Revisiting the debriefing debate: Does psychological debriefing reduce PTSD symptomology following work-related trauma? A meta-analysis

and

Assessing alexithymia in forensic practice: Psychometric properties of the 20-item Toronto Alexithymia Scale and the Perth Alexithymia Scale in a sample of adult male prisoners

by

Harry Stileman

A thesis submitted to the University of Birmingham in part completion of the degree of DOCTOR OF CLINICAL PSYCHOLOGY

Centre for Applied Psychology
School of Psychology
College of Life and Environmental Sciences
University of Birmingham
May 2022

Thesis overview

This thesis comprises four chapters. The first chapter is a meta-analysis reviewing the effectiveness of psychological debriefing, an early post-trauma intervention, following work-related trauma. The overall synthesis did not find consistent evidence that psychological debriefing helps to prevent or reduce symptoms of PTSD. Shortcomings in the methodology and reporting of many of the studies included within the review highlight the need for more rigorous research in the future to ensure that organisations can provide trauma-exposed employees with the effective support they both need and deserve.

The original plan for the empirical paper was to conduct a randomised controlled trial to evaluate whether Trauma Risk Management, a peer-support system which originated in Royal Marines, was more effective than psychological debriefing in maintaining the psychological wellbeing of mental health practitioners following exposure to a potentially traumatic event. Due to the constraints brought on my the Covid-19 pandemic, it was not possible to collect enough data within the necessary timeframe and so this project was suspended. Instead, the second chapter is an empirical research project which tested the psychometric properties of two self-report measures of alexithymia within a sample of male prisoners: the well-established 20-item Toronto Alexithymia Scale and the recently developed Perth Alexithymia Questionnaire. This paper offers preliminary support for the Perth Alexithymia Questionnaire as a measure of alexithymia within both forensic practice and research.

The third and fourth chapters of this thesis are the press releases for the meta-analysis and empirical paper, respectively.

Acknowledgements

The completion of this thesis would not have been possible without the advice, guidance, and support of my supervisor Dr. Chris Jones, to whom I am very grateful.

Special thanks are due to the prisoners whose permission to use their data for research purposes made the empirical paper possible.

I would also like to thank the psychology team at the prison, both for proposing the research topic and for their involvement in data collection.

Finally, I would like to thank my wife, family and friends for their continuous love and understanding.

Table of contents

Chapter 1: Literature Review - Revisiting the debriefing debate: Does psychological debriefing reduce PTSD symptomology following work-related trauma? A meta-analysis .1

1.1. Abstract 1

1.2. Introduction 2

1.2.1. Psychological debriefing 2

1.2.2. Cochrane Review of psychology debriefing 4

1.2.3. Criticisms of the Cochrane review 6

1.2.4. Public Health England scoping review 7

1.2.5. Aims of the current meta-analysis 7

1.3. Methods 8

1.3.1 Identifying primary studies 8

1.3.2. Data extraction 12

1.3.3. Defining problematic variance 15

1.3.4. Risk of bias assessment 16

1.4. Results 24

1.4.1. Selection of the meta-analytic model 24

1.4.2. Omnibus test of total score on PTSD measures 25

1.4.3. The impact of study design on effect size 27

1.4.4. The impact of time on effect size 29

1.4.5. The impact of risk of bias in the primary studies 30

1.4.6. Differences in avoidance, hyperarousal and intrusion symptom outcomes 31

1.4.7. Difference attributable to characteristics of psychological debriefings.....32

1.4.8. Differences attributable to trauma characteristics 35

1.4.9. Subgroup analyses that were not possible to conduct 36

1.4.10. Publication bias and small-study effects 36

1.5. Discussion.....38

1.5.1. Limitations 41

1.5.2. Conclusion.....43

1.6. References 45

Chapter 2: Empirical Paper - Assessing alexithymia in forensic practice: Psychometric properties of the 20-item Toronto Alexithymia Scale and the Perth Alexithymia Scale in a sample of adult male prisoners.....54

2.1. Abstract.....54

2.2. Introduction 55

2.2.1. Aetiology of alexithymia 55

2.2.2. Attention-appraisal model of alexithymia.....56

2.2.3. Assessing alexithymia 57

2.2.4. Alexithymia and offending behaviour 60

2.2.5. Current study 61

2.3. Methods.....62

2.3.1. Participants, materials and procedure 62

2.3.2. Analytic strategy 63

2.4. Results.....68

2.4.1. Descriptive statistics and internal consistency 68

2.4.2. Convergent validity 69

2.4.3. Factorial validity 69

2.5. Discussion 74

2.5.1. Limitations 76

2.5.2. Implications for forensic practice and future research 77

2.5.3. Conclusion 79

2.6. References 80

Chapter 3: Press release for the literature review 92

Chapter 4: Press release for the empirical paper 95

Appendices 97

Appendix A. Letter from ethics committee granting ethical approval for the research 97

Appendix B. Perth Alexithymia Questionnaire 98

Appendix C. 20-item Toronto Alexithymia Scale 100

List of figures

Figure 1.1. Process of study selection: PRISMA diagram. 12

Figure 1.2. QQ plot of the distribution of standardised mean differences within the primary studies 24

Figure 1.3. Subgroup plot on the impact of study design on estimated effect size. 28

Figure 1.4. Forest plot of the standardised mean difference of PTSD symptoms between participants who did and did not receive psychological debriefing following a potentially traumatic event.....28

Figure 1.5. Forest plot of the standardised mean difference of PTSD symptoms between participants who did and did not receive psychological debriefing at different timepoints after debriefing. 30

Figure 1.6. Forest plot of the standardised mean difference of specific PTSD symptom clusters between participants who did and did not receive psychological debriefing following a potentially traumatic event. 32

Figure 1.7. Subgroup plot of differences between studies that did and did not adhere to established seven phase models. 33

Figure 1.8. Subgroup plot of differences between studies offered single session debriefing versus those that offered multiple debriefing sessions. 34

Figure 1.9. Subgroup plot of differences between individual and group debriefings 35

Figure 1.10. Subgroup plot of differences between single or multiple traumatic incidences. 36

Figure 1.11. Funnel plot of the standardised mean difference for all PTSD symptom outcomes. 37

Figure 2.1. TAS-20 confirmatory factor analysis models.....66

Figure 2.2. PAQ confirmatory factor analysis models. 67

List of tables

Table 1.1. Mitchell's (1983) and Dyregrov's (1989) seven phase models.....3

Table 1.2. Inclusion and exclusion criteria 10

Table 1.3. Study design hierarchy.....16

Table 1.4. Criteria for ratings of low, unclear or high risk across seven domains. 18

Table 1.4. Criteria for ratings of low, unclear or high risk across seven domains. 19

Table 1.5. Ratings of risk of bias.. 20

Table 1.6. Treatment effects described in the primary studies. 26

Table 1.7. Standard Mean Differences for studies with a 'low risk' of bias and studies with 'any risk' of bias within each of the seven risk domains. 31

Table 2.1. Demographic and offence characteristics of participants.....63

Table 2.2. Descriptive statistics and Cronbach's α reliability coefficients for the TAS-20 and PAQ. 68

Table 2.3. Pearson bivariate correlations between the TAS-20 and PAQ total scores and subscale scores. 69

Table 2.4. Goodness-of-fit index values for the different confirmatory factor analysis models of the TAS-20 and PAQ.....72

Table 2.5. Standardised factor loadings from confirmatory analyses of the TAS-20 (three factor model) and PAQ (valenced three-factor model).....73

Chapter 1: Literature Review - Revisiting the debriefing debate: Does psychological debriefing reduce PTSD symptomology following work-related trauma? A meta-analysis

1.1. Abstract

Psychological debriefing is an early post-trauma intervention which aims to prevent the development of PTSD and accelerate normal recovery through discussing, validating, and normalising group members responses to trauma. While originally designed in the 1980s for groups of emergency service personnel, the scope of psychological debriefing extended to individual primary victims of trauma. A Cochrane review in 2002 concluded that psychological debriefing was ineffective, yet some authors have argued that many of the studies that informed the Cochrane review did not adhere to key elements of psychological debriefing. This meta-analysis sought to re-examine the effectiveness of psychological debriefing in preventing or reducing PTSD symptoms following work-related trauma. Appropriate studies were selected from three databases (MEDLINE, Embase and PsycINFO). Inclusion criteria was intentionally broad so that features of psychological debriefing that may determine its effectiveness could be explored through a series of subgroup analyses. The overall synthesis did not find consistent evidence that psychological debriefing helps to prevent or reduce PTSD symptoms following work-related trauma. Shortcomings in the methodology and reporting of many of the studies meant that several important subgroup analyses could not be conducted. Further well-designed studies in this field are warranted to ensure that employees exposed to potentially traumatic events receive the effective support they need and deserve.

1.2. Introduction

Occupational groups such as military personnel, emergency service workers and healthcare workers are routinely exposed to potentially traumatic events (PTEs), increasing their risk of developing mental health difficulties such as post-traumatic stress disorder (PTSD; Petereit-Haack et al., 2020; Skogstad et al., 2013). The World Health Organisation's International Classification of Diseases 11th Revision (ICD-11; WHO, 2018) notes that PTSD "may develop following exposure to an extremely threatening or horrific event or series of events" (ICD-11; WHO, 2018) and consists of three clusters of symptoms: (1) re-experiencing of the trauma through intrusive memories, flashbacks and nightmares, (2) avoidance of reminders of the trauma, and (3) hyperarousal and hyperreactivity associated with the traumatic event.

PTSD and other trauma-related mental health difficulties can have far-reaching consequences for the individual, including adverse effects upon health, productivity at work and the quality of relationships with those close to them (Brooks, Rubin & Greenberg, 2019; Lee et al., 2020). It is therefore important that organisations in which the likelihood of exposure to trauma is high have effective management strategies in place to support their employees. This is both a moral responsibility and a legal obligation. The Health and Safety at Work Act (1974) states that employers have a duty of care "to ensure, so far as is reasonably practicable, the health, safety and welfare at work of all employees" (p.4). One management strategy that has been widely used for decades is 'psychological debriefing'.

1.2.1. Psychological debriefing

Psychological debriefing has its origins in World War I (Litz et al., 2002). Following a battle, commanders would 'debrief' their soldiers. The rational was that sharing stories

would help boost the morale of soldiers and help prepare them for future conflict. Military psychiatrists also developed strategies to support soldiers who were experiencing traumatic stress reactions. Underlying these strategies were the principles of proximity, immediacy, and expectancy (Grinker & Spiegal, 1945). Soldiers were supported near the battlefield, soon after the onset of difficulties, and with the expectation of a quick return to combat.

In the 1980s, a psychologist and former firefighter called Jeffrey Mitchell noted similarities between the stress of combat and the stress of emergency services and developed the most widely used method of psychological debriefing - Critical Incident Stress Debriefing (CISD) - as part of his Critical Incident Stress Management Programme (Mitchell, 1983). CISD is a seven phase intervention which was specifically designed for groups of emergency service workers following exposure to a PTE, or what Mitchell termed a 'critical incident'. Mitchell went on to collaborate with another psychologist, Atle Dyregrov, who developed a seven phase model similar to CISD and coined the alternative term Psychological Debriefing (Dyregrov, 1989; see Table 1.1). The term 'psychological debriefing' will be used to refer collectively to these two models hereon in.

Table 1.1. *Mitchell's (1983) and Dyregrov's (1989) seven phase models.*

Mitchell's seven phase model of CISD	**Dyregrov's seven phase model of PD**
1. Introduction	1. Introduction
2. Facts	2. Facts
3. Thoughts	3. Thoughts (including expectations)
4. Reactions	4. Reactions (and sensory impressions)
5. Symptoms	5. Normalisation
6. Teaching	6. Future planning and coping
7. Re-entry	7. Disengagement

Psychological debriefing aims to prevent the development of PTSD and accelerate normal recovery through discussing, validating, and normalising group members responses to trauma (Mitchell & Everly, 1996). This aim is in keeping with the cognitive model of PTSD (Ehlers & Clark, 2000) which proposes that misconceptions and negative appraisals relating to a traumatic event and its sequalae play a role in the development and maintenance of PTSD symptoms. Further aims of psychological debriefing include enhancing group cohesion, providing information about coping strategies, screening for individuals who need further support and referring on for further assessment or intervention if required (Mitchell & Everly, 1996).

Psychological debriefings as described by Mitchell & Everly (1996) are typically led by two facilitators, although for larger groups there can be up to four facilitators. Facilitators should include a mental health professional and a specially trained peer support worker from the same profession as the group members. Debriefings usually involve a single session, lasting between one and three hours. They are typically facilitated twenty-four to seventy-two hours after the PTE, although significant delays can often occur.

Following Michell's (1983) seminal paper, the scope of psychological debriefing extended beyond groups of emergency service personnel to other occupations, including the military and healthcare. Furthermore, it was employed for individual primary traumas outside of an occupational setting, including burns (Bisson et al., 1997), violent crime (Rose et al., 1999), childbirth (Priest et al., 2003) and road traffic accidents (Hobbs et al., 1996).

1.2.2. Cochrane Review of psychology debriefing

In 2002, the Cochrane Collaboration for Evidence-based Practice published a review of the effectiveness of single-session psychological debriefing in preventing PTSD, which was

updated in 2010 (Rose at al., 2002). Fifteen randomised controlled trials met inclusion criteria. No consistent and substantive evidence was found that psychological debriefing reduces the risk of developing PTSD symptoms compared to no intervention and two trials which included longer follow up periods (Bisson et al., 1997; Hobbs et al., 1996) reported adverse effects. Consequently, Rose et al. (2002) concluded that "psychological debriefing is either equivalent to, or worse than, control or educational interventions in preventing or reducing the severity of PTSD" (p.2).

As a result of this Cochrane Review, the National Institute for Health and Care Excellence (NICE) completed its own systematic review of seven RCTs in this field which consisted of many of the same studies as the Cochrane Review, including both the studies by Bisson et al. (1997) and Hobbs et al. (1996). It also concluded that "single-session debriefing may be at best ineffective" (NICE, 2005, p.84).

NICE guidance for PTSD has since been unequivocal in its recommendation to "not offer psychologically-focused debriefing for the prevention or treatment of PTSD" (NICE, 2018a, p.15). Consequently, organisations have been left with limited guidance on suitable strategies to maintain the psychological wellbeing of their staff following exposure to PTEs. In some organisations, psychological debriefing continues to be offered, sometimes under different names (e.g. 'Powerful Event Group Support'; Hawker et al., 2011). The UK military now use Trauma Risk Management (TRiM), which shares many of the same objectives and practices as psychological debriefing (Greenberg et al., 2008). Other organisations offer 'psychological first aid', which broadly involves the provision of information, comfort, emotional care and practical support (Shultz & Forbes, 2014). However, there is currently a lack of evidence on its effectiveness (Dieltjens et al., 2014).

1.2.3. Criticisms of the Cochrane review

As with psychological debriefing itself, the Cochrane review of psychological debriefing (Rose et al., 2002) has provoked controversy. Two independent review papers (Hawker et al., 2011; Tamrakar et al., 2019) note some of the alternative explanations for the two negative outcomes reported by Bisson et al. (1997) and Hobbs et al. (1996). Firstly, debriefed participants had been more severely injured than those who were not debriefed. When this was controlled for, the negative outcomes of debriefing of trauma symptoms were either eliminated (Bisson et al., 1997) or reduced to marginal significance (Hobbs et al., 1996; Mayou et al., 2000).

Secondly, the scope and nature of the interventions evaluated by these two RCTs were inconsistent with key features of psychological debriefing: some of the debriefings were too short (under an hour); the facilitators often lacked adequate training; debriefings included a detailed review of the PTE rather than a brief overview; and the participants were individual victims of trauma among the general public, rather than groups of professionals for whom the intervention was originally developed. This was recognised by the follow-up review by NICE, which stated that "no trial on critical incident stress debriefing as it was originally conceived by Mitchell and colleagues (i.e. as a group intervention for teams of emergency workers, military personnel or others who are used to working together)… met our methodological inclusion criteria" (NICE, 2005, p.84).

These criticisms of the Cochrane review have led to more recent suggestions that psychological debriefing may have been dismissed to quickly and calls for further investigation to clarify the potential benefits of psychological debriefing (O'Toole & Eppich, 2022; Tamraker et al., 2019). Hawker & Hawker (2015) outline four lessons that can be

learnt from the Cochrane Review findings: (1) don't offer debriefing too soon after a traumatic event; (2) don't offer debriefing lasting less that one hour; (3) don't use insufficiently trained or inappropriate facilitators; (4) don't probe too hard for details.

1.2.4. Public Health England scoping review

A scoping review was recently undertaken by Public Health England's Behavioural Science Research Team (Richins et al., 2020) to identify research evaluating early interventions in occupations in which there is a high risk of exposure to PTEs. The review included 50 studies of mixed quality and method and included both quantitative and qualitative data. Qualitative outcomes were assessed using meta-ethnography. However, a meta-analysis was not conducted which is likely because of the wide range of interventions included within the review such as exposure therapy, cognitive behavioural therapy and compassion focused therapy, in addition to psychological debriefing. Nevertheless, most of the interventions included within the review were based on psychological debriefing and Richins et al. (2020) note that most of these led to a reduction in symptom severity. Furthermore, in the 12 studies where severity scores did not change, half were still evaluated as being helpful by the participants. Richins et al. (2020) concluded that psychological debriefing can be an effective support in emergency responders (for which psychological debriefing was originally intended) when they adhere to key components of established models and are: (a) informed by the organisational culture, (b) have the support of management, and (c) utilise existing peer support systems within teams.

1.2.5. Aims of the current meta-analysis

This meta-analysis aims to re-evaluate the evidence-base into the effectiveness of psychological debriefing in preventing or reducing PTSD symptoms following work-related

PTEs. In contrast to the Cochrane review of psychological debriefing, this review extended the scope of studies beyond RCTs to include other non-randomised or uncontrolled designs. The rationale for this came from the recognition that there are implicit difficulties in conducting methodologically robust RCTs when evaluating psychological debriefing (Deahl, 2000). Trauma generally occurs in unpredictable and chaotic circumstances. As a result, researchers are often required to work opportunistically within strict time constraints and in line with operational processes. Furthermore, there are ethical dilemmas with employing randomised non-intervention controls for participants who may want, and benefit from, psychological debriefing. Consequently, a lot of the research on the effectiveness of psychological debriefing would not meet the criteria insisted upon by the Cochrane Library.

It was hoped that including a wider range of study designs would result in a larger number of studies within the review. This would allow for subgroup analyses to identify key components that may determine the effectiveness of psychological debriefing, such as those proposed by Hawker & Hawker (2015) and Richins et al. (2020).

1.3. Methods

1.3.1 Identifying primary studies

Search of electronic databases. A systematic search of the literature was initially carried out on 28th May 2021 using MEDLINE, Embase and PsycINFO. The aim of the search was to obtain a comprehensive overview of the literature into the effectiveness of psychological debriefing in preventing the development of trauma reactions in individuals exposed to work-related PTEs. The search terms that were used to identify these studies are outlined below.

Generic Search Terms

(Early adj3 intervention* ‘OR’ debrief* ‘OR’ psychological intervention ‘OR’ crisis intervention ‘OR’ critical incident stress debrief* ‘OR’ critical incident stress management) ‘AND’ (PTSD ‘OR’ posttrauma* ‘OR’ post trauma* ‘OR’ post-trauma* ‘OR’ traumatic stress ‘OR’ stress disorder* ‘OR’)

Specific MEDLINE Search Terms

(Early adj3 intervention* ‘OR’ debrief* ‘OR’ psychological intervention ‘OR’ crisis intervention ‘OR’ critical incident stress debrief* ‘OR’ critical incident stress management) ‘AND’ (stress disorders, traumatic/ or combat disorders/ or psychological trauma/ or stress disorders, post-traumatic/ or stress disorders, traumatic, acute/)

Specific PsycINFO Search Terms

(Early adj3 intervention* ‘OR’ debrief* ‘OR’ psychological intervention ‘OR’ crisis intervention ‘OR’ critical incident stress debrief* ‘OR’ critical incident stress management) ‘AND’ (posttraumatic stress disorder/ or exp "stress and trauma related disorders"/ or exp acute stress disorder/ or exp posttraumatic stress/)

Specific Embase Search Terms

(Early adj3 intervention* ‘OR’ debrief* ‘OR’ psychological intervention ‘OR’ crisis intervention ‘OR’ critical incident stress debrief* ‘OR’ critical incident stress management) ‘AND’ (exp posttraumatic stress disorder/)

Inclusion Criteria. Full inclusion and exclusion criteria are described in Table 1.2.

Table 1.2. *Inclusion and exclusion criteria.*

Inclusion criteria	Justification
Nature of intervention	
Studies that have referred to their intervention as a 'debriefing' and involve some recollection of the trauma and subsequent reactions	While there are a range of different terms to refer to psychological debriefing (e.g., stress debriefing, critical incident stress debriefing, crisis intervention), to ensure internal validity of the meta-analysis it was important that there is homogeneity between the content of psychological debriefings included in this review.
Exclude: psychological therapies (e.g., CBT, EMDR, CFT)	These therapies are outside of the scope of this review.
Participant Characteristics	
Employees who have experienced a work-related traumatic event.	Psychological debriefing was originally intended for work-related trauma, and this remains the scope of this review.
Outcome data	
Studies include a measure of PTSD symptoms.	To ensure internal validity of the meta-analysis, only studies with validated measures of PTSD symptoms (either self-report or structured assessment) were included.
The studies are required to report either means and standard deviations, or F- Test statistics, or Cohen's *d* effect size.	This was to ensure that outcomes can be calculated into an effect size for the purpose of the meta-analysis.
Type of article	
Studies published in English language	English is the first language of the author (HS)
Articles published in peer-reviewed journals	This was to ensure methodological rigour in the articles included.
The following article types were excluded: meta-analysis, reviews, theoretical pieces, commentaries, clinical guidance, study protocols, opinion pieces.	These articles do not provide the outcome data needed for this meta-analysis.
Study design	
The following study designs were excluded: single-case designs, case series, samples where n<10)	This was to ensure that an effect size reported by the included studies could be calculated with methodological rigour. While previous reviews in this area have only include randomised controlled trials (e.g., Rose et al., 2002), it was recognised that RCTs represent only a small proportion of the research evidence and so a broader range of study designs were included.

The results of the systematic search are presented in Figure 1.1. The search yielded a total of 5,942 articles and 3,824 once duplications were removed. Sensitivity in the search strategy was privileged over specificity as any further grouping of search terms to narrow articles down to work-related traumas resulted in known papers being lost. The inclusion criteria were used to screen these 3,824 articles by title and abstract. The three most common reasons articles were excluded at this stage were that they either did not relate to psychological debriefing, they did not relate to work-related trauma, or they did not provide outcome data (i.e. review papers). The remaining 184 articles were sought for retrieval; however, it was not possible to retrieve 15 of these articles due to lack of availability as hard or electronic copies despite contacting the British Library. The full text of the remaining 169 articles were then reviewed in more detail against the exclusion criteria. 24 articles met the full inclusion/exclusion criteria. Three articles which met the inclusion criteria could not ultimately be included in the synthesis: Shalev et al. (1998) used a measure of PTSD symptoms pre-intervention but not post-intervention; Söndergaard (2008) included traumatic events that occurred outside of the workplace; and Deahl (2000) did not specify the number of participants who completed the outcome measures.

The same search strategy was implemented again on 15th February 2022, limiting articles to those published since 2021, to determine whether any further primary studies should be included within the meta-analysis. This search yielded 483 articles and 325 once duplications were removed. However, no further articles met the inclusion criteria.

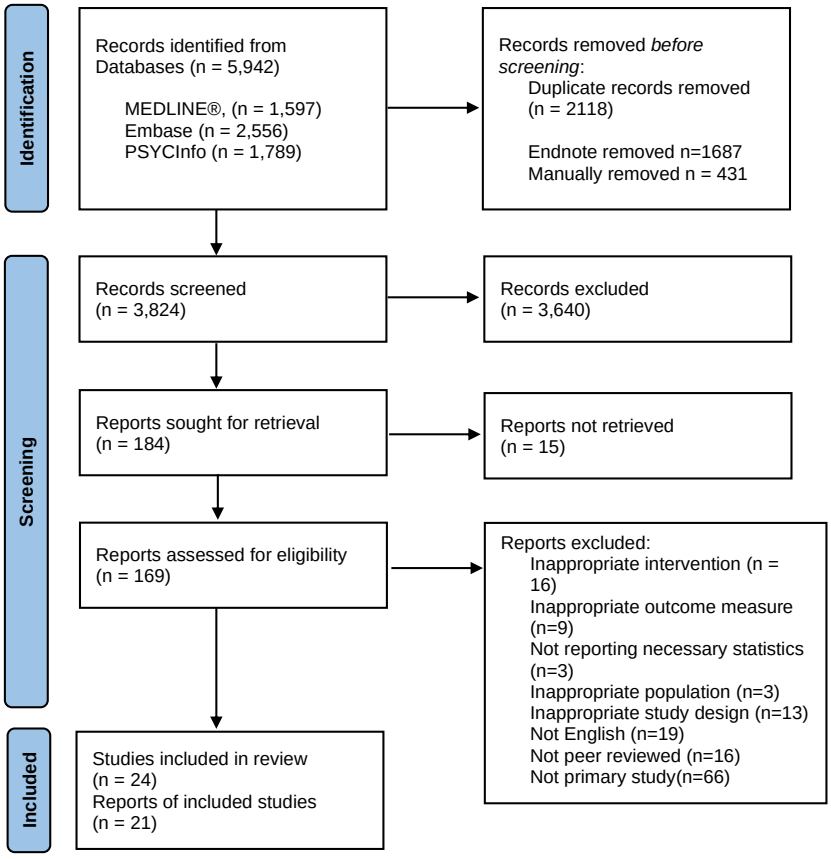


Figure 1.11. *Process of study selection: PRISMA diagram (Page et al., 2021).*

1.3.2. Data extraction

All data was extracted by a single individual (HS). Some studies reported PTSD symptom cluster subscale scores on outcome measures (avoidance, hyperarousal, intrusion), while other studies only reported overall scores. Three studies (Carlier et al., 2000; Harris et al., 2002; Tehrani et al., 2001) only reported cluster subscale scores. However, as these subscales included all the items on the PTSD measures used, the total mean score could be calculated and ultimately transformed into an estimate of Cohen's *d*.

15 studies used independent-group designs and reported means, standard deviations, and sample sizes of both a group who received psychological debriefing and a control group who either received no intervention, lower-level support such as stress

education, or were on a waiting list for an intervention. Tehrani et al. (2001) did not report standard deviations for each group so pooled standard deviations were substituted. Carlier et al. (1998) did not report means and standard deviations, instead reporting the percentage of participants in both the experimental and control groups that met a threshold for a PTSD diagnosis. In this case, percentages were converted into log ratios and then into estimates of Cohen's *d* using the sample sizes reported.

Two studies (Campfield et al., 2001; Richards et al., 2001) included a comparator group rather than a control group, in which participants also received a form of psychological debriefing. In the study by Campfield et al. (2001), participants either received an 'immediate' (<10 hours) or 'delayed' (>48 hours) debriefing. Data for both groups was extracted but treated separately as two before-and-after studies. In the study by Richard et al. (2001), one group received CISD, and the other group received the more extensive package of critical incident stress management. Again, this study was treated as a before-and-after study and only data from the CISD group was extracted. One between-group study (Ruck et al., 2013) reported significantly different baselines scores of PTSD symptoms between the experimental group and control group. As this was not controlled for in the statistical analysis (e.g., by using a treatment x timepoint ANCOVA), this study was also treated as a before-and-after study and only the data from the experimental group was extracted.

Both Carlier et al. (2000) and Matthews (1998) used two independent control groups in their studies. Carlier et al. (2000) included an 'external control group' of participants who had experienced trauma before debriefing was introduced in the workplace and an 'internal control group' of participants who had declined the offer of debriefing. As a different

outcome measure was used with the external control group, only data from the experimental group and internal control group were used. Matthews (2000) included one control group consisting of participants who did not request debriefing and another control group consisting of participants from a different area to the other two groups who did not receive debriefing because it was not available. In this case, both control group outcomes were combined into a single quantitative outcome using the procedure described by Borenstein et al. (2009). The combined mean was computed as the weighted mean (by sample size) across the two groups,

$$\bar{X}_1 = \frac{n_{11}\bar{X}_{11} + n_{12}\bar{X}_{12}}{n_{11} + n_{12}}$$

and the combined standardised deviation was computed as

$$S_1 = \sqrt{\frac{(n_{11} - 1)\, S_{11}^2 + (n_{12} - 1)\, S_{12}^2 + \frac{n_{11}n_{12}}{n_{11} + n_{12}} (\bar{X}_{11} - \bar{X}_{12})^2}{n_{11} + n_{12} - 1}}$$

where $\bar{X}_{11}$ and $\bar{X}_{12}$ were the means of the two control groups, S_{11} and S_{12} were the standard deviations and n_{11} and n_{12} were the sample sizes of the two control groups.

Adler et al. (2009) presented adjusted means and standard deviations comparing the experimental and control groups by combat exposure levels. In this instance, to ensure participants had all been exposed to trauma, data from the top-third exposure level (n=326) was extracted.

Several studies reported group means across multiple timepoints. In these cases, data was extracted for each timepoint. Timepoints were then grouped into the following categories: ‘short-term’ when outcome measures were collected 0-3 months after debriefing; ‘medium-term’ when outcome measures were collected 4-6 months after

debriefing; ‘long-term’ when outcome measures were collected 7 months or more after debriefing. In both Kenardy et al. (1996) and Wu et al. (2012), more than one timepoint fitted into the same time category and so only one of these datasets was extracted. When studies had included outcomes from multiple timepoints, unless the impact of time on outcome scores was being directly analysed, scores from the first data collection timepoint following intervention was used in analysis to avoid replication.

1.3.3. Defining problematic variance

As well as reporting a mean effect size, this meta-analysis sought to quantify and analyse the between-study heterogeneity. High levels of heterogeneity may arise between studies due to differences in interventions, participant characteristics, outcome measures or methodology (von Hippel, 2015).

Higgins I^2 (Higgins & Thompson, 2002) is a commonly used statistic to measure to amount of dispersion between studies. It is expressed as a percentage (0 to 100%) and provides an indication of the proportion of variation which is attributable to between-study variance rather than differences in precision of measurement due to sample size differences. In line with the benchmarks set by Higgin et al. (2003) and recognising the considerable variation in methodologies of the primary studies included within the synthesis, problematically high heterogeneity was defined as a Higgins I^2 value of more than 75%. Where problematic heterogeneity was observed, analyses were conducted to identify the source of heterogeneity between the effect sizes of the primary studies.

While standardised effect sizes from both repeated measures and independent-groups designs can be combined in a meta-analysis (Borenstein et al., 2009), it must be determined that potential sources of bias are not impacting the effect size estimates of

certain study designs (Morris & DeShon, 2002). Consequently, a subgroup analysis was conducted to determine whether these study outcomes differed in substantive ways.

1.3.4. Risk of bias assessment

A study hierarchy (see Table 1.3) was implemented to assess the contribution of each of the study designs to the overall quality score.

Table 1.3. *Study design hierarchy.*

Study Design	Quality Score	Description
Randomised controlled trial/experiment (including cluster randomisation)	30	An experimental study comparing two (or more) groups to establish the effectiveness of a specific intervention. An experimental group receives the intervention, while a comparison or control group receives either an alternative intervention or no intervention. Participants (or groups of participants) are randomly assigned to a group to minimise bias.
Non-randomised controlled trial/experiment	20	An experimental study in which people are allocated to either experimental or comparison/control groups using methods that are not random. As a result, there is an increased risk of biases being introduced into the research.
Repeated measures design (before-and-after studies without a separate control group)	10	A study in which observations are made before and after the implementation of an intervention. Data is collected at baseline and one or more times after the procedure. Uncontrolled before-and-after studies are an intrinsically weak evaluative design as they are unable to rule out alternative explanations for observed effects.

A set of quality criteria were developed to assess any risk of bias within this literature. The quality criteria were adapted from existing risk of bias frameworks, particularly 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). Risk of bias was assessed in seven domains: selection bias, performance bias, treatment fidelity, detection

bias, statistical bias, reporting bias and generalisation. The criteria for low, unclear, and high risk of bias within these seven domains is described in Table 1.4.

A quality index score was calculated for all papers included within the meta-analysis (see Table 1.5). This score was calculated using the study's overall design as assessed by the study design hierarchy (see Table 1.3) and the risk of bias ratings (see Table 1.4).

Table 1.4. *Criteria for ratings of low, unclear or high risk across seven domains.*

Domain	Details	Low risk of bias	Unclear risk of bias	High risk of bias
Selection Bias	Systematic differences between baseline characteristics of the groups that are compared.	Non-response rate is reported and of an acceptable level (< 50%). The source population is well described, and the study reports the characteristics of the sample e.g. the study details subgroups. The recruitment method is clearly reported and well defined. The article provides some reassurance that there is no selection bias (e.g. allocation concealment).	Non-response rate is not reported. The characteristics of the study population are not clearly reported. For example, the country, setting, location, population demographics are not adequately reported. The recruitment process/ sampling method of individuals are unclear or has not been reported.	Non-response rate is at unacceptable level (>50%) There are clear differences between groups being compared (e.g. experimental and control arms are from different populations). The characteristics of the study population are not reported.
Performance Bias	Systematic differences between/within groups in the participants motivation to complete the study or in exposure to factors other than the interventions of interest.	Study reports level of confidentiality and anonymity. Participants were not rewarded for their participation in the study. Information and procedures are provided in a way that does not differentially motivate participants. Participants were blinded where self-report measures are used.	The study does not report levels of confidentiality and anonymity. It is not clear if participants were rewarded for their participation. It is unclear how much information was provided to the participant prior to taking part in the study. Self-report measures are used but there is no evidence that participants were blinded.	Responses are not confidential or anonymous. Participants were rewarded for their participation in the study.
Treatment Fidelity	The extent to which the treatment is delivered competently and as intended and is representative of the class of treatments to which the study intends to generalise.	Treatment is sufficiently well described that it could be replicated. Treatment corresponded to intended treatment described in the methodology and established psychological debriefing protocols Procedures were in place to assess the fidelity of administered treatment.	Treatment protocol is unclear or has not been reported. There is no evidence that procedures are in place to assess the fidelity of administered intervention.	The treatment provided was different than the intended treatment. Treatment is provided inconsistently between participants. Treatment is not in line with established psychological debriefing protocols.

Table 1.4. *Criteria for ratings of low, unclear or high risk across seven domains.*

Domain	Details	Low risk of bias	Unclear risk of bias	High risk of bias
Detection Bias	Systematic differences between participants in how outcomes are determined. The extent to which the study design is optimised to detect the effect in questions.	The outcome measures are clearly defined, valid and reliable, and are implemented consistently across all participants. Outcomes were blindly rated by assessors (when an alternative to self-report measures have been used).	Information regarding the outcome measures are either not reported or not clearly reported e.g. definition, validity, reliability. The outcome measure(s) used has questionable psychometric properties (e.g. Cronbach's Alpha is between 0.6 and 0.7) It is not clear if the measure was implemented consistently across all participants.	The outcome measures were implemented differently across participants. The outcome measures used had poor reliability and validity reported e.g Cronbach's Alpha < 0.6. Only one dimension/subscale of the scale is used.
Statistical Bias	Bias resulting from the inappropriate statistical treatment of the data. This includes using completer-only analysis rather than intention-to-treat or other methods for inputting missing data.	Appropriate statistical testing was used. Confidence intervals or exact p-values for effect estimates were given or possible to calculate. Attrition rate – data loss is reported at analysis at an acceptable level (<5%) and appropriate method is used for inputting missing data.	It is unclear what statistical test was used. Confidence intervals or exact p-values for effect estimates were not reported and could not be calculated. Attrition rate – data loss is not reported at analysis or is at 10-20%	Statistics were not reported. Wrong statistical test was used which was not appropriate for the study design. Attrition rate – data loss is reported at analysis at an unacceptable level (>30%)
Reporting Bias	Systematic differences between reported and unreported findings (e.g., selective reporting of statistically significant findings).	Study has reported all results of measures as outlined in the method. Reasons for attrition or exclusions are reported.	Not all descriptive and/or summary statistics are presented. There is a description (narrative) in the results, but statistics are not recorded.	Study has not reported full outcome measures that are stated in the method section/ reported only a subsample of results/only significant results/ not reported the measure as it should be.
Generalisation	The extent to which the sample represents the target population from which it was drawn.	Sufficient sample size (35+ per arm) and representative of target population. A sample size justification, estimate or power analysis is provided.	20-30 participants per arm. Idiosyncratic features in sample. A sample size justification, estimate or power analysis are not provided.	Small sample (10-20 per arm) with or without idiosyncratic feature. Sample is not representative of wider profession.

Table 1.5. *Ratings of risk of bias. Red indicates high risk of bias, amber marks an unclear risk of bias and green is a low risk of bias.*

Study	*Study Design*	*Selection Bias*	*Performance Bias*	*Treatment Fidelity*	*Detection Bias*	*Statistical Bias*	*Reporting Bias*	*Generalisability*	*Quality Index*
Adler et al. (2008)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	Low risk	High risk	Low risk	Low risk	93%
Adler et al. (2009)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	Low risk	High risk	Low risk	Low risk	93%
Campfield et al. (2001)	Before-and-after study	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	High risk	24%
Carlier et al. (1998)	Non-randomised controlled trial/experiment	Low risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Unclear risk	Low risk	68%
Carlier et al. (2000)	Non-randomised controlled trial/experiment	High risk	Unclear risk	Unclear risk	Unclear risk	Unclear risk	High risk	Unclear risk	57%
Chemtob et al. (1997)	Non-randomised controlled trial/experiment	High risk	Unclear risk	High risk	Low risk	Unclear risk	Low risk	High risk	59%
Deahl et al. (1994)	Non-randomised controlled trial/experiment	Unclear risk	Unclear risk	Unclear risk	Low risk	Unclear risk	Low risk	Unclear risk	66%
Eid et al. (2001)	Non-randomised controlled trial/experiment	High risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	High risk	64%
Grundlingh et al. (2017)	Randomised controlled trial/experiment	Low risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	Unclear risk	93%
Harris et al. (2002)	Non-randomised controlled trial/experiment	Unclear risk	Unclear risk	High risk	Low risk	Low risk	Low risk	Low risk	68%
Humphries et al. (2001)	Non-randomised controlled trial/experiment	High risk	Unclear risk	Unclear risk	Low risk	Low risk	Low risk	High risk	64%
Kenardy et al. (1996)	Non-randomised controlled trial/experiment	Unclear risk	Unclear risk	High risk	Low risk	Unclear risk	Low risk	Low risk	66%
Matthews (1998)	Non-randomised controlled trial/experiment	High risk	Unclear risk	Unclear risk	Unclear risk	High risk	Low risk	High risk	57%
Regehr & Hill (2000)	Non-randomised controlled trial/experiment	Unclear risk	Unclear risk	High risk	Low risk	Low risk	Low risk	Low risk	68%
Richards (2001)	Before-and-after study	Low risk	Unclear risk	Unclear risk	Low risk	High risk	Low risk	Low risk	45%
Ruck et al. (2013)	Before-and-after study	High risk	Unclear risk	Unclear risk	Low risk	High risk	Low risk	Low risk	41%
Shoval-Zuckerman et al. (2015)	Non-randomised controlled trial/experiment	Low risk	Unclear risk	Unclear risk	Unclear risk	Low risk	Low risk	Unclear risk	68%
Tehrani et al. (2001)	Before-and-after study	Low risk	Unclear risk	High risk	Unclear risk	Low risk	Low risk	High risk	21%
Tuckey et al. (2014)	Randomised controlled trial/experiment	Low risk	Unclear risk	Unclear risk	Low risk	High risk	Low risk	Unclear risk	89%
Wee et al. (1999)	Non-randomised controlled trial/experiment	Unclear risk	Unclear risk	High risk	Unclear risk	Unclear risk	Low risk	Low risk	64%
Wu et al. (2012)	Randomised controlled trial/experiment	Low risk	Unclear risk	Low risk	Low risk	Unclear risk	Low risk	Low risk	95%

Selection bias. Selection bias was mixed within the studies. Ten studies were rated as low risk of bias due to reasons such as providing clear descriptions of the study population and recruitment methods, finding no significant differences in baseline characteristics between groups and acceptable levels of non-response rates. Five studies were rated as unclear. Four of these studies (Harris et al., 2002; Kenardy et al., 1996; Regehr & Hill, 2000; Wee at al., 1999) adopted a naturalistic design in which they approached participants who had or had not attended a psychological debriefing following a PTE at work retrospectively. As a result, these studies could not discount systematic differences between participants who attended psychological debriefing and those that did not. The remaining eight studies were rated as high-risk of bias, primarily due to clear differences between the groups being compared, including different occupations (Chemtob et al., 1997; Eid et al., 2001; Humphries et al., 2001) or different geographical areas (Matthew, 1998). In two studies (Carlier et al., 2000; Ruck et al., 2013), the intervention and control groups were formed through self-selection, with the control group consisting of those who had declined debriefing. As a result of this self-selection, the debriefed groups may have consisted of people more negatively impacted who sought out help (Tuckey, 2007).

Performance bias. All studies were rated as unclear risk of performance bias. This was primarily due to the studies being unable to blind participants to the intervention they were receiving. All but two of the studies collected self-report measures of PTSD symptoms. In these cases, participants' awareness of the intervention they were receiving, rather than the intervention itself, may have influenced their self-reported scores. The remaining two studies were rated as unclear due to a lack of clarity surrounding the information given to participants prior to taking part in the study, meaning that it was not possible to determine whether participants were differentially motivated (Carlier et al., 1998; Wu et al., 2012).

Treatment fidelity. Treatment fidelity was mixed within the studies. While most studies reported adhering to a seven phase model of psychological debriefing, only three studies provided evidence of treatment fidelity being appropriately assessed through the independent scoring of protocol adherence (Adler et al., 2008; Adler et al., 2009; Wu et al., 2012). Consequently, all the other studies were rated as either unclear risk or high risk. Six studies were rated as high risk either due to there being no assurances that facilitators were trained in delivering psychological debriefing (Chemtob et al., 1997; Tehrani et al., 2001) or researchers having no control over the intervention provided to participants (Harris et al., 2002; Kenardy et al., 1996; Regehr & Hill., 2000; Wee at al., 1999).

Detection bias. The majority of studies were rated as low-risk of detection bias as they used well established outcome measures of PTSD symptoms with good psychometric properties such as the Impact of Event Scale (IES; Horowitz et al., 1979), IES-revised version (Weiss, 2007) or PTSD Checklist (PCL; Weathers et al., 1993) and implemented these measures consistently across participants. In the two studies which used assessor ratings rather than self-rating, these assessors were blinded to the debriefing status of participants (Carlier et al., 1998; Wu et al., 2012). The remaining five studies were rated as unclear risk. In three cases, this was due to the study using a less well-established measure devised by an author of the paper without sufficient justification for this decision (Carlier et al., 2000; Tehrani et al., 2001; Shoval-Zuckerman et al., 2015). Other reasons for studies being rated as unclear risk were not reporting the psychometric properties of the measures used (Wee et al., 2012) or reporting a total score based on the combination of two separate outcome measures (Matthews, 1998).

Statistical bias. Eight studies were rated as low risk for statistical bias, with seven as unclear and six as high. Seven of the eight studies rated as low risk used appropriate statistical testing and reported no data loss, while one had an attrition marginally above 5%, but used intention-to-treat analysis (Grundlingh et al., 2017). Studies were primarily rated as unclear due to a lack of clarity regarding the statistical testing used or attrition rates between 10% and 20%, while the six high-risk studies had attrition rates above 30%.

Reporting bias. Overall, the full reporting of the outcome within studies was good, with 19 of the studies being rated as low risk of reporting bias. One study was rated as unclear risk as statistics were not reported for most of the data and, instead, presently solely as percentages (Carlier et al., 1998). One study was rated as high risk as the six-month follow-up data was not reported, with only a statement provided that "no significant difference" was found between the experimental and control groups (Carlier et al., 2000).

Generalisability. The majority of studies included within this meta-analysis were looking at the effectiveness of psychological debriefings within a specific occupation and demonstrated no intention to extrapolate these findings outside of this population. Consequently, ratings for generalisability were mostly determined by the sample sizes in studies. Ten studies were rated as low risk, with some of these studies, particularly those in military research, using very high sample sizes (e.g. Adler et al., 2008; Adler et al., 2009; Wu et al., 2012). However, the other eleven studies were rated as either unclear or high risk of generalisability due to the small sample sizes used and no evidence of power analysis being conducted, or other justifications provided, for the sample size utilised.

Summary. Overall, there was a mixed level of bias across the 21 studies included in the meta-analysis. However, due to the difficulties in conducting randomised controlled

trials with trauma, poorer quality studies with medium to high risk of bias were included. Consequently, sensitivity analysis was used to empirically assess the impact of methodological variations.

1.4. Results

1.4.1. Selection of the meta-analytic model

The distribution of primary study effects is shown in Figure 1.2. The between-study variance (tau^2) was calculated using the restricted maximum-likelihood estimator (Banks et a., 1985).

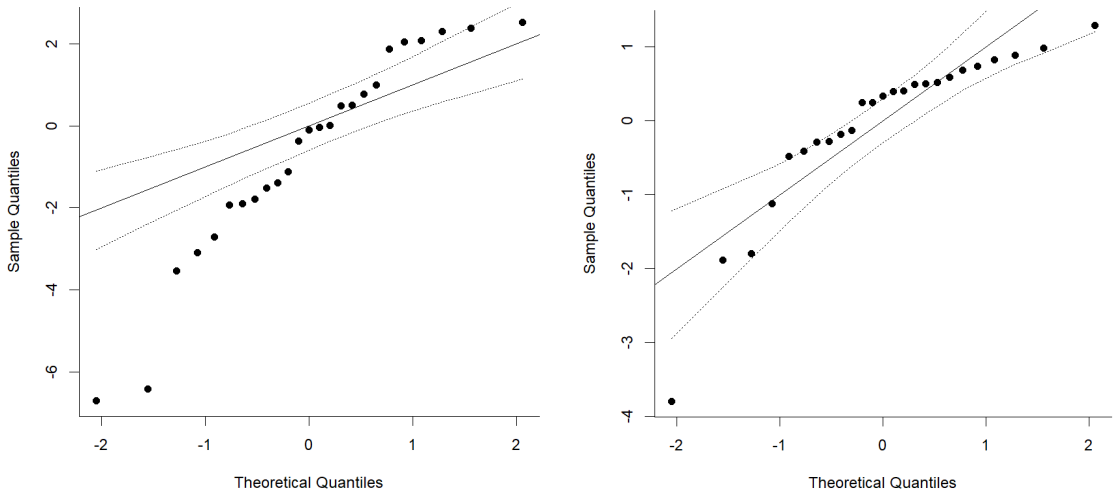


Figure 1.2. *QQ plot of the distribution of standardised mean differences within the primary studies. Chart A shows the fixed effects model and Chart B depicts the random effect model using the restricted maximum-likelihood estimator.*

As can be seen from Figure 1.2, the fixed effects model (Chart A) shows clear evidence of non-normality in the distribution of standard mean differences within the primary studies. While the random effects model using the restricted maximum-likelihood estimator also shows some evidence of non-normality, 90% of the primary study effects fall

within the 95% confidence intervals for the expected normal values. This indicates that the use of random effect model using the restricted maximum-likelihood estimator estimate was an appropriate method for the calculation of the variation of the true effect.

1.4.2. Omnibus test of total score on PTSD measures

The standardised mean differences described in the primary studies are reported in Table 1.6. There were 21 studies reporting a total of 3744 participants. Participants were recruited from a variety of occupations including military, emergency services, healthcare, prison and care sectors, as well as occupations where there is a lower risk of work-related PTEs such as financial and retail sectors. The reasons for the psychological debriefings taking place were predominantly due to a single, discrete event such as a robbery, assault, or road traffic accident (15 studies). However, for studies using military samples, psychological debriefings were predominantly offered due to multiple PTEs occurring during a deployment. In 18 of the studies, a single debriefing session was offered, with only two studies offering more than one debriefing session (Carlier et al., 2000; Grundlingh et al., 2017) and Kenardy et al. (1996) including in their sample both participants who had attended a single session and those that had attended multiple sessions. Studies took place in a variety of geographical locations including the UK, USA, Australia, Netherlands, Norway, Uganda, Ireland, Israel, and China. Most of the studies included mixed gender samples, although studies with participants from the military or emergency services consisted of predominantly male or all-male samples.

Table 1.6. *Treatment effects reported in the primary studies (using first [or only] data collection time point for each study).*

Study name	Year	Cohens *d*	SE	N	Study Design	Single or multiple incident	Single or multiple debrief	Area of employment
Campfield et al. 2001 (delayed debrief)	2001	-0.71	0.32	41	Before-and-after study	Single	Single	Fast food, hotel, petrol service station, rail, video store
Campfield et al. 2001 (immediate debrief)	2001	-3.89	0.57	36	Before-and-after study	Single	Single	Fast food, hotel, petrol service station, rail, video store
Richards 2001	2001	-1.86	0.28	75	Before-and-after study	Single	Single	Finance
Ruck et al. 2013	2013	-0.61	0.28	55	Before-and-after study	Single	Single	Prison staff
Tehrani et al. 2001	2001	2.17	0.67	12	Before-and-after study	Single	Single	Supermarket
Carlier et al. 1998	1998	0.00	0.20	105	Non-randomised controlled trial/experiment	Single	Single	Emergency services
Carlier et al. 2000	2000	0.16	0.14	168	Non-randomised controlled trial/experiment	Single	Multiple	Emergency services
Chemtob et al. 1997	1997	-1.29	0.34	43	Non-randomised controlled trial/experiment	Single	Single	Disaster workers
Deahl et al. 1994	1994	-0.19	0.27	62	Non-randomised controlled trial/experiment	Multiple	Single	Military
Eid et al. 2001	2001	-0.64	0.48	18	Non-randomised controlled trial/experiment	Single	Single	Military and emergency services
Harris et al. 2002	2002	0.04	0.07	660	Non-randomised controlled trial/experiment	Single	Single	Emergency services
Humphries et al. 2001	2001	-0.79	0.39	34	Non-randomised controlled trial/experiment	Single	Single	Finance, retail, hospital emergency
Kenardy et al. 1996	1996	0.22	0.15	195	Non-randomised controlled trial/experiment	Single	Both	Emergency services and disaster workers
Matthews 1998	1998	-0.12	0.30	63	Non-randomised controlled trial/experiment	Single	Single	Care workers
Regehr et al. 2000	2000	0.35	0.20	127	Non-randomised controlled trial/experiment	Single	Single	Emergency services
Shoval-Zuckerman et al. 2015	2015	-0.52	0.16	166	Non-randomised controlled trial/experiment	Multiple	Single	Military
Wee et al. 1999	1999	-0.49	0.26	65	Non-randomised controlled trial/experiment	Multiple	Single	Emergency services
Adler et al. 2008	2008	-0.10	0.10	382	Randomised controlled trial/experiment	Multiple	Single	Military
Adler et al. 2009	2009	-0.21	0.09	514	Randomised controlled trial/experiment	Multiple	Single	Military
Grundlingh et al 2017	2017	0.62	0.28	52	Randomised controlled trial/experiment	Multiple	Multiple	Violence researchers
Tuckey et al. 2014	2014	0.15	0.32	39	Randomised controlled trial/experiment	Single	Single	Emergency services
Wu et al. 2012	2012	-0.03	0.07	832	Randomised controlled trial/experiment	Single	Single	Military

1.4.3. The impact of study design on effect size

A random effects models was calculated using the generic inverse variance method to compare the effect size estimates of the three different study designs included within the meta-analysis (see Figure 1.3). The weighted average standardised mean difference for before-and-after studies (SMD = -1.78, 95% CI = 2.93 to -0.64) was significantly different (χ^2 = 9.24, p < .01) to the SMD for both non-randomised controlled trials (SMD = -0.19, 95% CI = -0.45 to 0.06) and randomised controlled trials (SMD = -0.05, CI = -0.20 to 0.10). The magnitude of the effect size estimate in the before-and-after studies is likely to have been inflated by the maturational biases inherent in this study design. Consequently, all uncontrolled before-and-after studies were removed from the meta-analysis.

There was no significant difference (χ^2 = 0.94, p = .33) between the effect size estimates of non-randomised controlled trials and randomised controlled trials, so these study designs were combined for the subsequent analyses. Furthermore, when before-and-after studies were excluded, heterogeneity went from being unacceptably high (I^2 = 86%) to below the 75% threshold (I^2 = 69%) and so a 'leave-one-out' analysis was not required.

The random effects model was recalculated following the removal of the before-and-after studies and the combining of both the non-randomised controlled trials and randomised controlled trials (see Figure 1.4). An overall effect favouring psychological debriefing was found (SMD=-0.11). However, this effect was statistically non-significant (-0.28 to 0.07).

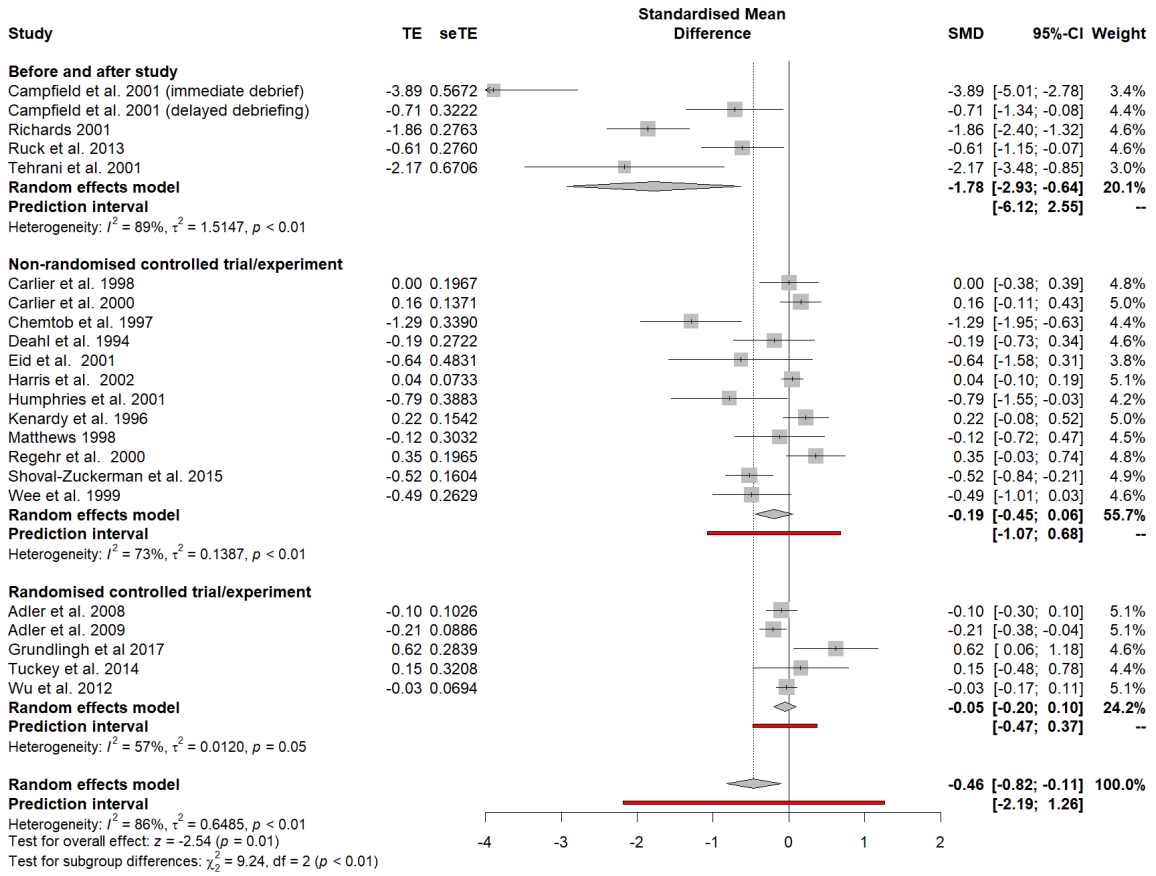


Figure 1.3. *Subgroup plot on the impact of study design on estimated effect size.*

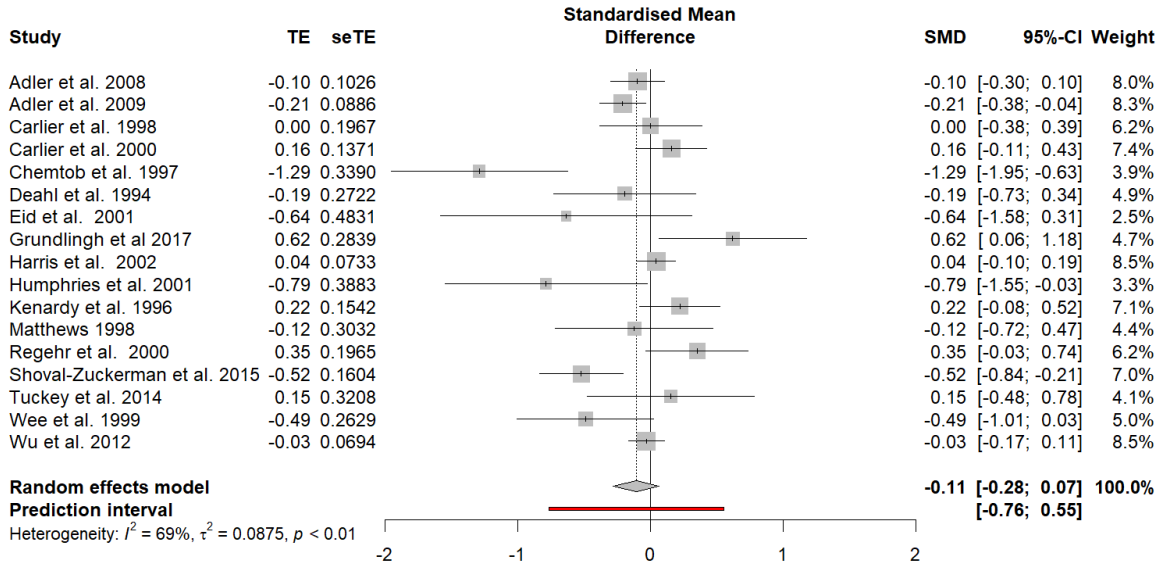


Figure 1.4. *Forest plot of the standardised mean difference of PTSD symptoms between participants who did and did not receive psychological debriefing following a potentially traumatic event.*

1.4.4. The impact of time on effect size

To examine the impact of time of time on estimate effect size, a subgroup analysis was conducted to compare studies which collected short-term outcomes (0-3 months after debriefing), medium-term outcomes (4-6 months after debriefing) and long-term outcomes (7 months or more after debriefing; see Figure 1.5). For short-term outcomes, an effect favouring the intervention was reported (SMD=-0.24), but this effect was statistically non-significant (95% CI -0.70 to 0.22). For medium-term outcomes, an effect favouring the intervention was reported (SMD=-0.14), but again this was non-significant (CI -0.33 to 0.06). For long-term outcomes, a treatment effect close to zero was observed, although this did favour non-intervention (SMD=0.07, 95% CI -0.12 to 0.25). Four studies could not be included in these subgroup comparisons because the timeframe between the PTE and outcome collected was either unspecified or varied between participants. As there was no significant different between these subgroups ($\chi^2 = 3.12$, $p = .37$), they were combined for all subsequent analyses.

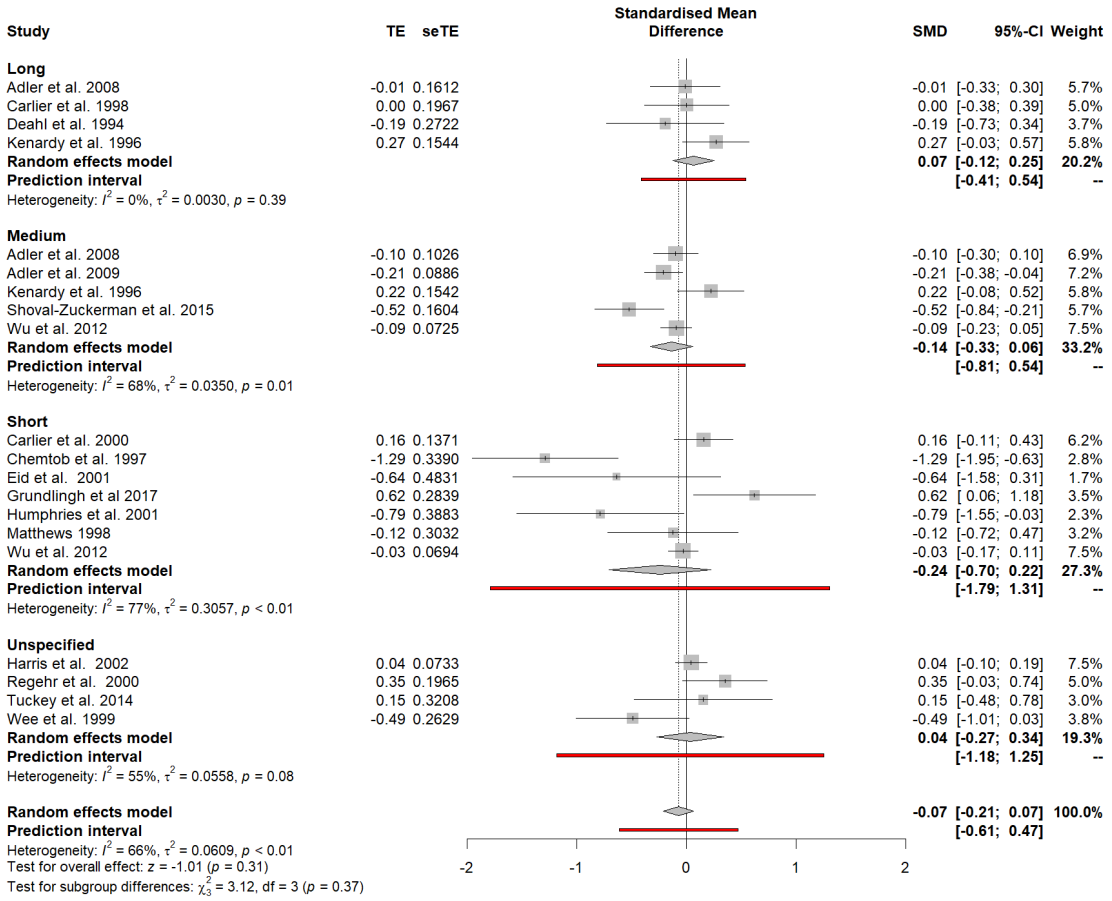


Figure 1.5. *Forest plot of the standardised mean difference of PTSD symptoms between participants who did and did not receive psychological debriefing at different timepoints after debriefing.*

1.4.5. The impact of risk of bias in the primary studies

To assess the impact of study-level risk of bias upon heterogeneity, a series of subgroup analysis were conducted on the estimates of SMD for the risk of bias ratings of 'low risk' and 'any risk' (unclear risk and high risk of bias combined) for each of the seven domains of methodological bias (see Table 1.7). No statistically significant differences in effect size estimates between studies with 'low risk' of bias and 'any risk' of bias were observed in any of the seven domains.

Table 1.7. *Standard mean differences for studies with a 'low risk' of bias and studies with 'any risk' of bias within each of the seven risk domains.*

	Low Risk			Any Risk				
	EFFECT	**95% CI**	**k**	**EFFECT**	**95% CI**	**k**	**X2**	**P**
Short Term								
Selection bias	-0.08	-0.28 to 0.11	7	-0.19	-0.49 to 0.11	10	0.35	0.55
Performance bias	----	----	----	-0.11	-0.28 to 0.07	17	----	----
Treatment fidelity	-0.10	-0.20 to 0.01	4	-0.14	-0.40 to 0.11	13	0.12	0.73
Detection bias	-0.07	-0.27 to 0.13	13	-0.23	-0.59 to 0.13	4	0.58	0.45
Statistical bias	-0.11	-0.52 to 0.29	6	-0.10	-0.28 to 0.07	11	<0.01	0.96
Reporting bias	-0.14	-0.35 to 0.06	15	0.11	-0.11 to 0.33	2	2.66	0.10
Generalisability bias	-0.02	-0.14 to 0.09	8	-0.25	-0.62 to 0.12	9	1.35	0.25

1.4.6. Differences in avoidance, hyperarousal and intrusion symptom outcomes

Outcomes were grouped in the three PTSD symptom clusters: avoidance, hyperarousal and intrusion. Studies which only reported total PTSD scores were excluded from this subgroup analysis. The difference between avoidance, arousal and intrusion symptoms was assessed in the subgroup plot shown in Figure 1.6.

No significant difference was found between outcomes on the three symptoms clusters ($\chi^2 = 3.36$, $p = .19$) and no significant treatment effects were observed for avoidance symptoms (SMD=-0.28, 95% CI 0.63 to 0.07), hyperarousal (SMD=0.12, 95% CI -0.15 to 0.39) or intrusion symptoms (SMD = -0.14, 95% CI -0.52 to 0.23).

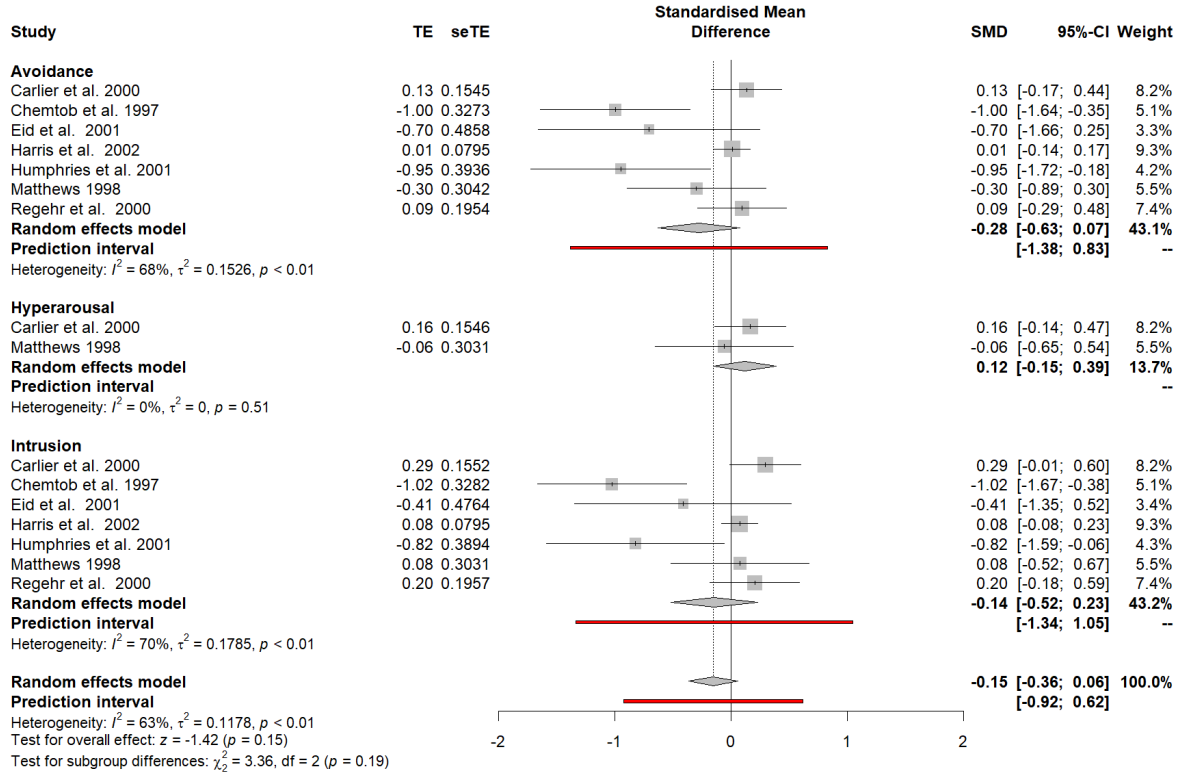


Figure 1.6. *Forest plot of the standardised mean difference of specific PTSD symptom clusters between participants who did and did not receive psychological debriefing following a potentially traumatic event.*

1.4.7. Difference attributable to characteristics of psychological debriefings

Adherence to an established model of psychological debriefing. Studies were categorised according to whether or not assurances were given that the psychological debriefing adhered to the seven phase models outlined by Mitchell (1983) or Dyregrov (1989; see Figure 1.7). There was no significant difference observed between those who did and did not adhere to the models (χ^2 = 0.02, p = .88), although there was markedly less heterogeneity between studies that adhered to a seven phase model (I^2 = 34%, p = .14) compared to those who did not (I^2 = 83%, p < .01).

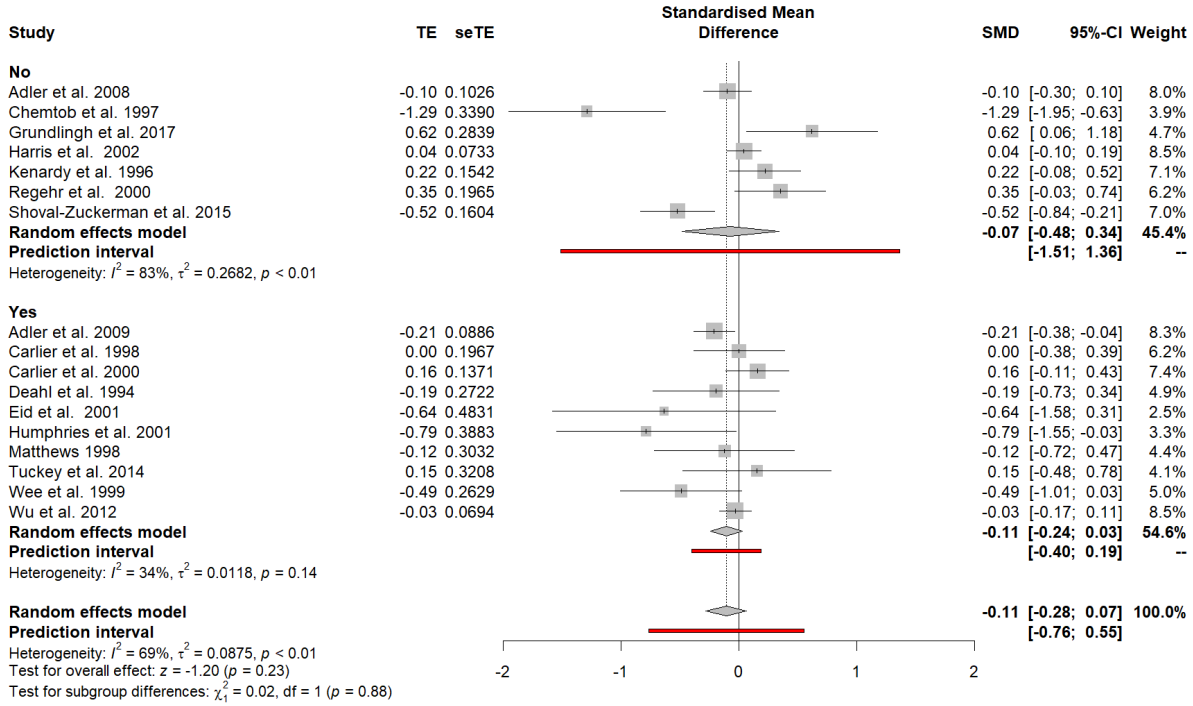


Figure 1.7. *Subgroup plot of differences between studies that did and did not adhere to established seven phase models.*

Single or multiple session debriefings. Studies that evaluated a single-session debriefing were compared with studies that provided multiple debriefing sessions (see Figure 1.8). Kenardy et al. (1996) included data from both single and multiple session debriefings, so was excluded from this subgroup analysis. A significant difference (χ^2 = 4.64, p = .03) favouring single session debriefing was observed and when only single-session debriefings were included in the analysis, a significant effect was found (SMD=-0.19, 95% CI -0.37 to -0.02).

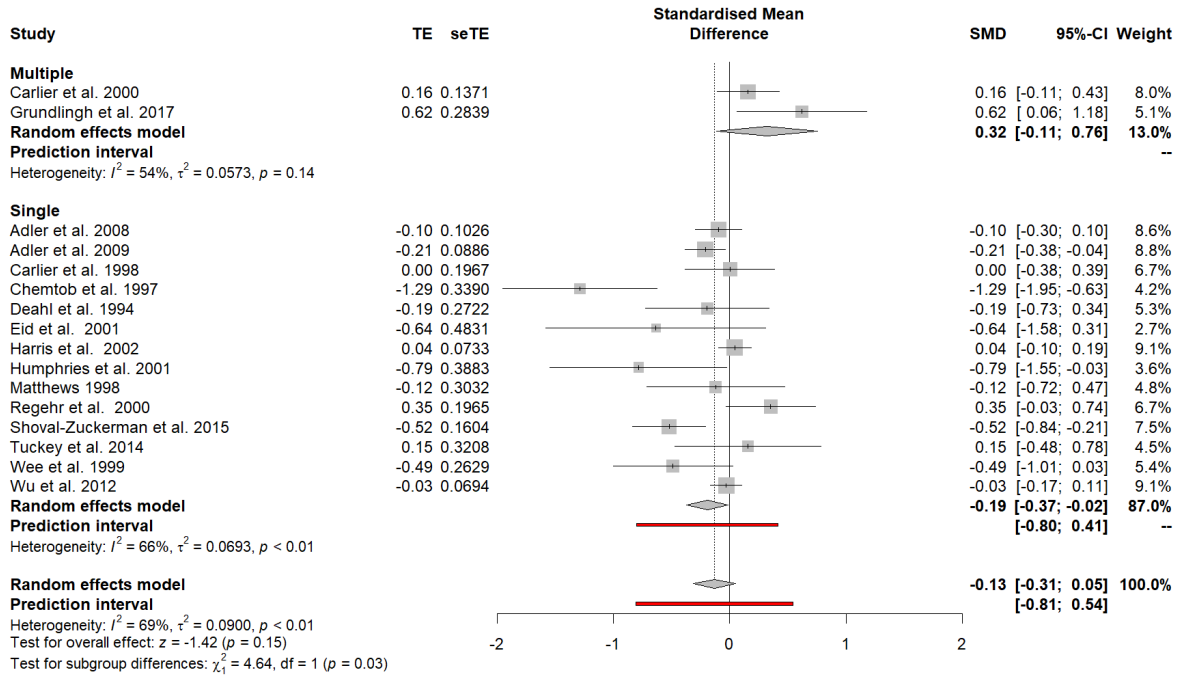


Figure 1.8. *Subgroup plot of differences between studies offered single session debriefing versus those that offered multiple debriefing sessions.*

Individual or group debriefings. Studies that evaluated group debriefings were compared with studies that evaluated individual debriefings (see Figure 1.9). Three studies did not specify whether debriefings were done with groups or individuals (Humphries et al., 2003; Kenardy et al., 1996; Wee et al., 2012) and so they were excluded from this subgroup analysis. No statistically significant difference was observed between the two subgroups (χ^2 = 2.24, p = .13)

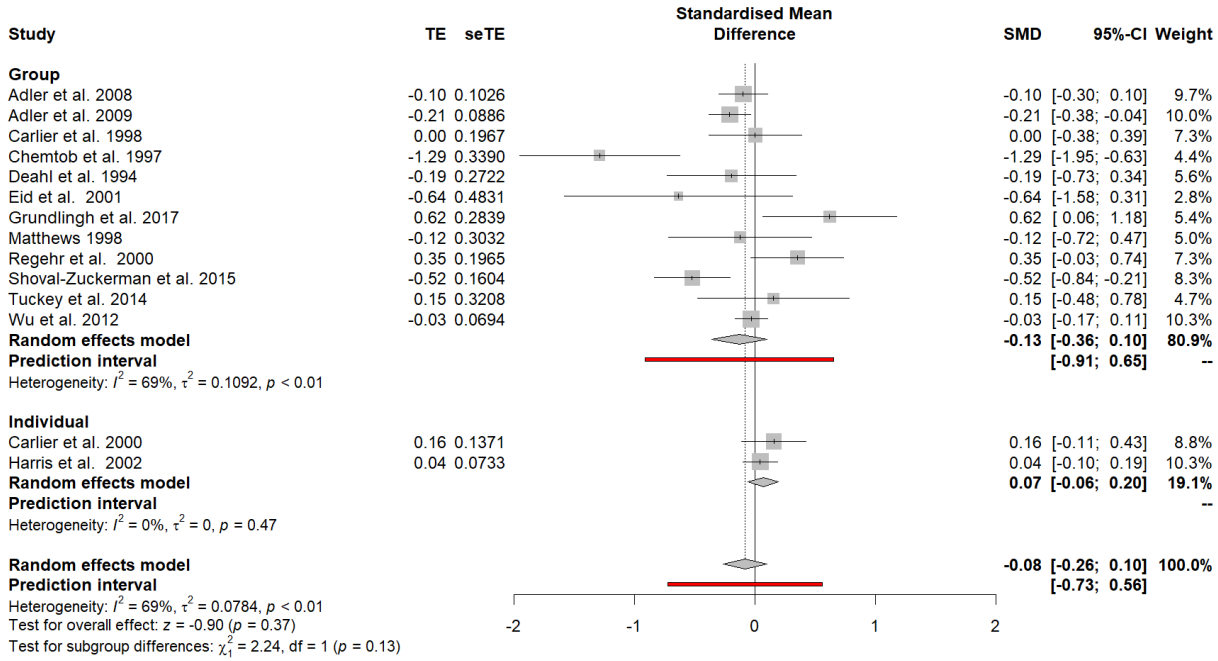


Figure 1.9. *Subgroup plot of differences between individual and group debriefings*

1.4.8. Differences attributable to trauma characteristics

Single or multiple traumatic incidences. Outcomes were compared for participants who had experienced a single PTE versus participants who were reported to have experienced multiple PTEs (see Figure 1.10). There was no significant difference observed in effect sizes between participants exposed to a single PTE and participants exposed to multiple PTEs (χ^2 = 0.35, p = .55)

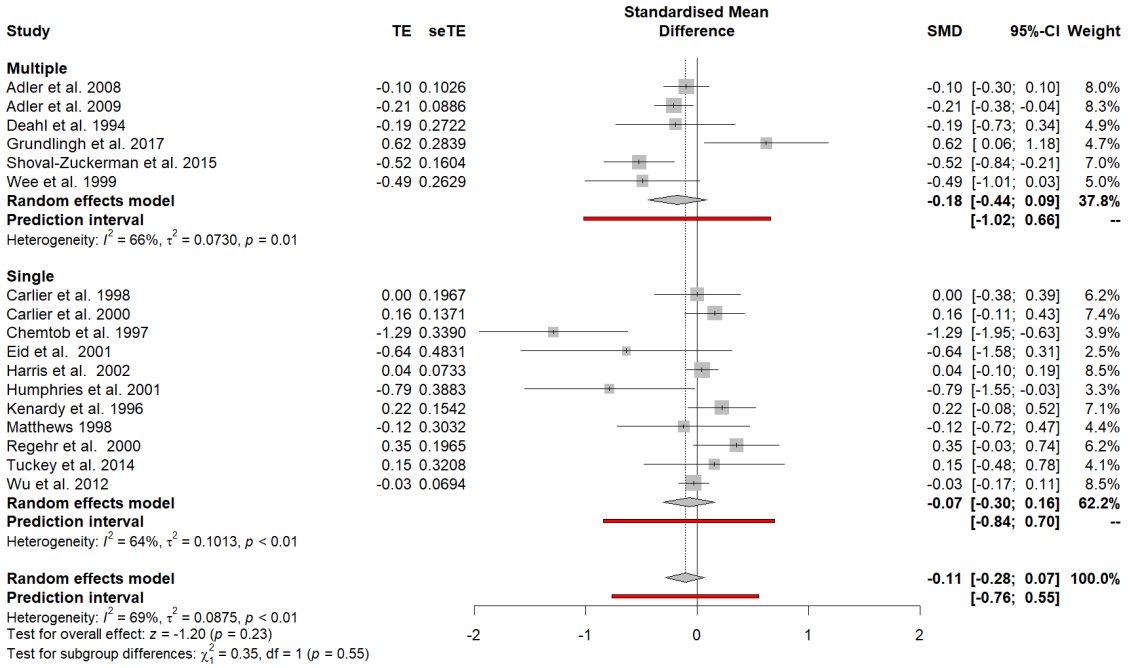


Figure 1.10. *Subgroup plot of differences between single or multiple traumatic incidences.*

1.4.9. Subgroup analyses that were not possible to conduct

Data was organised so that subgroup analyses could also be conducted in other areas including the length of debriefing, the length of time between the PTE and debriefing and the extent of debriefers training. However, in several of the studies, this information was not reported and so these subgroup analyses could not be conducted.

1.4.10. Publication bias and small-study effects

Small-study effects refers to the tendency for studies with smaller sample sizes to show different and often larger treatment effects than studies with larger sample sizes (Rücker et al., 2011). One possible reason for this is publication bias, whereby statistically significant results are more likely to be published than non-significant results (Rothstein et al., 2006). Firstly, in smaller studies, larger treatment effects are needed for a result to be statistically significant. Secondly, due to the higher levels of resource and often higher

methodological quality of larger studies, non-significant results in larger studies are more likely to be published than non-significant results in smaller studies (Sterne et al., 2000).

Bias in a meta-analysis may be assessed visually using a funnel plot; a simple scatter plot of the treatment effect estimates from each primary study against a measure of study size (Rothstein et al., 2006). If there is an absence of bias, the plot will resemble a symmetrical inverted funnel as the effects from the smaller studies at the bottom of the plot show greater variability than the larger studies at the top of the plot, which will lie closer to the overall meta-analytic effect. However, if there is an absence of studies in the area of the plot associated with small sample sizes and non-significant results, it is likely that publication bias is resulting in an overestimation of the true effect size.

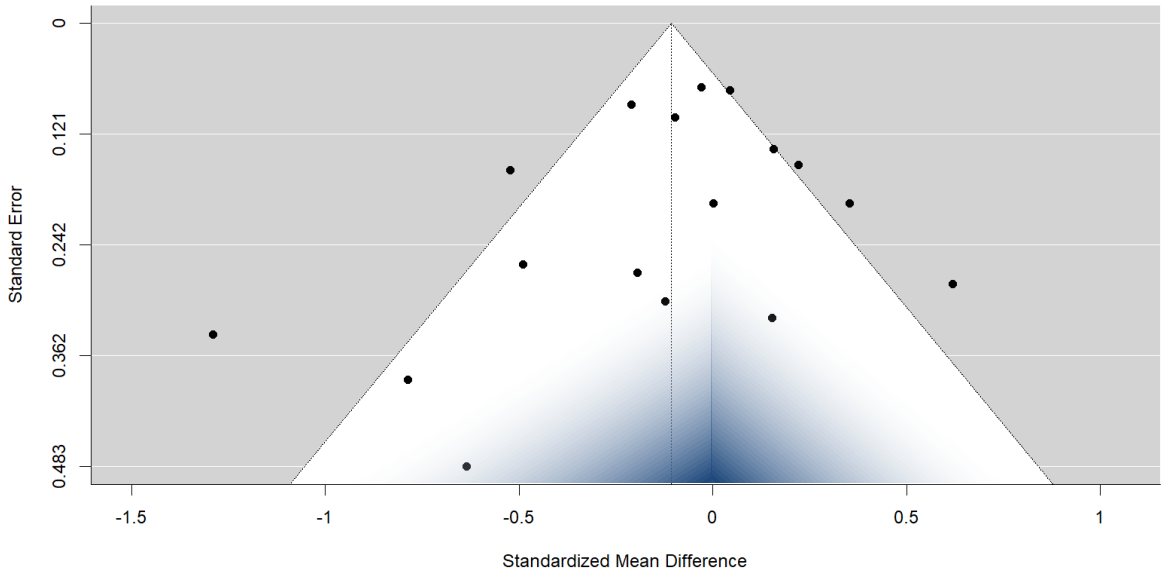


Figure 1.11. *Funnel plot of the standardised mean difference for all PTSD symptom outcomes. The 95% confidence interval of the expected distribution of treatment effects is shown as an inverted 'funnel'. The highlighted area in blue is that associated with publication bias.*

Visual inspection of the funnel plot (see Figure 1.11) would suggest the presence of publication bias as there appears to be an absence of studies with higher standard errors (i.e., smaller samples) around the area of the funnel plot consistent with null results (standardised mean difference = 0). In addition, the heterogeneity of this data is evident in the number of SMD outside of the expected 95% confidence interval.

The trim-and-fill method (Duval & Tweedle, 2000a, 2000b) was used to detect and adjust for the publication bias evident in the funnel plots asymmetry. The trim-and-fill method involves iteratively removing the most extreme small studies from the positive side of the funnel plot and re-computing the effect size at each iteration until the funnel plot is symmetrical about a corrected effect size. The omitted studies are then added back into the analysis and a mirror image for each of these studies is imputed. The trim-and-fill procedure did not identify statistically significant funnel plot asymmetry and therefore did not result in any corrections to the current analysis.

1.5. Discussion

The aim of this meta-analysis was to evaluate the effectiveness of psychological debriefing in preventing or reducing PTSD symptoms following a work-related PTE and identify factors that appear to impact the effectiveness of psychological debriefing though a series of subgroup analyses.

It was recognised that the unpredictable nature of trauma means that most trauma research cannot meet the gold-standard of study design insisted upon by the Cochrane Library and so a variety of study designs were initially included within this meta-analysis, including uncontrolled before-and-after studies or studies which lacked a suitable control group so were treated as before-and-after studies. However, in the absence of a control

group in these studies, it was not possible to determine whether psychological debriefing resulted in an improvement over and beyond that of natural recovery and the particuarly high effect sizes in these studies suggested that maturational effects and other potential biases influenced outcomes. Consequently, little could be inferred from these study's results, and they were removed from the meta-analysis.

While four of the controlled studies included in the meta-analysis found a statistically significant positive effect of psychological debriefing, with only one finding a significant negative effect (Grundlingh et al. 2017), the overall synthesis did not find consistent and substantive evidence that psychological debriefing helps to prevent or reduce PTSD symptoms following a work-related PTE.

While heterogeneity of the controlled trials was below the pre-defined 75% threshold, there was still substantial variation between studies. The subgroup analysis on adherence to the seven phase models outlined by Mitchell (1983) and Dyregrov (1989) explains much of this variation. There was markedly less heterogeneity in effect sizes between the studies which adhered to a seven phase model compared to the studies that evaluated interventions referred to as psychological debriefing but that were either significantly modified or did not offer any assurances a standardised seven phase model was used. The apparent confusion and inconsistency in the literature regarding use of the term 'psychological debriefing' has been previously recognised (Tuckey, 2007). This lack of clarity has hampered research progress and increased the likelihood of misapplication of research findings. Future research in this area should ensure that the psychological debriefing being evaluated adheres to an established standardised seven phase model. This will improve the robustness of the evidence base in this field.

Only one of the subgroup analyses conducted produced a statistically significant finding. Single-session debriefings were found to produce better outcomes than multiple-session debriefings. Furthermore, when analysis was limited to studies that solely evaluated single-session debriefing, a significant effect favouring psychological debriefing was found. This result contrasts with the non-significant finding of the Cochrane review (Rose et al., 2002) which only evaluated single-session psychological debriefings. It is important to note that heterogeneity in outcomes between studies was high and, while the overall effect was significant, the overall effect size was small (*d* = -0.19; Cohen, 1992). Nevertheless, the finding brings into question the assertion that "single-session debriefing may be at best ineffective" (NICE, 2005, p.84).

Many other subgroup analyses were unable to be conducted due to unreported information within studies, including some directly linked to the recommendations made about psychological debriefing by Hawker & Hawker (2015). These included the impact of the timing of psychological debriefing following a PTE, the length of debriefing sessions and the qualifications and training of facilitators. For some studies, this absence of this information was simply due to poor reporting. For other studies, it was due to methodological shortcomings. For example, the four studies relying on naturalistic methods (Harris et al., 2002; Kenardy et al., 1996; Regehr & Hill, 2000; Wee at al., 1999) had no influence on the provision of debriefings and only limited knowledge about the nature of the interventions they were evaluating.

Tuckey (2007) makes a number of recommendations regarding the clarity of reporting in psychological debriefing research. These recommendations include clearly and accurately reporting the level of training and experience of the debriefers, the timing of the

debriefing sessions relative to the potentially traumatic events, and the size of the group debriefing sessions. Following these recommendations would, again, improve the robustness of the evidence base into the effectiveness of psychological debriefing.

While recommendations have been made to improve the robustness of the evidence base, it appears that research into psychological debriefing has reduced in recent years. Most of the 21 articles included within this meta-analysis were publish before or around the turn of the millennia, with only seven published since 2002 when the Cochrane review on psychological debriefing was published. Hawker & Hawker (2011) note the difficulties in obtaining ethical approval and funding for research in this area in the present day due to the widespread belief that psychological debriefing is harmful. Yet this meta-analysis suggests that future studies, which both adhere to a standardised models and clearly and accurate report on the nature of the psychological debriefing being offered, are warranted.

1.5.1. Limitations

There are some limitations to this meta-analysis which must be acknowledged. There were significant methodological shortcomings in many of the studies included in the synthesis. As previously noted, some studies had no control over the nature of the psychological debriefings provided to participants. In other studies, attrition rates were very high. One of the most noticeable methodological limitations to several of the studies related to the recruitment of control groups. In four studies, control groups were taken from a different occupational group or geographical area. In two studies, intervention and control groups were formed through self-selection, with the control group comprising those who had declined debriefing. These approaches are likely to have introduced selection bias.

There are inherent difficulties in establishing appropriate control groups within trauma research. It is important that psychological debriefing is optional rather than mandatory and, conversely, that available interventions are not intentionally withheld from people. Furthermore, given the early nature of the intervention, waiting-list control groups are often not practicable. Consequently, it is to be expected that studies resort to self-selection methods to form intervention and control groups. However, in these cases it is important baseline assessments are administered to ensure there is no differences in symptom severity between the groups prior to intervention or, if there is, that this is accounted for using an interaction effect between group and time.

A second limitation is that all but two of the studies included relied on self-report outcome measures. While subjective experience of symptomology is important, the psychoeducational component of debriefing may have increased participant's awareness of symptoms and, therefore, increased their self-reported scores on outcome measures (Grundlingh et al., 2017).

A third limitation is that outcomes were restricted to PTSD symptoms. While the addition of further outcomes would have resulted in an unwieldy analysis, Richins et al. (2020) note that additional outcome measures in primary studies may uncover other benefits. For example, Tuckey & Scott (2014) found that emergency service personnel who had been debriefed following a PTE consumed less alcohol as a means of coping and reported better quality of life. Furthermore, Richins et al. (2020) note the high proportion of studies evaluating group-based early interventions where peer support was reported to facilitate recovery or improve experience. These identified social benefits of psychological

debriefing may not captured by measures of PTSD symptoms but could still make psychological debriefing a worthwhile intervention.

Finally, while the meta-analysis focused specifically on work-related PTEs, the scope of studies included was still large. One of the greatest variations between studies was the length of time between a PTE and the psychological debriefing. This ranged from 24 hours (Carlier et al., 2000) to 6 months (Chemtob et al., 1997). Mitchell & Everly (1996 p.87) only caution against the use of psychological debriefing "several months" after a PTE and so this timeframe does not necessarily go against established recommendations. Nevertheless, interventions at different timeframes are likely to serve different functions. Unfortunately, it was not possible to explore the impact of timing of psychological debriefing on PTSD symptomology and so studies using markedly different timeframes were combined throughout the analysis.

1.5.2. Conclusion

It appears that, for now, the debriefing debate will continue. While the overall synthesis in this meta-analysis did not provide any consistent and substantive evidence that psychological debriefing improves natural psychological recovery after a traumatic event, the findings also suggest that Rose et al.'s (2002) conclusion that "psychological debriefing is either equivalent to, or worse than, control or educational interventions in preventing or reducing the severity of PTSD" (p.2) may have been premature. The widespread belief that psychological debriefing is harmful appears that have hindered the progress of research in this field. It is hoped that further well-designed studies that account for the methodological limitations inevitable in trauma research are conducted. This will help to inform

organisations' provision of intervention following work-related PTEs and ultimately ensure that employees receive the effective support they need and deserve.

1.6. References

Adler, A. B., Bliese, P. D., McGurk, D., Hoge, C. W., & Castro, C. A. (2009). Battlemind debriefing and battlemind training as early interventions with soldiers returning from Iraq: Randomization by platoon. *Journal of consulting and clinical psychology*, *77*(5), 928-940.

Adler, A. B., Litz, B. T., Castro, C. A., Suvak, M., Thomas, J. L., Burrell, L., McGurk, D., Wright, K. M., & Bliese, P. D. (2008). A group randomized trial of critical incident stress debriefing provided to US peacekeepers. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, *21*(3), 253-263.

Banks, B. D., Mao, I. L., & Walter, J. P. (1985). Robustness of the Restricted Maximum Likelihood Estimator Derived Under Normality as Applied to Data with Skewed Distributions. *Journal of Dairy Science*, *68*(7), 1785–1792.

Bisson, J. I., Jenkins, P. L., Alexander, J., & Bannister, C. (1997). Randomised controlled trial of psychological debriefing for victims of acute burn trauma. *The British Journal of Psychiatry*, *171*(1), 78-81.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Brooks, S. K., Rubin, G. J., & Greenberg, N. (2019). Traumatic stress within disaster-exposed occupations: overview of the literature and suggestions for the management of traumatic stress in the workplace. *British Medical Bulletin, 129,* 25-34.

Campfield, K. M., & Hills, A. M. (2001). Effect of timing of critical incident stress debriefing (CISD) on posttraumatic symptoms. *Journal of Traumatic Stress*, *14*(2), 327-340.

Carlier, I. V., Lamberts, R. D., Van Uchelen, A. J., & Gersons, B. P. (1998). Disaster-related post-traumatic stress in police officers: a field study of the impact of debriefing. *Stress Medicine*, *14*(3), 143-148.

Carlier, I. V., Voerman, A. E., & Gersons, B. P. R. (2000). The influence of occupational debriefing on post-traumatic stress symptomatology in traumatized police officers. *British Journal of Medical Psychology*, *73*(1), 87-98.

Chemtob, C. M., Tomas, S., Law, W., & Cremniter, D. (1997). Postdisaster psychosocial intervention: a field study of the impact of debriefing on psychological distress. *American Journal of Psychiatry*, *154*(3), 415-417.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Deahl, M. (2000). Psychological debriefing: controversy and challenge. *Australian & New Zealand Journal of Psychiatry*, *34*(6), 929-939.

Deahl, M. P., Gillham, A. B., Thomas, J., Searle, M. M., & Srinivasan, M. (1994). Psychological sequelae following the Gulf War. *The British Journal of Psychiatry*, *165*(1), 60-65.

Deahl, M., Srinivasan, M., Jones, N., Thomas, J., Neblett, C., & Jolly, A. (2000). Preventing psychological trauma in soldiers: The role of operational stress training and psychological debriefing. *British Journal of Medical Psychology*, *73*(1), 77-85.

Dieltjens, T., Moonens, I., Van Praet, K., De Buck, E., & Vandekerckhove, P. (2014). A systematic literature search on psychological first aid: lack of evidence to develop guidelines. *PloS One*, *9*(12), e114714.

Duval, S., & Tweedie, R. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463.

Duval, S. & Tweedie, R. (2000b). A Nonparametric 'Trim and Fill' Method of Accounting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Dyregrov, A. (1989). Caring for helpers in disaster situations: Psychological debriefing. *Disaster Management*, *2*(1), 25-30.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*(4), 319-345.

Eid, J., Helge Johnsen, B., & Weisæth, L. (2001). The effect of group psychological debriefing on acute stress reactions following a traffic accident: A quasi-experimental approach. *International Journal of Emergency Mental Health*.

Greenberg, N., Langston, V., & Jones, N. (2008). Trauma risk management (TRiM) in the UK Armed Forces. *BMJ Military Health*, *154*(2), 124-127.

Grinker, R. & Spiegel, J. (1944). Brief Psychotherapy in War Neuroses. *Psychosomatic Medicine, 6,* 123-131.

Grundlingh, H., Knight, L., Naker, D., & Devries, K. (2017). Secondary distress in violence researchers: a randomised trial of the effectiveness of group debriefings. *BMC psychiatry*, *17*(1), 1-14.

Harris, M. B., Baloğlu, M., & Stacks, J. R. (2002). Mental health of trauma-exposed firefighters and critical incident stress debriefing. *Journal of Loss &Trauma*, *7*(3), 223-238.

Hawker, D. M., Durkin, J., & Hawker, D. S. (2011). To debrief or not to debrief our heroes: that is the question. *Clinical Psychology & Psychotherapy*, *18*(6), 453-463.

Hawker, D., & Hawker, D. (2015). What can be learned from the debriefing controversy? In W. Yule, D. Hawker, D. Alexander, H. Holowenko, N. Tehrani, J. Durkin, & J. Hacker Hughes (Eds.), *Early Interventions for Trauma* (pp. 6–17). British Psychological Society.

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savović, J. Shultz, K. F., Weeks, L., & Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, *343*.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558.

Hobbs, M., Mayou, R., Harrison, B., & Worlock, P. (1996). A randomised controlled trial of psychological debriefing for victims of road traffic accidents. *BMJ*, *313*(7070), 1438-1439.

Horowitz, M., Wilner, N., & Alvarez, W. (1979). Impact of Event Scale: A measure of subjective stress. *Psychosomatic Medicine*, *41*(3), 209-218.

Humphries, C. & Carr, A. (2001). The short term effectiveness of critical incident stress debriefing. *Irish Journal of Psychology,* 22, 188-197.

Kenardy, J. A., Webster, R. A., Lewin, T. J., Carr, V. J., Hazell, P. L., & Carter, G. L. (1996). Stress debriefing and patterns of recovery following a natural disaster. *Journal of Traumatic Stress*, *9*(1), 37-49.

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.

Lee, W., Lee, Y. R., Yoon, J. H., Lee, H. J., & Kang, M. Y. (2020). Occupational post-traumatic stress disorder: an updated systematic review. *BMC Public Health*, *20*, 1-12

Litz, B. T., Gray, M. J., Bryant, R. A., & Adler, A. B. (2002). Early intervention for trauma: Current status and future directions. *Clinical Psychology: Science and Practice*, *9*(2), 112-134.

Matthews, L. R. (1998). Effect of staff debriefing on posttraumatic stress symptoms after assaults by community housing residents. *Psychiatric Services*, *49*(2), 207-212.

Mayou, R. A., Ehlers, A., & Hobbs, M. (2000). Psychological debriefing for road traffic accident victims: Three-year follow-up of a randomised controlled trial. *The British Journal of Psychiatry*, *176*(6), 589-593.

Mitchell, J. T. (1983). When disaster strikes: The critical incident stress debriefing process, *Journal of Emergency Medical Services, 8,* 36-39.

Mitchell, J. T., & Everly, G. S. (1996). *Critical incident stress debriefing: An operations manual for the prevention of traumatic stress among emergency services and disaster workers.* Chevron.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods*, *7*(1), 105-125.

National Institute for Health and Care Excellence (2018b). *Post-traumatic stress disorder: management Consultation on draft guideline - Stakeholder comments table*. Retrieved from https://www.nice.org.uk/guidance/ng116/documents/consultation-comments-and-responses-2

National Institute for Health and Care Excellence (2018a). *Post-traumatic stress disorder. (Clinical Guideline No. 116).* Retrieved from https://www.nice.org.uk/guidance/ng116/resources/posttraumatic-stress-disorder-pdf-66141601777861

NICE. (2005a). *Post-traumatic stress disorder (PTSD): The management of PTSD in adults and children in primary and secondary care.* (Full Clinical Guideline 26 Developed by the National Collaborating Centre for Mental Health). Retrieved from https://www.nice.org.uk/guidance/ng116/evidence/march-2005-full-guideline-pdf-6602623598

O'Toole, M., & Eppich, W. (2022). In support of appropriate psychological debriefing. *Medical Education*, *56*(2), 229-229.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*.

Petereit-Haack, G., Bolm-Audorff, U., Romero Starke, K., & Seidler, A. (2020). Occupational risk for post-traumatic stress disorder and trauma-related depression: a systematic review with meta-analysis. *International Journal of Environmental Research and Public Health*, *17*(24), 9369.

Priest, S. R., Henderson, J., Evans, S. F., & Hagan, R. (2003). Stress debriefing after childbirth: a randomised controlled trial. *Medical journal of Australia*, *178*(11), 542-545.

Regehr, C., & Hill, J. (2001). Evaluating the efficacy of crisis debriefing groups. *Social Work with Groups*, *23*(3), 69-79.

Richards, D. (2001). A field study of critical incident stress debriefing versus critical incident stress management. *Journal of Mental Health*, *10*(3), 351-362.

Richins, M. T., Gauntlett, L., Tehrani, N., Hesketh, I., Weston, D., Carter, H., & Amlôt, R. (2020). Early post-trauma interventions in organizations: a scoping review. *Frontiers in Psychology*, 1176.

Rose, S. C., Bisson, J., Churchill, R., & Wessely, S. (2002). Psychological debriefing for preventing post traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, 2, CD000560.

Rose, S., Brewin, C. R., Andrews, B., & Kirk, M. (1999). A randomized controlled trial of individual psychological debriefing for victims of violent crime. *Psychological Medicine*, *29*(4), 793-799.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Ruck, S., Bowes, N., & Tehrani, N. (2013). Evaluating trauma debriefing within the UK prison service. *Journal of Forensic Practice*, *15*(4), 281-290.

Rücker, G., Carpenter, J. R., & Schwarzer, G. (2011). Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal*, *53*(2), 351-368.

Shalev, A. Y., Peri, T., Rogel-Fuchs, Y., Ursano, R. J., & Marlowe, D. (1998). Historical group debriefing after combat exposure. *Military Medicine*, *163*(7), 494-498.

Shoval-Zuckerman, Y. (2015). The effectiveness of early group intervention for military reserves soldiers: the role of the repressive coping style. *Israel Journal of Psychiatry*, *52*(3), 49-59.

Shultz, J. M., & Forbes, D. (2014). Psychological first aid: rapid proliferation and the search for evidence. *Disaster Health*, *2*(1), 3-12.

Skogstad, M., Skorstad, M., Lie, A., Conradi, H. S., Heir, T., & Weisæth, L. (2013). Work-related post-traumatic stress disorder. *Occupational medicine*, *63*(3), 175-182.

Söndergaard, H. P. (2008). The work environment, critical incidents, debriefing and psychological functioning-a study of trade union members in Sweden. *Scandinavian Journal of Work, Environment & Health*, *6,* 111-116.

Sterne, J. A., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, *53*(11), 1119-1129.

Tamrakar, T., Murphy, J., & Elklit, A. (2019). Was Psychological Debriefing Dismissed Too Quickly?. *Crisis, Stress, and Human Resilience: An International Journal*, *1*(3), 146-155.

Tuckey, M. R. (2007). Issues in the debriefing debate for the emergency services: Moving research outcomes forward. *Clinical Psychology: Science and Practice*, *14*(2), 106-116.

Tuckey, M. R., & Scott, J. E. (2014). Group critical incident stress debriefing with emergency services personnel: a randomized controlled trial. *Anxiety, Stress & Coping*, *27*(1), 38-54.

von Hippel, P. T. (2015). The heterogeneity statistic I 2 can be biased in small meta-analyses. *BMC Medical Research Methodology*, *15*(1), 1-8.

Weathers, F. W., Litz, B. T., Herman, D. S., Huska, J. A., & Keane, T. M. (1993, October). The PTSD Checklist (PCL): Reliability, validity, and diagnostic utility. In *annual convention of the international society for traumatic stress studies, San Antonio, TX* (Vol. 462).

Wee, D. F., Mills, D. M., & Koehler, G. (1999). The effects of critical incident stress debriefing (CISD) on emergency medical services personnel following the Los Angeles civil disturbance. *International Journal of Emergency Mental Health*, *1,* 33-37.

Weiss, D. S. (2007). The impact of event scale: revised. In *Cross-cultural assessment of psychological trauma and PTSD* (pp. 219-238). Springer.

Wu, S., Zhu, X., Zhang, Y., Liang, J., Liu, X., Yang, Y., Yang, H. & Miao, D. (2012). A new psychological intervention: “512 Psychological Intervention Model” used for military rescuers in Wenchuan Earthquake in China. *Social Psychiatry and Psychiatric Epidemiology*, *47*(7), 1111-1119.

Chapter 2: Empirical Paper - Assessing alexithymia in forensic practice: Psychometric properties of the 20-item Toronto Alexithymia Scale and the Perth Alexithymia Scale in a sample of adult male prisoners

2.1. Abstract

Alexithymia is a trans-diagnostic construct comprised of a cluster of difficulties relating to emotional processing: difficulties identifying feelings, difficulties describing feelings and externally oriented thinking. Alexithymia is positively associated with several risk factors for offending behaviour and higher levels of alexithymia have been found in violent offenders. Consequently, it has become an area of interest in criminological research and there have been recommendations to consider alexithymia in risk assessments and interventions among violent offenders. This study sought to test and compare the psychometric properties of two measures of alexithymia in a forensic sample: the well-established 20-item Toronto Alexithymia Scale (TAS-20) and the recently developed Perth Alexithymia Questionnaire (PAQ). The internal consistency, convergent validity, and factorial validity of both the TAS-20 and PAQ were examined using a sample of 78 adult male prisoners in the UK. Both measures were found to assess a similar multifaceted alexithymia construct. However, the externally oriented thinking subscale of the TAS-20 had poor internal consistency and factor loadings. Conversely, the PAQ was found to have robust psychological properties across all subscales, while also offering the added benefit of valence-specific measurement. These findings provide preliminary support for the PAQ as a measure of alexithymia within both forensic practice and research.

2.2. Introduction

The term alexithymia was coined by psychiatrist and psychoanalyst Peter Sifneos in the 1970s (from Greek stems a = lack, lexis = word, and thymos = emotion*;* Sifneos, 1973). Its origin is rooted in psychosomatic medicine. Sifneos (1996) described how, during his time working with psychosomatic patients in a psychiatric clinic, he was struck by these patients' "marked difficulty in finding appropriate words to describe how they felt" (p. 137).

Alexithymia is not a psychiatric diagnosis, but rather a trans-diagnostic construct comprised of a cluster of difficulties relating to emotional processing (Preece et al., 2020a). While there has been continuing disagreement regarding the conceptualisation of alexithymia (Lane et al., 2015; Taylor et al., 2016; Taylor & Bagby, 2021), there is general consensus on three interrelated facets of the construct: a difficulty identifying feelings (DIF) and distinguishing between feelings; a difficulty describing feelings (DDF) to others; and an externally oriented thinking (EOT) style in which there is a lack of attention towards inner experience (Preece at al., 2017). While the original conceptualisation of alexithymia also included a fourth component - reduced fantasising and other imaginal activities (Nemiah et al., 1976) – a lack of psychometric support has led to suggestions that this is not a salient component of alexithymia (Sekely et al., 2018b; Watters et al., 2016a; Watters et al., 2016b).

2.2.1. Aetiology of alexithymia

While some authors consider alexithymia to be a stable personality trait (e.g. Luminet et al., 2007; Martínez-Sánchez, 2003; Parker et al., 2008; Salminen et al., 2006), other authors view it as a coping reaction to distress and, therefore, state-dependent (e.g. De Vente et al., 2006; Honkalampi et al., 2000). There is evidence to support both positions

(Schimmenti & Caretti, 2018). Alexithymia appears to be genetically influenced (Jørgensen et al., 2007; Kano et al., 2012, Picardi et al., 2011). Furthermore, adverse childhood experiences such as abuse and neglect appear to contribute to the development of high levels of alexithymia persisting into adulthood (Evren et al., 2009; Güleç et al., 2013). However, there is also evidence to suggest that high levels of alexithymia can occur later in life due to psychological stress (Yehuda et al., 1997; Zeitlin et al., 1993). Consequently, the contemporary view is that alexithymia is a multifaceted construct that includes both trait and state components (Karukivi & Saarijärvi, 2014; Lumley et al., 2007; Messina et al., 2014).

2.2.2. Attention-appraisal model of alexithymia

The attention-appraisal model of alexithymia (Preece at al., 2017) maps the three facets of alexithymia (DIF, DDF, EOT) onto the process model of emotion regulation (Gross, 2015). The attention-appraisal model theorises that there are four sequential stages of evaluation through which emotions can be understood and regulated: situation, attention, appraisal, and response. In stage one (situation), an emotional response becomes a stimulus. In stage two (attention), awareness is brought towards the emotion. In stage three (appraisal), the emotional response is assessed in terms of what it is and what it means. In stage four (response), an action may be taken to modify the emotion. EOT is conceptualised as a deficit at the attention stage, whereby there are difficulties focusing attention on the emotional response that has arisen. DIF and DDF are conceptualised as deficits at the appraisal stage, whereby there are difficulties accurately appraising the emotional response (Preece at al., 2017).

2.2.3. Assessing alexithymia

Toronto Alexithymia Scale-20. One of the early assessment measures that sought to operationalise the alexithymia construct was the Toronto Alexithymia Scale (Taylor et al., 1985), which was then revised to become the 20-item Toronto Alexithymia Scale (TAS-20; Bagby et al., 1994). The TAS-20 is a self-report measure which uses a five-point Likert scale ranging from 1 *(strongly disagree)* to 5 *(strongly agree).* Higher scores indicate higher levels of alexithymia. The TAS-20 comprises 20 items, five of which are negatively keyed and therefore reverse-scored. It has three subscales: Difficulty Identifying Feelings (DIF; e.g. "I am often confused about what emotion I am feeling"), Difficulty Describing Feelings (DDF; e.g. "it is difficult for me to find the right works for my feelings") and Externally Orientated Thinking (EOT; e.g. "I find examination of my feelings useful in solving personal problems"). While two items enquire about negatively valenced emotions (e.g. "I often don't know why I am angry"), all other items do not specify an emotional valence (e.g. "I am often puzzled by sensations in my body"). The TAS-20 is the most widely and frequently used measure in alexithymia research (Sekely et al., 2018a) and has been translated into over 25 different languages (Bagby et al., 2020).

The TAS-20 has been found to have good construct validity, with confirmatory factor analysis supporting the three-factor (DIF, DDF, EOT) model (Schroeders et al., 2021). The TAS-20 also demonstrated good internal reliability both in clinical and non-clinical samples (Bagby et al., 1994), although the Cronbach's alpha coefficients of the EOT subscale were below the generally accepted standard of ≥.70 (Nunnally & Bernstein, 1994). Low internal consistency of the EOT subscale has been found in many subsequent studies, both in the English and translated versions of the TAS-20 (Meganck et al., 2008; Taylor et al., 2003). This

appears to be, in part, due to response biases to the four negatively keyed items within this subscale. A method factor has been added in two factor analytic studies (Meganck et al., 2008; Preece et al., 2018) to test for method-specific variance associated within the negatively keyed items. Both studies found that the three-factor model with an additional method factor provided a better fit than the model without a method factor, raising questions about the effectiveness of the negatively keyed items (Sonderen et al., 2013).

Perth Alexithymia Scale. The Perth Alexithymia Questionnaire (PAQ; Preece et al., 2018) is a recently developed 24-item self-report measure, based on the attention-appraisal model of alexithymia. It uses a seven-point Likert scale ranging from 1 *(strongly disagree)* to 7 *(strongly agree),* again with higher scores indicating higher levels of alexithymia. It is freely available for use. In contrast to the TAS-20, it has no negatively keyed items. The PAQ was developed as a result of the inadequate internal consistency of the EOT subscale in the TAS-20, alongside an identified need to include valence-specific items to assess alexithymia across both positive and negative affectivity.

The PAQ has five subscales: Negative-Difficulty Identifying Feelings (N-DIF; e.g. "When I'm feeling *bad*, I can't tell whether I'm sad, angry or scared"); Positive-Difficulty Identifying Feelings (P-DIF; e.g. "When I'm feeling *good*, I can't make sense of those feelings"); Negative-Difficulty Describing Feelings (N-DDF; e.g. "When I'm feeling *bad*, I can't talk about those feelings in much depth or detail"); Positive-Difficulty Describing Feelings (P-DDF; e.g. "When I'm feeling *good*, if I try to describe how I'm feeling I don't know what to say"); and General-Externally Orientated Thinking Style (G-EOT; e.g. "I prefer to just let my feelings happen in the background, rather than focus on them").

The original PAQ development study (Preece at al., 2018) found good internal consistency for all five subscales (α≥.80) within an Australian community sample. Furthermore, support was found for the five-factor structure (N-DIF, P-DIF, N-DDF, P-DDF, G-EOT), which was superior to a three-factor structure which did not have valence-specific factors.

Preece et al. (2020a) compared the psychometric properties of the PAQ and the TAS-20 within a community sample in the United States. The PAQ and TAS-20 correlated highly with each other on both their total scores ($r = .76$, $p < .001$) and corresponding subscales ($r = .47$-$.74$, $p < .001$) suggesting they both assess a similar alexithymia construct. All subscales on the PAQ were found to have good internal consistency ($\alpha = .85$-$.88$). The subscales on the TAS-20 were more mixed, with the EOT subscale showing poor internal consistency (DIF $\alpha = .89$, DDF $\alpha = .77$, EOT $\alpha = .52$), leading Preece et al. (2020a) to conclude that the PAQ appears to provide a more comprehensive facet-level profile of alexithymia. In support of the initial findings by Preece et al. (2018), confirmatory factor analysis showed the PAQ five-factor model to be superior to the non-valenced three-factor model, suggesting that valence is an important consideration when assessing alexithymia. Participants generally reported more difficulties identifying and describing negatively valenced emotions.

While the psychometric properties of the TAS-20 have been broadly tested across both clinical and non-clinical samples within a variety of countries (Schroeders et al., 2021), the PAQ currently has a has a much more limited evidence base. Consequently, Preece et al. (2020a) recommend testing the replicability of their findings in other populations, including forensic samples (Preece at al., 2021).

2.2.4. Alexithymia and offending behaviour

Since it's conceptualisation, the scope of interest in alexithymia has extended far beyond psychosomatics. Alexithymia has now been found to be positively associated with dynamic risk factors for offending behaviour such as impulsivity (Shishido et al., 2013), aggression (Velotti at al., 2016), poor problem-solving skills (Christopher & McMurran, 2009), lack of empathy (Jonason & Krause, 2013) and emotional dysregulation (Roberton et al., 2014; Andrews & Bonta, 2010). It has, therefore, become an area of interest in criminological research.

While alexithymia does not appear to be a criminogenic factor for non-violent offenders (Parry et al., 2020), violent offenders have been found to score higher on the TAS-20 than community samples (Gillespie et al., 2018; Strickland et al., 2017). Furthermore, Howells & Day (2006) suggest that difficulties in accessing and expressing emotions may impact the effectiveness of insight-oriented therapeutic interventions for offenders. This is supported by research which has found that patients who score higher on measures of alexithymia tend to have poorer therapeutic outcomes (Ogrodniczuk et al., 2018). This evidence for alexithymia as both a potential moderator for risk factors for offending behaviour and a responsivity factor for potential interventions prompted Leshem et al. (2019) to recommend that "alexithymia should be considered in risk assessment and rehabilitation programmes among violent offenders" (p.100).

The extent to which alexithymia is modifiable parallels the debate about whether alexithymia is a trait- or state-dependent construct (Ogrodniczuk et al., 2018). Yet there is evidence to suggest that alexithymia can be modified with psychological interventions that specifically targets the construct (Cameron et al., 2014), leading Ogrodniczuk et al. (2011) to

propose that “alexithymic patients can at least partly develop some capacity to recognize their feelings and to communicate them to other people, thus enhancing their ability to use emotional information to guide adaptive behaviour” (p.46). This further emphasises the importance of assessing alexithymia within forensic settings so that support tailored to the needs of prisoners with high levels of alexithymia can be provided (Hemming et al., 2020).

The psychometric properties of the TAS-20 were recently examined in a prison sample (Preece et al., 2021). While the DIF and DDF subscales had adequate internal consistency, the EOT subscale had poor internal consistency. Preece et al. (2021) note that this limits the utility of the TAS-20 for facet-level assessment in forensic practice and research and recommend testing the psychometrics properties of the PAQ within a prison population.

2.2.5. Current study

The current study replicated the analytic strategy used in Preece et al. (2020a) for a community sample, to see if their results held true in a UK prison population. By examining the factorial validity, internal consistency, and convergent validity of the TAS-20 and PAQ with a prison sample, the study aimed to answer two questions. The first was whether the TAS-20 and PAQ are psychometrically robust measures for use within the prison population. The second was whether valence is an important consideration when assessing alexithymia within this population. If psychometrically robust, these self-report measures of alexithymia could be used to inform risk assessments and help guide and tailor interventions within forensic practice.

2.3. Methods

2.3.1. Participants, materials and procedure

The study sample consisted of 78 male prisoners within a UK prison, all of whom had completed the PAQ as part of a psychometric assessment pack upon arrival at the prison and 75 of whom had also completed the TAS-20. One of the participants had a missing item on the TAS-20 so only the full 74 datasets of the TAS-20 were used in certain analyses. At the start of the psychometric pack, a consent form explains that the information from assessments is collated by the psychology team and that the anonymous data may be used in confidence by the research and development unit or approved external researchers. All participants provided written consent for their data to be used for research purposes.

Demographic information of this sample can be seen in Table 2.1. The age and ethnicity distribution of the sample were reasonably representative of the UK prison population; however, the proportion of serious violent offences was markedly higher than the UK prison population overall (Sturge, 2021). The three most common offence categories were violence against the person (n = 48), sexual offences (n = 15) and robbery (n = 11).

Table 2.1. *Demographic and offence characteristics of participants.*

Demographic variables		**Sample**	
		n	**%**
Age bracket	21-30	27	34.6%
	31-40	22	28.2%
	41-50	13	16.7%
	51+	16	20.5%
Ethnicity	White	57	73.1%
	Black or Black British	10	12.8%
	Asian or Asian British	4	5.1%
	Mixed	6	7.7%
	Not recorded	1	1.3%
Offence category	Violence against the person	48	61.5%
	Sexual offences	15	19.2%
	Robbery	11	14.1%
	Burglary	2	2.6%
	Possession of weapons	2	2.6%

Ethical approval for this study was granted by the University of Birmingham's Science, Technology, Engineering and Mathematics Ethical Review Committee (see Appendix A).

2.3.2. Analytic strategy

Descriptive statistics and internal consistency. Cronbach's coefficients were calculated for both the TAS-20 and PAQ total scores and subscale scores, where values ≥.70 were judged as acceptable, ≥.80 as good, and ≥.90 as excellent (George & Mallery, 2019).

Comparisons of means between corresponding negatively and positively valenced subscales on the PAQ were also analysed to determine whether there was a significant difference in participants' scores in these areas.

Convergent validity. Pearson correlations were calculated to assess the association between the TAS-20 and PAQ total scores and subscale scores.

Factorial validity. A series of confirmatory factor analyses (CFA) were conducted using maximum likelihood estimation with robust standard errors. The factor structure of both the TAS-20 and PAQ were examined using the same models as Preece et al. (2020; see Figures 2.1 and 2.2). For the TAS-20, models included a single-factor model comprising all 20 items (Model 1), a two-factor model comprising the EOT subscale and the DIF and DDF subscales combined (Model 2), a three-factor model comprising the three intended subscales (Model 3) and the same three-factor model with the addition of a method factor for the reverse-scored items (Model 4). For the PAQ, models included a single-factor model (Model 5) comprising all 24 items, a two-factor model comprising the EOT subscale and a general-DIF/general-DDF factor which combined all the other four subscales (Model 6), a non-valenced three-factor model similar to that of the TAS-20 (Model 7), a valenced three-factor model which distinguished between negatively and positively valenced items but not DIF and DDF (Model 8), and a five-factor model which comprises the five intended subscales (Model 9).

The Satorra-Bentler scaled chi-squared test ($SB\chi^2$; Satorra & Bentler, 1994) was used to statistically assess the goodness-of-fit of these models as this has been found to be a robust estimator across different levels of non-normality, and model complexity (Brown, 2015; Curran et al., 1996). $SB\chi^2$ is used to test the null hypothesis of perfect model fit, i.e., that there is no difference between the covariance predicted by the candidate model and the sample data. A nonsignificant result for this test ($p \geq .05$) indicates good model fit. However, chi-squared significance tests are highly susceptible to sample size (Matsunaga, 2010). Furthermore, it has been argued that this perfect-fit hypothesis is unnecessarily stringent and unrealistic (e.g. Brown et al., 2015; Steiger, 2007). Consequently, approximate fit indices, which quantify the degree of fit of a model, were used to supplement $SB\chi^2$ (Hu &

Bentler, 1999). In line with the guidelines of Kline et al. (2015), the comparative fit index (CFI), standardised root mean square residual (SRMR) and root mean square error of approximation (RMSEA) indices were calculated. Hu & Bentler (1999) provide approximate cut-off values needed to conclude acceptable fit between a candidate model and the observed data when using maximum-likelihood estimation methods: CFI values close to ≥ .95, SRMR values close to ≤ .08 and RMSEA values close to ≤ .06.

The second-order 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 AIC_c penalises model complexity and, therefore, seeks to resolve the trade-off between model fit and complexity. A smaller AIC_c value indicates a more parsimonious model (Brown, 2015). Delta AIC_c (ΔAIC_c) is the difference in AICc value between a candidate model and the model with the lowest AIC_c value. When $\Delta \mathrm{AIC_c} => 10$, the candidate model is not supported (Burnham & Anderson, 2004).

Factor loadings ≥ .40 were considered to be meaningful (Kline, 2014).

Descriptive statistics, Cronbach's alpha coefficients and Pearson correlations were calculated using IBM SPSS Statistics, Version 28.0.1.0 (IBM Corp, 2021). The confirmatory factor analyses were performed using the Lavaan package (Rosseel, 2012) for R, Version 4.1.0.

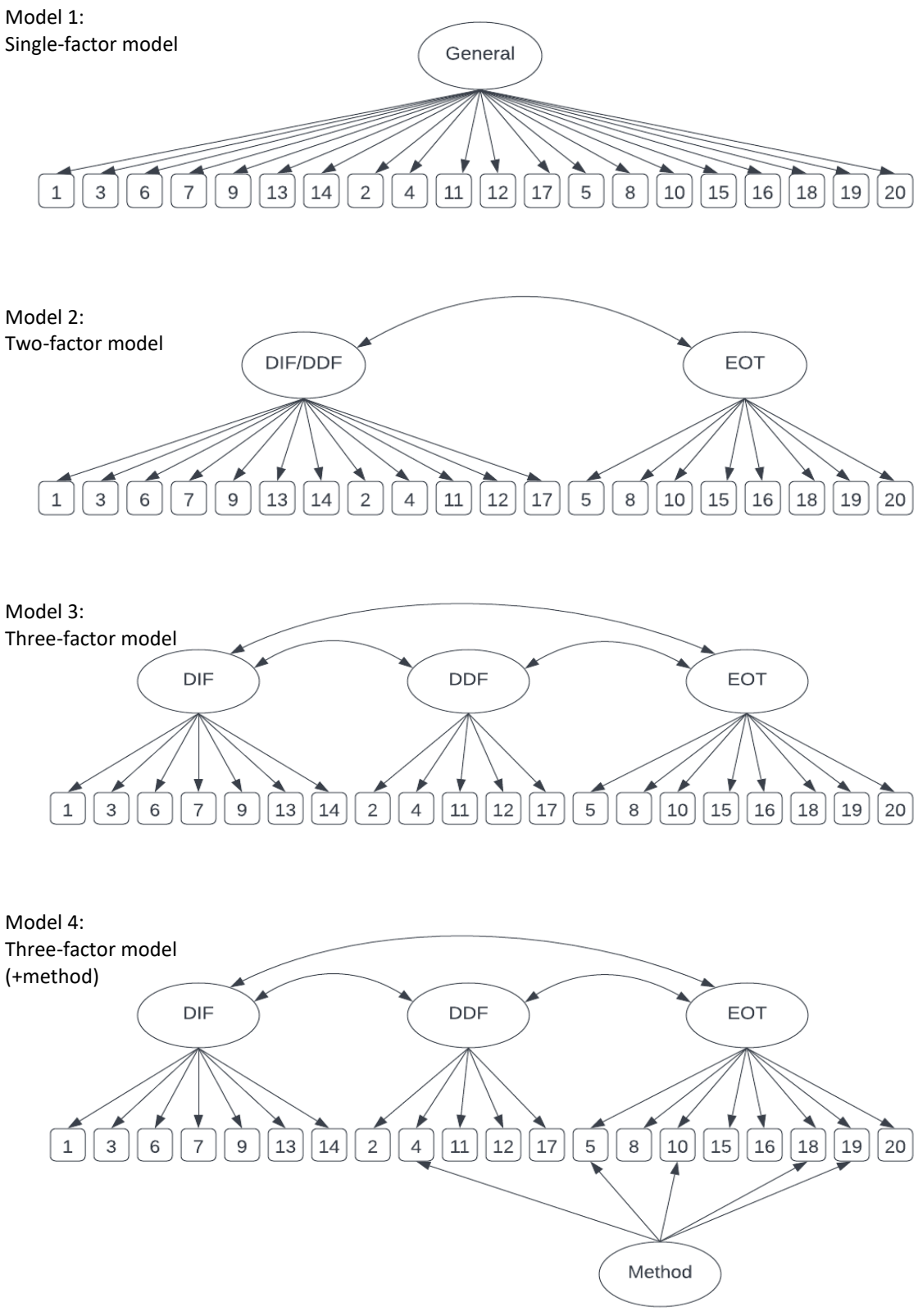
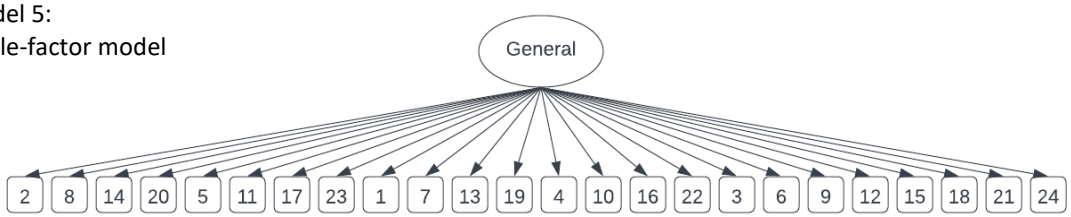
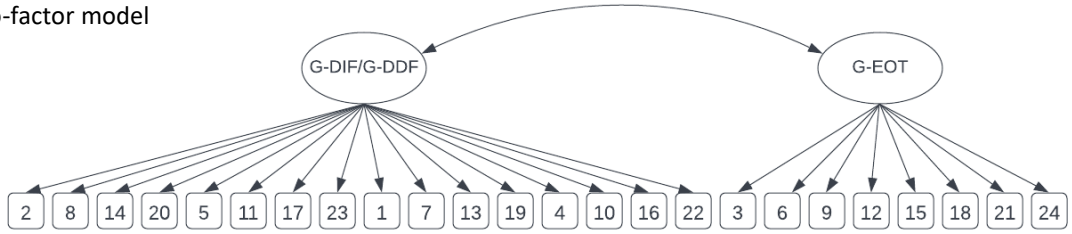


Figure 2.1. *TAS-20 confirmatory factor analysis models (taken from Preece et al., 2020a).*

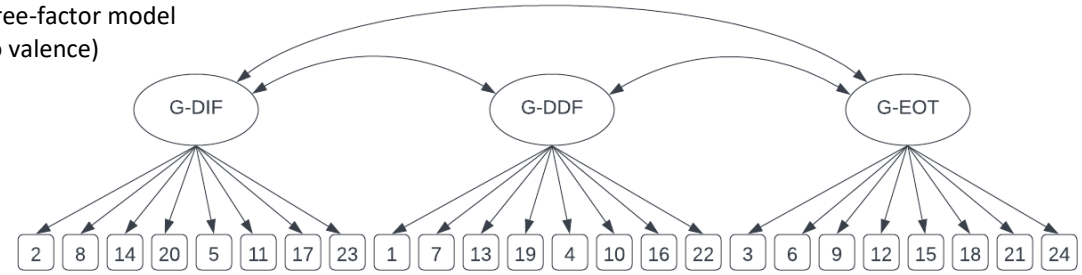
Model 5:
Single-factor model



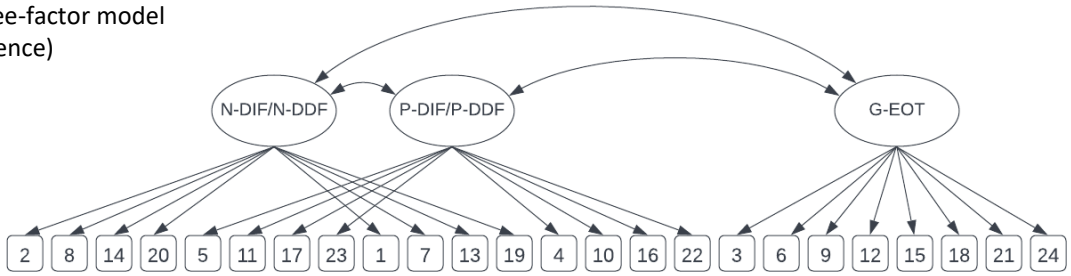
Model 6:
Two-factor model



Model 7:
Three-factor model
(no valence)



Model 8:
Three-factor model
(valence)



Model 9:
Five-factor model

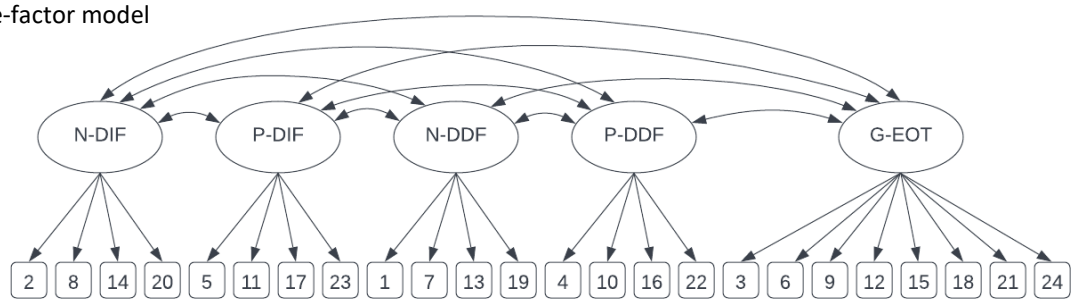


Figure 2.2. *PAQ confirmatory factor analysis models (taken from Preece et al., 2020a).*

2.4. Results

2.4.1. Descriptive statistics and internal consistency

Descriptive statistics and Cronbach's alpha coefficients for the TAS-20 and PAQ total scores and subscale scores are presented in Table 2.2. The TAS-20 total score (α = .86) and DIF (α = .89) and DDF subscales (α = .79) demonstrated adequate to good reliability. However, the TAS-20 EOT subscale demonstrated poor reliability (α = .38). The PAQ total score (α = .95) and all five subscales (α = .82-.91) demonstrated good to excellent reliability.

Table 2.2. *Descriptive statistics and Cronbach's α reliability coefficients for the TAS-20 and PAQ.*

Measure	Total sample TAS-20 (n=75) PAQ (n=78)			
	M	SD	Range	α
Toronto Alexithymia Scale-20				
Difficulty Identifying Feelings (DIF)	17.76	6.98	7-33	.89
Difficulty Describing Feelings (DDF)	15.24	4.55	5-24	.79
Externally Orientated Thinking (EOT)	20.80	3.90	8-28	.38
Total Scale	54.03	12.72	20-77	.86
Perth Alexithymia Questionnaire				
Negative-Difficulty Identifying Feelings (N-DIF)	13.95	6.14	4-28	.82
Positive- Difficulty Identifying Feelings (P-DIF)	12.23	5.89	4-27	.82
Negative-Difficulty Describing Feelings (N-DDF)	15.62	6.62	4-28	.83
Positive Difficulty Describing Feelings (P-DDF)	12.68	6.14	4-25	.86
General-Externally Orientated Thinking (G-EOT)	27.73	11.75	8-56	.91
Total Scale	82.21	31.56	24-161	.95

Normality was checked using the Shapiro-Wilk test. The results suggested significant deviation from the normality assumption in several of the subscales (TAS DIF, PAQ N-DIF, P-DIF, PAQ P-DDF). Consequently, the Wilcoxon Signed-Ranks Test was used to analyse differences in mean scores between the valence-specific subscales of the PAQ. Participants reported significantly more difficulties identifying negative emotions compared to positive

emotions (Z = -3.251, p = .001) and significantly more difficulties describing negative emotions compared to positive emotions (Z = -4.846, p < .001).

2.4.2. Convergent validity

Pearson correlations between the TAS-20 and PAQ total scores and subscale scores are presented in Table 2.3. The TAS-20 and PAQ total scores correlated strongly with each other (r = .70, p < .001). The TAS-20 DIF and DDF subscales correlated more strongly with the respective negatively valenced subscales of the PAQ (r = .61-.76, p < .001) than with the positively valenced subscales (r = .46-.63, p < .001).

Table 2.3. *Pearson bivariate correlations between the TAS-20 and PAQ total scores and subscale scores.*

		TAS-20				**PAQ**					
		DIF	DDF	EOT	Total	N-DIF	P-DIF	N-DDF	P-DDF	G-EOT	Total
TAS-20	DIF	1	.62**	.45**	.90**	.61**	.46**	.52**	.43**	.28*	.50**
	DDF		1	.46**	.83**	.72**	.60**	.76**	.63**	.59**	.76**
	EOT			1	.72**	.39**	.40**	.41**	.35**	.36**	.44**
	Total					.70**	.57**	.67**	.55**	.46**	.70**
PAQ	N-DIF					1	.77**	.83**	.77**	.60**	.87**
	P-DIF						1	.66**	.90**	.54**	.85**
	N-DDF							1	.72**	.61**	.86**
	P-DDF								1	.63**	.90**
	EOT									1	.84**
	Total										1

**. Correlation is significant at the .01 level (2-tailed).
*. Correlation is significant at the .05 level (2-tailed).

2.4.3. Factorial validity

Results from the series of CFAs are presented in Table 2.4. The TAS-20 factor structure was found to be best represented by the intended three-factor model ($SB\chi^2$ = 360.665(251), CFI = .820, SRMR = .106, RMSEA = .083, AIC_c = 4444.873), although these indices values did not meet the pre-defined thresholds for goodness-of-fit. Adding the

method factor to this three-factor model further improved the values of model fit (233.247 (159), CFI = .844, SRMR = .096, RMSEA = .079); however, the added complexity of this model meant that the AIC_c value (4459.121) was substantially higher than the three-factor model with no method factor ($\Delta AIC_c = 14.248$) and therefore considered to be a poorer fit to the data. Within the three-factor model, all DIF and DDF items loaded well on their intended subscales factor (.565-.847). However, the EOT items had poor factor loading, with five of the items < .40 (see Table 2.5).

The PAQ factor structure was found to be best represented by the valenced three-factor model. Although the $SB\chi^2$ statistic of this model was statistically significant ($SB\chi^2$ = 313.392 (249), $p < .05$) which suggested imperfect fit, values of CFI (.947), SRMR (.078) and RMSEA (.058) all indicated acceptable fit. While the five-factor model marginally improved the values of model fit ($SB\chi^2$ = 304.191 (242), CFI = .949, SRMR = .077, RMSEA = .057), the added complexity of the five-factor model meant that the AIC_c value (6700.179) was substantially higher than the valenced three-factor model ($\Delta AIC_c = 14.757$) and therefore considered to be a poorer fit to the data. All N-DIF/N-DDF, P-DIF/P-DDF and G-EOT items loaded well on their intended factors (.501-.904; see Table 2.5).

There was also very little difference in goodness-of-fit values between the PAQ two-factor model ($SB\chi^2$ = 360.665(251), CFA = .909, SRMR =.082, RMSEA = .081, RMSEA = .075) and non-valenced three-factor model ($SB\chi^2$ = 359.601 (249), CFA = .909, SRMR = .081; RMSEA = .075). Furthermore, the AIC_c value of the non-valenced three factor model (AIC_c = 6751.056) demonstrated a poorer fit than the two-factor model (AIC_c = 6745.929). These findings show that the DIF and DDF items of the PAQ effectively grouped onto the same

factor, suggesting that the DIF and DDF subscales may not be measuring two distinct facets of alexithymia.

Table 2.4. *Goodness-of-fit index values for the different confirmatory factor analysis models of the TAS-20 and PAQ.*

		Goodness-of-fit indices			Model comparisons	
Measure	$SB\chi^2(df)$	CFI	SRMR	RMSEA (90% CI)	AIC_c	ΔAIC_c
Toronto Alexithymia Scale 20						
1. Single-factor model (general factor)	272.287 (170)	.785	.105	.090 (0.071-0.108)	4466.182	21.309
2. Two-factor model (DIF/DDF, EOT)	271.863 (169)	.784	.106	.091 (0.072-0.109)	4468.902	24.029
3. Three-factor model (DIF, DDF, EOT)	252.748 (167)	.820	.106	.083 (0.063-0.102)	4444.873	0
4. Three factor model + method (DIF, DDF, EOT, method)	233.247 (159)	.844	.096	.079 (0.058-0.099)	4459.121	14.248
Perth Alexithymia Questionnaire						
5. Single factor model (general factor)	432.650 (252)	.851	.103	.096 (0.084-0.108)	6887.331	201.906
6. Two factor model (G-DIF/G-DDF, EOT)	360.665 (251)	.909	.082	.075 (0.061-0.088)	6745.929	60.505
7. Three-factor model - no valence (G-DIF, G-DDF, EOT)	359.601 (249)	.909	.081	.075 (0.061-0.089)	6751.056	65.634
8. Three-factor model – valence (N-DIF/N-DDF, P-DIF/P-DDF, G-EOT)	313.392 (249)	.947	.078	.058 (0.040-0.073)	6685.424	0
9. Five factor model (N-DIF, N-DDF, P-DIF, P-DDF, G-EOT)	304.191 (242)	.949	.077	.057 (0.039-0.073)	6700.179	14.757

$SB\chi^2$ – Satorra-Bentler scaled chi-squared test. All $SB\chi^2$ $p < .05$
CFI - Comparative fit index – value close to ≥ .95 indicates acceptable fit; SRMR - Standardised root mean square residual – value close to ≤ .08 indicates acceptable fit; RMSEA (90% CI) - Root mean square error of approximation – value close to ≤ .06 indicates acceptable fit.
AIC_c – Second-order Akaike information Criteria. The AIC_c penalises for model complexity and lower values indicate a more parsimonious model; ΔAIC_c – Delta AIC_c – difference between the AIC_c with the minimum value and the candidate model. $\Delta AIC_c > 10$ indicates there is no support for the candidate model.

Table 2.5. *Standardised factor loadings from confirmatory analyses of the TAS-20 (three factor model) and PAQ (valenced three-factor model).*

Item	Toronto Alexithymia Scale-20			Item	Perth Alexithymia Questionnaire		
	F1	F2	F3		F1	F2	F3
	DIF	DDF	EOT		N-DIF/N-DDF	P-DIF/P-DDF	G-EOT
1	**.759**	-	-	1	**.760**	-	-
2	-	**.847**	-	2	**.681**	-	-
3	**.603**	-	-	3	-	-	**.705**
4	-	**.681**	-	4	-	**.669**	-
5	-	-	.280	5	-	**.501**	-
6	**.797**	-	-	6	-	-	**.80**
7	**.710**	-	-	7	**.576**	-	-
8	-	-	**-.401**	8	**.677**	-	-
9	**.778**	-	-	9	-	-	**.809**
10	-	-	-.283	10	-	**.702**	-
11	-	**.593**	-	11	-	**.777**	-
12	-	**.565**	-	12	-	-	**.753**
13	**.775**	-	-	13	**.781**	-	-
14	**.669**	-	-	14	**.834**	-	-
15	-	-	**-.536**	15	-	-	**.807**
16	-	-	**-.403**	16	-	**.904**	-
17	-	**.587**	-	17	-	**.892**	-
18	-	-	-.084	18	-	**-**	**.761**
19	-	-	-.020	19	**.819**	-	-
20	-	-	-.393	20	**.748**	-	-
				21	-	-	**.621**
				22	-	**.886**	-
				23	-	**.813**	-
				24	-	-	**.711**

Note. Values in bold meet the > .40 threshold.

2.5. Discussion

The aims of this study were to determine whether the TAS-20 and PAQ are psychometrically robust measures for use within forensic practice and whether valence is an important consideration when assessing alexithymia within a prison population.

The high convergent validity between the TAS-20 and PAQ suggests these measures assess a similar alexithymia construct. However, the psychometric properties of the PAQ were found to be superior to the TAS-20. In line with previous research (Bagsby et al., 1994; Meganck et al., 2008; Preece at al., 2020; Taylor et al., 2003), the EOT subscale of the TAS-20 had poor internal consistency and item factor loadings, demonstrating inconsistency in relationship between the items in this subscale. In contrast, the G-EOT subscale of the PAQ had good internal consistency and factor loadings. These findings add further support to the suggestion by Preece et al. (2020a) that the PAQ provides a more comprehensive profile of alexithymia than the TAS-20, which may be particuarly useful for researchers and clinicians interested in assessing alexithymia at a facet-level. However, this is somewhat obfuscated by the observation that there was no meaningful difference in goodness-of-fit between models of the PAQ which combined the DIF and DDF subscales and models which separated these two subscales. The attention-appraisal model of alexithymia (Preece et al., 2017) theorises that both DIF and DDF are the result of deficits in the appraisal stage of emotional processing and so significant overlap in participant scores in these subscales would be expected. Nevertheless, more research is needed to determine whether the PAQ effectively distinguishes between these two facets of alexithymia.

The valenced three-factor model of the PAQ, which distinguished between positively and negatively valenced items, performed better in the confirmatory factor analysis than

the more traditional three-factor model which distinguished between the three facets of alexithymia (DIF, DDF, EOT). This supports the conclusion by Preece et al. (2020a) that valence is an important consideration when assessing alexithymia. Previous research has found that there are different neural correlates of alexithymia depending on whether emotions being processed are positive or negative (van der Velde et al., 2013). Notably, recent measures of other affective phenomena such as emotion regulation and reactivity have sought to account for both positive and negative emotions (e.g. Becerra et al., 2019; Weiss et al., 2015; Zou et al., 2019).

While the importance of valence in the context of alexithymia requires further investigation, distinguishing between emotional valence in assessment could allow a more nuanced profile of a person's difficulties and, in turn, help to guide psychological interventions (Taylor & Bagby, 2021). For example, individuals who score highly on negatively-valenced scales of the PAQ may benefit from interventions focused primarily on building the recognition and communication of negative emotions to help guide adaptive behaviour (Ogrodniczuk et al., 2018). In the case of rehabilitation programmes for violent offenders, seeking to enhance an individual's capacity to attend to emotions such as anger may help reduce risk of violent recidivism (Garaigordobil & Peña-Sarrionandia, 2015; Moroń & Biolik-Moroń, 2021; Roberton et al., 2015). When clients score highly on positively-valenced scales, this may be an indication for therapists to pay particuarly close attention to the therapist-client relationship, as there is evidence that a lack of expression of positive emotions by clients with high levels of alexithymia elicits negative reactions in therapists (Ogrodniczuk et al., 2008), which may in turn increase the likelihood of poorer therapeutic outcomes (Ogrodniczuk et al., 2005).

2.5.1. Limitations

To the author's knowledge, this study is the first to test the psychometric properties of the PAQ within a forensic setting and offers initial support for the use of the PAQ within forensic practice and research. However, some limitations to the study should be noted.

Firstly, the sample used in this study is small for CFA. Many rules of thumb have been offered regarding minimum sample size for CFA investigations, such as a sample size of 100 or above (Boomsma, 1985), 5-10 cases per estimated parameter (Betler & Chou, 1987) or 10 cases per variable (Nunnally & Bernstein, 1994). The sample in this study falls short of these thresholds. A small sample size increases the likelihood of sampling error, whereby the data collected is not representative of the larger population. The sample of 78 was determined by the number of men who had completed the PAQ on arrival at the prison at the point of data collection and so it was not possible to increase the sample size to fit with guidelines. However, appropriate statistical methods were employed to correct for small sample size and the results from this study generally support those found in previous studies with larger sample sizes (Preece et al., 2018, 2020, 2021).

Secondly, the sample was treated as a general forensic sample, rather than separated into violent and non-violent offending groups as has been done in previous alexithymia research in the forensic field (Gillespie et al., 2018; Parry et al., 2020; Strickland et al., 2017). Grouping the sample by offence type was not feasible due to the limited sample size. Future studies testing the replicability of this study's findings in larger forensic samples and in different categories of offences would be beneficial.

Thirdly, this study did not examine test-retest reliability of the TAS-20 and PAQ. While the original development also did not examine test-retest reliability (Preece et al.,

2018), two recent studies testing the psychometric properties of the Farsi and Iranian versions of the PAQ both found the PAQ and all five subscales had good test-retest reliability (Asl et al., 2020; Lashkari et al., 2021).

Finally, some authors have questioned the validity of self-report measures of alexithymia due to the difficulties with introspection inherent in individuals with high levels of alexithymia (e.g. Lane et al., 2015; Leising et al., 2009; Marchesi et al., 2014). Furthermore, the TAS-20 appears to be is sensitive to general distress and, therefore, has been proposed to be more a measure of general negative affect than alexithymia (Leising et al., 2009; Marchesi et al., 2014). Only one study to date (Preece et al., 2020b) has compared participant scores on the PAQ and a measure of general psychological distress and it was found that the PAQ measures an alexithymia construct that is distinct from psychological distress. Nevertheless, more research is required to determine whether alexithymia, as measured by the TAS-20 and PAQ, is a sufficiently distinct psychological construct. An informant form of the TAS-20 has recently been developed (Bagby et al., 2021) and future research into self-informant agreement between these measures would help to evaluate the validity of self-report measures of alexithymia in more detail. At present, a multi-method approach is recommended to ensure an effective assessment of alexithymia (Montebarocci & Surcinelli, 2018; Taylor et al., 2016).

2.5.2. Implications for forensic practice and future research

Although both conceptual and interpretative limitations remain within the alexithymia construct (Lumley et al., 2007), the higher levels of alexithymia found in violent offenders (Gillespie et al., 2018; Strickland et al., 2017), the association between alexithymia and several risk factors for violent offending (Lesham et al., 2019) and the association

between alexithymia and poorer therapeutic outcomes (Ogrodniczuk et al., 2018) all suggest that assessments of alexithymia could be used to inform risk assessments and help to guide and tailor interventions among violent offenders (Lesham et al., 2019).

Recent literature on offender rehabilitation has highlighted a shift towards an individualised approach (Sicard & Birch, 2020). As Lesham et al. (2019) note, tailoring offender rehabilitation programmes for those with high levels of alexithymia may increase their effectiveness, thereby further promoting desistance and improving reintegration into society. The results of this study suggest that, when assessing alexithymia in forensic practice, the PAQ can confidently be used, both to obtain overall alexithymia scores and to conduct more in-depth facet-level assessments.

While there continues to be a growing research interest in alexithymia within the forensic field, there are still a number of areas where further research is needed. For example, the author could only find one study which explored whether alexithymia is a predictor of recidivism (Romero Martínez & Moya-Albiol, 2019). This study found that perpetrators of intimate partner violence with higher levels of alexithymia were both at higher risk of dropout and recidivism during the early stages of intervention. However, the study was correlational and did not use official reoffending rates. Higher quality, longitudinal studies are required. Furthermore, while there are promising indications that alexithymia can be modified through therapy (Cameron et al., 2014), more research is required to both develop and evaluate the effectiveness of interventions that aim to reduce levels of alexithymia (Ogrodniczuk et al., 2018; Taylor & Bagby, 2012).

2.5.3. Conclusion

In summary, this study demonstrates that, when used in a prison population, the PAQ appears to be a more psychometrically robust measure of alexithymia than the TAS-20, particuarly when comparing the two measures at facet-level. The PAQ also offers the added potential benefit of valence-specific measurement. Use of the TAS-20 EOT subscale in both forensic practice and research should be avoided. The PAQ is a brief and free resource which could be used as a measure of alexithymia within prison settings, ideally as part of a multi-method assessment.

2.6. References

Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.

Asl, E. M., Mahaki, B., Khanjani, S., & Mohammadian, Y. (2020). The assessment of alexithymia across positive and negative emotions: The psychometric properties of the Iranian version of the Perth alexithymia questionnaire. *Iran Journal of Psychiatry and Behavioural Science*, *14*(4), e102317.

Bagby, R. M., Parker, J. D., & 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.

Bagby, R. M., Parker, J. D., & Taylor, G. J. (2020). Twenty-five years with the 20-item Toronto Alexithymia Scale. *Journal of Psychosomatic Research*, *131*, 109940.

Bagby, R. M., Parker, J. D., Onno, K. A., Mortezaei, A., & Taylor, G. J. (2021). Development and psychometric evaluation of an informant form of the 20-item Toronto alexithymia scale. *Journal of Psychosomatic Research*, *141*, 110329.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, *15*(6), 713-724.

Becerra, R., Preece, D., Campitelli, G., & Scott-Pillow, G. (2019). The assessment of emotional reactivity across negative and positive emotions: Development and validation of the Perth Emotional Reactivity Scale (PERS). *Assessment*, *26*(5), 867-879.

Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, *16*(1), 78-117.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, *50*(2), 229-242.

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

Cameron, K., Ogrodniczuk, J., & Hadjipavlou, G. (2014). Changes in alexithymia following psychological intervention: a review. *Harvard Review of Psychiatry*, *22*(3), 162-178.

Christopher, G., & McMurran, M. (2009). Alexithymia, empathic concern, goal management, and social problem solving in adult male prisoners. Psychology, Crime & Law, 15(8), 697-709.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16-29.

De Vente, W., Kamphuis, J. H., & Emmelkamp, P. M. G. (2006). Alexithymia, risk factor or consequence of work-related stress?. *Psychotherapy and Psychosomatics*, *75*(5), 304-311.

Evren, C., Evren, B., Dalbudak, E., Ozcelik, B., & Oncu, F. (2009). Childhood abuse and neglect as a risk factor for alexithymia in adult male substance dependent inpatients. *Journal of Psychoactive Drugs*, *41*(1), 85-92.

Garaigordobil, M., & Peña-Sarrionandia, A. (2015). Effects of an emotional intelligence program in variables related to the prevention of violence. *Frontiers in Psychology*, *6*, 743-243.

George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

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.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1-26.

Güleç, M. Y., Altintaş, M., İnanç, L., Bezgin, Ç. H., Koca, E. K., & Güleç, H. (2013). Effects of childhood trauma on somatization in major depressive disorder: The role of alexithymia. *Journal of Affective Disorders*, *146*(1), 137-141.

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. *Frontiers in psychiatry*, 1385.

Honkalampi, K., Hintikka, J., Saarinen, P., Lehtonen, J., & Viinamäki, H. (2000). Is alexithymia a permanent feature in depressed patients?. *Psychotherapy and Psychosomatics*, *69*(6), 303-308.

Howells, K., & Day, A. (2006). Affective determinants of treatment engagement in violent offenders. *International Journal of Offender Therapy and Comparative Criminology*, *50*(2), 174-186.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, *6*(1), 1-55.

IBM Corp. (2021). IBM SPSS Statistics for Windows, Version 28.0. IBM Corp.

Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, *55*(5), 532-537.

Jørgensen, M. M., Zachariae, R., Skytthe, A., & Kyvik, K. (2007). Genetic and environmental factors in alexithymia: a population-based study of 8,785 Danish twin pairs. *Psychotherapy and Psychosomatics*, *76*(6), 369-375.

Kano, M., Mizuno, T., Kawano, Y., Aoki, M., Kanazawa, M., & Fukudo, S. (2012). Serotonin transporter gene promoter polymorphism and alexithymia. *Neuropsychobiology*, *65*(2), 76-82.

Karukivi, M., & Saarijärvi, S. (2014). Development of alexithymic personality features. *World Journal of Psychiatry*, *4*(4), 91-102.

Kline, P. (2014). *An Easy Guide to Factor Analysis*. Routledge.

Kline, R. B. (2015). *Principles and practice of structural equation modelling* (4th ed.). Guilford Press.

Lane, R. D., Weihs, K. L., Herring, A., Hishaw, A., & Smith, R. (2015). Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. *Neuroscience & Biobehavioral Reviews*, *55*, 594-611.

Lashkari, A., Dehghani, M., Sadeghi-Firoozabadi, V., Heidari, M., & Khatibi, A. (2021). Further Support for the Psychometric Properties of the Farsi Version of Perth Alexithymia Questionnaire. *Frontiers in psychology*, *12*, 657660.

Leising, D., Grande, T., & Faber, R. (2009). The Toronto Alexithymia Scale (TAS-20): A measure of general psychological distress. *Journal of Research in Personality*, *43*(4), 707-710.

Leshem, R., van Lieshout, P. H., Ben-David, S., & Ben-David, B. M. (2019). Does emotion matter? The role of alexithymia in violent recidivism: A systematic literature review. *Criminal Behaviour and Mental Health*, *29*(2), 94-110.

Luminet, O., Rokbani, L., Ogez, D., & Jadoulle, V. (2007). An evaluation of the absolute and relative stability of alexithymia in women with breast cancer. *Journal of Psychosomatic Research*, *62*(6), 641-648.

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: implications for understanding and treating health problems. *Journal of Personality Assessment*, *89*(3), 230-246.

Marchesi, C., Ossola, P., Tonna, M., & De Panfilis, C. (2014). The TAS-20 more likely measures negative affects rather than alexithymia itself in patients with major depression, panic disorder, eating disorders and substance use disorders. *Comprehensive Psychiatry*, *55*(4), 972-978.

Martínez-Sánchez, F., Ato-García, M., & Ortiz-Soria, B. (2003). Alexithymia—state or trait?. *The Spanish Journal of Psychology*, *6*(1), 51-59.

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International Journal of Psychological Research*, *3*(1), 97-110.

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.

Messina, A., Beadle, J. N., & Paradiso, S. (2014). Towards a classification of alexithymia: primary, secondary and organic. *Journal of Psychopathology*, *20*, 38-49.

Montebarocci, O., & Surcinelli, P. (2018). Correlations between TSIA and TAS-20 and their relation to self-reported negative affect: a study using a multi-method approach in the assessment of alexithymia in a nonclinical sample from Italy. *Psychiatry Research*, *270*, 187-193.

Moroń, M., & Biolik-Moroń, M. (2021). Emotional awareness and relational aggression: The roles of anger and maladaptive anger regulation. *Personality and Individual Differences*, *173*, 110646.

Nemiah, J. C., Freyberger, H., & Sifneos, P. E. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.), *Modern trends in psychosomatic medicine* (Vol. 3, pp. 430–439). Butterworths.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

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?. *Comprehensive Psychiatry*, *46*(3), 206-213.

Ogrodniczuk, J. S., Piper, W. E., & Joyce, A. S. (2008). Alexithymia and therapist reactions to the patient: Expression of positive emotion as a mediator. *Psychiatry: Interpersonal and Biological Processes*, *71*(3), 257-265.

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.

Ogrodniczuk, J., Kealy, D., Hadjipavlou, G., & Cameron, K. (2018). Therapeutic Issues. In O. Luminet, R. Bagby, & G. Taylor (Eds.), *Alexithymia: Advances in Research, Theory, and Clinical Practice* (pp. 190-206). Cambridge University Press.

Parker, J. D., Keefer, K. V., Taylor, G. J., & Bagby, R. M. (2008). Latent structure of the alexithymia construct: a taxometric investigation. *Psychological Assessment*, *20*(4), 385-396.

Parry, C. L., Preece, D. A., Allan, M. M., & Allan, A. (2020). Alexithymia in nonviolent offenders. *Criminal Behaviour and Mental Health*, *31*(1), 44-48.

Picardi, A., Fagnani, C., Gigantesco, A., Toccaceli, V., Lega, I., & Stazi, M. A. (2011). Genetic influences on alexithymia and their relationship with depressive symptoms. *Journal of Psychosomatic Research*, *71*(4), 256-263.

Preece, D. A., Becerra, R., Allan, A., Robinson, K., Chen, W., Hasking, P., & Gross, J. J. (2020). Assessing alexithymia: psychometric properties of the Perth Alexithymia Questionnaire and 20-item Toronto Alexithymia Scale in United States adults. *Personality and Individual Differences*, *166*, 110138.

Preece, D. A., Becerra, R., Boyes, M. E., Northcott, C., McGillivray, L., & Hasking, P. A. (2020). Do self-report measures of alexithymia measure alexithymia or general psychological distress? A factor analytic examination across five samples. *Personality and Individual Differences*, *155*, 109721.

Preece, D. A., Parry, C. L., Allan, M. M., & Allan, A. (2021). Assessing alexithymia in forensic settings: Psychometric properties of the 20-item Toronto Alexithymia Scale among incarcerated adult offenders. *Criminal Behaviour and Mental Health*, *31*(1), 31-43.

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.

Preece, D., Becerra, R., Robinson, K., Dandy, J., & Allan, A. (2018). The psychometric assessment of alexithymia: Development and validation of the Perth Alexithymia Questionnaire. *Personality and Individual Differences*, *132*, 32-44.

Roberton, T., Daffern, M., & Bucks, R. S. (2014). Maladaptive emotion regulation and aggression in adult offenders. *Psychology, Crime & Law*, *20*(10), 933-954.

Roberton, T., Daffern, M., & Bucks, R. S. (2015). Beyond anger control: Difficulty attending to emotions also predicts aggression in offenders. *Psychology of Violence*, *5*(1), 74.

Romero Martínez, Á., Lila, M., & Moya-Albiol, L. (2019). The importance of considering alexithymia during initial stages of intimate partner violence interventions to design adjuvant treatments. *International Journal of Environmental Research and Public Health*, *16*(19), 3695.

Rosseel, Y. (2012). Lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36.

Salminen, J. K., Saarijärvi, S., Toikka, T., Kauhanen, J., & Äärelä, E. (2006). Alexithymia behaves as a personality trait over a 5-year period in Finnish general population. *Journal of psychosomatic research*, *61*(2), 275-278.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243-248.

Schimmenti, A. & Caretti, V. (2018). Attachment, trauma, and alexithymia. In O. Luminet, R. M. Bagby & G. J. Taylor (Eds.), *Alexithymia: Advances in research, theory, and clinical practice* (p. 127–141). Cambridge University Press.

Schroeders, U., Kubera, F., & Gnambs, T. (2021). The structure of the Toronto Alexithymia Scale (TAS-20): A meta-analytic confirmatory factor analysis. *Assessment*, 10731911211033894.

Sekely, A., Bagby, R. M., & Porcelli, P. (2018a). Assessment of the alexithymia construct. In O. Luminet, R. M. Bagby & G. J. Taylor (Eds.), *Alexithymia: Advances in research, theory, and clinical practice* (p. 127–141). Cambridge University Press.

Sekely, A., Taylor, G. J., & Bagby, R. M. (2018b). Developing a short version of the Toronto Structured Interview for Alexithymia using item response theory. *Psychiatry Research*, *266*, 218-227.

Shishido, H., Gaher, R. M., & Simons, J. S. (2013). I don't know how I feel, therefore I act: alexithymia, urgency, and alcohol problems. *Addictive Behaviors*, *38*(4), 2014-2017.

Sicard, L. A., & Birch, P. (2020). Adopting a 'creative corrections' approach for offender treatment programs. In P. Birch & L Sicard (Eds.), *Prisons and community corrections: critical issues and emerging controversies* (pp. 203-215). Routledge.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics*, *22*(2-6), 255-262.

Sifneos, P. E. (1996). Alexithymia: Past and present. *The American Journal of Psychiatry, 153*, 137–142.

Sonderen, E. V., 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.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual differences*, *42*(5), 893-898.

Strickland, J., Parry, C. L., Allan, M. M., & Allan, A. (2017). Alexithymia among perpetrators of violent offences in Australia: Implications for rehabilitation. *Australian Psychologist*, *52*(3), 230-237.

Sturge, G. (2021). Prison population statistics. *House of Commons Library*. Retrieved from https://commonslibrary.parliament.uk/research-briefings/sn04334/

Taylor, G. J., & Bagby, R. M. (2012). The alexithymia personality dimension. In T. A. Widiger (Ed.), *The Oxford handbook of personality disorders* (pp. 648–673). Oxford University Press.

Taylor, G. J., & Bagby, R. M. (2021). Examining proposed changes to the conceptualization of the alexithymia construct: the way forward tilts to the past. *Psychotherapy and Psychosomatics*, *90*(3), 145-155.

Taylor, G. J., Bagby, R. M., & Parker, J. D. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, *55*(3), 277-283.

Taylor, G. J., Bagby, R. M., & Parker, J. D. (2016). What's in the name 'alexithymia'? A commentary on "Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy". *Neuroscience & Biobehavioral Reviews*, *68*, 1006-1020.

Taylor, G. J., Ryan, D., & Bagby, M. (1985). Toward the development of a new self-report alexithymia scale. *Psychotherapy and Psychosomatics*, *44*(4), 191-199.

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 and Biobehavioral Reviews, 37*, 1774–1785.

Velotti, P., Garofalo, C., Petrocchi, C., Cavallo, F., Popolo, R., & Dimaggio, G. (2016). Alexithymia, emotion dysregulation, impulsivity and aggression: A multiple mediation model. *Psychiatry Research*, *237*, 296-303.

Watters, C. A., Taylor, G. J., & Bagby, R. M. (2016a). Illuminating the theoretical components of alexithymia using bifactor modeling and network analysis. *Psychological Assessment*, *28*(6), 627-638.

Watters, C. A., Taylor, G. J., Quilty, L. C., & Bagby, R. M. (2016b). An examination of the topology and measurement of the alexithymia construct using network analysis. *Journal of Personality Assessment*, *98*(6), 649-659.

Weiss, N. H., Gratz, K. L., & Lavender, J. M. (2015). Factor structure and initial validation of a multidimensional measure of difficulties in the regulation of positive emotions: The DERS-Positive. *Behavior Modification*, *39*(3), 431-453.

Yehuda, R., Steiner, A., Kahana, B., Binder-Brynes, K., Southwick, S. M., Zemelman, S., & Giller, E. L. (1997). Alexithymia in Holocaust survivors with and without PTSD. *Journal of Traumatic Stress*, *10*(1), 93-100.

Zeitlin, S. B., McNally, R. J., & Cassiday, K. L. (1993). Alexithymia in victims of sexual assault: an effect of repeated traumatization?. *The American Journal of Psychiatry*, *150*(4), 661-663.

Zou, C., Plaks, J. E., & Peterson, J. B. (2019). Don't get too excited: Assessing individual differences in the down-regulation of positive emotions. *Journal of Personality Assessment*, *101*(1), 73-83.

Chapter 3: Press release for the literature review

In some jobs, exposure to potentially traumatic events is inevitable and with this comes an increased risk of developing post-traumatic stress disorder. For example, studies have found that up to 2 in every 10 paramedics have had PTSD and up to 2 in 10 military personnel have PTSD upon returning from deployment. Organisations need to support their employees following exposure to traumatic events, but the best way to do this is still unclear.

'Psychological debriefing' is an early post-trauma intervention which aims to prevent or reduce symptoms of PTSD. It was originally designed for groups of emergency service workers in the 1980s and involves discussing and normalising group members responses to trauma, usually over a single session. Psychological debriefing has remained a controversial intervention since a review in 2002 found it to be ineffective, and possibly even harmful. But some researchers have noted that many of the studies in this review evaluated interventions that were offered to individuals exposed to traumas unrelated to work, such as burns, road traffic accidents and miscarriage. They have argued that debriefing may have been dismissed too early and could still be effective when offered to groups of workers and conducted in line with the way debriefings were originally designed.

Authors from the University of Birmingham merged the findings of 17 studies which had compared the PTSD symptoms of workers who received psychological debriefing following a work-related trauma to those who had not received psychological debriefing. While the overall analysis found no conclusive evidence that psychological debriefing was effective at preventing or reducing PTSD symptoms, four of the studies found psychological

debriefing to be significantly more effective than no intervention and only one study found it to be harmful.

The lead author of the study said: “While the results of this review were mixed, the findings suggests that previous assertions that single-session psychological debriefing are, at best, ineffective, may have been premature. In recent years, less research has been conducted on the effectiveness psychological debriefing, which is likely to be due to the widespread belief that the intervention is harmful. Our review highlights that further research in this area is warranted”.

The authors had also planned to evaluate the impact that the length of time between a trauma and debriefing had on the effectiveness of the intervention. However, poor reporting in many of the studies meant that this was not possible. The authors emphasise that future research needs to clearly and accurately report the nature of the intervention being evaluated.

The authors acknowledge that some of the studies included in the review were of poor quality, but limiting the meta-analysis to only randomised controlled trials would have meant that most of the research in the field was excluded.

“Trauma research in often of poor quality. By its very nature, trauma usually occurs in chaotic and unpredictable circumstances and so researchers must often work opportunistically. It is also unethical to withhold interventions from people who may want or benefit from them and so establishing suitable control groups for research is challenging”.

The authors conclude that this review shows that further well-designed research into psychological debriefing for work-related trauma is needed. This will help to ensure that

employees receive the support they need and deserve following exposure to trauma. For now, however, it looks like that the debriefing debate will continue.

Chapter 4: Press release for the empirical paper

The term 'alexithymia' was coined in the 1970s and literally translates as "no words for emotions". People with high levels of alexithymia have difficulties identifying their emotions, difficulties describing their emotions to others and an externally orientated thinking style in which they don't attend to their inner feelings.

Prisoners are likely to attend offending behaviour programmes or therapeutic interventions. These require prisoners to access and express their feelings. It is therefore important that any difficulties prisoners may have in doing this are assessed so that they can be best supported. These difficulties seem to be more common in violent offenders, who have been found to have higher levels of alexithymia than the general population.

Authors from the University of Birmingham examined the validity and reliability of the Perth Alexithymia Questionnaire (PAQ), a new measure of alexithymia, using a sample of 78 prisoners. Validity refers to whether a questionnaire is measuring what it claims to measure. Reliability refers to how consistent the results of a questionnaire are. Recent research using samples from the general population have found that the PAQ is a reliable measure across all the aspects of alexithymia. However, no study has previously tested the validity and reliability of the PAQ in prisoners.

The study found that the PAQ was a valid and reliable measure of alexithymia amongst the sample of prisoners. The lead author of the study said: "There's promising research to suggest that levels of alexithymia can be reduced through psychological interventions. Effectively assessing for alexithymia means that offending behaviour programmes and other interventions could then be tailored for people with high levels of alexithymia. For example, in the case of rehabilitation programmes for violent offenders,

interventions to help people attend to emotions such as anger might help to reduce risk of future violence."

The authors note that the sample was small for the type of research being conducted and that larger samples should be used in future research. However, this study provides promising preliminary support for using the PAQ in assessments of alexithymia within prisons.

Appendices

Appendix A. Letter from ethics committee granting ethical approval for the research

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

BLACK SWAN
PSYCHOLOGICAL ASSESSMENT & THERAPY

Appendix C. 20-item Toronto Alexithymia Scale

A copy of the 20-item Toronto Alexithymia Scale has not been included within this appendix due to copyright restrictions. For a list of the scale's items, please read the original development study referenced below.

Bagby, R. M., Parker, J. D., & 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.