The relationship between challenge and threat states and performance: A systematic review

Adrian Hase, MSc[a]

Jessica O'Brien[b]

Lee J. Moore, PhD[c]

Paul Freeman, PhD[a]

Names for PubMed indexing: Hase, A., O'Brien, J., Moore, L. J., Freeman, P.

[a] School of Sport, Rehabilitation and Exercise Sciences, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK.

[b] School of Sport and Exercise, University of Gloucestershire, Oxstalls Campus, Gloucester, Gloucestershire, GL2 9HW, UK.

[c] Department for Health, University of Bath, Claverton Down, Bath, Somerset, BA2 7AY, UK.

Corresponding author:

Adrian Hase, MSc.

School of Sport, Rehabilitation and Exercise Sciences

University of Essex

Wivenhoe Park

Colchester CO4 3SQ

United Kingdom

Phone: +44 (0) 1206 872179

E-mail: ahase@essex.ac.uk

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Word Count (including abstract, excluding tables and references): 6,911

Number of Figures: 1

Number of Tables: 5

Abstract

The biopsychosocial model of challenge and threat states specifies that these states engender different physiological and behavioural responses in potentially stressful situations. This model has received growing interest in the sport and performance psychology literature. The present systematic review examined whether a challenge state is associated with superior performance than a threat state. Across 38 published studies that conceptualised challenge and threat states in a manner congruent with the biopsychosocial model, support emerged for the performance benefits of a challenge state. There was, however, significant variation in the reviewed studies in terms of the measures of challenge and threat states, tasks, and research designs. The benefits of a challenge state on performance were largely consistent across studies using cognitive, physiological, and dichotomous challenge and threat measures, cognitive and behavioural tasks, and direct experimental, indirect experimental, correlational, and quasi-experimental designs. The results imply that sports coaches, company directors, and teachers might benefit from trying to promote a challenge state in their athletes, employees, and students, respectively. Future research could benefit from a greater consensus on how best to measure challenge and threat states to help synthesise the evidence across studies. Specifically, we recommend that researchers use both cognitive and physiological measures and develop stronger manipulations for experimental studies. Finally, future research should report sufficient information to enable risk of bias assessment.

Keywords: Motivated performance situation; biopsychosocial model; stress; cardiovascular reactivity; demand resource evaluations

The relationship between challenge and threat states and performance: A systematic review

Understanding individuals' responses to stress is key for optimising performance in contexts including business, medicine, education, and sport. Although some models explain individuals' successes and failures in terms of psychology or physiology, one increasingly popular theory combines these perspectives. The biopsychosocial model (BPSM; Blascovich & Mendes, 2000) of challenge and threat (CAT) states built on Lazarus and Folkman's (1984) transactional theory of stress and Dienstbier's (1989) theory of physiological toughness, and has been applied to contexts as diverse as sport, education, and medicine (Moore, Wilson, Vine, Coussens, & Freeman, 2013; Roberts, Gale, McGrath, & Wilson, 2015; Seery, Weisbuch, Hetenyi, & Blascovich, 2010). Across these contexts, CAT states have been associated with different performance outcomes (e.g., Allen & Blascovich, 1994; Blascovich, Seery, Mugridge, Norris, & Weisbuch, 2004), although some studies have found non-significant or contradictory results (e.g., Feinberg & Aiello, 2010; Laborde, Lautenbach, & Allen, 2015), and there is notable diversity in how CAT states have been measured and the research designs employed. To advance our understanding of the impact of CAT states on performance, the consistency of findings across different methods, and to highlight important directions for future research, the current article reports a systematic review of the published literature that utilised the BPSM as a theoretical framework.

Central to the BPSM is the assumption that CAT states only occur in motivated performance situations. Motivated performance situations are goal-relevant, evaluative, and potentially stressful, requiring adequate active performance in order to ensure wellbeing and personal growth (Blascovich & Mendes, 2000). Sport competitions, academic exams, and job interviews are typical examples of such situations. Importantly, according to the BPSM, CAT states represent opposite ends of a unidimensional continuum rather than two dichotomous states, allowing researchers to examine relative (rather than absolute) differences in challenge and threat (i.e., greater vs. lesser challenge or threat; Blascovich, 2008). This contrasts the earlier views of Lazarus and Folkman (1984), and other researchers (e.g., Skinner & Brewer, 2004), who considered CAT as independent cognitive appraisals that can occur simultaneously. Although these other frameworks offer useful insights, this review focused only on publications that examined CAT states in the unidimensional manner hypothesised in the BPSM.

CAT states differ in terms of underlying cognitive evaluations and resulting physiological responses, which are predicted to be linked (Blascovich & Mendes, 2000). According to the BPSM, challenge states are characterised by the largely subconscious evaluation that one's personal coping resources match or exceed situational demands. Physiologically, challenge states are marked by increases in heart rate (HR) and cardiac output (CO), and decreases in total peripheral resistance (TPR). This cardiovascular pattern is due to sympathetic adrenal medullary activation, which causes epinephrine release, and dilation of the blood vessels. In contrast, threat states are characterised by an evaluation that coping resources fall short of situational demands. Threat states are indexed by little change or small increases in HR, little change or minor decreases in CO, and little change or small increases in TPR (Tomaka, Blascovich, Kelsey, & Leitten, 1993). This physiological response is due to additional activation of the pituitary-adrenocortical pathway, which constricts blood vessels, causes cortisol release, and inhibits the effects of sympathetic-adrenomedullary activation (Blascovich & Mendes, 2000). Importantly, validation studies showed that: a) cognitive CAT evaluations and physiological CAT responses were significantly correlated, and b) cognitive CAT evaluations triggered physiological responses, not vice versa (Blascovich, 2008). These divergent CAT states are predicted to influence performance, with challenge states being related to superior performance than threat states.

The relevance of the BPSM to a range of contexts has led to considerable variation in the tasks and performance outcomes examined across the literature. For example, studies have examined the relationship between CAT states and cognitive performance in academic (Seery et al., 2010), GRE word problem (Chalabaev, Major, Cury, & Sarrazin, 2009), and mental arithmetic (Kelsey et al., 2000) tasks. Further, Blascovich et al. (2004) found that a cardiovascular CAT index, measured during a pre-season speech about athletes' sports, predicted batting performance during the season, with a challenge state linked to better performance than a threat state (i.e., more runs). This initial evidence provided impetus for subsequent research involving behavioural tasks as varied as simulated surgery (Vine et al., 2013) and cricket batting (Turner et al., 2013).

This early research also led to the development of new theories that extended the predictions of the BPSM (i.e., Theory of Challenge and Threat States in Athletes [TCTSA]; Jones, Meijen, McCarthy, & Sheffield, 2009; integrated framework of stress, attention, and visuomotor performance; Vine, Moore, & Wilson, 2016). These theories suggest that CAT states could influence performance through various mechanisms. For example, the TCTSA predicts that a threat state may lead to more negative emotions, unfavourable interpretations of emotions, impaired cognitive functioning, decision-making and anaerobic power, greater self-regulation, increased reinvestment and avoidance coping, and less effective attention, which may in turn impair performance (Jones et al., 2009). Further, Vine et al. (2016) argue that a threat state might deter performance by disrupting attentional and visuomotor control, causing individuals to become distracted by less relevant (and potentially negative) stimuli at the expense of more important task-relevant cues. This is in keeping with the original mechanism proposed by Blascovich et al. (2004), who speculated that attentional resources might be diverted from the task at hand towards the environment or themselves during a threat state. However, to date, relatively little research has tested these potential mechanisms (e.g., Moore, Vine, Wilson, & Freeman, 2012).

With increasing interest in the BPSM, there has been greater diversity in the conceptualisation and measurement of CAT states. Indeed, while some authors have used self-report measures of demand and resource evaluations (e.g., Gildea, Schneider, & Shebilske, 2007), others have used physiological indices computed from CO and TPR reactivity (i.e., change in CO and TPR from baseline to post-instruction/task exposure; e.g., Blascovich et al., 2004). Although both the cognitive evaluations and physiological responses accompanying CAT states are predicted to influence performance, it is not known which has the strongest effect. Even within these approaches, little consensus exists regarding standardised measurements. For example, both single-and multi-item self-report measures of cognitive evaluations have been used to calculate either a ratio (e.g., demands divided by resources), or a difference score (e.g., resources minus demands). Researchers have also differed in the timing and duration of baseline and post-instruction/task exposure periods when recording cardiovascular data, and have used different methods to calculate a single CAT index from CO and TPR reactivity (e.g., difference vs. residualised change scores).

In addition to the diversity in the measurement of CAT states and the tasks employed, studies have adopted different research designs. Some studies have employed experimental designs, directly manipulating individuals into CAT states and observing performance. For example, Moore and colleagues (2013) used verbal instructions to elicit CAT states before a golf putting task, and found that the golfers in the challenge group outperformed those in the threat group (Moore, Wilson et al., 2013). Other experimental studies have indirectly manipulated CAT states via an antecedent and then measured performance (e.g., resource appraisals; Turner, Jones, Sheffield, Barker, & Coffee, 2014). Correlational studies have also been employed, with CAT states observed before a task and subsequently related to performance (e.g., Turner et al., 2013). Finally, studies have used quasi-experimental designs, recording CAT states with continuous measures, and then splitting the sample into CAT groups before examining between-group differences in performance (e.g., via median split; Gildea et al., 2007).

Given the increasing adoption of the BPSM for understanding performance variation during stressful tasks, aligned with notable diversity in the conceptualisation of CAT states, performance outcomes, and research designs employed, the primary aim of this systematic review was to examine the pattern of associations between CAT states and performance outcomes. The secondary aim was to examine the consistency of this pattern across different conceptualisations of CAT states (i.e., cognitive evaluations vs. physiological responses vs. dichotomous groups), performance outcomes (i.e., cognitive vs. behavioural tasks), and research designs (i.e., direct experimental vs. indirect experimental vs. correlational vs. quasi-experimental designs). Synthesising the current evidence will provide crucial insight into the utility of the BPSM to explain performance variation under stress, the impact of employing different methods, and highlight important directions and methodological considerations for future research.

Method

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic reviews and Meta-Analyses guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). It involved four steps: (1) initial literature search (including selection of search terms, electronic databases, and inclusion criteria), (2) screening based on title, (3) screening based on abstract, and (4) screening based on full text. Two independent assessors completed each step, compared their records and discussed any disagreements. The assessors searched for relevant articles using the following databases: MedLine, PsycINFO, and SPORTDiscus (combined in one search) and Web of Science (in a separate search). The search terms were ("challenge and threat" AND "performance"). To be included, studies had to fulfil five inclusion criteria: (1) published in English in a peer-reviewed academic journal, (2) report at least one empirical study, (3) conducted with healthy human participants, (4) conceptualise CAT in terms of a unidimensional continuum, and (5) report at least one performance outcome and its association with at least one CAT measure, or dichotomous CAT groups that were compared on a CAT measure in a manipulation check.

To examine the consistency of the pattern of associations between CAT states and performance within different conceptualisations of CAT states, performance outcomes and research designs, we used Sallis, Prochaska, and Taylor's (2000) sum code classification. This classification focuses on the percentage of studies that demonstrate a statistically significant effect. Further, to assess the quality and risk of bias in experimental and non-experimental studies, respectively, the Cochrane Collaboration's tool for assessing risk of bias (Higgins & Altman, 2008) and the Risk of Bias Assessment Tool for Nonrandomised Studies (Kim et al., 2013) were used. For experimental studies, two independent assessors examined random sequence generation (were experimental conditions assigned randomly?), allocation concealment (could condition allocations have been foreseen before/during enrolment?), blinding of participants and personnel (were participants and researchers blind to the participants' allocated experimental condition?), blinding of outcome assessment (were outcome assessors blind to experimental condition?), incomplete outcome data (were attrition/exclusion rates and reasons reported?), selective reporting (was there a possibility of selective reporting?), and other sources of bias (Higgins & Altman, 2008). For non-experimental studies, two independent assessors examined blinding of outcome assessment, incomplete outcome data, selective reporting, selection of participants (how adequate was the selection of participants?), confounding variables (was there adequate consideration of confounders?), and intervention (exposure) measurement (was there performance bias caused by inadequate measurement of exposure?; Kim et al., 2013).

Results

The initial search (conducted in December 2017) yielded 1107 unique results. After reviewing titles, 155 records remained. After reading abstracts, 59 records remained. After reviewing full-texts, 30 articles reporting 38 studies with a total of 3257 participants were identified and included in the review. Figure 1 illustrates the search and screening process. Inter-rater agreements in the second, third, and fourth step were 96.6%, 84.4%, and 84.7%. Disagreements were resolved through discussion between the assessors and a third member of the research team.

General Study Characteristics

Table 1 presents the characteristics and main outcomes of the included studies. Sample sizes ranged from 16 to 238 with a mean sample size of 85.7 participants (SD = 54.4). Most samples contained both genders, but four samples were all male (Gildea et al., 2007; Laborde et al., 2015; Turner et al., 2013), and five samples were all female (Chalabaev et al., 2009; Chalabaev, Major, Sarrazin, & Cury, 2012; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007; Study 2, Scheepers, 2017; Turner, Jones, Sheffield, & Cross, 2012). The average age in the 28 studies that reported this statistic ranged from 11.0 to 36.3 years with an average mean of 22.5 years (SD = 4.9). The remaining studies reported a mode age of 18 years (Quigley, Barrett, & Weinstein, 2002), a median of 28 years (Roberts et al., 2015), or no age statistic (Blascovich et al., 2004; Chalabaev et al., 2009; Chalabaev et al., 2012; Feinberg & Aiello, 2010; Kelsey et al., 2000; Seery et al., 2010). Most studies sampled university students, but others incorporated athletes, doctors, adolescents, academic staff, and non-specified adults.

Risk of Bias in Individual Studies

Table 2 presents the risk of bias results. Interrater agreements were 84.1% and 85.8% for experimental and non-experimental studies, respectively. The assessors resolved disagreements in discussions with a third member of the research team. In experimental studies, the lowest risk of bias ratings emerged for “random sequence generation”, “incomplete outcome data”, and “other sources of bias”, as 88.9%, 77.8%, and 100% of studies received a “low risk of bias” rating, respectively. Unclear risk of bias was more apparent for “allocation concealment”, “blinding of participants and personnel”, “blinding of outcome assessment”, and “selective reporting”, with 88.9%, 88.9%, 55.6%, and 100% of studies rated as “unclear risk of bias” respectively. The assessors rated one study (5.6%) in the “incomplete outcome data” category as “high risk of bias”.

In non-experimental studies, a low risk of bias ratings emerged for “blinding of outcome assessment”, “incomplete outcome data”, “confounding variables”, and “intervention (exposure) measurement”, as 55.0%, 75.0%, 100%, and 100% of studies in these categories received a “low risk of bias” rating, respectively. “Selective reporting” and “selection of participants” received mostly “unclear risk of bias” ratings (100% and 90.0%, respectively). The assessors rated two studies (10.0%) in the “incomplete outcome data” category as “high risk of bias”.

Association between CAT States and Performance

Of the 38 included studies, 28 (74%) found an effect on performance favouring a challenge state, although three of the observed effects were contingent on an interaction with another variable. The three interaction effects depended on solo status (performing alone or not; Study 1, White, 2008), performance goals (performance-avoidance or approach goal; Chalabaev et al., 2012), and integrative task structure (whether concessions on less important aspects of a negotiation tasks led to gains on more important aspects or not; Study 2, O’Connor, Arnold, & Maurizio, 2010). Of the remaining 10 studies, one found an effect favouring a threat state (Study 1, Feinberg & Aiello, 2010), and nine found no significant effects (Chalabaev et al., 2009; Study 4, Feinberg & Aiello, 2010; Study 2, Gildea et al., 2007; Laborde et al., 2015; Mendes et al., 2007; Quigley et al., 2002; Rith-Najarian et al., 2014; Sammy et al., 2017; Turner et al., 2014). At least one effect size was reported in 24 studies, yielding 29 in total: 12 Cohen's *d* values ranging from 0.29 to 1.09, 15 R^2 values ranging from .06 to .61, one sr^2 of .04, and one η_p^2 of .12 (see Table 1). These reflected 11 small, 14 medium, and four large effect sizes (Cohen, 1992).

Effects of cognitive, physiological, and dichotomous CAT measures on performance. Table 3 lists the associations between CAT states and performance based on whether CAT was analysed as a continuous cognitive, continuous physiological, or dichotomous variable. The dichotomous category included studies that compared challenge and threat groups in the analysis, regardless of whether the groups were created by an experimental manipulation or by a median split of a continuous CAT measure. Studies that reported an association with performance of more than one CAT measure are included in each relevant category; thus, the number of effects is 43.

Sixteen studies reported 17 analyses that examined the association between a cognitive CAT measure and performance. Thirteen analyses (76%) found a statistically significant effect favouring a challenge state, with two effects contingent on interactions (Study 1, White, 2008; Chalabaev et al., 2012). Four analyses found no significant effect (Chalabaev et al., 2009; Laborde et al., 2015; Quigley et al., 2002; Rith-Najarian et al., 2014). Of the six effect sizes reported, three were small (Chalabaev et al., 2012; Moore, Young, Freeman, & Sarkar, 2017; Study 1, Moore, Wilson et al., 2013), two were medium (Study 1, O'Connor et al., 2010; Schneider, 2004), and one was large (Vine et al., 2015). The majority of the cognitive CAT indices used self-report items from Tomaka and colleagues' (1993) cognitive appraisal ratio or Schneider's (2008) stressor appraisal scale to create demand and resource evaluation scores. These scores were combined into a ratio (i.e., demands divided by resources; e.g., Quigley et al., 2002) or a difference score (i.e., resources minus demands; e.g., Chalabaev et al., 2012). However, some studies used single-item measures that assessed the degree to which participants felt challenged or threatened (e.g., Turner et al., 2012).

Eleven studies reported 12 analyses that examined the association between a physiological CAT measure and performance. Eight (67%) found that a challenge cardiovascular response was associated with better performance than the threat response (Blascovich et al., 2004; Moore et al., 2017; Scheepers, 2017; Scholl, Moeller, Scheepers, Nuerk, & Sassenberg, 2015; Seery et al., 2010; Turner et al., 2013; Studies 1 and 2, Turner et al., 2012). Four analyses found no significant effect (Mendes et al., 2007; Rith-Najarian et al., 2014; Seery et al., 2010; Vine, Freeman, Moore, Chandra-Ramanan, & Wilson, 2013). Of the 10 effect sizes reported, five were small (Blascovich et al., 2004; Moore et al., 2017; Scheepers, 2017; Scholl et al., 2015; Seery et al., 2010), and five were medium (Scholl et al., 2015; Studies 1 and 2, Turner et al., 2012). The physiological CAT index comprised a sum score of the changes in CO and TPR from baseline to a post-instruction (or manipulation) period. These changes were determined by using difference scores in all studies in the "Physiological" group. However, two studies in the "Dichotomous" group used residualised change scores (i.e., standardised residuals of a regression of post-instruction on baseline values, to control for differences in baseline values) to create the index (e.g., Moore et al., 2015; Moore, Vine, Wilson, & Freeman, 2014). Both approaches typically weighted TPR reactivity negatively, so that a greater value on the summed CAT index was more reflective of a challenge state. Finally, the timing and duration of physiological data differed between studies. For example, some studies recorded five minutes of baseline data and one minute after giving task instructions, although they often only used the final minute of the baseline period in the analyses (e.g., Moore et al., 2014). Other studies measured five minutes of baseline data and two minutes of reactivity data during the task, using mean values of the entire time periods (e.g., Blascovich et al., 2004).

Only 11 studies included both physiological and cognitive CAT indices, and only three of these studies reported associations with performance for both indices[1] (Moore et al., 2017; Rith-

[1] Chalabaev et al.'s (2009) study is not listed here despite reporting performance analyses for the cognitive and physiological variables (i.e., CO and TPR reactivity). This is because the physiological CAT variables were not

Najarian et al., 2014; Vine et al., 2013). Moore and colleagues (2017) found that both the cognitive and physiological CAT measures were related to performance. Rith-Najarian and colleagues (2014) found that neither measure was related to performance. Vine and colleagues (2013) found that only the cognitive CAT measure was related to performance, with a challenge state linked with better performance. Further, only three of the studies that computed both cognitive and physiological CAT measures provided a correlation between the two indices[2] (Moore et al., 2017; Turner et al., 2013; Vine et al., 2013). Moore et al. (2017; r = .19) and Turner et al. (2013; r = .21) found no significant correlation, whereas Vine et al. (2013) found a significant correlation during the baseline test (r = .32), but not the pressurised test (r = -.11).

Fifteen studies created dichotomous groups, which were confirmed with a manipulation check using a cognitive and/or physiological CAT measure. Ten (67%) studies found that the challenge group significantly outperformed the threat group (Study 2, Feinberg & Aiello, 2010; Studies 1 and 3, Gildea et al., 2007; Moore et al., 2012; Moore et al., 2014; Moore et al., 2015; Study 2, Moore, Wilson et al., 2013; Study 2, O'Connor et al., 2010; Scheepers, 2017), with one effect contingent on an interaction (O'Connor et al., 2010). Furthermore, Feinberg and Aiello (2010) reported three significant interaction effects between CAT instructions and experimenter presence. However, they did not report whether challenge was related to better performance than threat in any of the two experimenter presence conditions, comparing challenge with challenge, and threat with threat across the two conditions instead. Four studies found no significant effect (Study 4, Feinberg & Aiello, 2010; Study 2, Gildea et al., 2007; Sammy et al., 2017; Turner et al., 2014), and one study found that participants in the threat condition outperformed those in the challenge condition, although it should be noted that the manipulation check in this study was only marginally significant (Study 1, Feinberg & Aiello, 2010). Of the 16 effect sizes reported, six were small

combined into a single CAT index, which violated the inclusion criteria. However, it is noteworthy that this analysis did find challenge reactivity to be associated with better performance, supporting the contentions of the BPSM.

[2] Two other studies provided associations between cognitive and physiological variables, but did not use a single physiological CAT index (Turner et al., 2012; Quigley et al., 2002). Turner et al. (2012) did not find any significant correlations, although the coefficients were consistent with the BPSM in terms of direction. Quigley et al. (2002) found a marginally significant association between cognitive CAT and CO, but not between cognitive CAT and TPR.

(Study 2, Gildea et al., 2007; Moore et al., 2014; Moore et al., 2017; Study 2, O'Connor et al., 2010; Scheepers, 2017), seven were medium (Study 3, Gildea et al., 2007; Moore et al., 2012; Study 2, Moore, Wilson et al., 2013; Schneider, 2004; Turner et al., 2014), and three were large (Study 1, Feinberg & Aiello, 2010; Study 1, Gildea et al., 2007; Moore et al., 2015).

Effects of CAT states on cognitive and behavioural task performance. The performance tasks varied across studies, but could be placed into two main categories: Cognitive and behavioural. Table 4 lists the studies in each category and their corresponding results.

Twenty studies reported 23 effects involving cognitive performance outcomes, of which eight were mathematical (e.g., serial subtraction task; Kelsey et al., 2000). Examples of other tasks included Stroop (Study 1, Turner et al., 2012), and word-finding (Mendes et al., 2007) tasks. Fifteen (65%) analyses found that a challenge state was associated with superior performance, although two of these effects were contingent on an interaction with another variable (Chalabaev et al., 2012; Study 1, White, 2008). Seven effects were not significant, and one analysis found that participants performed significantly better in the threat condition (Study 1, Feinberg & Aiello, 2010). Of the 15 effect sizes, four were small (Chalabaev et al., 2012; Scholl et al., 2015; Seery et al., 2010), nine were medium (Study 3, Gildea et al., 2007; Schneider, 2004; Scholl et al., 2015; Studies 1 and 2, Turner et al., 2012), and two were large (Study 1, Feinberg & Aiello, 2010; Study 1, Gildea et al., 2007).

Nineteen effects involved behavioural tasks such as golf putting (Moore et al., 2012; Moore et al., 2015; Study 2, Moore, Wilson et al., 2013), cricket batting (Turner et al., 2013), flight simulation (Vine et al., 2015), and a medical selection practical (Roberts et al., 2015). Sixteen (84%) effects favoured a challenge state, with one effect qualified by an interaction with another variable (Study 2, O'Connor et al., 2010). Three effects were not significant (Rith-Najarian et al., 2014; Sammy et al., 2017; Turner et al., 2014). Of the 15 effect sizes reported, six were small (Blascovich et al., 2004; Moore et al., 2014; Study 1, Moore, Wilson et al., 2013; Moore et al., 2017; Study 2, O'Connor et al., 2010), seven were medium (Moore et al., 2012; Study 2, Moore, Wilson et al., 2013; Study 1, O'Connor et al., 2010; Turner et al., 2014; Studies 1 and 2, Turner et al., 2012), and two were large (Moore et al., 2015; Vine et al., 2015).

Effects of CAT states on performance within different research designs. Four types of research designs were used: (1) experiments that directly manipulated CAT states (explicitly targeting CAT states), (2) experiments that indirectly manipulated CAT states (targeting another variable, including putative CAT antecedents), (3) correlational studies, and (4) quasi-experiments. Table 5 lists the studies grouped by research design. Although the "dichotomous" group in Table 3 shares some studies with the "experimental (direct)" and "quasi-experimental" groups, the research questions pertaining to Table 3 and Table 5 are different. Table 3 is about the type of CAT measure and analysis, whereas Table 5 is about the type of research design.

Six studies reported experiments that directly manipulated participants into CAT states by framing the task instructions consistent with either a challenge or threat state (i.e., perceptions of task demands and personal coping resources). Four (67%) studies found that participants in the challenge group performed significantly better than those in the threat group (Study 2, Feinberg & Aiello, 2010; Moore et al., 2012; Study 2, Moore, Wilson et al., 2013), although one effect was qualified by an interaction (Study 2, O'Connor et al., 2010). One study found no significant effect (Study 4, Feinberg & Aiello, 2010), and one study found that the threat group outperformed the challenge group (Study 1, Feinberg & Aiello, 2010). Of the five effect sizes, one was small (Study 2, O'Connor et al., 2010), three were medium (Moore et al., 2012; Study 2, Moore, Wilson et al., 2013), and one was large (Study 1, Feinberg & Aiello, 2010).

Twelve studies reported experiments that indirectly manipulated CAT states by manipulating another variable such as resource appraisals (Turner et al., 2014), perceived effort and support (Moore et al., 2014), or interpretations of physiological arousal (Moore et al., 2015), and obtained different CAT responses between groups. Eight (67%) studies found that a challenge state was associated with superior performance, although one effect was contingent on an interaction (O'Connor et al., 2010). Four studies found no significant effect (Chalabaev et al., 2009; Mendes et al., 2007; Sammy et al., 2017; Turner et al., 2014). Of the six effect sizes reported, three were small (Chalabaev et al., 2012; Moore et al., 2014; Scheepers, 2017), two were medium (Study 1, O'Connor et al., 2010; Turner et al., 2014), and one was large (Moore et al., 2015).

Sixteen studies used a correlational design, correlating either a cognitive or physiological CAT measure with performance. Of the 18 effects in this group, 14 (78%) showed a significant association between CAT and performance, with a challenge state related to better performance. Four analyses found no significant association (Laborde et al., 2015; Quigley et al., 2002; Rith-Najarian et al., 2014; Seery et al., 2010). Of the 12 effect sizes reported, five were small (Blascovich et al., 2004; Moore et al., 2017; Scholl et al., 2015; Seery et al., 2010), six were medium (Study 2, Moore, Wilson et al., 2013; Scholl et al., 2015; Studies 1 and 2, Turner et al., 2012), and one was large (Vine et al., 2015).

Finally, four studies used a quasi-experimental approach by dividing the sample into CAT groups based on scores on a cognitive CAT measure. All four (100%) studies found that participants in the challenge group performed significantly better than those in the threat group (Gildea et al., 2007; Schneider, 2004). Of the six effect sizes reported, one was small (Study 2, Gildea et al., 2007), four were medium (Study 3, Gildea et al., 2007; Schneider, 2004), and one was large (Study 1, Gildea et al., 2007).

Discussion

For over two decades, the BPSM of CAT states has been used as a framework to understand variations in cognitive, physiological, and behavioural responses in motivated performance situations (Blascovich & Mendes, 2000). The aim of this systematic review was to examine the relationship between CAT states and performance, and the consistency of this relationship across different CAT measures, performance tasks, and research designs. In 28 (74%) of the 38 studies, a challenge state was related to better performance. Based on statistical significance, the relationship between CAT states and performance was relatively consistent across different measures of CAT states (cognitive vs. physiological vs. dichotomous), performance outcomes (cognitive vs.

behavioural), and research designs (direct experimental vs. indirect experimental vs. correlational vs. quasi-experimental), although there were few studies in the direct experimental group. The common finding that individuals who exhibited a challenge state outperformed individuals who displayed a threat state, supports the predictions of the BPSM and holds relevance for sports psychologists, coaches, business managers, educators, and other professionals interested in optimising human performance.

The beneficial effect of a challenge state was generally consistent across different CAT measures (i.e., cognitive vs. physiological vs. dichotomous). As such, the findings support the prediction of the BPSM that CAT states occur on both a cognitive (i.e., underlying demand/resource evaluations) and physiological (i.e., accompanying cardiovascular responses) level, and influence performance. However, it is noteworthy that studies including the relationships between both CAT measures and performance found an inconsistent pattern (e.g., Moore et al., 2017; Rith-Najarian et al., 2014; Turner et al., 2013), implying that more research is needed to compare the two measures as predictors of performance. In addition, although the BPSM predicts that different demand and resource evaluations lead to distinct physiological responses (Blascovich, 2008), only three studies included both cognitive and physiological CAT measures and reported correlations among these variables (Moore et al., 2017; Turner et al., 2013; Vine et al., 2013). Weak to moderate correlations were reported in these studies, raising questions about whether demand and resource evaluations trigger distinct cardiovascular responses, as proposed by the BPSM (Blascovich, 2008). Indeed, the wider BPSM literature has also demonstrated weak to moderate links between cognitive and physiological markers of CAT (e.g., Zanstra, Johnston, & Rasbash, 2010).

Studies that used a single cognitive measure of CAT states to dichotomise individuals into CAT groups (e.g., via a median split) also tended to support the superiority of a challenge state (e.g., Gildea et al., 2007). However, dichotomising CAT states is incongruent with the notion that they represent opposite ends of a single bipolar continuum (Blascovich & Mendes, 2000). Further, dichotomising a sample with a median split could lead to problems like loss of statistical power and difficulty in comparing results between studies due to the different cut-off points employed (Altman & Royston, 2006). Researchers should therefore consider whether it is appropriate to dichotomise CAT measures and, if so, ensure that the study has sufficient power.

This review revealed notable diversity in the recording and calculation of cognitive and physiological CAT measures. For instance, both single and multiple self-report items assessed demand and resource evaluations (Schneider, 2008; Tomaka et al., 1993; Turner et al., 2013). In addition, responses to these items were used to calculate a ratio (i.e., demands divided by resources; e.g., Moore et al., 2012), or difference (i.e., resources minus demands; e.g., Moore et al., 2013) score. Moreover, CO and TPR were reported as reactivity (e.g., Blascovich et al., 2004) or residualised change scores (e.g., Moore et al., 2012). These values were often calculated by averaging across different durations and time periods (e.g., final minute of baseline and first minute after receipt of task instructions, Moore et al., 2014; or final two minutes of baseline and first two minutes of the task itself, Blascovich et al., 2004). The justifications for these variations were not always clearly articulated and should be made more explicit in future research.

Although these variations did not appear to impact the findings, future research would benefit from adopting a more consistent approach in CAT measurement to facilitate the synthesis of evidence across studies. If studies adopt different methods to measure CAT states, it is unclear whether the observed relationships are due to CAT states themselves or the idiosyncratic measurement processes (e.g., because self-report was employed rather than cardiovascular indices or a ratio vs. a difference score). Although we encourage future research to contrast the different ways of measuring CAT states to empirically identify the optimal approach, we make the following recommendations based on the justifications provided in the current literature. Researchers should use both cognitive evaluations and cardiovascular responses to measure CAT states, and further examine their relationship and respective effects on performance. Given the limitations associated with single-item scales (e.g., lower relative precision than multi-item scales; McHorney, Ware, Rogers, Raczek, & Lu, 1992), multi-item measures of demand and resource evaluations should be employed (e.g., Schneider, 2008). The scores from these items should then be used to calculate a difference score, as ratio scores have been discouraged due to their highly nonlinear distribution (Vine et al., 2013). When measuring the physiological indices of CAT states (i.e., CO and TPR reactivity), researchers should use comparable time periods and indices. To ensure true resting values are obtained, researchers should use the final minute of the baseline period (Sherwood, Allen, Kelsey, Lovallo, & van Doornen, 1990). Further, given the dynamic nature of CAT states (i.e., reappraisal; Blascovich, 2008), researchers should utilise the first minute after task instructions or of task exposure. While most research has employed difference scores rather than residualised change scores, we recommend that researchers consult guidelines and use the approach most suitable for their data (e.g., Burt & Obradovic, 2013). Finally, CO and TPR reactivity should be combined into a single CAT index, which is more in keeping with the unidimensional nature of CAT states, increases reliability, and simplifies analyses (Seery et al., 2010).

The risk of bias assessment showed that random sequence generation, incomplete outcome data, other sources of bias, blinding of outcome assessment, incomplete outcome data, confounding variables, and intervention (exposure) measurement exhibited a low risk of bias across most studies. Allocation concealment, blinding of participants and personnel, blinding of outcome assessment, selection of participants, and selective reporting often exhibited an unclear risk of bias. As only three studies were rated as high risk of bias, the body of evidence appears to be of adequate quality overall, but the findings highlight the importance of considering and reporting potential risks in future studies. For example, researchers should minimise missing physiological and outcome data, ensure that performance assessors are naive to CAT data, and provide information about allocation concealment, blinding of participants, personnel and outcome assessment, and selective reporting.

Based on statistical significance, there was a relatively consistent relationship between CAT states and performance on behavioural and cognitive tasks. The notable difference in support for cognitive vs. behavioural tasks (see Table 4) could have been influenced by the included and excluded studies. First, although Chalabaev et al. (2009) found that greater CO reactivity and lower

TPR reactivity were associated with better cognitive performance separately, the review excluded this study as no single physiological CAT index was reported. Second, Feinberg and Aiello's (2010) three studies that manipulated participants into CAT groups using verbal instructions, found inconsistent effects for CAT states on performance, one of which involved an only marginally significant manipulation check. As well as being inconsistent with the notion that CAT states are a continuum (Blascovich & Mendes, 2000), this approach averages data across CAT groups and individuals who were not successfully manipulated into the required state might have attenuated the results (i.e., individuals in the challenge group displaying a threat state, and vice versa; Turner et al., 2013). As such, the weaker effect on cognitive outcomes might have been caused by other confounding statistical and methodological issues.

Studies that directly manipulated CAT states provided support for the superiority of a challenge state, although only six studies utilised such a design. Four studies found that the challenge group outperformed the threat group (Study 2, Feinberg & Aiello, 2010; Moore et al., 2012; Moore, Wilson et al., 2013; O'Connor et al., 2010), and two studies reported null or contradictory results (Studies 1 and 4, Feinberg & Aiello, 2010). Issues such as the strength and effectiveness of the CAT manipulation instructions (as well as the limitations noted above) might explain the heterogeneous results among Feinberg and Aiello's (2010) studies. For example, Feinberg and Aiello read instructions aloud to participants, whereas Moore et al. (2012, 2013) delivered standardised instructions from memory more directly to participants. Researchers employing experimental designs should report the methods used to manipulate participants into CAT states and use both cognitive and physiological CAT measures as manipulation checks, as the two measures could yield divergent results.

Although two theoretical models (Jones et al., 2009; Vine et al., 2016) have proposed several potential mechanisms through which CAT states might influence performance, only three studies included in the review explicitly tested mediation (Moore et al., 2012; Moore, Wilson et al., 2013 study 2; Vine et al., 2013). Of these studies, only one study reported statistically significant mediation (Moore et al., 2012), with the findings suggesting that CAT states influenced golf-putting performance primarily via kinematic variables and not through emotional, attentional, or physiological pathways. Despite this limited evidence for significant mediating processes, studies have reported that CAT states are associated with different emotional, attentional, and physiological responses, with a challenge state linked with less cognitive anxiety, more optimal visual attention, and less muscle activity (Moore et al., 2012; Moore, Wilson et al., 2013 study 2; Vine et al., 2013). It is vital for research to continue exploring these and other potential underlying mechanisms to better understand how a challenge state facilitates performance. In particular, research should test the attentional mechanisms outlined by Vine et al. (2016), and examine whether a threat state increases the influence of the stimulus-driven system and draws attention away from task-relevant to less relevant (and potentially negative) stimuli, resulting in suboptimal performance.

Several issues emerged as limitations to the present review. First, a meta-analysis may have provided additional information about the strength of the relationship between CAT states and performance. However, this was not feasible due to the substantial variability in methodologies adopted across studies. The variability across studies also hindered the ability to clearly delineate how strongly the effects were influenced by the CAT measure, task, or research design. Second, as this review only included published studies, publication bias might have influenced its results. Third, the sum codes used in Tables 3, 4, and 5 (adopted from Sallis et al., 2000) use arbitrary cut-off points and refer to patterns of statistical significance, which do not take into account effect sizes. Finally, while the research team categorised tasks as either cognitive or behavioural, many tasks required both cognitive input and behavioural execution. For example, golf putting requires cognition to determine the optimal direction and behavioural control to execute the motor skill.

This review highlights key directions for future research. Given that a challenge state facilitates performance, it is important to identify factors that elicit a challenge state to aid the development of theory and effective interventions. While some antecedents proposed by the BPSM (e.g., required effort and support; Moore et al., 2014) and TCTSA (e.g., control, self-efficacy, and achievement goals, Turner et al., 2014) have been investigated, research should examine other possible antecedents (e.g., danger, uncertainty, familiarity, knowledge, skills, abilities; Blascovich, 2008). Further, although some interventions have received attention (e.g., arousal reappraisal, Moore et al., 2015), research should examine other interventions aimed at promoting a challenge state. Finally, the longitudinal (and likely reciprocal) relationship between CAT states and performance should be explored.

Conclusion

To conclude, a challenge state was related to better performance than a threat state in 74% of studies. The quality of the included studies was generally good, although the risk of bias assessment identified some areas for improvement (e.g., minimise data loss). This association between CAT states and performance was relatively consistent across cognitive, physiological, and dichotomous CAT variables; cognitive and behavioural tasks; and direct experimental, indirect experimental, correlational, and quasi-experimental designs. Future research would benefit from a more consistent approach to CAT measurement (e.g., multi-item self-report measures of cognitive evaluations), to reduce ambiguity and aid the synthesis of results across studies. Furthermore, researchers should develop challenge-promoting interventions to optimise the performance of individuals across a range of domains (e.g., sport, academia, business, and medicine).

References

Allen, K. M., & Blascovich, J. (1994). Effects of music on cardiovascular reactivity among surgeons. *Journal of American Medical Association, 272*, 882-884. doi: 10.1001/jama.1994.03520110062030.

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ (Clinical Research Ed.), 332*, 1080. doi: 10.1136/bmj.332.7549.1080.

Blascovich, J. (2008). Challenge and threat. In A.J. Elliot (Ed.), *Handbook of approach and avoidance motivation* (pp. 431-445). New York: Psychology Press.

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas, J. P. Forgas (Eds.) , *Feeling and thinking: The role of affect in social cognition* (pp. 59-82). New York: Cambridge University Press.

Blascovich, J., Seery, M. D., Mugridge, C. A., Norris, R. K., & Weisbuch, M. (2004). Predicting athletic performance from cardiovascular indexes of challenge and threat. *Journal of Experimental Social Psychology*, *40*, 683-688. doi: 10.1016/j.jesp.2003.10.007.

Burt, K. B., & Obradović, J. (2013). The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review*, *33*, 29-57. doi:10.1016/j.dr.2012.10.002

Chalabaev, A., Major, B., Cury, F., & Sarrazin, P. (2009). Physiological markers of challenge and threat mediate the effects of performance-based goals on performance. *Journal of Experimental Social Psychology*, *45*, 991-994. doi: 10.1016/j.jesp.2009.04.009.

Chalabaev, A., Major, B., Sarrazin, P., & Cury, F. (2012). When avoiding failure improves performance: Stereotype threat and the impact of performance goals. *Motivation and Emotion*, *36*, 130-142. doi: 10.1007/s11031-011-9241-x.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159. doi: 10.1037/0033-2909.112.1.155.

Cronbach, L. J., & Furby, L. (1970). How should we measure ‘‘change’’— or should we? *Psychological Bulletin*, *74*, 68-80. doi: 10.1037/h0029382.

Dienstbier, R. A. (1989). Arousal of physiological toughness: Implications for physical and mental health. *Psychological Review, 96*, 84-100. doi: 10.1037/0033-295X.96.1.84.

Feinberg, J. M., & Aiello, J. R. (2010). The effect of challenge and threat appraisals under evaluative presence. *Journal of Applied Social Psychology*, *40*, 2071-2104. doi: 10.1111/j.1559-1816.2010.00651.x.

Gildea, K. M., Schneider, T. R., & Shebilske, W. L. (2007). Stress appraisals and training performance on a complex laboratory task. *Human Factors*, *49*, 745-758. doi: 10.1518/001872007X215818.

Higgins, J. P., & Altman, D. G. (2008). Assessing risk of bias in included studies. In J. P. Higgins & S. Green (Ed.), *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*. Retrieved from http://handbook-5-1.cochrane.org.

Jones, M., Meijen, C., McCarthy, P. J., & Sheffield, D. (2009). A theory of challenge and threat states in athletes. *International Review of Sport and Exercise Psychology, 2*, 161-180. doi:10.1080/17509840902829331.

Kelsey, R. M., Blascovich, J., Leitten, C. L., Schneider, T. R., Tomaka, J., & Wiens, S. (2000). Cardiovascular reactivity and adaptation to recurrent psychological stress: The moderating effects of evaluative observation. *Psychophysiology, 37*, 748-756. doi: 10.1111/1469-8986.3760748.

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

Laborde, S., Lautenbach, F., & Allen, M. S. (2015). The contribution of coping-related variables and heart rate variability to visual search performance under pressure. *Physiology & Behavior*, *139*, 532-540. doi: 10.1016/j.physbeh.2014.12.003.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.

McHorney, C. A., Ware, J. E., Rogers, W., Raczek, A. E., & Lu, J. R. (1992). The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP Charts: Results from the medical outcomes study. *Medical Care*, *30*, 253-265. doi:10.1097/00005650-199205001-00025

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: physiological responses during social interactions with expectancy-violating partners. *Journal of Personality and Social Psychology, 92*, 698-716. doi: 10.1037/0022-3514.92.4.698.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Plos Medicine*, *6*, e1000097. doi: 10.1016/j.jclinepi.2009.06.005.

Moore, L. J., Vine, S. J., Freeman, P., & Wilson, M. R. (2013). Quiet eye training promotes challenge appraisals and aids performance under elevated anxiety. *International Journal of Sport and Exercise Psychology, 11*, 169-183. doi: 10.1080/1612197X.2013.773688.

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2015). Reappraising threat: How to optimize performance under pressure. *Journal of Sport & Exercise Psychology*, *37*, 339-343. doi: 10.1123/jsep.2014-0186.

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2012). The effect of challenge and threat states on performance: An examination of potential mechanisms. *Psychophysiology, 49*, 1417-1425. doi: 10.1111/j.1469-8986.2012.01449.x.

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2014). Examining the antecedents of challenge and threat states: the influence of perceived required effort and support availability. *International Journal of Psychophysiology*, *93*, 267-273. doi: 10.1016/j.ijpsycho.2014.05.009.

Moore, L. J., Wilson, M. R., Vine, S. J., Coussens, A. H., & Freeman, P. (2013). Champ or chump? Challenge and threat states during pressurized competition. *Journal of Sport & Exercise Psychology*, *35*, 551-562. doi: 10.1123/jsep.35.6.551.

Moore, L. J., Young, T., Freeman, P., & Sarkar, M. (2018). Adverse life events, cardiovascular responses, and sports performance under pressure. *Scandinavian Journal of Medicine & Science in Sports, 28,* 340-347. doi:10.1111/sms.12928

O'Connor, K. M., Arnold, J. A., & Maurizio, A. M. (2010). The prospect of negotiating: Stress, cognitive appraisal, and performance. *Journal of Experimental Social Psychology*, *46*, 729-735. doi:10.1016/j.jesp.2010.04.007.

Quigley, K. S., Barrett, L. F., & Weinstein, S. (2002). Cardiovascular patterns associated with threat and challenge appraisals: a within-subjects analysis. *Psychophysiology*, *39*, 292-302. doi: 10.1017/S0048577201393046.

Rith-Najarian, L. R., McLaughlin, K. A., Sheridan, M. A., & Nock, M. K. (2014). The biopsychosocial model of stress in adolescence: Self-awareness of performance versus stress reactivity. *Stress: The International Journal on the Biology of Stress*, *17*, 193-203. doi: 10.3109/10253890.2014.891102.

Roberts, M. J., Gale, T. E., McGrath, J. S., & Wilson, M. R. (2015). Rising to the challenge: Acute stress appraisals and selection centre performance in applicants to postgraduate specialty training in anaesthesia. *Advances in Health Sciences Education, 21*, 323-339. doi: 10.1007/s10459-015-9629-6.

Sallis, J. F., Prochaska, J. J., & Taylor, W. C. (2000). A review of correlates of physical activity of children and adolescents. *Medicine and Science in Sports and Exercise, 32,* 963-975. doi: 10.1097/00005768-200005000-00014.

Sammy, N., Anstiss, P. A., Moore, L. J., Freeman, P., Wilson, M. R., & Vine, S. J. (2017). The effects of arousal reappraisal on stress responses, performance and attention. *Anxiety, Stress, &Coping*, *30*, 619-629. doi:10.1080/10615806.2017.1330952

Scheepers, D. (2017). Intergroup status differences as challenge or threat: The role of legitimacy. *Group Processes & Intergroup Relations, 20*, 75-90. doi: 10.1177/1368430215595108

Schneider, T. R. (2004). The role of neuroticism on psychological and physiological stress responses. *Journal of Experimental Social Psychology*, *40*, 795-804. doi: 10.1016/j.jesp.2004.04.005.

Schneider, T. R. (2008). Evaluations of stressful transactions: what's in an appraisal? *Stress and Health*, *24*, 151-158. doi: 10.1002/smi.1176.

Schneider, T. R., Rench, T. A., Lyons, J. B., & Riffle, R. R. (2012). The influence of neuroticism, extraversion and openness on stress responses. *Stress and Health*, *28*, 102-110. doi: 10.1002/smi.1409.

Scholl, A., Moeller, K., Scheepers, D., Nuerk, H., & Sassenberg, K. (2015). Physiological threat responses predict number processing. *Psychological Research*, *81*, 278-288. doi: 10.1007/s00426-015-0719-0.

Seery, M. D., Weisbuch, M., Hetenyi, M. A., & Blascovich, J. (2010). Cardiovascular measures independently predict performance in a university course. *Psychophysiology*, *47*, 535-539. doi: 10.1111/j.1469-8986.2009.00945.x.

Sherwood, A., Allen, M.T., Kelsey, R.M., Lovallo, W.R., & van Doornen, L.J.P. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, *27*, 1-23. doi:10.1111/j.1469-8986.1990.tb02171.x.

Skinner, N., & Brewer, N. (2004). Adaptive approaches to competition: Challenge appraisals and positive emotion. *Journal of Sport & Exercise Psychology*, *26*, 283-305. doi: 10.1123/jsep.26.2.283.

Tomaka, J., Blascovich, J., Kelsey, R., & Leitten, C. (1993). Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, *65*, 248-260. doi: 10.1037/0022-3514.65.2.248.

Turner, M. J., Jones, M. V, Sheffield, D., Barker, J. B., & Coffee, P. (2014). Manipulating cardiovascular indices of challenge and threat using resource appraisals. *International Journal of Psychophysiology*, *94*, 9-18. doi: 10.1016/j.ijpsycho.2014.07.004.

Turner, M. J., Jones, M. V, Sheffield, D., Slater, M. J., Barker, J. B., & Bell, J. J. (2013). Who thrives under pressure? Predicting the performance of elite academy cricketers using the cardiovascular indicators of challenge and threat states. *Journal of Sport & Exercise Psychology*, *35*, 387-397. doi: 10.1123/jsep.35.4.387.

Turner, M. J., Jones, M. V., Sheffield, D., & Cross, S. L. (2012). Cardiovascular indices of challenge and threat states predict competitive performance. *International Journal of Psychophysiology*, *86*, 48-57. doi: 10.1016/j.ijpsycho.2012.08.004.

Vine, S. J., Freeman, P., Moore, L. J., Chandra-Ramanan, R., & Wilson, M. R. (2013). Evaluating stress as a challenge is associated with superior attentional control and motor skill performance: Testing the predictions of the biopsychosocial model of challenge and threat. *Journal of Experimental Psychology: Applied*, *19*, 185-194. doi: 10.1037/a0034106.

Vine, S. J., Moore, L. J., & Wilson, M. R. (2016). An integrative framework of stress, attention, and visuomotor performance. *Frontiers in Psychology*, *7*. doi: 10.3389/fpsyg.2016.01671.

Vine, S. J., Uiga, L., Lavric, A., Moore, L. J., Tsaneva-Atanasova, K., & Wilson, M. R. (2015). Individual reactions to stress predict performance during a critical aviation incident. *Anxiety, Stress, & Coping*, *28*, 467-477. doi: 10.1080/10615806.2014.986722.

White, J. B. (2008). Fail or flourish? Cognitive appraisal moderates the effect of solo status on performance. *Personality and Social Psychology Bulletin*, *34*, 1171-1184. doi: 10.1177/0146167208318404.

Zanstra, Y. J., Johnston, D. W., & Rasbash, J. (2010). Appraisal predicts hemodynamic reactivity in a naturalistic stressor. *International Journal of Psychophysiology*, *77*, 35-42. doi: 10.1016/j.ijpsycho.2010.04.00.

Table 1

Summary of Included Studies

Reference Number	Authors, Year	N	Design	Population	Mean age (years)	CAT	Main Performance Measures	Results	Effect Sizes
1	Blascovich, Seery, Mugridge, Norris & Weisbuch, 2004	27	CR	Baseball and softball student athletes	N/A	P	Baseball and softball season performance (runs created)	CAT index related to runs created during season; (challenge > threat)	R^2 =.11
2	Chalabaev, Major, Cury & Sarrazin, 2009	27	EX - performance goal	Female undergraduates	N/A	P, C	Multiple-choice score on GRE word problems	Self-reported challenge was unrelated to performance CO and TPR were related to performance, but only examined separately (no CAT index)	N/A
3	Chalabaev, Major, Sarrazin & Cury, 2012	58	EX - Performance goal (approach, avoidance, control)	Female psychology undergraduates	N/A	C	Score on math word problems from GRE practice book	For those participants who received a performance avoidance goal, challenge was associated with better performance than threat	R^2 =.06
4	Feinberg & Aiello, 2010[3]	91	EX - CAT appraisal	Undergraduates	N/A	C, DC	Mental arithmetic score	Threat group outperformed challenge group	*d* = 0.85
		238	EX - CAT appraisal		N/A	C, DC	Mental arithmetic score	Challenge group outperformed threat group	N/A
		54	EX - CAT appraisal		N/A	C, DC	Anagram task score	No significant difference between groups	N/A
5	Gildea, Schneider & Shebilske, 2007	54 154 48	QE QE QE	Adults and adolescents (all male in studies 1 and 3)	22.5 19.9 24.1	C, DC C, DC C, DC	Space Fortress (total scores; used in all studies)	Challenge associated with higher scores than threat across three experiments (not significant in experiment 2)	*d* = 1.09 *d* = 0.29 *d* = 0.65

[3] Studies 1, 2, and 4 from this publication were included in the systematic review. Study 3 was not included because it did not report the results of the main effect comparison between the CAT conditions.

6	Kelsey et al., 2000	162	CR	Psychology undergraduates	N/A	C	Three arithmetic tasks (number of responses, arithmetic errors)	Number of responses inversely correlated with pre-task evaluations (challenge > threat)	N/A
								Arithmetic errors positively correlated with pre-task evaluations	N/A
7	Laborde, Lautenbach & Allen, 2015	96	CR	Male sport science students	24.8	C	Concentration grid exercise (consecutive numbers clicked in two minutes)	CAT not significantly related to visual search task performance	N/A
8	Mendes, Blascovich, Hunter, Lickel & Jost, 2007	47	EX - 2x2 (confederate ethnicity x confederate accent)	Female students	19.6	P	Word-finding task (number and accuracy of responses)	No significant effect of CAT index on performance in a mediation model (marginally significant trend was found)	N/A
9	Moore, Vine, Freeman & Wilson, 2013	30	EX - training (quiet eye, technical)	Undergraduates without golf putting experience	19.7	C	Golf putting (mean radial error)	Evaluations mediated the relationship between group and mean radial error (challenge associated with smaller radial error than threat)	N/A
10	Moore, Vine, Wilson & Freeman, 2012	127	EX – CAT appraisal	Undergraduates without golf putting experience	19.5	P, C, DC	Golf putting (mean radial error)	Lower mean radial error in challenge group	*d* = 0.69
11	Moore, Vine, Wilson & Freeman, 2014	120	EX - 2x2 (effort x support)	Undergraduates	21.6	P, C, DC	Laparoscopic surgery completion time	Low effort group (challenged) outperformed high effort group (threatened)	η^2_p = .12
12	Moore, Vine, Wilson & Freeman, 2015	50	EX - Arousal reappraisal	Participants without golf putting experience	20.2	P, DC	Golf putting (mean radial error)	Arousal reappraisal group was more challenged and performed more accurately (lower error)	*d* = 0.93
13	Moore, Wilson, Vine, Coussens & Freeman, 2013	199	CR	Competitive golfers	36.3	C	Golf competition performance	Challenge evaluations were associated with superior competition performance than threat evaluations	R^2 = .09
		60	EX – CAT appraisal	Experienced golfers	22.9			Challenge group holed higher percentage	*d* = 0.63

						P, C, DC	Golf putting (putts holed, performance error)	of putts than threat group Challenge group had lower error than threat group	*d* = 0.70
14	Moore, Young, Freeman & Sarkar, 2017	100	CR	Participants engaging in club or university level sports	21.9	P, C	Dart-throwing task	Physiological CAT index and cognitive CAT evaluations related to dart throwing performance (challenge > threat)	R^2 = 0.08 R^2 = 0.11
15	O'Connor, Arnold & Maurizio, 2010	138	EX - academic focus	Undergraduates	24.8	C	Negotiation task score	Threat associated with lower negotiation outcomes than challenge	R^2 = .16
		196	EX - 2x2 (CAT appraisal x task structure)	Undergraduates	22.2	C, DC	Negotiation task score	Challenge group scored better negotiation outcome than threat group in the integrative task structure condition only – no main effect	*d* = 0.32
16	Quigley, Barrett & Weinstein, 2002	74	CR	Psychology undergraduates	18 (mode)	P, C	Four verbal mental arithmetic tasks (attempts, number correct)	No relation between cognitive evaluations and performance (number of attempts made, percentage correct responses) No analysis reported for physiological data	N/A
17	Rith-Najarian, McLaughlin, Sheridan & Nock, 2014	79	CR	Adolescents	14.70	P, C	Independently rated speech performance	No relation between physiological and cognitive measures of CAT and performance before task	N/A
18	Roberts, Gale, McGrath & Wilson, 2015	94	CR	Doctors	28 (median)	C	Overall station performance score	CAT predicted station performance (threat < challenge)	N/A
19	Sammy et al., 2017	54	EX – Arousal reappraisal	Undergraduates	21.7	P, C, DC	Dart throwing task	Arousal reappraisal group more challenged on physiological index and evaluations, but not better on dart throwing task	N/A
20	Scheepers, 2017	103	EX – 2x2 (Group status x group legitimacy)	Female undergraduates	21	P, DC	Pattern recognition task	CAT index negatively correlated with performance (higher challenge – lower response times)	R^2 = 0.07
									N/A

								High status group was more challenged and outperformed low status group	
21	Schneider, 2004	59	QE	Undergraduates	21	C, DC	Mental arithmetic performance (responses, errors)	Threat group gave fewer responses Threat group made more errors CAT predicted percent correct (threat < challenge)	*d* = -0.78 *d* = 0.53 *r* = -.33
22	Schneider, Rench, Lyons & Riffle, 2012	152	CR	Psychology undergraduates	20.3	C	Mental arithmetic score (responses and accuracy)	Cognitive evaluations were negatively related with performance (threat < challenge)	N/A
23	Scholl, Moeller, Scheepers, Nuerk & Sassenberg, 2015	50	CR	Undergraduates	20.0	P	Number bisection task[4] errors made	Physiological CAT index was negatively related with number of errors made in all task conditions (challenge associated with less errors than threat)	R^2 = .21 R^2 = .20 R^2 = .11 R^2 = .16
24	Seery, Weisbuch, Hetenyi & Blascovich, 2010	95	CR	Undergraduates	N/A	P	University course grades	Cardiovascular CAT (academic interests speech) predicted course grades (challenge > threat) No association found for general test taking speech	sr^2 = .04 N/A
25	Turner, Jones, Sheffield, Barker & Coffee, 2014	46	EX - resource appraisals	Undergraduates and academic staff	21.7	P, DC	Bean bag throwing score	Performance not significantly higher in challenge group	*d* = 0.50
26	Turner, Jones, Sheffield & Cross, 2012	25	CR	Academic staff members	34.0	P, C	Modified Stroop accuracy and latency	Cardiovascular challenge responses predicted superior performance over threat responses in both studies	R^2 = .16
		21	CR	Female netball players	21.1	P, C	Netball shooting score		R^2 = .14
27	Turner et al., 2013	42	CR	Male elite-level cricketers	16.5	P, C	Cricket batting task (runs awarded by coaching staff)	Physiological CAT associated with batting performance (challenge > threat) Cognitive evaluations not associated with performance	N/A N/A

[4] Analyses were only provided for each of the four sub-conditions of the number bisection task. The authors did not report on a total performance score. Thus, four values are reported in the "Effect Sizes" column.

28	Vine, Freeman, Moore, Chandra-Ramanan & Wilson, 2013	52	CR	Final-year medical students	20.5	P, C	Laparoscopic surgery task completion time	Cognitive evaluations associated with performance under pressure (challenge > threat)	N/A
								Relationship not mediated by physiological CAT index	N/A
29	Vine et al., 2015	16	CR	Active pilots	34.8	C	Flight simulator metrics	Challenge evaluation associated with better performance than threat	$R^2 = .61$
30	White, 2008	128	EX - Solo status manipulation	Undergraduates	19.1	C	Math test scores	Challenge associated with higher math test scores than threat	N/A
							Recall task score	Challenge was only associated with better performance than threat under solo status.	N/A
		90	EX - Solo status manipulation		19.5	C	Math test score	Challenge associated with higher math test scores than threat	N/A

Note. CAT = Challenge and threat variables recorded, CR = Correlational, DC = Dichotomous (challenge group vs. threat group), EX = Experimental, QE = Quasi-experimental, C = Cognitive, P = Physiological.

Table 2

Risk of Bias Assessment Results

Experimental Studies

Reference Number		Random Sequence Generation	Allocation Concealment	Blinding of Participants and Personnel	Blinding of Outcome Assessment	Incomplete Outcome Data	Selective Reporting	Other Sources of Bias
2		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
3		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
4	Study 1	Low	Unclear	Unclear	Low	Low	Unclear	Low
	Study 2	Low	Unclear	Unclear	Low	Low	Unclear	Low
	Study 3	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
8		Low	Low	Low	Unclear	Unclear	Unclear	Low
9		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
10		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
11		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
12		Low	Unclear	Unclear	Unclear	Unclear	Unclear	Low
13	Study 2	Low	Unclear	Unclear	Unclear	Low	Unclear	Low
15	Study 1	Low	Unclear	Unclear	Low	Low	Unclear	Low
	Study 2	Low	Unclear	Unclear	Low	Low	Unclear	Low
19		Low	Unclear	Unclear	Unclear	Low	Unclear	Low
20		Low	Low	Low	Low	Unclear	Unclear	Low
25		Low	Unclear	Unclear	Unclear	High	Unclear	Low
30	Study 1	Unclear	Unclear	Unclear	Low	Low	Unclear	Low
	Study 2	Low	Unclear	Unclear	Low	Low	Unclear	Low

Non-experimental Studies

Reference Number		Blinding of Outcome Assessment	Incomplete Outcome Data	Selective Reporting	Selection of Participants	Confounding Variables	Intervention (Exposure) Measurement
1		Low	Unclear	Unclear	Unclear	Low	Low
5	Study 1	Low	Low	Unclear	Unclear	Low	Low
	Study 2	Low	Low	Unclear	Unclear	Low	Low
	Study 3	Low	Low	Unclear	Unclear	Low	Low
6		Unclear	Low	Unclear	Unclear	Low	Low
7		Low	Low	Unclear	Unclear	Low	Low
13	Study 1	Low	Low	Unclear	Unclear	Low	Low
14		Low	Low	Unclear	Unclear	L ow	Low
16		Unclear	Low	Unclear	Unclear	Low	Low
17		Unclear	Low	Unclear	Unclear	Low	Low
18		Low	Low	Unclear	Low	Low	Low
21		Unclear	High	Unclear	Unclear	Low	Low
22		Unclear	Low	Unclear	Unclear	Low	Low
23		Low	Unclear	Unclear	Unclear	Low	Low
24		Unclear	High	Unclear	Unclear	Low	Low
26	Study 1	Low	Low	Unclear	Unclear	Low	Low
	Study 2	Unclear	Low	Unclear	Unclear	Low	Low
27		Unclear	Unclear	Unclear	Unclear	Low	Low
28		Unclear	Low	Unclear	Unclear	Low	Low
29		Low	Low	Unclear	Low	Low	Low

Note. For the “Reference Number” column coding, please consult the corresponding column in Table 1.

Table 3

Effects on Performance of Cognitive, Physiological, and Dichotomous CAT Variables

CAT Variable	Reference Number	Number of Effects	Percentage of Effects Supporting the Association: Positive	Negative	None	Sum Code
Cognitive	- 2, 3, 6, 7, 9, 13, 14, 15, 16, 17, 18, 22, 27, 28, 29, 30	17	76	0	24	++
Physiological	- 1, 8, 14, 17, 20, 23, 24, 26, 27, 28	12	67	0	33	++
Dichotomous	- 4, 5, 10, 11, 12, 13, 15, 19, 20, 21, 25	15	67	7	27	++

Note. Percentages are rounded to integers so do not always total 100. The "Sum Code" was adapted from Sallis, Prochaska, and Taylor (2000): "0" indicates that 0 – 33% of the supported an association, "?" indicates that 34 – 59% of the studies supported the association, and "+" indicates that 60% or more of the studies supported the association. Codes are doubled ("??", "00", or "++" when four or more studies supported the association/lack of association). For the "Reference Number" column coding, please consult the corresponding column in table 1.

Table 4

Effects of CAT States on Cognitive and Behavioural Task Performance

			Percentage of Effects Supporting the Association			
Performance Outcome	Reference Number	Number of Effects	Positive	Negative	None	Sum Code
Cognitive	- 2, 3, 4, 5, 6, 7, 8, 16, 20, 21, 22, 23, 24, 26, 30	23	65	4	30	++
Behavioural	- 1, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 25, 26, 27, 28, 29	19	84	0	16	++

Note. Percentages are rounded to integers so do not always total 100. The "Sum Code" was adapted from Sallis et al. (2000): "0" indicates that 0 – 33% of the supported an association, "?" indicates that 34 – 59% of the studies supported the association, and "+" indicates that 60% or more of the studies supported the association. Codes are doubled ("??", "00", or "++" when four or more studies supported the association/lack of association). For the "Reference Number" column coding, please consult the corresponding column in table 1.

Table 5

Effects of CAT States on Performance Within Different Research Designs

Research Design	Reference Number	Number of Effects	Percentage of Effects Supporting the Association: Positive	Negative	None	Sum Code
Experimental (direct)	- 4, 10, 13, 15	6	67	17	17	++
Experimental (indirect)	- 2, 3, 8, 9, 11, 12, 15, 19, 20, 25, 30	12	67	0	33	++
Correlational	- 1, 6, 7, 13, 14, 16, 17, 18, 22, 23, 24, 26, 27, 28, 29	18	78	0	22	++
Quasi-Experimental	- 5, 21	4	100	0	0	++

Note. Percentages are rounded to integers so do not always total 100. The "Sum Code" was adapted from Sallis et al. (2000): "0" indicates that 0 – 33% of the supported an association, "?" indicates that 34 – 59% of the studies supported the association, and "+" indicates that 60% or more of the studies supported the association. Codes are doubled ("??", "00", or "++" when four or more studies supported the association/lack of association). For the "Reference Number" column coding, please consult the corresponding column in table 1.

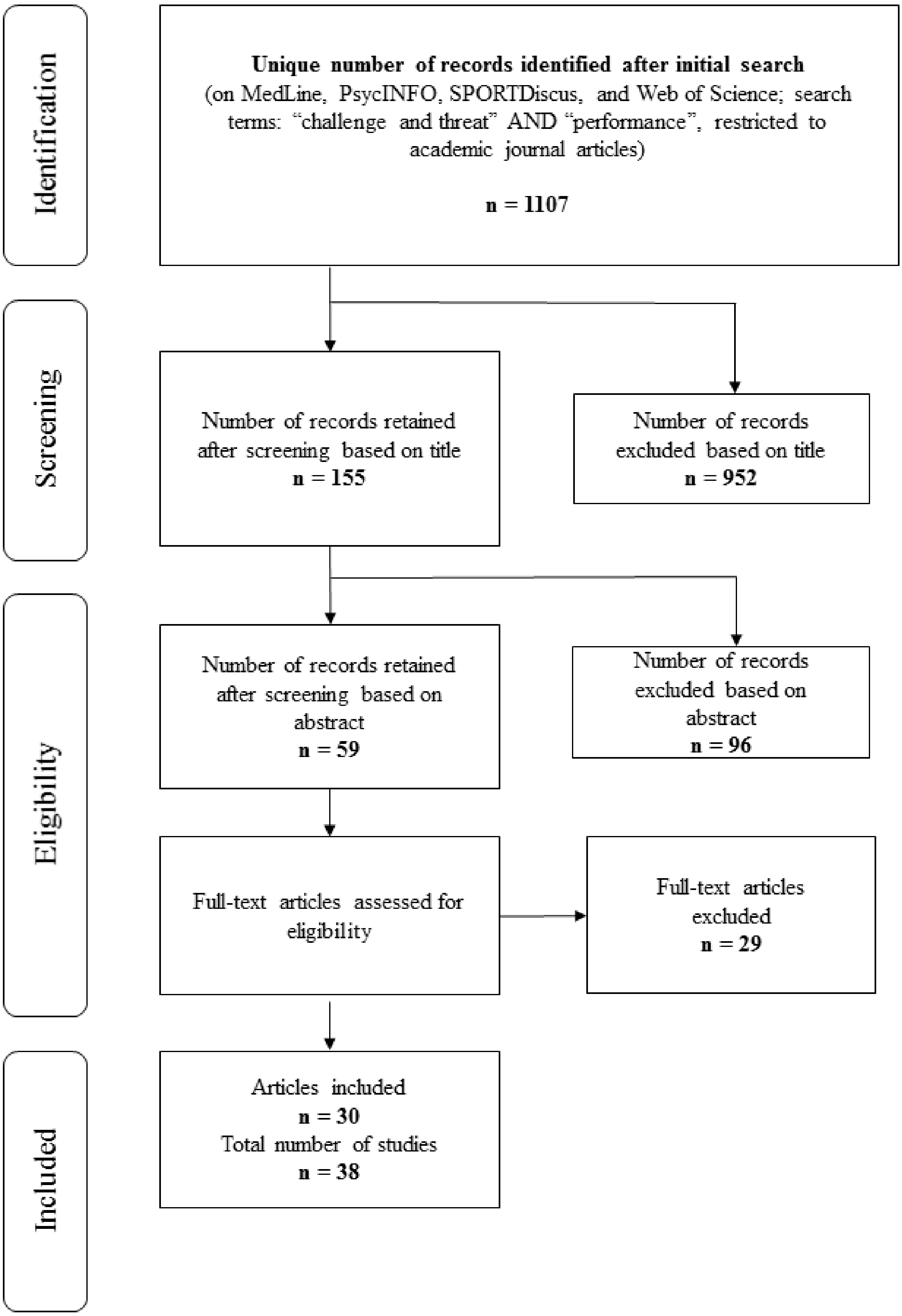

Figure 1. Systematic review search and screening procedure.