
		95% Confidence Interval	Lower	-0.64	-0.73	-0.62	-0.66	-0.69	-0.36	-0.71	-0.68	-0.72	-0.70	-0.71
			Upper	0.10	-0.09	0.10	0.01	0.00	0.43	-0.02	0.03	-0.02	0.00	-0.02
CFSEI Personal se	Pearson Correlation			-0.29	-.44**	-0.30	-.39*	-0.18	-0.02	-.44**	-0.24	-.37*	-.40*	-0.32
	Sig. (2-tailed)			0.10	0.01	0.09	0.02	0.30	0.90	0.01	0.17	0.03	0.02	0.07
	N			34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00
	Bootstrap[c]	Bias		0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		Std. Error		0.21	0.16	0.18	0.18	0.19	0.22	0.17	0.19	0.18	0.17	0.20
			Lower	-0.65	-0.71	-0.61	-0.71	-0.54	-0.42	-0.73	-0.59	-0.68	-0.68	-0.66

				ALEXI (total)	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
		95% Confidence Interval	Upper	0.16	-0.08	0.10	0.02	0.20	0.42	-0.04	0.15	0.05	-0.02	0.10
CFSEI lie scale on rec	Pearson Correlation			-0.20	-0.31	-0.16	-0.10	-0.15	-0.13	-0.21	-0.16	-0.20	-0.25	-0.14
	Sig. (2-tailed)			0.25	0.07	0.38	0.58	0.38	0.46	0.23	0.36	0.26	0.15	0.43
	N			34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00
	Bootstrap[c]	Bias		0.02	0.01	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.02
		Std. Error		0.16	0.13	0.15	0.17	0.18	0.16	0.15	0.17	0.16	0.14	0.17
		95% Confidence Interval	Lower	-0.48	-0.54	-0.46	-0.40	-0.47	-0.43	-0.48	-0.46	-0.48	-0.50	-0.44
			Upper	0.14	-0.03	0.16	0.27	0.21	0.19	0.11	0.17	0.14	0.05	0.24
CFSEI General se	Pearson Correlation			-.41*	-.55**	-.39*	-.50**	-0.27	-0.12	-.55**	-0.34	-.48**	-.51**	-.42*
	Sig. (2-tailed)			0.02	0.00	0.02	0.00	0.12	0.52	0.00	0.05	0.00	0.00	0.01
	N			34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00
	Bootstrap[c]	Bias		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		Std. Error		0.15	0.11	0.15	0.15	0.16	0.18	0.13	0.15	0.12	0.11	0.14
		95% Confidence Interval	Lower	-0.68	-0.73	-0.64	-0.75	-0.58	-0.45	-0.76	-0.61	-0.69	-0.69	-0.68
			Upper	-0.08	-0.30	-0.07	-0.18	0.06	0.28	-0.28	-0.01	-0.20	-0.26	-0.12
CFSEI Total reception self esteem	Pearson Correlation			-.39*	-.55**	-.38*	-.49**	-0.31	-0.06	-.55**	-.35*	-.48**	-.51**	-.44*
	Sig. (2-tailed)			0.02	0.00	0.03	0.00	0.08	0.74	0.00	0.04	0.00	0.00	0.01
	N			34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00	34.00
	Bootstrap[c]	Bias		0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
		Std. Error		0.17	0.12	0.17	0.16	0.18	0.20	0.14	0.17	0.15	0.13	0.16
		95% Confidence Interval	Lower	-0.68	-0.75	-0.66	-0.75	-0.62	-0.42	-0.77	-0.65	-0.72	-0.72	-0.71
			Upper	-0.01	-0.28	-0.02	-0.15	0.08	0.38	-0.23	0.02	-0.16	-0.21	-0.08

Blame Attribution

PAQ association with blame attribution

The Gudjonsson Blame Attribution Inventory-Revised (Gudjonsson & Singh, 1989) measures three elements: mental element attribution (i.e. blame attributed to mental illness or poor self-control), external attribution (blame attributed to social circumstances, the victim, or society), and guilt feeling attribution (feelings of regret and remorse). Table 23 presents Pearson correlations between blame attribution scores and alexithymia scores on the PAQ. There was no significant relationship found between PAQ scores and external or guilt elements of blame. Significant relationships were found between the mental element attribution and alexithymia scores on the PAQ, indicating that those who scored higher on the PAQ were more likely to attribute blame to mental illness or poor self-control.

Table 23. Correlations between PAQ and GBAI-R

		ALEXI (total)	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	N-DAF	P-DAF	G-DAF	G-DIF	G-DDF
	Pearson Correlation	.390*	.453**	.357*	0.262	.371*	0.218	.372*	.376*	.403*	.438**	.348*
GBAI	Sig. (2-tailed)	0.023	0.007	0.038	0.134	0.031	0.216	0.030	0.028	0.018	0.010	0.043
Mental	N	34	34	34	34	34	34	34	34	34	34	34

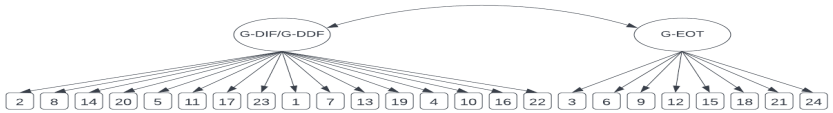
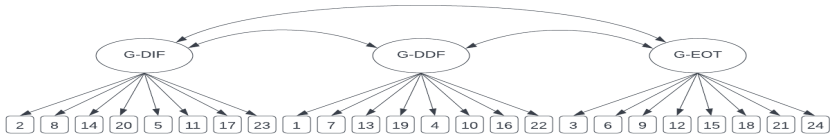
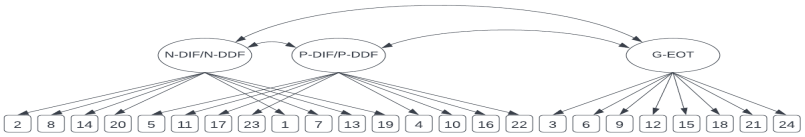
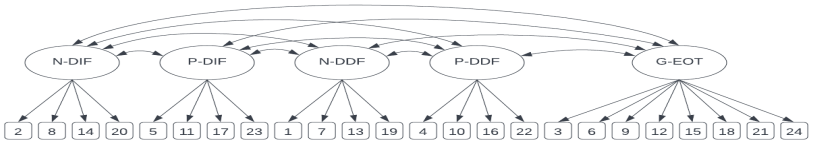
		Bias		-0.007	-0.011	-0.011	-0.011	-0.002	-0.005	-0.012	-0.007	-0.010	-0.013	-0.006
		Std. Error		0.157	0.126	0.151	0.160	0.159	0.174	0.144	0.153	0.153	0.140	0.164
		95% Confidence	Lower	0.061	0.178	0.016	-0.056	0.047	-0.138	0.061	0.048	0.073	0.123	0.017
	Bootstrapc	Interval	Upper	0.674	0.664	0.622	0.564	0.668	0.543	0.641	0.639	0.669	0.671	0.651
	Pearson Correlation			0.153	0.074	0.207	0.100	0.056	0.163	0.092	0.131	0.120	0.151	0.086
	Sig. (2-tailed)			0.387	0.679	0.241	0.573	0.754	0.357	0.604	0.459	0.497	0.395	0.630
	N			34	34	34	34	34	34	34	34	34	34	34
		Bias		0.001	0.003	0.004	0.005	0.001	0.003	0.004	0.002	0.004	0.005	0.004
		Std. Error		0.146	0.156	0.159	0.156	0.166	0.147	0.154	0.161	0.156	0.159	0.155
		95% Confidence	Lower	-0.137	-0.220	-0.085	-0.181	-0.251	-0.137	-0.192	-0.173	-0.175	-0.145	-0.201
GBAI External	Bootstrapc	Interval	Upper	0.428	0.378	0.519	0.410	0.393	0.437	0.400	0.456	0.428	0.471	0.401
	Pearson Correlation			0.181	0.262	0.104	0.160	0.218	0.056	0.220	0.170	0.210	0.199	0.208
	Sig. (2-tailed)			0.305	0.135	0.556	0.365	0.216	0.754	0.211	0.338	0.234	0.260	0.237
	N			34	34	34	34	34	34	34	34	34	34	34
		Bias		0.005	0.005	0.006	0.002	0.006	0.002	0.003	0.007	0.006	0.007	0.004
		Std. Error		0.148	0.145	0.157	0.130	0.162	0.159	0.134	0.158	0.141	0.149	0.140
		95% Confidence	Lower	-0.091	-0.016	-0.194	-0.086	-0.099	-0.236	-0.040	-0.126	-0.057	-0.073	-0.069
GBAI Guilt	Bootstrapc	Interval	Upper	0.492	0.546	0.428	0.410	0.541	0.388	0.481	0.492	0.495	0.508	0.495

As the correlations presented here were used on a smaller sub-sample of participants (as these measures were voluntarily completed), a post-hoc power calculation was conducted. Using G-Power (https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower) to calculate the post-hoc power of a correlation when calculated on 34 participants, then a small correlation of r=0.1 would return a power of 0.134, a medium correlation of r=0.3 would return a power of 0.514 and a large correlation of r=0.5 would return a power of 0.925. Accordingly, from the sample size used to calculate the correlations reported in this section (n=34), correlations greater than r=0.42 can be reported with a power greater than 0.8. The failure to find statistical significance in correlations smaller than 0.42 may reflect a type 2 bias.

Factorial Validity

The PAQ factor structure was found to be best represented by the five-factor model, with the CFI, SRMR and RMSEA values all indicating good fit (CFI = .937, SRMR = .073, RMSEA = .091) (see Table 24). With respect to the model AIC values, the five-factor model showed a clear advantage in that the AIC values associate with the other models were all substantially higher ($\Delta AIC_c > 10$) and the relative likelihood of models 1 to 4 were all lower than 0.05. Therefore, models 1 to 4 should be considered a poorer fit to the data than the five factor model, which contains separate factors for positive and negatively valenced items.

Table 24. Goodness-of-fit index values for the different confirmatory factor analysis models of the PAQ

	Goodness-of-fit Indices					Model comparison		
Measure	**SBχ2(df)**	**p**	**CFI**	**SRMR**	**RMSEA (90% CI)**	**AIC_c**	**ΔAIC_c**	**Relative likelihood**
 **1. Single factor model (General factor)**	561.138 (249)	<0.001	0.868	0.090	0.131 (0.117 to 0.146)	11030.528	305.606	<0.00001
 **2. Two factor model (G-DIF/G-DDF, EOT)**	463.046 (251)	<0.001	0.857	0.077	0.082 (0.073 to 0.090)	10825.372	100.45	<0.00001
 **3. Three-factor model - no valence (G-DIF, G-DDF, EOT)**	461.600 (249)	<0.001	0.909	0.075	0.107 (0.092 to 0.122)	10821.734	96.812	<0.00001
 **4. Three-factor model – valence (N-DIF/N-DDF, P-DIF/P-DDF, G-EOT)**	400.608 (249)	<0.001	0.935	0.074	0.092 (0.075 to 0.108)	10744.263	19.341	0.00006
 **5. Five factor model (N-DIF, N-DDF, P-DIF, P-DDF, G-EOT)**	389.411 (242)	<0.001	0.937	0.073	0.091 (0.074 to 0.108)	10724.922	Minimum model	

SBχ2 – Satorra-Bentler scaled chi-squared test. CFI - Comparative fit index – value close to ≥ .95 indicates good fit; SRMR - Standardised root mean square residual – value close to ≤ .08 indicates good fit; RMSEA (90% CI) - Root mean square error of approximation – value close to ≤ .06 indicates good fit. AICc – Second-order Akaike information Criteria. The AICc penalises for model complexity and lower values indicate a more parsimonious model; ΔAICc – difference between the AICc with the minimum value and the candidate model. ΔAICc > 10 indicates a meaningful difference between the model and the candidate model. Relative likelihood values less than 0.05 suggest little support for the model with the higher AIC value

Discussion

This study sought to explore the psychometric properties and utility of the Perth Alexithymia Questionnaire (PAQ) in a forensic population. This study looked at the distribution of scores on the PAQ, the prevalence of alexithymia compared to the normal population, and compared PAQ scores with other key variables within this population, such as blame and self-esteem.

This study conducted a series of confirmatory factor analyses, replicating those of the original authors, in order to assess the factor structure of the PAQ in a forensic population. The factor structure of the PAQ was consistent with the original authors' findings (Preece et al., 2017; Preece, Becerra, Robinson, Dandy, et al., 2018), with all factors found to be significantly correlated. The five-factor model (model 5) produced the best fit indexes overall, suggesting that the PAQ is best represented by the intended structure of the measure with 5 subscales and a general alexithymia total composite. This suggests that the PAQ does accurately measure the multidimensional construct of alexithymia and provides further support for the measurement of emotional valence, which had been overlooked by other measures of alexithymia.

The subscales and composite scales of the PAQ were found to be reliable measures of their intended constructs, consistent to those of the original authors. This study replicated previous findings when testing the reliability of the TAS-20, which was found to have poor reliability within its EOT (*externally oriented thinking*) subscale (Bermond et al., 2007; D. Preece et al., 2017). The PAQ was developed with the intention of creating a reliable EOT subscale, and the present study demonstrates the reliability of the PAQ's G-EOT component. In the present study, all PAQ subscales and composite scales scored alpha coefficients of higher than 0.8. G-EOT, and all composite scales, had alpha coefficients of higher than 0.9. Preece and colleagues (2018) had previously purported that no known previous studies had reported a reliability of 0.9 for the total alexithymia composite score, which is the desired value for clinical decision making.

In terms of convergent validity, correlations between subscales of the TAS-20 and PAQ demonstrate high convergent validity for both measures, demonstrating that the PAQ and TAS-20 are measuring a highly similar construct. All subscales highly correlated with one another, though relationships were weakest with the TAS-20 EOT subscale. Although still significant,

this finding was expected considering EOT's lower internal reliability, which provides further support for the validity of the PAQ's G-EOT subscale, and as a whole.

This study compared mean scores from the PAQ subscales and composite scales from this forensic population with those from a normal population (presented by Preece et al., 2018). Alexithymia scores in the forensic population were significantly higher on both P-DIF and N-DIF subscales, and G-DIF and G-DAF composites. These findings suggest that those in a prison population have significantly greater difficulty identifying their feelings compared to the normal population. When considering the attention-appraisal model of alexithymia (Preece et al., 2017), difficulty identifying feelings is conceptualised at the *appraisal* stage of the emotional valuation system, along with difficulty describing feelings. In essence, an emotional response has occurred (*attention* stage), but a person has difficulty appraising what the emotional response is and what it means. Previous studies have reported that rates of alexithymia in a prison population are higher than in the general population (Chen et al., 2017; Parker et al., 2005; Zimmermann, 2006). This study shows no difference in overall alexithymia between the two samples, but a marked difficulty specific to identifying emotional states. As no studies have examined the relationship between DIF and DDF within the framework of the attention-appraisal model of alexithymia, future studies may wish to explore the potential differing contributions of DIF and DAF to the appraisal of emotional experiences in the context of the attention-appraisal model of alexithymia.

No significant relationship was found between alexithymia scores on the PAQ and offence type, therefore those with high levels of alexithymia were no more likely to commit a violent crime than those with low levels of alexithymia. Moreover, no significant relationships were found between alexithymia scores on the PAQ and substance misuse or parental separation in childhood. Although previous studies have found a relationship between alexithymia and developmental experiences such as childhood trauma and emotional neglect (Chen et al., 2017), this study did not formally measure symptoms of trauma so cannot confirm whether parental separation reported here was experienced as traumatic or not. This could explain why no relationship was found here, however more research is needed to understand the relationship between adverse developmental experiences and alexithymia.

This study identified a significant negative relationship between alexithymia and self-rated self-esteem in a prison population. Essentially, those with higher levels of alexithymia had lower self-reported self-esteem, while those with lower levels of alexithymia reported higher

levels of self-esteem. This corroborates findings from previous literature (Ersöğütçü & Kargin, 2022; Yelsma, 1995). As alexithymia is characterised by a difficulty identifying and describing emotional states, it is hypothesised that this leads to inadequate internal measures of attitudes and personality-related variables, which includes self-esteem (Dentale et al., 2010). Self-esteem is based on intuitive interpretations of the emotional reactions towards oneself; so those with alexithymia are likely to experience self-judgements informed by poorly appraised implicit feelings. Notably, this study found higher negative correlations between self-esteem and negatively-valenced items. This means that those with greater difficulty appraising negative affect are more likely to report lower levels of self-esteem. As previous research on self-esteem and alexithymia has not used the PAQ to measure alexithymia, they have not been able to differentiate DIF and DAF in terms of positive and negative emotions. Future research might wish to explore further the trend found in this study in order to explain why those with greater difficulty appraising negative emotions have lower self-reported self-esteem. One such explanation could be that emotional regulation acts as a mediator between alexithymia and self-esteem. There is a growing body of literature that suggests an important role for emotional regulation for self-esteem (Bajaj et al., 2016), so those with alexithymia (and, thus, difficulty in appraising negative affect specifically) are more likely to have difficulties with emotional regulation, which can lead to affective disorder symptoms (such as low self-esteem). Moreover, those with affective disorder symptoms are more likely to engage in offending behaviours (Gross, 2015).

This study also identified a significant relationship between alexithymia scores on the PAQ and the mental attribution subscale of the GBAI-R, so those with higher scores on the PAQ were more likely to attribute blame to mental illness or poor self-control. It is interesting that those with significant difficulties identifying and describing emotional states are more likely to attribute blame for a crime to an internal deficit (such as mental illness) rather than an external source, such as the victim or target of the offending behaviour. There could be a number of reasons for this. It might be that those with alexithymia have some level of insight into their condition, or perhaps their mental health difficulties more broadly, in order for them to attribute blame to mental illness or poor self-control. Although alexithymia is characterised by a difficulty in emotional awareness, it could be argued that individuals with such difficulties can still be aware of their own alexithymic traits. Essentially, those with an inability to appraise emotional states can identify that they have this deficit, because that knowledge does not require an ability to appraise emotions. Little literature exists exploring individuals' awareness

of alexithymia, so this could be an area for future research. Another explanation could simply be that those who are currently incarcerated for committing a serious crime are more likely to claim that they behaved as a result of mental illness or poor self-control in order to benefit from the well-known legal system's reduced punishments around diminished responsibility. No known studies have explored the link between alexithymia and self-blame in a forensic population, and further research is needed in order to explore this relationship.

The findings from this study are tempered by a number of limitations. Firstly, this study contained a relatively small sample of male inmates from one particular prison in the UK. Although all inmates completed the alexithymia measures, only a small sample completed the measures on blame and self-esteem. Broader generalisation of the findings from this study would also be difficult considering the data was collected from a specific population with a lack of diversity. Ethnic diversity is a potentially important factor when considering alexithymic traits as it has been found that individuals of Asian heritage score more highly for alexithymia than those from a European background (Le et al., 2002). This is most likely explained by cultural differences in values and beliefs relating to individualism and the expression of inner experience, rather than biomedical factors or differences in interoception (Lo, 2014). In relation to this, it would be interesting for future research to explore the differences within cultural values over time and how differences in alexithymia might be mediated by current social narratives and how these are endorsed (for example in popular culture). Moreover, this study conducted retrospective analyses on data collected routinely in the prison. As such, the authors of this research had very little control over the data that was collected, as well as the administrative procedures of the measures in this study. As such, there is a possibility that measures, including the PAQ and TAS-20, were not administered as intended by their original authors. With no process to measure fidelity, there is a chance the results of this study were subject to administration bias. Another limitation of this study is the lack of measure for emotional states. When exploring factors associated with alexithymia, and considering the PAQ includes emotionally-valenced items, it would have been useful to have measured individuals' experiences of a range of emotional states to aid in the interpretation of these data. Future research could consider utilising validated symptom checklists to describe experiences of emotional states in a forensic population.

References

Ates, M. A., Algul, A., Gulsun, M., Gecici, O., Ozdemir, B., Basoglu, C., Semiz, U. B., Ebrinc, S., & Cetin, M. (2009). The relationship between alexithymia, aggression and psychopathy in young adult males with antisocial personality disorder/Antisosyal kisilik bozuklugu olan genc erkeklerde aleksitimi, saldirganlik ve psikopati iliskisi. *Archives of Neuropsychiatry*, *46*(4), 135–140.

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/https://doi.org/10.1016/0022-3999(94)90005-1

Bajaj, B., Gupta, R., & Pande, N. (2016). Self-esteem mediates the relationship between mindfulness and well-being. *Personality and Individual Differences*, *94*, 96–100. https://doi.org/https://doi.org/10.1016/j.paid.2016.01.020

Battle, J. (1981). *Culture-free self-esteem inventories for children and adults*. Special child publications.

Berenbaum, H. (1996). Childhood abuse, alexithymia and personality disorder. *Journal of Psychosomatic Research*, *41*(6), 585–595. https://doi.org/https://doi.org/10.1016/S0022-3999(96)00225-5

Bermond, B., Clayton, K., Liberova, A., Luminet, O., Maruszewski, T., Ricci Bitti, P. E., Rimé, B., Vorst, H. H., Wagner, H., & Wicherts, J. (2007). A cognitive and an affective dimension of alexithymia in six languages and seven populations. *Cognition and Emotion*, *21*(5), 1125–1136. https://doi.org/10.1080/02699930601056989

Brown, T. A. (2015). Confirmatory factor analysis for applied research, 2nd ed. In *Confirmatory factor analysis for applied research, 2nd ed.* The Guilford Press.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Chen, L., Xu, L., You, W., Zhang, X., & Ling, N. (2017). Prevalence and associated factors of alexithymia among adult prisoners in China: a cross-sectional study. *Bmc Psychiatry*, *17*, 1–13.

Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: a generic method with applications. *Neuropsychology*, *21*(4), 419–430. https://doi.org/10.1037/0894-4105.21.4.419

Dentale, F., San Martini, P., De Coro, A., & Di Pomponio, I. (2010). Alexithymia increases the discordance between implicit and explicit self-esteem. *Personality and Individual Differences*, *49*(7), 762–767. https://doi.org/https://doi.org/10.1016/j.paid.2010.06.022

Ersöğütçü, F., & Kargin, M. (2022). The relationship between self-esteem and alexithymia in individuals with substance use disorders. *Journal of Substance Use*, 1–6. https://doi.org/10.1080/14659891.2022.2098840

Evren, C., Cinar, O., Evren, B., Umut, G., Can, Y., & Bozkurt, M. (2015). Relationship between alexithymia and aggression in a sample of men with substance dependence. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology*, *25*(3), 233–242.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., Arshanapalli, B. G., & Hoechstoetter, M. (2014). *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*. Wiley. https://books.google.co.uk/books?id=3NrjAgAAQBAJ

Franz, M., Popp, K., Schaefer, R., Sitte, W., Schneider, C., Hardt, J., Decker, O., & Braehler, E. (2008). Alexithymia in the German general population. *Social Psychiatry and Psychiatric Epidemiology*, *43*, 54–62.

Garofalo, C., Velotti, P., & Zavattini, G. C. (2017). Emotion Regulation and Aggression: The Incremental Contribution of Alexithymia, Impulsivity, and Emotion Dysregulation Facets. *Psychology of Violence*, *8*, 470–483.

Gillespie, S. M., Garofalo, C., & Velotti, P. (2018). Emotion regulation, mindfulness, and alexithymia: Specific or general impairments in sexual, violent, and homicide offenders? *Journal of Criminal Justice*, *58*, 56–66. https://doi.org/https://doi.org/10.1016/j.jcrimjus.2018.07.006

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford University Press.

Gross, J. J. (2015). Emotion Regulation: Current Status and Future Prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Gudjonsson, G. H., & Singh, K. K. (1989). The revised Gudjonsson blame attribution inventory. *Personality and Individual Differences*, *10*(1), 67–70. https://doi.org/https://doi.org/10.1016/0191-8869(89)90179-7

Hemming, L., Bhatti, P., Shaw, J., Haddock, G., & Pratt, D. (2020). Words don't come easy. How male prisoners difficulties identifying and discussing feelings relate to suicide and violence. *Front Psychiatry, Forensic Psychiatry, 11.* https://doi.org/10.3389/fpsyt.2020.581390

Honkalampi, K., Hintikka, J., Laukkanen, E., & Viinamäki, J. L. H. (2001). Alexithymia and Depression: A Prospective Study of Patients With Major Depressive Disorder. *Psychosomatics*, *42*(3), 229–234. https://doi.org/https://doi.org/10.1176/appi.psy.42.3.229

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hudson, S. M., Ward, T., & McCormack, J. C. (1999). Offense Pathways in Sexual Offenders. *Journal of Interpersonal Violence*, *14*(8), 779–798. https://doi.org/10.1177/088626099014008001

Kline, R. B. (2011). Principles and practice of structural equation modeling, 3rd ed. In *Principles and practice of structural equation modeling, 3rd ed.* Guilford Press.

Krakowski, M. I., & Czobor, P. (2014). Depression and impulsivity as pathways to violence: implications for antiaggressive treatment. *Schizophrenia Bulletin*, *40*(4), 886–894.

Le H. N., Berenbaum H., Raghavan C. (2002). Culture and alexithymia: Mean levels, correlates, and the role of parental socialization of emotions. *Emotion*, 2, 341-360.

Lo, C. (2014). Cultural Values and Alexithymia. *SAGE Open*, *4*(4). https://doi.org/10.1177/2158244014555117

Mason, O., Tyson, M., Jones, C., & Potts, S. (2005). Alexithymia: its prevalence and correlates in a British undergraduate sample. *Psychology and Psychotherapy: Theory, Research and Practice*, *78*(1), 113–125.

Mattila, A. K., Salminen, J. K., Nummi, T., & Joukamaa, M. (2006). Age is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research*, *61*(5), 629–635.

Meganck, R., Vanheule, S., & Desmet, M. (2008). Factorial Validity and Measurement Invariance of the 20-Item Toronto Alexithymia Scale in Clinical and Nonclinical Samples. *Assessment*, *15*(1), 36–47. https://doi.org/10.1177/1073191107306140

Norlander, B., & Eckhardt, C. (2005). Anger, hostility, and male perpetrators of intimate partner violence: A meta-analytic review. *Clinical Psychology Review*, *25*(2), 119–152.

Parker, J. D. A., Shaughnessy, P. A., Wood, L. M., Majeski, S. A., & Eastabrook, J. M. (2005). Cross-cultural alexithymia: Validity of the 20-item Toronto Alexithymia Scale in North American aboriginal populations. *Journal of Psychosomatic Research*, *58*(1), 83–88.

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (1989). The alexithymia construct: relationship with sociodemographic variables and intelligence. *Comprehensive Psychiatry*, *30*(5), 434–441.

Polaschek, D. L. L., & Ward, T. (2002). The implicit theories of potential rapists: What our questionnaires tell us. *Aggression and Violent Behavior*, *7*(4), 385–406. https://doi.org/https://doi.org/10.1016/S1359-1789(01)00063-5

Preece, D. A., Mehta, A., Becerra, R., Chen, W., Allan, A., Robinson, K., Boyes, M., Hasking, P., & Gross David A.; (2022). Why is alexithymia a risk factor for affective disorder symptoms? The role of emotion regulation. *Journal of Affective Disorders*, *296*, 337–341. https://doi.org/https://dx.doi.org/10.1016/j.jad.2021.09.085

Preece, D., Becerra, R., Allan, A., Robinson, K., & Dandy, J. (2017). Establishing the theoretical components of alexithymia via factor analysis: Introduction and validation of the attention-appraisal model of alexithymia. *Personality and Individual Differences*, *119*, 341–352. https://doi.org/10.1016/j.paid.2017.08.003

Preece, D., Becerra, R., Robinson, K., & Dandy, J. (2018). Assessing Alexithymia: Psychometric Properties and Factorial Invariance of the 20-Item Toronto Alexithymia Scale in Nonclinical and Psychiatric Samples. *Journal of Psychopathology and Behavioral Assessment*, *40*(2), 276–287. https://doi.org/10.1007/s10862-017-9634-6

Preece, D., Becerra, R., Robinson, K., Dandy, J., & Allan David; (2018). The psychometric assessment of alexithymia: Development and validation of the Perth Alexithymia Questionnaire. *Personality and Individual Differences*, *132*, 32–44. https://doi.org/https://dx.doi.org/10.1016/j.paid.2018.05.011

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/https://doi.org/10.1016/S0001-6918(99)00050-5

Roberton, T, Daffern, M., & Bucks, R. S. (2014). Maladaptive emotion regulation and aggression in adult offenders. *Psychology, Crime and Law*, *20*(10), 933–954. https://doi.org/10.1080/1068316X.2014.893333

Roberton, Terri, Daffern, M., & Bucks, R. S. (2015). Beyond anger control: Difficulty attending to emotions also predicts aggression in offenders. *Psychology of Violence*, *5*, 74–83. https://doi.org/10.1037/a0037214

Satorra, A. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *American Statistical Association 1988 Proceedings of Business and Economics Sections*, 308–313.

Seroczynski, A. D., Bergeman, C. S., & Coccaro, E. F. (1999). Etiology of the impulsivity/aggression relationship: genes or environment? *Psychiatry Research*, *86*(1), 41–57.

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*, 255–262. https://doi.org/10.1159/000286529

Taylor, G. J., Parker, J. D. A., Bagby, R. M., & Bourke, M. P. (1996). Relationships between alexithymia and psychological characteristics associated with eating disorders. *Journal of Psychosomatic Research*, *41*(6), 561–568. https://doi.org/https://doi.org/10.1016/S0022-3999(96)00224-3

Teten, A. L., Miller, L. A., Bailey, S. D., Dunn, N. J., & Kent, T. A. (2008). Empathic deficits and alexithymia in trauma-related impulsive aggression. *Behavioral Sciences & the Law*, *26*(6), 823–832.

Thorberg, F. A., Young, R. M., Sullivan, K. A., & Lyvers, M. (2009). Alexithymia and alcohol use disorders: A critical review. *Addictive Behaviors*, *34*(3), 237–245. https://doi.org/https://doi.org/10.1016/j.addbeh.2008.10.016

van der Velde, J., Servaas, M. N., Goerlich, K. S., Bruggeman, R., Horton, P., Costafreda, S. G., & Aleman, A. (2013). Neural correlates of alexithymia: A meta-analysis of emotion processing studies. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1774–1785. https://doi.org/https://doi.org/10.1016/j.neubiorev.2013.07.008

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: let's learn from cows in the rain. *PloS One*, *8*(7), e68967. https://doi.org/10.1371/journal.pone.0068967

Yelsma, P. (1995). Self-esteem and alexithymia. *Psychological Reports*, *77*(3 Pt 1), 735–738. https://doi.org/10.2466/pr0.1995.77.3.735

Zeitlin, S. B., & McNally, R. J. (1993). Alexithymia and anxiety sensitivity in panic disorder and obsessive-compulsive disorder. *The American Journal of Psychiatry*, *150*(4), 658–660. https://doi.org/10.1176/ajp.150.4.658

Zimmermann, G. (2006). Delinquency in male adolescents: The role of alexithymia and family structure. *Journal of Adolescence*, *29*(3), 321–332.

Chapter 3: Press releases

Meta Analysis press release

Title: Can we effectively measure someone's inability to recognise and describe their own emotions, and what impact does it have for Psychology practice?

There is a psychological phenomenon whereby people cannot adequately identify or describe the way that they are feeling, which is called alexithymia. Those with alexithymia find it increasingly difficult to notice whether they're feeling happy, sad, or angry as they struggle to interpret cues from their body. Alexithymia is gaining increasing interest in mental health research as it has been identified as a transdiagnostic risk factor for many psychological difficulties, including depression, anxiety, personality and eating disorders, as well as substance abuse and criminal behaviour. As such, it's become increasingly more important to measure and potentially diagnose alexithymia.

A recent study at the University of Birmingham has conducted a meta-analysis on one such measure of alexithymia – the Perth Alexithymia Questionnaire (PAQ; original authors are David Preece and colleagues from Australia). "Our study basically brings together all available studies that have investigated the effectiveness of the PAQ and combines their data together to create one large overall dataset that we can analyse to ultimately conclude whether the PAQ does accurately measure alexithymia" said the lead author of the research.

A meta-analysis is typically used in academic research to provide robust evidence on a given research questions; in this case whether professionals actually can accurately measure and diagnose people with alexithymia. "The PAQ is a relatively new measure for alexithymia, having only ben developed in 2018. It's becoming increasingly more important to have an adequate measure for alexithymia as it can have quite a profound impact on individuals mental health, and the way that we would treat those difficulties."

The study found that the PAQ is the most reliable measure of alexithymia for use at present and can accurately measure alexithymia generally and in terms of both positive and negative emotions. The authors recommend that the PAQ could be utilised for diagnostic purposes, which would be useful for determining the type of therapeutic interventions chosen for people scoring high in alexithymia, as their inability to identify emotions would make it difficult to engage in therapy that relies heavily on identifying and discussing emotions.

The University of Birmingham continues to produce high-quality research as part of their Doctoral programme in Clinical Psychology.

Empirical Paper press release

Title: Alexithymia – the inability to adequately identify and describe emotions – significantly associated with criminal behaviour.

People who struggle to know how they are feeling are more likely to engage in criminal behaviour, new study finds. Alexithymia is the term for the psychological phenomena in which people find it difficult to identify and describe the way they're feeling, whether that be happy or angry or anything in between. Researchers at the University of Birmingham have conducted a new study investigating the rates of alexithymia in a UK prison and how alexithymia differs between inmates depending on a number of different personal characteristics including the type of crime they've committed, what their upbringing was like, and how old they are.

The study found that rates of alexithymia in a prison are higher than in wider society. One explanation for this might be that those with alexithymia are less able to regulate their emotions (accept and then deal with them appropriately) because they find it difficult to identify them; therefore when feeling a negative emotion, like anger, frustration or stress, are more likely to act in an antisocial way. The study also found that those with higher levels of alexithymia were more likely to attribute blame of the incarcerating incident to mental health difficulties or a lack of self-control. This is an interesting finding as it could be that those who are more alexithymic commit crime due to their alexithymia, and are subsequently more likely to attribute responsibility to their difficulties that alexithymia presents, or it could be that alexithymia is getting in the way of attributing blame at the time of completing the blame measures due to ongoing difficulties appraising emotions.

"Alexithymia is a really interesting area, and one that's definitely gaining a lot of traction in clinical research at the moment" said the lead researcher for the study. "Alexithymia is a construct that has quite profound clinical relevance because it's a marker for a whole bunch of psychological pathology, including depression and anxiety and, like we found here, criminal behaviour". Many people that experience psychological difficulties or need some sort of behavioural rehabilitation, are recommended to receive psychological therapy. "The preface of

almost all psychological or behavioural therapies is to be able to identify and express your thoughts and feelings. However, those with alexithymia can find that a real struggle, which has implications for the kinds of treatments we typically offer and how effective they can be when someone has difficulty expressing emotions. Thus, it's become important to firstly assess people for alexithymia, and then to think more carefully about our approach to psychological therapy with people that are alexithymic."

The authors concede that this study was conducted with a small sample and further research is needed to corroborate their findings.

Appendices

Appendix A: Original ethical approval letter for this study

*This study was granted ethical approval as an extension to an existing research study

Dear Dr Jones

Re: "Psychometric Properties of the Perth Alexithymia Questionnaire and Toronto Alexithymia Scale in a UK prison population"

Application for Ethical Review ERN_22-0097

Thank you for your application for ethical review for the above project, which was reviewed by the Science, Technology, Engineering and Mathematics Ethical Review Committee.

On behalf of the Committee, I confirm that this study now has full ethical approval.

I would like to remind you that any substantive changes to the nature of the study as described in the Application for Ethical Review, and/or any adverse events occurring during the study should be promptly brought to the Committee's attention by the Principal Investigator and may necessitate further ethical review.

Please also ensure that the relevant requirements within the University's Code of Practice for Research and the information and guidance provided on the University's ethics webpages (available at https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Links-and-Resources.aspx) are adhered to and referred to in any future applications for ethical review. It is now a requirement on the revised application form (https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Ethical-Review-Forms.aspx) to confirm that this guidance has been consulted and is understood, and that it has been taken into account when completing your application for ethical review.

Please be aware that whilst Health and Safety (H&S) issues may be considered during the ethical review process, you are still required to follow the University's guidance on H&S and to ensure that H&S risk assessments have been carried out as appropriate. For further information about this, please contact your School H&S representative or the University's H&S Unit at healthandsafety@contacts.bham.ac.uk.

Kind regards

Mrs Susan Cottam
Research Ethics Manager
Research Support Group
University of Birmingham

Appendix B: Perth Alexithymia Questionnaire

PAQ

Name: **Date:**

This questionnaire asks about how you perceive and experience your emotions. Please score the following statements according to **how much you agree or disagree that the statement is true of you**. Circle one answer for each statement.

Some questions mention *bad* or *unpleasant* emotions, this means emotions like sadness, anger, or fear. Some questions mention *good* or *pleasant* emotions, this means emotions like happiness, amusement, or excitement.

		Strongly disagree	---	---	Neither agree nor disagree	---	---	Strongly agree
1	When I'm feeling *bad* (feeling an unpleasant emotion), I can't find the right words to describe those feelings.	1	2	3	4	5	6	7
2	When I'm feeling *bad*, I can't tell whether I'm sad, angry, or scared.	1	2	3	4	5	6	7
3	I tend to ignore how I feel.	1	2	3	4	5	6	7
4	When I'm feeling *good* (feeling a pleasant emotion), I can't find the right words to describe those feelings.	1	2	3	4	5	6	7
5	When I'm feeling *good*, I can't tell whether I'm happy, excited, or amused.	1	2	3	4	5	6	7
6	I prefer to just let my feelings happen in the background, rather than focus on them.	1	2	3	4	5	6	7
7	When I'm feeling *bad*, I can't talk about those feelings in much depth or detail.	1	2	3	4	5	6	7
8	When I'm feeling *bad*, I can't make sense of those feelings.	1	2	3	4	5	6	7
9	I don't pay attention to my emotions.	1	2	3	4	5	6	7
10	When I'm feeling *good*, I can't talk about those feelings in much depth or detail.	1	2	3	4	5	6	7
11	When I'm feeling *good*, I can't make sense of those feelings.	1	2	3	4	5	6	7
12	Usually, I try to avoid thinking about what I'm feeling.	1	2	3	4	5	6	7

		Strongly disagree	---	---	Neither agree nor disagree	---	---	Strongly agree
13	When something *bad* happens, it's hard for me to put into words how I'm feeling.	1	2	3	4	5	6	7
14	When I'm feeling *bad*, I get confused about what emotion it is.	1	2	3	4	5	6	7
15	I prefer to focus on things I can actually see or touch, rather than my emotions.	1	2	3	4	5	6	7
16	When something *good* happens, it's hard for me to put into words how I'm feeling.	1	2	3	4	5	6	7
17	When I'm feeling *good*, I get confused about what emotion it is.	1	2	3	4	5	6	7
18	I don't try to be 'in touch' with my emotions.	1	2	3	4	5	6	7
19	When I'm feeling *bad*, if I try to describe how I'm feeling I don't know what to say.	1	2	3	4	5	6	7
20	When I'm feeling *bad*, I'm puzzled by those feelings.	1	2	3	4	5	6	7
21	It's not important for me to know what I'm feeling.	1	2	3	4	5	6	7
22	When I'm feeling *good*, if I try to describe how I'm feeling I don't know what to say.	1	2	3	4	5	6	7
23	When I'm feeling *good*, I'm puzzled by those feelings.	1	2	3	4	5	6	7
24	It's strange for me to think about my emotions.	1	2	3	4	5	6	7