Anxiety and Working Memory Capacity: A Meta-Analysis and Narrative Review

Tim P. Moran
Georgia Institute of Technology and Michigan State University

Cognitive deficits are now widely recognized to be an important component of anxiety. In particular, anxiety is thought to restrict the capacity of working memory by competing with task-relevant processes. The evidence for this claim, however, has been mixed. Although some studies have found restricted working memory in anxiety, others have not. Within studies that have found impairments, there is little agreement regarding the boundary conditions of the anxiety/WMC association. The aim of this review is to critically evaluate the evidence for anxiety-related deficits in working memory capacity. First, a meta-analysis of 177 samples (N = 22,061 individuals) demonstrated that self-reported measures of anxiety are reliably related to poorer performance on measures of working memory capacity (g = −.334, p < 10−29). This finding was consistent across complex span (e.g., OSPAN; g = −.342, k = 30, N = 3,196, p = .000001), simple span (e.g., digit span; g = −.318, k = 127, N = 17,547, p < 10−17), and dynamic span tasks (e.g., N-Back; g = −.437, k = 20, N = 1,318, p = .000003). Second, a narrative review of the literature revealed that anxiety, whether self-reported or experimentally induced, is related to poorer performance across a wide variety of tasks. Finally, the review identified a number of methodological limitations common in the literature as well as avenues for future research.

Keywords: anxiety, meta-analysis, review, working memory capacity

Supplemental materials: http://dx.doi.org/10.1037/bul0000051.supp

Outside of the laboratory, national etiological studies have found that individuals diagnosed with anxiety disorders are 1.4 times more likely to drop out of high school (Kessler, Foster, Saunders & Stang, 1995), that approximately half of anxious patients fail to graduate high school because of anxiety (Van Ameringen, Mancini, & Farvolden, 2003), and that subclinical anxiety negatively impacts academic performance (e.g., Crozier & Hostetller, 2003; Gumora & Arsenio, 2002).

In attempting to explain the widespread relationships between anxiety and cognition, researchers have long posited that anxiety restricts the capacity of working memory (WM; e.g., Lewinski, 1945; Moldawsky & Moldawsky, 1952; Rashkis & Welsh, 1946) and have recently begun to formally include WM as a key explanatory mechanism in models of anxiety (e.g., Eysenck et al., 1992, 2007; Outim et al., 2009; Robinson et al., 2013; Shackman et al., 2006; also see Posner & Rothbart, 2000). However, as the quotes presented above imply, several inconsistent findings have accumulated in the literature. A number of studies have indeed documented impaired WMC in anxiety (e.g., Ashcraft & Kirk, 2001; Darke, 1988; Maloney et al., 2010; Shackman et al., 2006); among these studies, however, there is little agreement regarding whether anxiety is related to WMC in a domain-specific way (e.g., phonological storage vs. domain-general attention) and which domains are most closely related to anxiety (e.g., phonological, spatial, visual; cf. Ashcraft & Kirk, 2001; Markham & Darke, 1991; Moriya & Sugiura, 2012; Shackman et al., 2006). In contrast, several studies have found near-zero associations between anxiety and WMC (e.g., Johnson & Gronlund, 2009; Moritz et al., 2002; Salthouse, 2012) or have found improved WMC in anxiety (e.g., Moriya & Sugiura, 2012, 2013).

“Anxiety . . . interferes with processing only in the articulatory loop.”
—Ikeda et al. (1996)

“Anxiety impairs central executive functioning . . . and does not impair functioning of the “slave” systems.”
—Eysenck et al. (2005)

“Anxiety selectively disrupts performance of spatial WM.”
—Shackman et al. (2006)

“Anxiety was uncorrelated with WM capacity.”
—Johnson and Gronlund (2009)

“Anxiety is positively associated with visual working memory.”
—Moriya & Sugiura, 2012

Over the last 60 years, there has been a great deal of work detailing the relationship between anxiety and performance on measures of cognitive abilities. For example, measures of anxiety are reliably associated with increased interference from distractors in search tasks (e.g., Bar-Haim et al., 2007; Moser et al., 2012), poorer performance on measures of reading comprehension (Calvo et al., 1992), and mathematical problem solving (e.g., Ashcraft & Kirk, 2001), as well as lower scores on standardized tests of intelligence and general aptitude/achievement (Hembree, 1988; see Eysenck et al., 2007 for a review).

Correspondence concerning this article should be addressed to Tim P. Moran, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332. E-mail: timothy.moran@psych.gatech.edu
The aim of this review was to examine the evidence for impaired working memory capacity in anxiety and provide a quantitative assessment of this association. To provide some context, this review will begin with brief overviews of the measurement of WMC as well as theories of anxiety and cognition. This is followed by a critical review of the literature on anxiety and working memory capacity.

Working Memory and Its Measurement

Working memory is generally described as a system that retains access to a limited amount of information in the service of more complex cognitive operations. Although there exist many models of WM, the model most often identified in anxiety research is the influential multiple component model (Baddeley, 1986, 2007). This model partitions the WM system into separable domain-specific stores—that is, the phonological loop and visuospatial sketchpad—as well as a domain-general central executive. The central executive acts as more of an attention system that is utilized during controlled processing—particularly in maintaining task goals and reducing interference from distraction and prepotent responses. As will be discussed below, anxiety is most commonly associated with attentional control aspects of the central executive (although the multicomponent model has been the most widely adopted model of WM by anxiety researchers, links between anxiety and other models of WM will become apparent throughout the review because of the focus on attentional control; e.g., Engle, Kane, & Tuholski, 1999; Engle, Tuholski, Laughlin, & Conway, 1999; Engle, 2002).

A large body of research has been dedicated to studying individual differences in the capacity of working memory (WMC). The earliest attempts to measure WMC in anxiety utilized short-term memory tasks or “simple” measures of span—most often the digit span—which only require the storage and rehearsal of the to-be-remembered items. For example, a participant might be required to recall lists of digits in either the same order as presented or in the reverse order. Span is typically quantified as the average number of items correctly recalled or as the longest list that was recalled perfectly. More recently, studies have begun using measures of “complex” span in which the presentation of the memoranda is interleaved with a demanding secondary task. For example, in Daneman and Carpenter’s (1980) reading span task participants were required to recall a list of serially presented words. Between each word, participants were required to confirm the veracity of a sentence (also see Turner & Engle, 1989). The crucial distinction between simple and complex span tasks appears to be that the secondary task removes the memoranda from immediate awareness and controlled attention is necessary to maintain access to it outside of awareness (e.g., Kane et al., 2007; Shipstead et al., 2014; see Unsworth & Engle, 2007a, 2007b for a related account).

Research on individual differences in WMC implies both a practical and a structural distinction between measures of simple and complex span. First, whereas complex span tasks reliably predict abilities such as reading comprehension, emotion regulation, and resistance to distraction (among others; Kane et al., 2001; Schmeichel, Volokhov, & Demaree, 2008; Redick & Engle, 2006; Turner & Engle, 1989), relationships between simple span and higher cognition are much less consistent (Carpenter, Miyake, & Just, 1995; Daneman & Merikle, 1996). Second, Kane and colleagues (2004) conducted a large psychometric investigation of simple and complex WMC tasks. They found (a) that complex span tasks loaded on a domain-general “WMC” factor and (b) that simple span tasks loaded on domain-specific phonological and visuospatial STM factors which were correlated with, but distinct from, the WMC factor.

Although it is clear that measures of simple and complex span are psychometrically separable and differ in their ability to predict complex cognition, simple span measures have figured prominently in research on anxiety due to the fact that they are included on standard neuropsychological batteries (e.g., the Wechsler Intelligence Scale). The current review will therefore include findings from both complex and simple spans. Referring to simple span measures as “working memory” should not be interpreted as endorsing a particular theoretical view; rather, it reflects the current state of how anxiety researchers view the WM system. When these tasks must be distinguished, they will be referred to separately as measures of “simple” and “complex” span.

Recently, researchers have begun to favor more dynamic tasks as an index of individual differences in WMC. In such tasks, participants are required to monitor a continuous stream of to-be-remembered items and update the contents of WM in real time by adding new items as potential targets and dropping old items as they become unnecessary for task performance. For example, in the N-Back, participants monitor a continuous stream of items (typically letters or numbers) while maintaining the most recent “n” items (where n typically ranges between 0 and 4). The job of the participant is to determine whether the current item matches the one that was presented n items ago thereby requiring participants to continuously update representations of the most recent n items.

Because the level of load can be parametrically varied and that phonological and spatial versions of the N-Back appear to have equivalent psychometric properties (e.g., Shackman et al., 2006), the N-Back has recently become a popular alternative to simple and complex WMC tasks among anxiety researchers (e.g., Shackman et al., 2006; Vytal et al., 2012). However, the relationship between the dynamic span tasks and other measures of WMC appears to be somewhat tenuous. For example, a recent meta-analysis (Redick & Lindsey, 2013) found that the correlations between the N-Back and measures of complex span (r = .2) and simple span (r = .25) were quite modest suggesting that they are unlikely to be measuring the same underlying WM operations. Similarly, Shipstead and colleagues (2014) conducted an analysis of potential mechanisms underlying individual differences in WMC. This study found that performance on complex span tasks was related to performance on measures of attentional control whereas running span tasks—that is, tasks in which a continuous stream of memoranda are presented and participants must report the most recent n items when the list is terminated—were much more strongly related to storage capacity (a latent variable derived from simple span tasks). Although attentional control and storage capacity were correlated, these results imply that individual differences in complex span and running span tasks rely on separable mechanisms.

As with simple span measures, these dynamic span measures will also be included in the current review as they have been used by many researchers to draw conclusions regarding WMC and...
anxiety. When distinctions are drawn between types of tasks, these will be separated from simple and complex spans and referred to as “dynamic” spans.

Finally, a number of studies have imported the dual-task methodology from experimental studies on working memory (e.g., Baddeley, 1986, 2007). In these studies, participants simultaneously perform combinations of tasks that are thought to involve the phonological loop (e.g., articulatory suppression), visuospatial sketchpad (e.g., tapping simple spatial patterns), and central executive (e.g., random generation) under the assumption that, if two tasks rely on the same underlying mechanisms, performance in those tasks will suffer. In these studies, anxiety-related deficits in a WM process are then inferred from more pronounced dual-task effects for anxious individuals.

**Theories of Anxiety and WMC**

Cognitive theories of anxiety differ with respect to the degree of domain specificity of the anxiety/WMC relationship, the hypothesized level of analysis (e.g., psychological vs. biological) at which anxiety and WMC are related, and the direction of the causal arrow. However, they tend to share two features. First, “anxiety” is considered to be a heterogeneous construct that can be psychometrically and physiologically decomposed into at least two dimensions: worry/apprehension and arousal/emotionality (e.g., Andrews & Borkovec, 1988; Barlow et al., 1996; Borkovec et al., 1983; Gorsuch, 1966; Heller & Nitschke, 1998; Heller et al., 1997a; Moris, 1967; Nitschke et al., 2001; Sassenrath, 1964). “Worry” refers to verbal rumination regarding the potential negative outcomes of future events and is characteristic of generalized anxiety disorder. “Arousal”, on the other hand, refers to physiological hyperarousal and somatic tension (i.e., dizziness, high heart rate, sweaty palms, hypervigilance etc.) and is characteristic of stressful circumstances and panic (e.g., Watson et al., 1995). Although most theories accept this distinction, there are differences with respect to the relative importance placed on worry and arousal.

Second, most theories propose that the anxiety/WMC relationship can be attributed to interference or competition between anxiety-related processes and task-related processes. For example, given that worry involves verbal rumination, it is often thought to interfere with the maintenance of phonological information (e.g., Eysenck & Calvo, 1992; Sarason, 1988). Along these lines, theories differ with respect to the domain-specificity of the interference as well as at what level the interference occurs.

According to several theories, worry acts as a type of dual-task that interferes with cognitive task performance. For example, Sarason (1988) proposed that anxiety is characterized by task-irrelevant thoughts and “proneness to self-preoccupation.” These worrisome thoughts were assumed to reduce the attention that could be allocated to task performance. Perhaps the most influential accounts of the relationship between anxiety and performance have been offered by Eysenck and colleagues’ (Eysenck, 1979, 1982; Eysenck & Calvo, 1992; Eysenck et al., 2007) processing efficiency theory (PET) and attentional control theory (ACT). PET/ACT agree with Sarason (1988) insofar as worry is thought to affect attentional processes; however, ACT further suggests that anxiety affects specific executive processes such as inhibition. In addition to these effects on domain-general executive processes, PET/ACT proposes that worry also has domain-specific effects. That is, based on research indicating that worry is primarily represented as a phonological code (e.g., Borkovec, 1994; Borkovec et al., 1983), PET/ACT proposes that worry also preempts the capacity of phonological storage.

Shackman and colleagues (2006) proposed a somewhat different view of the anxiety/WMC association. Previous electroencephalographic research conducted in healthy adults (Heller et al., 1997a, b; Muri et al., 2000) and neuropsychological research conducted in psychiatric patients (Davidson et al., 2002), lesion patients (Bechara et al., 2000; Hornak et al., 2004), and nonhuman primates (Kalin et al., 1998) implicates regions in the right prefrontal cortex and right posterior-parietal cortex in supporting anxious arousal (see Murphy et al., 2003 for a meta-analysis). That is, arousing states, as well as the trait tendency to experience physiological arousal under stress, are associated with increased activation in right PFC and PPC regions (Keller et al., 2000; Murphy et al., 2003; Nitschke et al., 2000). In addition to anxious arousal, the right mid- and ventrolateral-PFC have been associated with vigilance and spatial working memory processes (Lawrence et al., 2003; Manoach et al., 2004; Sturm & Willmes, 2001). Similarly, activity in the right PPC has been associated with spatial attention and working memory (Bender, 1952; Corballis et al., 2002; Heller et al., 1997a, b; Muri et al., 2000). Based on these findings, Shackman et al. (2006) propose that anxious arousal competes with other processes localized to right PFC and PPC, such as spatial working memory, for access to limited neural resources. Additionally, given that arousal is difficult to downregulate as it aids in the survival of an organism by preparing the organism to deal with potential threats (e.g., scanning the environment for threats; Cornwell et al., 2011), it is assumed that this competition will generally be resolved in favor of processes supporting anxious arousal rather than cognitive task performance. Thus, anxious arousal is assumed to interfere with the maintenance of spatial information.

More recent models (Robinson et al., 2013; Vytal et al., 2012, 2013) have proposed an interaction between dimensions of anxiety (worry vs. arousal), working memory modality (phonological vs. spatial) and task difficulty. First, they note that both worry and phonological WMC engage regions in the dorsal, medial and prefrontal cortex. Additionally, anxious arousal and spatial working memory both engage medial and ventral prefrontal regions (Clark et al., 2003; Dalton et al., 2005; D’Esposito et al., 1998; Engels et al., 2007; Kalisch et al., 2006; Paulesu et al., 2010; Silk et al., 2010). Thus, like Shackman et al. (2006), Vytal and colleagues propose domain-specific interference from anxiety (e.g., that arousal interferes with spatial processes specifically and worry interferes with phonological processes specifically). Second, they note that worrisome thoughts are more amenable to downregulation by top-down control than arousal (e.g., Kalisch et al., 2006; Ochsner & Gross, 2005; Pessoa, Padmala, & Morland, 2005). Given these findings, Vytal and colleagues propose that worry interferes with phonological processes at low levels of task difficulty; at high levels of difficulty, however, task-relevant processes take precedence and high-worry individuals engage in some compensatory effort or strategic shift to improve performance. Third, anxious arousal is thought to prime defensive mechanisms (e.g., hypervigilance) meant to protect the organism in the face of potential threats (Lang, Bradley, & Cuthbert, 1998). Importantly,
these arousal-related processes that subserve survival are less amendable to top-down regulation than worry (Cornwell et al., 2011). Thus, arousal is expected to disrupt spatial WM across all levels of task difficulty. This account is somewhat unique in that it predicts domain-general effects to be produced in domain-specific ways. That is, both spatial and phonological storage are expected to be impacted but by different mechanisms.

Whereas the preceding discussion has tended to focus on the detrimental effects of anxiety on cognition, models emerging from the clinical literature have tended to hold the opposite assumption—namely, that cognitive deficits and biases are causal factors in the development of anxiety (e.g., Beck, 1995; Beck & Clark, 1997; Mathews & MacLeod, 2005; McNally et al., 1994; Ouimet et al., 2009; van den Hout et al., 1995). To test this, MacLeod et al. (2002) developed versions of a spatial cueing paradigm in which neutral and threatening words preceded a target. For one group, the target always appeared following the threatening word; for the other group, the target always followed the neutral word. Following the spatial cueing task, the researchers induced stress by requiring participants to attempt to solve unsolvable anagrams while being recorded. Participants in the attend-threat group reported greater increases in negative mood in response to the stress induction.

Similarly, it is believed that negative interpretation biases (i.e., the tendency to interpret ambiguous events as threatening), difficulty inhibiting intrusive thoughts, and difficulty regulating one’s emotions contribute to the development of anxious pathology (e.g., Bar-Haim et al., 2007; Campbell-Sills & Barlow, 2007; Ehlers & Clark, 2000; Mathews & MacLeod, 2005). Importantly, many of these processes are known to rely on WM. For example, lower working memory has been shown to predict greater distractor interference (Lavie, 2005; Lavie et al., 2004), reduced ability to regulate one’s emotions (Schmeichel, 2007; Schmeichel et al., 2008), and more frequent intrusive thoughts (Bomyea et al., 2012; Brewin & Smart, 2005). This has led some researchers to postulate that restricted WM itself may be a risk factor for later anxiety (e.g., Ouimet et al., 2009; also see Posner & Rothbart, 2000). At present, no theoretical account makes clear predictions regarding domain-general or domain-specific WM systems or which dimension(s) of anxiety should be most sensitive to individual differences in WM.

Overall, there appears to be a great deal of agreement that anxiety should be associated with impaired performance on working memory tasks. Additionally, many theoretical approaches propose that this relationship results from some type of competition for access to limited resources. However, there appears to be considerable disagreement regarding which dimension(s) of anxiety should be associated with which working memory domains (e.g., phonological vs. spatial) as well as the direction of the causal arrow. Additionally, the empirical findings have proven to be somewhat equivocal insofar as many studies have failed to find effects and, of the studies that have found effects, there is little agreement as to the circumstances in which task performance relates to anxiety.

The review of the literature on anxiety and working memory will be split into two broad sections. The first section will focus on correlational studies in which scores on measures of working memory were compared between high- and low-anxious individuals, as defined by scores on self-report measures of anxiety and diagnostic assessments. The second section will focus on experimental studies in which either anxiety or working memory was experimentally manipulated. Sections will be further divided based on methodology.

**Review of the Literature on Anxiety and Working Memory Capacity**

**Correlational Studies**

**Studies of bivariate associations.** The majority of studies have involved comparing WM task performance between high- and low-anxious individuals or presenting bivariate correlations between measures of anxiety and WM. Given the large number of studies that were identified (>100), these studies were synthesized meta-analytically rather than via a narrative review. The meta-analytic methods are described in detail in the Supplemental Materials.

The focal effect size for this analysis was Hedges’ g (Hedges, 1981). Hedges’ g is computed from the more commonly used Cohen’s d and can be interpreted in the same way. Effect sizes were coded such that a negative g value indicates that anxious individuals show reduced WM whereas a positive value indicates that anxious individuals show greater capacities. A total of 177 samples, consisting of 22,061 individuals, met criteria for inclusion. Effect sizes were pooled within the framework of a random-effects model.

In addition to examining the overall relationship between measures of anxiety and WM, several potential moderators were also examined in order to evaluate the generality of the anxiety/WM relationship. Table 1 lists the moderator variables and describes how they were defined.

**Meta-analytic results.** Across all 177 samples, measures of anxiety were significantly associated with lower scores on measures of WM (g = −.334, k = 177, N = 22,061, 95% CI: −.392; −.276, p < 10−28). Analyses of the primary moderators of interest are shown in Table 2. Most population-related moderators (i.e., age, trait vs. state, worry vs. arousal) were not significant and all subcategories within those moderators produced significant associations with WM. Symptom severity did significantly moderate the relationship between anxiety symptoms and WM. Although effect sizes were significant in both clinical and subclinical samples, effect sizes for clinical populations were significantly larger.

To examine dimensions of anxiety, studies were grouped based on whether they included measures of worry or arousal (note that few studies examined arousal directly; most of these studies examined syndromes characterized by arousal symptoms such as panic disorder). Interestingly, both worry and arousal predicted performance on both phonological and spatial tasks. This finding would seem to contradict the more domain-specific predictions of existing theories. However, it should be borne in mind that, although worry and arousal tend to form separate dimensions, they are not uncorrelated (e.g., Nitschke et al., 2001). Thus, even when one dimension of anxiety is examined directly, the measure may be
contaminated by another dimension (e.g., worry may influence scores on an arousal measure).

With respect to procedural moderators, the results were somewhat more complex. The type of span used (i.e., simple, complex or dynamic) was not a significant moderator; effect sizes were significant for all subcategories. However, span domain (i.e., visual vs. phonological vs. spatial) was a significant moderator; specifically, anxiety was negatively associated with scores on measures of phonological and spatial working memory but not significantly positively associated with measures of visual working memory.

A more detailed examination of studies of visual WMC further complicates these findings. The majority of the studies of visual WMC were conducted in socially anxious students attending East Asian universities thus hinting at possible cultural differences, differences between types of anxiety, and/or a culture × anxiety interaction. A series of additional analyses were conducted to attempt to test these possibilities. First, within studies of visual WMC, studies conducted in East Asia showed a significant positive association (g = .704, k = 5, N = 201, p < .001), whereas studies conducted within the United States and Europe showed a significant negative association (g = -.317, k = 6, N = 1,029, p = .001; Q(I) = 29.05, p < .001). However, culture and type of anxiety were nearly completely confounded in this analysis as the American/European studies generally included state/trait anxiety whereas 4 of 5 Asian studies were conducted with socially anxious college students. Within Asian studies of visual WMC, studies involving socially phobic individuals showed significant positive effect sizes (g = .850, k = 4, N = 164, p < .001), whereas the effect size for the study of trait anxiety was considerably smaller and nonsignificant (g = .190, N = 37, p = .56). Although these results are based on a very small number of studies and should be interpreted cautiously, it suggests that this cultural influence may not hold across the board; rather, it appears that it may be specific to social anxiety.

The possibilities of a main effect of type of anxiety and a main effect of culture were tested. First, to test the possibility that social anxiety, regardless of culture, is associated with greater working memory capacity, the effect sizes for Asian social phobics (g = .850, k = 4, N = 164, p < .001) was compared against that for European/American social phobics (g = -.269, k = 7, N = 341, p = .01). Contrary to this notion, these effect sizes were significantly different (Q(I) = 30.47, p < .001). Second, the possibility that anxiety predicts greater WMC across all modalities in Asian populations was tested. However, the results did not support this possibility. Anxiety predicted poorer performance for both spatial (g = -.502, k = 1, N = 65, p = .19) and phonological (g = -.523, k = 4, N = 258, p = .01) material and these were

<table>
<thead>
<tr>
<th>Moderator and subcategories</th>
<th>Coding definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>The study included participants age 18 and over</td>
</tr>
<tr>
<td>Adults</td>
<td>The study included participants under 18</td>
</tr>
<tr>
<td>Anxiety dimension</td>
<td>The study included an explicit measure of anxious arousal (e.g. the Mood and Anxiety Symptom Questionnaire) or included clinically diagnosed disorders characterized by arousal symptoms (e.g. panic, phobia etc.)</td>
</tr>
<tr>
<td>Arousal</td>
<td>The study included an explicit measure of worry (e.g. the Penn State Worry Questionnaire) or included clinically diagnosed disorders characterized by worry symptoms (e.g. generalized anxiety disorder)</td>
</tr>
<tr>
<td>Worry</td>
<td>The study did not include an explicit measure of arousal or worry</td>
</tr>
<tr>
<td>Other</td>
<td>The study did not include an explicit measure of arousal or worry</td>
</tr>
<tr>
<td>Trait vs. State</td>
<td>The Study included an explicit measure of trait anxiety (e.g. the STAI-Trait)</td>
</tr>
<tr>
<td>Trait State</td>
<td>The Study included an explicit measure of state anxiety (e.g. the STAI-State)</td>
</tr>
<tr>
<td>Severity</td>
<td>The study included clinically diagnosed anxiety disorders using the DSM or ICD classification systems</td>
</tr>
<tr>
<td>Nonclinical</td>
<td>The study included self-reported anxiety using a questionnaire</td>
</tr>
<tr>
<td>WM Span type</td>
<td>The study used a WM task that combined storage with an interleaved processing task (e.g. OSPAN)</td>
</tr>
<tr>
<td>Complex</td>
<td>The study included a WM task that included temporary storage only (e.g. Digit Span)</td>
</tr>
<tr>
<td>Simple</td>
<td>The study included a WM task that required continually updating WM representations (e.g. N-Back)</td>
</tr>
<tr>
<td>WM Span domain</td>
<td>The study included a WM task that requires storing spatial locations</td>
</tr>
<tr>
<td>Spatial</td>
<td>The study included a WM task that requires storing phonologically-rehearsable material (e.g. letters, digits)</td>
</tr>
<tr>
<td>Visual</td>
<td>The study included a WM task that requires storing non-spatial visual features (e.g. color)</td>
</tr>
<tr>
<td>Manuscript type</td>
<td>The study was published in a peer-reviewed journal</td>
</tr>
<tr>
<td>Peer-reviewed</td>
<td>The study was conducted as part of a thesis or dissertation but was not published in a peer-reviewed journal</td>
</tr>
</tbody>
</table>

Table 1

Coding Definitions for Moderators in the Meta-Analysis
significantly different than the effect sizes reported for visual material, $Q(2) = 20.12, p < .001$. Thus, these results point to a specific culture × anxiety type × WM domain interaction (see Supplemental Materials for a reanalysis with outlying studies removed).

Finally, to determine whether published studies systematically produced larger effect sizes than unpublished work, studies published in peer-reviewed journals were compared to otherwise unpublished theses and dissertations. As shown in Table 2, peer-reviewed studies and theses/dissertations produced very similar results.

The results of this meta-analysis provide evidence for a robust association between anxiety symptoms and performance on measures of WMC. Although it appears to hold under a variety of types of tasks, anxiety subtypes as well as levels of symptom severity, this effect appears to be of moderate magnitude. Interestingly, the results of this analysis suggest that both worry and arousal symptoms predict impaired performance on tasks using both phonological and spatial material. This may point to more domain-general deficits than some theorists have suggested, although more careful analyses are required to partition the effects of anxiety domain and WM domain.

The most unexpected result of this meta-analysis was the finding that anxiety predicted greater capacity for nonspatial visual features in East Asian populations. These results hint at a culture × type of anxiety × WM domain interaction. This will be returned to in the discussion.

**Dual-task studies.** Studies of anxiety and dual-tasking have proceeded under the assumption that if a given task relies on a component of WM, then performance on that task will suffer when the participant is required to simultaneously perform an additional task which also relies on that WM component. Importantly, if anxiety is associated with restricted working memory capacity, the additional load should impact anxious individuals’ performance to a greater degree than nonanxious individuals.

Early studies of dual-tasking (Macleod & Donnellan, 1993; Derakshan & Eysenck, 1998) tended to find preserved WMC in anxious participants. In these studies, participants completed a grammatical reasoning task while under conditions of varying load. For example, in Macleod and Donnellan (1993), participants

### Table 2

Meta-Analytic Results for High-Anxiety Versus Low-Anxiety Comparisons

<table>
<thead>
<tr>
<th>Moderator</th>
<th>$g$ (95% CI)</th>
<th>$k$</th>
<th>$N$</th>
<th>$p$</th>
<th>$Q^2 (df)$</th>
<th>$p^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>–.351 (–.414; –.287)</td>
<td>152</td>
<td>19,989</td>
<td>&lt;.001</td>
<td>1.82 (1)</td>
<td>.18</td>
</tr>
<tr>
<td>Anxiety dimension</td>
<td>.89 (1)</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arousal</td>
<td>–.306 (–.436; –.175)</td>
<td>52</td>
<td>2,711</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Worry</td>
<td>–.446 (–.707; –.185)</td>
<td>11</td>
<td>1,105</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Severity</td>
<td>–.292 (–.424; –.161)</td>
<td>26</td>
<td>10,967</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>WM Span domain</td>
<td>.89 (1)</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td>–.410 (–.519; –.300)</td>
<td>54</td>
<td>4,310</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Phonological</td>
<td>–.336 (–.411; –.261)</td>
<td>110</td>
<td>16,350</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Visual</td>
<td>.008 (–.222; .238)</td>
<td>11</td>
<td>1,230</td>
<td>.95</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Anxiety dimension × Domain: Arousal</td>
<td>.89 (1)</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial</td>
<td>–.388 (–.644; –.132)</td>
<td>15</td>
<td>793</td>
<td>.003</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Phonological</td>
<td>–.376 (–.549; –.204)</td>
<td>31</td>
<td>1,650</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Visual</td>
<td>.271 (–.130; .672)</td>
<td>6</td>
<td>268</td>
<td>.19</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Citation type</td>
<td>.89 (1)</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer-reviewed</td>
<td>–.335 (–.397; –.273)</td>
<td>155</td>
<td>18,808</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
<tr>
<td>Thesis/dissertation</td>
<td>–.324 (–.479; –.168)</td>
<td>22</td>
<td>3,253</td>
<td>&lt;.001</td>
<td>.89 (1)</td>
<td>.35</td>
</tr>
</tbody>
</table>

$a$ $Q$ statistic and $p$ value for comparison between subcategories of a moderator.
were presented with a letter pair (e.g., YX) and a sentence describing the pair (e.g., Y is preceded by X) and participants were required to judge the veracity of the sentence. This was interleaved with a secondary task in which participants maintained either six zeros (i.e., low load) or six random digits (i.e., high load). In these studies, the high load condition reduced accuracy, but this effect was not modulated by anxiety. Richards et al. (2000), on the other hand, found that test-anxious individuals performed less accurately on the reasoning task. However, this finding was independent of load suggesting it did not result from restricted WMC.

Interestingly, all of those studies paired a fairly demanding task (e.g., grammatical reasoning) with a relatively undemanding task (e.g., rehearsing digits) that relied on phonological storage. This type of secondary task probably does not require the involvement of attentional processes (e.g., Baddeley, 2007; Cowan, 2005). Indeed, Eysenck, Payne and Derakshan (2005) examined this issue by varying task difficulty as well as task modality (i.e., phonological vs. visuospatial). In this study, participants completed the Corsi Blocks task (Corsi, 1972) which requires participants to reproduce a spatial sequence that has been recently produced by the experimenter. Given that Corsi Block performance is impaired when it is presented concurrently with a random number generation task (Vecchi & Richardson, 2001), Eysenck et al. (2005) assumed that performing this task would require executive involvement. The Corsi Block task was paired with secondary tasks that varied in difficulty and involved either phonological or spatial processes. When the secondary task was articulatory suppression (easy/phonological) which, like the string of digits used by MacLeod and Donnellan (1993), presumably relied on phonological storage, high- and low-trait-anxious individuals performed equally well. Similarly, high- and low-trait-anxious participants showed equivalent performance during both spatial-task conditions. In the difficult/phonological condition, however, high-anxious individuals committed more errors on both tasks than did low-anxious individuals.

Another set of studies have taken an approach to dual-tasking that might be considered more ecologically valid. In all of these studies, participants completed a version of a random generation task—a task in which participants produce random sequences of items (e.g., digits). This task is thought to be highly attentionally demanding and reliant on frontal lobe-mediated executive processes (Baddeley et al., 1998; Baddeley et al., 1984; Milner, 1982) given that it requires avoiding stereotyped and repetitious responses. Importantly, the secondary task in each of these studies was worry itself. That is, participants were instructed to complete the random generation task while either worrying about a self-relevant topic or under some control condition. The advantage of this method is that it maps onto some of the predictions of theory (e.g., Sarason, 1988) far more directly than other dual-task methodologies. This methodology also allows for a more direct test of the causal role of worry than dual-task studies in which worry is not explicitly manipulated.

In Hayes et al. (2008), participants completed a random generation task while either worrying about an important negative topic or thinking about a positive topic. Trait-worriers in the worry condition produced less random responses (Towsie & Neil, 1998) than both trait worriers in the positive-thought condition and nonworriers. Stefanopoulou et al. (2014) replicated these findings in a sample of patients suffering from generalized anxiety disorder, a condition primarily characterized by persistent and uncontrollable worry (APA, 2013). Interestingly, these findings appear to be specific to worry in a phonological form as negative visual imagery does not seem to have these effects (Leigh & Hirsch, 2011).

Although, given that worry appears to be inherently phonological (Borkovec, 1994; Borkovec et al., 1983), one could argue that participants in this study were simply less capable of engaging in “visual worry.”

Rapee (1993), like Eysenck et al. (2005), took a more comprehensive approach by evaluating the generality of these findings to more domain-specific stores. In this study, participants completed two worry sessions wherein they were asked to worry about a personally relevant topic. In the single task condition, worrying was accomplished in isolation. In the dual-task condition, participants worried while simultaneously completing either easy (articulatory suppression/finger tapping) or difficult (spoken random generation/random key pressing) phonological and spatial tasks. Participants reported fewer thoughts judged to be “worry-related” in the spoken random number generation condition than any other condition.

Studies of dual-tasking seem to suggest that anxiety is most closely related to deficits in domain-general attentional processes. That is, when studies pair a primary task with a secondary loading task that primarily relies on phonological storage, anxious individuals are equally as accurate as nonanxious individuals (although they are slowed; Derakshan & Eysenck, 1998; MacLeod & Donnellan, 1993). However, when two attentionally demanding tasks are paired (i.e., Eysenck et al., 2005), anxious individuals begin to show impaired performance. Note that, in Eysenck et al. (2005), the largest performance decrements were observed when the Corsi Blocks were paired with counting backward—that is, when a visuospatial task was paired with a phonological task. Thus, this finding cannot be attributed to interference produced by superficially similar tasks or “overloading” a storage buffer. Additionally, a number of studies (Hayes et al., 2008; Leigh & Hirsch, 2011; Stefanopoulou et al., 2014) have found that worry interferes with performance in a random generation task. Given the relative homogeneity of methods used in random generation tasks, effect sizes were pooled to yield a moderate, but significant, decrease in randomness while engaging in worry (g = −.377, k = 3, N = 127, p < .001).

Finally, dual-tasking studies have provided some evidence of domain-specificity as well. For example, Leigh and Hirsch (2011) found that phonological worry, but not negative visual imagery, affected random generation performance. Similarly, Rapee (1993) found that only spoken random generation affected worry, which involved speaking aloud, whereas random key pressing did not. Given that the studies showing the strongest evidence for domain-specificity examined worry specifically, it is possible that worry has both domain-general and domain-specific effects on the WM system. However, this has not been shown unambiguously. For example, the meta-analysis found that anxiety in general, and worry in particular, predicted (a) both simple and complex spans and (b) memory for both phonological and spatial material. Given that tasks measuring domain-general processes as well as domain-specific stores are highly correlated (e.g., Kane et al., 2004), it is possible that worry also predicts the variance that is shared between these types of tasks. Alternately, these differences may be attributable to the tasks themselves. As noted earlier, given that
worry tends to take a phonological form (Borkovec, 1994; Borkovec et al., 1983), participants may have found it difficult to comply with the task demands when asked to engage in negative visual imagery.

**Prospective studies.** The obvious limitation of these cross-sectional studies is that one cannot draw clear causal inferences. Although most studies have assumed that interference caused by anxiety is the causal agent, this need not be the case. Deficits in WMC could represent a risk factor for later anxiety (e.g., Ouijmet, Gawronski, & Dozois, 2009). A small number of studies have attempted to address this limitation by employing longitudinal designs under the assumption that if one variable is the causal factor, it should predict the other variable at a later date.

Along these lines, Macklin et al. (1998) examined general intelligence and the development of subversive posttraumatic stress disorder. In this study, Wechsler intelligence scale scores were recorded from U.S. army recruits before and after tours of duty in Vietnam. Individuals who developed posttraumatic stress disorder performed worse on the intelligence measures both prior to and following deployment ($d_s = -.84$ and $-.64$, respectively). Interestingly, symptom severity did not predict postdeployment intelligence and differences in postdeployment intelligence were nonsignificant when premorbid test scores were controlled suggesting that developing PTSD did not produce any further declines in intelligence scores. The findings of this study make it one of the few to provide evidence that individual differences in cognitive abilities can act as a risk factor for later emotional disturbances. It should be noted, however, that premorbid intelligence also predicted the amount and severity of combat exposure. This leaves open the possibility that trauma severity can account for this relationship. Additionally, although measures of intelligence and measures of working memory are highly related, they are not isomorphic (Ackerman et al., 2005).

To the author’s knowledge, there are only two studies that have examined WMC, specifically, and anxiety prospectively. First, Bredemeier and Berenbaum (2013) used a phonological/spatial N-Back composite to predict self-reported worry at a later time point. In this study, Time 1 N-Back significantly predicted Time 2 worry scores; interestingly, this effect persisted even when Time 1 worry was controlled suggesting that developing PTSD did not produce any further declines in intelligence scores. The findings of this study make it one of the few to provide evidence that individual differences in cognitive abilities can act as a risk factor for later emotional disturbances. It should be noted, however, that premorbid intelligence also predicted the amount and severity of combat exposure. This leaves open the possibility that trauma severity can account for this relationship. Additionally, although measures of intelligence and measures of working memory are highly related, they are not isomorphic (Ackerman et al., 2005).

The work by Macklin et al. (1998) and Bredemeier and Berenbaum (2013) has provided some of the most direct evidence that cognitive abilities can act as a risk factor for later anxiety that currently exists in the literature. However, the inferences that can be drawn from these studies are limited by the fact that neither presented fully crossed lagged correlations. For example, measures of worry and WM are often highly reliable over short periods of time (Hockey & Geffen, 2004; Meyer et al., 1990; Redick et al., 2012; Stöber, 1998; however, see Jaeggli et al., 2010); although Bredemeier and Berenbaum (2013) showed that Time 1 N-Back predicted Time 2 worry, it is quite plausible that Time 1 worry would have predicted Time 2 N-Back had it been measured. Indeed, Visu-Petra et al. (2014) found that Time 1 anxiety predicted Time 2 digit span. A stronger test of the causal relationship would involve cross-lagged correlations. In such designs, anxiety and working memory would be measured at two disparate time points. If anxiety is the causal factor, anxiety at Time 1 would be expected to predict working memory at Time 2 but working memory at Time 1 should be significantly less predictive of anxiety at Time 2 (or vice versa if WMC is the causal factor; Campbell & Stanley, 1963; Kenny, 1975). Unfortunately, it does not appear as if any such studies have been conducted.

**WMC as a mediator.** A handful of studies have gone beyond demonstrating bivariate associations between anxiety and measures of WM and have begun to examine the mediational role of WM in linking anxiety to real-world outcomes such as academic performance and treatment response. Although many of these studies were included in the meta-analysis, their findings are worth reviewing here.

**Academic performance.** As noted earlier, anxious individuals experience academic difficulties such as lower GPAs, scores on standardized tests, and graduation rates. Recently, research has begun to link these difficulties to anxiety-related impairments in cognitive abilities. For example, in Owens et al. (2008), 11-year-old children completed measures of trait anxiety, backward digit and spatial spans, and the National Curriculum Standard Assessment Test—a standardized academic achievement test in the U.K. Trait anxiety predicted poorer performance on both the simple spans and the standardized test. Importantly, the association between anxiety and academic achievement was fully accounted for by individual differences in simple span scores. Similar findings were reported by Vukovic et al. (2013) who examined mathematical problem solving in math anxious children. A follow-up study by Owens and colleagues (2014) has provided the most thorough exploration of the interrelationships between anxiety, WMC, and academic performance. They conducted path models including anxiety/depression, worry, the composite score computed from backward digit and spatial span tasks, and academic performance. The indirect path leading from anxiety/depression to academic performance through worry and working memory was significant but the direct path was not (also see Ganley & Vasilyeva, 2014). In other words, the tendency to worry mediated the association between trait anxiety and impaired WMC and this impaired WMC mediated the association between worry and academic difficulties.

Interestingly, each of the child studies in this section found that anxiety predicts performance on backward simple span measures and those that examined forward span measures found near-zero correlations (often < .1). Although work conducted with adults has typically found that backward simple spans load with forward simple span tasks and not with complex span tasks (e.g., OSPAN; Engle, Tuholski, et al., 1999), this may not be the case for children. Some work suggests that the requirements of backward span tasks are sufficiently difficult to challenge the limits of attention in children (Alloway et al., 2004). It is also interesting to note that Owens et al. (2014) found that the composite computed from phonological and spatial measures of simple span correlated with worry and mediated the association with academic performance—suggesting fairly domain-general deficits. These results, then, are consistent with the idea that anxiety is most closely coupled with attentionally demanding tasks and that this deficit can account for anxiety’s association with academic deficits.
Treatment response. In addition to mediating the link between anxiety and academic performance, some work suggests that domain-general abilities may be an important determinant of one’s response to treatment. For example, Mohlman (2005) proposed that the successful completion of CBT may be reliant on executive functions due to the fact that patients are often required to increase the extent to which they regulate their negative emotions and well as increase their use of logic and reasoning to combat negative moods and beliefs.

To test this, Mohlman and Gorman (2005) examined the effectiveness of CBT for GAD in patients with high and low executive functioning (EF; defined by a composite score of several EF tasks). This study produced several results worth noting: first, although high- and low-EF individuals completed the same amount of CBT-related homework, the high-EF individuals’ produced significantly higher quality homework. Second, high-EF individuals showed significantly greater gains on self-report measures of anxiety. Third, the rate of GAD was not significantly different between high- and low-EF individuals at the posttreatment assessment or at the 12-month follow-up. However, this latter finding may be the result of low statistical power (37% vs. 80%; d = 1.05) and at the follow-up (29% vs. 80%; d = 1.27). Similarly, D’Alcante and colleagues (2012) examined the effectiveness of both CBT and fluoxetine, a common SSRI, for treating OCD. This study found that pretreatment intelligence predicted greater reductions in the Y-BOCS for patients receiving CBT but not for patients receiving fluoxetine.

Overall, these studies, though there are few of them, suggest that domain-general abilities play a role in the successful completion of CBT. Although the underlying mechanism is currently unclear—perhaps high-EF individuals already possess many of the skills taught by CBT, or they are more compliant with the treatment protocols (e.g., Mohlman & Gorman, 2005)—it appears that EF is an important determinant of treatment response. It should be noted, however, that neither study examined WMC directly. Mohlman and Gorman (2005) used a composite of several executive function tasks rather than measures of WM span. Executive functions refer to a heterogeneous set of abilities which generally include control functions (e.g., inhibition), updating, maintenance of relevant information, planning, and reasoning as well as others. Thus, findings regarding EFs may be highly dependent on the particular abilities that were measured. This is not to say that these findings would not extend to measures of WMC. Indeed, McCabe and colleagues (2010) found that EF tasks and complex spans shared considerable variance—which they considered to reflect domain-general attention. However, these findings should be replicated with measures of WMC.

Experimental Studies

This section of the review focuses on studies that have experimentally manipulated either anxiety or working memory capacity. As with the correlational studies, this section is divided into subsections based on methodology.

Ego-threat. One of the earliest lines of research investigating the effects of anxiety on working memory has been characterized as “ego-threat.” In studies of ego-threat, the cognitive tests themselves are used as a threat to participants’ self-image and self-esteem. Participants are typically informed that the task is an important measure of intelligence or that they are doing poorly on a task at which most participants excel.

Studies by Moldawsky and Moldawsky (1952) and Hodges and Spielberger (1969) were among the first to explore working memory under ego-threat conditions. Undergraduates were first assessed with the digit span. Following this baseline assessment, participants were assigned to either an ego-threat condition or a control condition and readministered the digit span. Although performance for the control participants did not change between assessments, participants in the ego-threat condition recalled fewer correct digit series during the second assessment. Similar findings were reported by Hodges and Durham (1972) who recorded digit spans at baseline and then again after participants were informed that they appeared to be experiencing much more difficulty than other participants (however, see Steyaert & Snyder, 1985). Walker, Sannito, & Firetto (1970) and Firetto and Davey (1971) attempted to confirm that the ego-threat effect was due to an increase in anxiety. As these studies were conducted before the development of statistical mediation techniques, they rated participants on the extent to which they “appeared anxious” and demonstrated that the ego-threat manipulation reduced digit spans only for those participants who appeared anxious.

Coy et al. (2011) took a slightly more detailed approach by examining the effect of ego-threat on both phonological and visuospatial storage. Participants in the ego-threat condition performed worse on phonological measures but not on the spatial measures. Interestingly, path analytic results indicated that self-evaluative negative thoughts (similar to worry) was a mediator between threat condition and digit span performance which is consistent with notions of domain-specific interference.

Another set of studies of have discussed ego-threat in terms of threats to one’s social image. Perhaps the most well-known manipulation is the Trier Social Stress Task in which participants complete a videotaped speech and a mental arithmetic task in front of an evaluative audience (Kirschbaum et al., 1993). Using this method, Schoofs and colleagues (2008) found that participants in the ego-threat condition were less accurate on a phonological 2- and 3-Back tasks relative to controls. Oei and colleagues (2006) used a similar anxiety induction but required participants to complete the Sternberg memory task (Sternberg, 1966). In this task, participants must memorize a set of 1 to 4 target items. Then, after a short delay, a new set of items are presented and participants must determine if one of the target items is an element in this new display. This study found that performance was impaired in the stress condition but only for high levels of load (i.e., 4 items). Although the number of studies is quite small, these results suggest that the anxiety elicited by the Trier Social Stress Task can impair performance in WM measures that are superficially quite dissimilar (i.e., N-Back and Sternberg); however, it is interesting that these studies differed with respect to the condition × load interaction. Whereas Schoofs and colleagues (2008) found group effects for both 2- and 3-back trials, Oei found that the group effect was specific to high load trials. One could posit that maintaining and continually updating even 2 items is more demanding that than simply storing 1 to 3 items. This difference could, thus, reflect differences in task demands.
Overall, the ego-threat manipulation appears to reliably affect WM scores—simple span scores, in these studies. To get a sense of the extent to which the ego-threat manipulation affects WM performance, effect sizes were pooled across known studies. This resulted in a significant aggregate effect \( (g = -0.618, k = 9, N = 493, p < .001) \) such that participants under ego-threat conditions tended to perform worse on simple span measures.

**Affective video clips.** The manipulations described in this section involve presenting participants with film clips that are meant induce an anxious state (e.g., clips from *Scream, Halloween* etc.) or are meant to be affectively neutral. Following the induction, participants in all studies completed a variant of the N-Back task.

Results from this line of research have been quite mixed. Some studies have found no effect on accuracy or RTs (e.g., Fales et al., 2008; Qin et al., 2009; however, if one examines Qin et al.’s data, there appears to be a notable, albeit nonsignificant, dip in accuracy for the anxiety condition). As with ego-threat studies, it is possible that changes in N-Back performance were only apparent in participants for whom the manipulation was effective. For example, Qin et al. (2009) measured self-report and physiological measures of anxiety and found that these measures predicted the degree to which the manipulation affected N-Back RTs \( (r^2 = \text{ranges between} .29 \text{ and } .42) \); correlations with accuracy were not reported.

On the other hand, some work has found that anxiety-inducing film clips impair phonological 3-Back performance (Gray et al., 2002) and may improve spatial 3-Back performance (Gray, 2001). One should note, however, that Gray (2001, Gray, Braver, & Raichle, 2002) also included pleasant film clips. The analyses that were presented make it difficult to determine whether the anxiety condition differed from the neutral baseline or only from the pleasant conditions.

In evaluating these studies, it would be useful to pool effect sizes to determine whether any overall effect is present. However, in general, there was not enough information available to do so. Possible explanations for these mixed results will be returned to toward the end of the section on experimental studies.

**Threat-of-shock studies.** The logic underlying threat-of-shock studies is quite straightforward. In these studies, participants complete a measure of working memory in one of two conditions. In the threat condition, a randomly timed electrical shock is delivered via an electrode affixed to the participant’s arm. In the safe condition, participants are informed that no shock will be delivered.

The first to use this methodology were Pyke and Agnew (1963), who recorded digit spans under threat and safe conditions. The threat condition reduced spans relative to the safe condition. However, this effect was notably larger in participants who completed the threat condition first (relative to those who completed the safe condition first) possibly suggesting that anxiety disrupted task performance when the task was more novel and less so when it had been learned.

Shackman and colleagues (2006) used threat-of-shock to examine the relationship between anxiety and performance in phonological and spatial versions of the 3-Back. Although phonological 3-Back accuracy was unaffected by the threat-of-shock manipulation, there was a decrease in accuracy for the spatial 3-Back. Lavric et al. (2003) used the same methodology with one exception—shocks were threatened but never actually delivered. The findings of this study were consistent with those of Shackman et al. (2006). This ensures that the results of Shackman et al. (2006) cannot be attributed to some distraction caused by the delivery of the shock.

In a series of similar studies, Vytal and colleagues (Vytal et al., 2012, 2013; see also Robinson et al., 2013) have complicated the view presented by Shackman and Lavric. These studies also examined the effect of threat-of-shock on the N-Back; however, they employed 1-, 2-, and 3-Back designs. Vytal et al. (2012), using a phonological N-Back, replicated the findings of Shackman et al. (2006; i.e., that threat-of-shock did not affect 3-back performance). In the 1- and 2-Back conditions, however, accuracy was reduced under threat.

Using both a phonological and spatial N-Back, Vytal et al. (2013) have provided the most comprehensive examination of the effect of shock on N-Back performance. Participants completed 1-, 2-, and 3-back conditions for both a phonological and spatial N-Back. Results were consistent with previous studies. For the phonological N-Back, the threat condition reduced accuracy in the 1- and 2-back but not the 3-back. For the spatial N-Back, the threat condition impaired performance at all levels of load. Thus, this line of work suggests a modality × load interaction. That is, anxiety, as induced by threat-of-shock, restricts spatial storage at all levels of load but only restricts phonological storage at lower levels of load. This finding, however, has not been completely consistent. Kalisch et al. (2006) examined phonological 2-Back performance using a threat-of-shock design as well. Unlike Vytal et al. (2012, 2013), 2-Back performance was unaffected (also see Hansen et al., 2009). As the methodology of this study was quite similar to the other studies in this section and there are no clear flaws in the design, it is unclear why this study failed to find an effect. One could posit that low power \( (N = 14) \) resulted in the null finding but this is not clear. More work will need to be conducted in order to determine if there are boundary conditions on when the modality × load interaction is expected—this issue will be returned to later.

Overall, the threat-of-shock appears to reliably reduce accuracy \( (g = -0.681, k = 3, N = 115, p < .001) \) for spatial 3-Back tasks. However, the pooled effect sizes for phonological 3-Back performance \( (g = .010, k = 4, N = 141, p = .92) \) and phonological 2-Back performance \( (g = -.200, k = 4, N = 125, p = .42) \) were nonsignificant. One-Backs and spatial 2-Backs were not pooled as they have been less frequently studied; however, the extant evidence suggests that accuracy is reliably reduced by threat-of-shock in these tasks (e.g., Vytal et al., 2012).

**Treatment studies.** Although not explicitly intended to test the causal relationship between anxiety and WMC, another class of studies may nonetheless provide additional insights. These studies focused on the effectiveness of anxiety treatments while also recording measures from neuropsychological batteries, such as working memory. The logic of these designs is straightforward; if anxiety causally influences working memory capacity, then the successful alleviation of anxiety symptoms would be expected to increase scores on measures of working memory.

Mataix-Cols and colleagues (2002) compared phonological and spatial simple spans between medicated and unmedicated obsessive–compulsive patients. However, no reliable differences emerged on any tests. Similarly, Butters et al. (2011) examined the effects of a pharmacological treatment for geriatric generalized anxiety disorder. In this study, digit span scores did not differ at
baseline and showed no improvement over the course of treatment. In a third study (Nielen & Den Boer, 2003) obsessive–compulsive patients were treated with fluoxetine. Participants completed a visual and spatial version of the change detection task (Luck & Vogel, 1997) before and after treatment. In the change detection task, participants view a memory array consisting of several (typically 2–12) discrete shapes. After a brief delay (~1 sec), participants view a test array and must determine if it is identical to the memory array. In visual version of the task, participants saw a pattern of colored items and had to determine whether the memory array and the test array included the same colors. In the spatial task, participants had to determine whether the spatial locations of the items on the screen matched the previous presentation. In the visual task, patients were equally accurate as controls at baseline and did not improve over the course of treatment. With respect to the spatial task, obsessive–compulsive patients performed worse at baseline, relative to controls, and performed worse following treatment, relative to pretreatment. However, controls who did not receive treatment also performed worse at the posttest suggesting that this is unlikely to be attributable to the treatment.

At an initial glance, the results of these studies fail to provide any support for a causal role for anxiety. Of the three studies reviewed above, none found evidence of increased WMC following treatment. However, the interpretation of these results is complicated by several factors. First, it should be noted that Mataix-Cols et al. (2002) did not randomly assign patients to treated and untreated groups; patients were grouped based on preexisting treatment regimens. It is thus difficult to rule out any number of preexisting group differences which may have masked an effect of treatment. Second, the efficacy of the treatment regimens appears to be somewhat dubious. In Butters et al. (2011) the authors failed to provide evidence that the treatment successfully reduced anxiety symptoms. In Nielen and Den Boer (2003), the authors did present group-level evidence that the fluoxetine treatment reduced symptoms relative to baseline; however, only 63% of patients responded to the treatment. Thus, it is possible that either the treatment regimens were ineffective at reducing anxiety symptoms (i.e., Butters et al., 2011) or that gains made by responsive patients were masked patients who did not respond to treatment (Nielen & Den Boer, 2003).

Work by van der Wee and colleagues (2007) has helped clear up issues surrounding treatment response. In this study, obsessive–compulsive patients were treated with paroxetine and venlafaxine, a selective serotonin reuptake inhibitor and serotonin-norepinephrine reuptake inhibitor, respectively. After treatment, participants were grouped based on whether they showed evidence of treatment response. Results revealed that responders committed fewer errors on a spatial 3-Back posttreatment, relative to pretreatment. For nonresponders, performance did not significantly improve.

Overall, studies of treatment response have produced mixed results. The combined effect size for the four pharmaceutical studies suggests modest gains from treatment but is statistically unreliable ($g = .280$, $k = 4$, $N = 292$, $p = .27$). The study by van der Wee and colleagues (2007) has lent some credence to the notion that anxiety causally influences WMC by demonstrating that $N$-Back scores increased in treatment-responders but not non-responders. It should be noted, however, that the division of participants into “responders” and “nonresponders” was somewhat questionable. The criteria by which this distinction was made was not provided; additionally, nonresponders showed significant improvement on one of the three measures of obsessive–compulsive symptoms. In conjunction with the small sample size (7 responders and 7 nonresponders), this would seem to warrant some caution when interpreting these findings. It is also worth noting that these studies represent a very limited sample of possible anxiety treatments. That is, it is possible that cognitive/behavioral interventions would be more effective than pharmacological ones; however, it does not seem that this has been tested. Recently, studies have taken a different approach to the treatment methodology by examining whether adaptive working memory training (e.g., Cogmed, 2006) improves scores on measures of anxiety. In adaptive training programs, the set size of the to-be-remembered material gradual increases as the participants achieve a certain level of accuracy. This ensures that the training program constantly challenges the boundaries of the participant’s WMC. The results of these studies have been quite mixed. In Åkerlund et al. (2013), patients with acquired brain injuries completed an adaptive training regimen as well as a measure of anxiety developed for hospital use (Zigmond & Snaith, 1983). Although anxiety scores improved for the control group, there were no significant effects for the trained group at the 6- or 18-week follow-up sessions. Roughan and Hadwin (2011) examined adaptive training in young adolescents. This study provided somewhat more mixed results. Although there was some evidence of improvement on measures of test and trait anxiety at the 3-week follow-up, these gains were not maintained at the 3-month follow-up. Although there was not enough information to pool effect sizes, these findings suggest that adaptive WMC training does not causally affect anxiety. However, as with the pharmacological studies, WMC training may be more effective for “responders”—that is, individuals who are more engaged by the task. Indeed, Sari and colleagues (in press) examined adaptive WMC training in individuals high in self-reported trait anxiety. This study found that training-related gains in WMC predicted reduced anxiety at the posttest session—that is, individuals who benefited from the WMC training showed reduced anxiety following training. It should be noted, however, that the efficacy of adaptive WMC training programs is currently disputed (e.g., Shipstead et al., 2010; Redick et al., 2013; see Au et al. [2015] and Melby-Lervåg & Hulme [2013] for conflicting meta-analyses). For example, Åkerlund et al. (2013) found that the trained group did not significantly differ from controls at the 6 or 18 week follow-up on any measure of working memory (forward/backward digit and spatial span and WAIS-III WM scale). For Roughan and Hadwin (2011), trained participants did show increased WMC scores (as measured by a composite of digit and spatial spans) relative to controls at the 3-week and 3-month follow-up sessions. However, this study also included a measures of response inhibition and general intelligence. Although no effects were found for inhibition, general intelligence showed increases relative to controls at the 3-week follow-up but not the 3-month follow-up. These findings suggest that, although performance improved on superficially similar tasks following training, far-transfer effects were not achieved (e.g., Shipstead et al., 2010). Finally, Sari et al. (in press) found that WMC training transferred to the flanker task but transfer to the antisaccade task was less clear despite the fact that both tasks ostensibly measured attentional control/inhibition. Overall then, the evidence suggests that WMC training may reduce anxiety for
individuals who strongly engage in the task; however, given some of these limitations as well as the broader debate regarding the efficacy of WMC training, these results warrant future replication attempts.

**Other experimental approaches.** Finally, this section will describe manipulations that have only appeared in a small number of studies. The first of these is the set of studies that used the Cold Stressor task. This involves immersing the participant’s hand in painfully cold water or in water at a comfortable temperature. Following the stressor, participants then complete a version of the Sternberg task. The results of these studies have been quite mixed. Porcelli et al. (2008; Exp. 1) found a nearly significant condition × load interaction such that the manipulation reduced accuracy for low load trials only. Their second study, however, did not replicate this effect. Duncko et al. (2009), on the other hand, found that the stress condition impaired Sternberg performance only at moderate levels of load although it should be noted that the overall condition × load interaction was nonsignificant. It is not quite clear why the findings of these studies diverge (also see Oei et al., 2006, described above, which used the same WM task but found effects only at high levels of load). It does not appear that task difficulty can explain these findings. The high load used by Porcelli et al. (2008) included 6 items whereas Duncko and Oei only used 4. One possibility is that all of these studies differed with respect to their outcome measures (e.g., overall accuracy vs. hit rate vs. false alarm rate). It may be that anxiety differentially affects processes that contribute to these measures (see below); however this is currently unclear. When these results are considered in conjunction with threat-of-shock studies, there appears to be an important, albeit inconsistent, role played by the level of cognitive load used in a study. This issue will be returned to later.

Sorg and Whitney (1992) took a somewhat unique approach by requiring high- and low-trait-anxious participants to play video games for 10 min before completing simple and complex measures of working memory. For the simple span task, no effects were significant. For the more demanding complex span task, trait-anxious individuals in the stress condition showed impaired span scores relative to anxious controls.

Finally, the study by Robinson and colleagues (2008) involved participants engaging in a simulated underwater helicopter evacuation. In their first study, OSPAN scores were recorded two days before the stress manipulation as well as immediately before it. Participants in the stress group scored lower on the OSPAN, relative to a control group, immediately before the simulated evacuation. However, they also scored lower at the baseline assessment. There was no within-subjects effect of time. The baseline difference makes these findings difficult to interpret but the nonsignificant effect of time makes it unlikely that OSPAN scores were affected by the stressor. In their second study, the stress group again performed more poorly—this time immediately following the stressor; however no baseline assessment was presented.

**Summary of experimental findings.** Overall, experimental studies of anxiety and WMC have painted a very mixed picture. Some inductions, such as threat-of-shock and ego-threat seem to reliably affect WM performance. Other manipulations, such as affective film clips, have failed to provide consistent results. There seem to be two main factors responsible for these inconsistent findings: (a) many studies seemed to suffer from problems of internal validity insofar as it is unclear whether the manipulation had the desired effect, and (b) even when anxious affect was elicited, it may not have lasted throughout the task.

For example, fatigue seems to be a viable alternative explanation for the findings of Sorg and Whitney (1992). That is, the stress manipulation (i.e., video games) consisted of relatively a demanding task. Participants in the experimental condition later performed worse on the demanding complex span task but not on the simple span task. Similarly, although the stressor used by Robinson et al. (2008) was likely successful at eliciting lasting affect, the sample consisted of students at a naval academy for whom this stressor was a required training exercise. Thus, this manipulation may have been interpreted as “challenge” rather than as a stressor by this sample.

Perhaps less intuitively, ego-threat studies also suffer from internal validation issues. While all of these studies utilized some type of ego-threat manipulation, there were subtle differences between studies that made it difficult to determine what was induced by the manipulation and whether the various manipulations were equivalent. For example, Hodges and Durham (1972) “directly” threatened participants’ self-image by informing them that they appeared to be experiencing more difficulty than average. Coy et al. (2011), on the other hand, informed participants that the measures used in that study were highly indicative of intelligence whereby rendering the threat to the participant’s self-image contingent on potential later failures (i.e., completing an intelligence test only threatens one’s ego if they perform poorly)—perhaps better characterized as the “threat-of-ego-threat.” In addition to this variation in how ego-threat is operationalized, these manipulations also conflate threats to the self-image with threats to their social image. That is, in most of these manipulations, participants knew that the experimenter was aware of their performance. These issues are not necessarily problematic assuming that all of these manipulations induce some manipulation-independent self-evaluative or test anxiety. However, this is not clear. Finally, some researchers (e.g., Chalus, 1976) have used the “indirect” form of ego-threat as a control condition under the assumption that the mere potential for failure would not have the same effect as informing participants that they had actually failed. Thus, the extent to which this task induces anxiety, which processes other than anxiety are impacted, and the conditions under which this manipulation is effective have not been well-documented.

This would be much less problematic if these studies had (a) provided evidence that manipulation had the intended effect (e.g., an increase on a measure of state anxiety) and (b) shown that the increase in anxiety mediated the group effect. In general, however, studies did not provide evidence that the manipulation successfully induced anxiety and no study ego-threat demonstrated mediation. Some studies (e.g., Walker et al., 1970; Fretto & Davey, 1971) rated participants’ anxiety; however, a fairly serious challenge to this manipulation-check was posed by Walker et al. (1970). As in the other studies, participants who were rated as anxious recalled fewer digit strings than those who did not appear anxious. What separates this study is that they collected scale scores from the Wechsler Intelligence test as well and found that participants who appeared anxious performed worse on almost all subscales. This finding allows for the possibility that the ego-threat manipulation increased anxiety in participants with lower baseline intelligence. If this is the case, then this manipulation-check actually demonstrates that digit span scores vary as a function of individual differences in general intelligence.
The second problem—that is, that the induced affect may have been short-lived—affects all of the studies in which the affect-induction ended prior to the WM task (e.g., Affective Film Clips, Trier Social Stress Task, Cold Stressor, etc.). Given that induced affect can be very short-lived (Davidson, Ekman, et al., 1990; Garrett & Maddock, 2001; Gross, 1998) and can be down-regulated by demanding tasks (such as a WM task; Erber & Erber, 2000), it is possible that the anxious affect did not persist through the WM task. Many studies did record self-reported and physiological measures of anxiety; however, these were often recorded directly before or after the induction rather than during the WM task (see Shackman et al., 2006 for a good discussion of this issue).

Among the experimental studies discussed in this review, the threat-of-shock studies have been the most thorough in terms of establishing the internal validity of their manipulation. In all cases, anxiety was manipulated within-subjects in a counterbalanced design and the authors demonstrated the presence of anxiety during task performance. For Lavric et al. (2003), participants wore heart rate monitors during N-Back performance. Both heart rate and self-reports indicated greater anxiety in the threat condition; additionally, these measures predicted N-Back performance in the threat condition. Shackman et al. (2006) took a different approach and presented acoustic startle probes (i.e., loud bursts of white noise), which is widely considered to be a valid measure of arousal and defensive motivation (e.g., Lang et al., 1998), at random intervals throughout the N-Back. Not only were startle reflexes larger during the threat condition but the change in startle reflex between the threat and control condition predicted the change in accuracy between threat and safety. Vytal and colleagues (2012, 2013) found the same results using startle and self-report.

However, it is worth noting that anxiety indicators only accounted for up to about 44% of the variance in the change in N-Back performance in these studies. Given the fact that most of the variance in the change in N-Back scores appears to be unaccounted for and that even effective manipulations are likely to have a few unintended effects (Shadish et al., 2001), it would be worth demonstrating that indices of anxiety fully mediate the difference in N-Back scores as a function of condition. Indeed, Shackman and colleagues (2006) found that changes in corrugator activity mediated the group effect on spatial N-Back scores and that changes in startle scores nearly did as well. Based on these findings, it seems safe to conclude that threat-of-shock can causally affect performance on the N-Back under certain circumstances and that this effect can be attributed to increases in anxiety.

**General Discussion**

The aim of the present review was to determine whether anxiety is related to reduced working memory capacity. Following both a meta-analysis and a narrative review, the main conclusion of the present review appears to be a qualified yes. The meta-analysis revealed a moderate, yet robust, association ($g = -0.334$) between measures of anxiety and measures of working memory capacity. Additionally, the studies reviewed above suggest both that anxiety can causally influence performance on WM tasks and that the most pronounced effects of anxiety are on measures tapping domain-general attentional processes rather than on domain-specific stores. The “yes” is qualified as the review also identified a number of methodological limitations that have made drawing firm conclusions difficult. The following sections review the primary findings of this review as well as discuss methodological limitations, their implications, and ways to move forward.

**Correlational Studies**

The findings that (a) anxiety is related to poorer performance on measures of complex span and (b) dual-task studies suggest that anxiety—and worry in particular—is most strongly associated with performance on attentionally demanding tasks are readily interpretable within attentional accounts of working memory and anxiety. According to the attentional account of WMC, individual differences in WMC—particularly complex span task performance—are driven by the ability to select relevant information, resist distraction and maintain access to relevant information outside of immediate awareness (e.g., Engle, 2002).

This dovetails nicely with attentional/inhibition accounts of anxiety which propose that worry is attentionally demanding and acts as a dual-task during cognitive test performance (e.g., Eysenck et al., 1992, 2007; Sarason, 1988). Specifically, Eysenck and colleagues (see Derakshan & Eysenck, 2009; Eysenck et al., 2007; Eysenck & Derakshan, 2011 for reviews) have proposed that anxiety interferes with the ability to inhibit irrelevant stimuli and responses. Indeed, a large body of research indicates that anxiety is associated with deficits in attentional control even in the absence of a storage component. For example, Moser and colleagues (Moser et al., 2012; Moran & Moser, 2015) have conducted a series of studies in which participants viewed circular search arrays consisting of 10 circles and diamonds. On half of trials, all shapes were presented in the same color (either red or green); on the other half of trials, a distractor item was presented in the opposing color (e.g., one green item among 9 red items). Trait anxiety was found to predict the degree of slowing produced by the discrepant item. Using the same search task (i.e., Moser et al., 2012; Moran & Moser, 2015), Esterman et al. (2013) replicated this finding in a sample of veterans suffering from PTSD. Given the relatively low informational load of these tasks, these results imply aberrant attentional control—in particular, inhibiting interference from the distractor—in anxiety rather than deficits in temporary storage alone (see Berggren & Derakshan, 2013 for a review of anxiety-related attentional impairments).

Recent work has extended this line of thinking to WM. In this view, anxiety impairs the ability to inhibit irrelevant information from gaining access to WM and, as a result, impairs the ability to maintain relevant information. Much of this research has been conducted using a variant of Vogel and colleagues’ (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001) change detection task that includes three conditions: two relevant items that need to be stored, four relevant items that need to be stored, or two relevant items along with two irrelevant distractors. Vogel, McCollough, and Machizawa (2005) were the first to examine “filtering efficiency” (the ability to filter or inhibit irrelevant information from gaining access to WM) using this method. This study found that, for high-span individuals, neural activity on distractor trials was identical to neural activity when two relevant items were presented without distractors. For low-span individuals, neural activity on distractor trials was identical to neural activity when four relevant items were presented without distractors. In other words, high-span individuals were able to successfully filter out...
the irrelevant distractors whereas low-span individuals stored irrelevant items along with relevant items. Thus, the logic of filtering efficiency studies is as follows: the extent to which irrelevant distractors items affects performance—both behavioral and physiological measures of filtering have been studied—can be taken as an indicator of the extent to which individuals fail to filter those distractors from WM.

To the author’s knowledge, Stout and Rokke (2010) were the first to examine the association between filtering efficiency and anxiety. This study found that self-report measures of state anxiety, rumination, and depression predicted poorer filtering efficiency in the change detection task as measured by behavioral performance. Interestingly, further analyses suggested that this was only true for low-span individuals. For high-span individuals, these measures were unrelated to filtering efficiency. Moriya and Sugiura (2012) extended this work by finding that trait and social anxiety also predicted poorer filtering efficiency. However, it is unclear whether anxiety interacted with WMC in this study. More recently, Qi and colleagues (2014) examined neural measures of filtering efficiency as a function of trait anxiety. This study reported poorer filtering efficiency for high trait anxious individuals and that WMC predicted filtering efficiency only for low trait anxious individuals (also see Stout, Shackman, Johnson, & Larson [2015] and Stout, Shackman, & Larson [2013] who examined filtering efficiency using threatening stimuli).

Effect sizes were pooled across the five studies examining the association between anxiety and filtering efficiency. Overall, anxiety predicted poorer gating of irrelevant information from WM ($r = −.700, k = 5, N = 229, p < .001$). This finding implies that anxiety is robustly associated with poorer inhibition of irrelevant information and may be a mechanism by which anxiety restricts available WM. Thus, broadly speaking, the findings of this review can be considered consistent with an inhibition account of anxiety. That is, anxiety appears to interfere with inhibitory control over attention. This impaired inhibition, in turn, leaves anxious individuals less capable of preventing irrelevant information from gaining access to WM and reducing the WMC available for task performance.

Interestingly, the current meta-analysis did not support the more domain-specific predictions of any theory. Anxiety, broadly defined, predicted WM performance for both phonological and spatial material. Additionally, the meta-analysis found that the relationship between worry and WM was statistically equivalent between phonological and spatial material and, if anything, was greater for spatial material.

How should these findings be accounted for? First, it is possible visual forms of worry play a larger role in disrupting cognition than believed by theorists. This would imply that worry is associated with separable deficits in phonological and spatial storage depending on which type of worry the individual is engaging in. However, this seems unlikely given the relatively large literature suggesting that worry is primarily phonological in nature (e.g., Borkovec, 1994; Borkovec et al., 1983; Rapee, 1993). A more likely possibility is that anxiety is associated with the variance shared between the phonological and spatial tasks. Latent variable analyses suggest that complex span tasks involving phonological and spatial material share substantial variance and load on a single, domain-general factor which has been interpreted as executive attention (e.g., Engle, Tuholski, et al., 1999; Kane et al., 2004). The results of this meta-analysis allow for the possibility that anxiety predicts this domain-general variance. In other words, anxiety would be associated with other processes to the extent that they require controlled attention.

Even within the context of this controlled attention/inhibition view of anxiety and WMC, one might wonder why the association between worry and WMC was much larger in the spatial domain ($r = −.676$ vs. $−.339$). There would seem to be two possibilities. First, it should be noted that the difference between these effect sizes was not significant ($p = .12$); it is possible that this difference would not persist in a larger sample of studies. However, given that the magnitude of the effect size for spatial WMC is nearly twice that of the effect size for phonological tasks, it is also possible that this nonsignificant finding is simply due to low power (only 10 studies contributed to this comparison). This leads to the second possibility—that phonological storage is more dissociable from executive processes than is visuospatial storage. For example, Kane et al. (2004) found that residual variance from visuospatial simple span tasks predicted general fluid intelligence whereas residual variance from phonological simple span tasks did not (also see Badeley, 2000, 2007; Engle, Kane, et al., 1999; Engle, Tuholski, et al., 1999; Logie, 1995 for similar findings). Badeley (2000, 2007) suggested that this is due to the fact that phonological memoranda (e.g., letters, words, digits etc.) tend to be highly overlearned and can be reproduced by the rehearsal process. Visuospatial material, on the other hand, is more likely to require continued attention in order to be maintained. This issue will need to await further research as no studies have examined associations between anxiety, phonological WMC and spatial WMC at the level of latent variables.

However, the inhibition interpretation should be considered in the light of two caveats. First, there is a larger disagreement in the literature regarding whether executive attention or inhibition is the more fundamental ability on which the other relays (cf. Engle, 2002; Engle, Kane, et al., 1999; Engle, Tuholski, et al., 1999; Hasher & Zacks, 1988; Lavie et al., 2004; Zacks & Hasher, 1994). For example, Hasher and Zacks (e.g., Hasher & Zacks, 1988; Zacks & Hasher, 1994) proposed that individual differences in WMC can be attributed to individual differences in the ability to inhibit the no-longer-relevant contents of WMC. In this view, then, inhibition is the more fundamental ability. Engle and colleagues, on the other hand, have proposed that the ability to inhibit interference is reliant on available WMC (e.g., Engle, 2002; Engle, Kane, et al., 1999; Engle, Tuholski, et al., 1999; see Lavie et al., 2004 for a related account). This view is based on findings indicating that (a) distractor interference is greater both for low-span individuals (Kane & Engle, 2003) and that WM load manipulations are capable of influencing distractor interference and (b) high- and low-span individuals appear to be equally able to resist interference when placed under conditions of WM load (Kane & Engle, 2000; Rosen & Engle, 1997). This latter finding indicates that high-span individuals’ ability to resist distraction is reliant on available WMC. It would, therefore, also be consistent with the existing literature to suggest that restricted WMC in anxiety results in impaired inhibition.

Second, some of the findings of the current review hint at more general deficits than inhibition. While the dual-tasking studies tended to find that anxiety interfered with demanding tasks (e.g., random generation) but not tasks that relied more on domain-
specific stores (e.g., digit span), the meta-analysis found that anxiety is associated with poorer performance on simple span tasks.\(^1\) Given that (a) simple span tasks tend primarily reflect domain-specific storage (e.g., Kane et al., 2004) and that (b) the variance shared between simple and complex spans seems to reflect storage whereas the variance that is unique to complex spans appears to reflect executive processes (e.g., Engle, Kane, et al., 1999; Engle, Tuholski, et al., 1999), the fact that anxiety predicts simple span performance could suggest that anxiety is also associated with impaired cognition in the absence of an inhibitory component.

Indeed, this would be consistent with a growing literature that suggests anxiety is related to very broad deficits in fluid cognition. For example, Hembree (1988) meta-analyzed over 500 studies on anxiety and cognition and found that anxiety predicted poorer performance on measures of general intelligence. Similarly, Salthouse (2012) examined inductive reasoning, spatial visualization, verbal episodic memory, perceptual speed and vocabulary at the level of latent variables. Trait anxiety was related to the variance shared between all five abilities and, once that shared variance was accounted for, did not relate to the individual abilities. That is, in these studies, anxiety predicted highly general abilities rather than more specific functions/abilities.

Thus, the existing literature also allows for another interpretation—namely, that anxiety is not related to specific executive functions, such as inhibition, but, instead, is related to general fluid cognition at a highly domain-general level. Figure 1 presents one potential model describing the relationship between anxiety and fluid cognition. In this model, a “cognitive hierarchy” is adapted from Salthouse and colleagues (2008, 2012; note that this is presented for demonstrative purposes and is not meant to be an exhaustive list of all relevant abilities). At the lower end of the hierarchy are more specific abilities such as memory, vocabulary and so forth (see Calvo et al. [1992] for evidence that anxiety predicts poorer performance on measures of vocabulary). These specific abilities are defined by individual cognitive tasks which are not depicted in Figure 1. At the top of the hierarchy is “fluid cognition” which is defined as the variance that is shared between the more specific abilities. In this model, anxiety is assumed to have its effects at the top of the hierarchy (solid black arrow). Once the shared variance is accounted for, anxiety is not expected to predict the individual abilities (broken gray arrows).

If this is the case, one could reasonably wonder why anxiety predicts certain abilities, such as inhibition, far more reliably than other abilities (e.g., Derakshan & Eysenck, 2009; Eysenck & Derakshan, 2011; Eysenck et al., 2007 for reviews). In accounting for this, at least two facts must be borne in mind. First, large scale studies of anxiety and cognition have tended to reveal modest effect sizes. For example, the current meta-analysis found an aggregate effect of \(g = -0.334\) for the relationship between anxiety and measures of WMC. Hembree (1988) reported aggregate correlations ranging between \(r = -0.10\) and \(r = -0.31\) in a meta-analysis on the relationship between test anxiety and intelligence and academic performance. Similarly, the association between trait anxiety and general fluid cognition in Salthouse (2012) was \(\beta = -0.11\) (\(N > 1,000\)). Given these modest effects, it may not be surprising that many studies fail to find associations between anxiety and cognitive performance—especially when considering the relatively small samples employed in many of these studies. Indeed, the median power to detect an effect of \(g = -0.334\) for studies in the current meta-analysis was 0.23.

Second, there is variability in terms of the extent to which individual tasks rely on specific abilities. For example, Miyake and colleagues (e.g., Miyake et al., 2000; Friedman et al., 2008) have conducted a number of latent variable analyses of tasks thought to measure executive functions (e.g., inhibition, shifting and updating). These studies reported factor loadings between .33 and .63 (Miyake et al., 2000) and .43 and .72 (Friedman et al., 2008). That is, certain tasks were much better representations of those functions (i.e., they were more reliant on those functions) than other tasks.

With this in mind, it is possible that some failures to find associations between anxiety and certain executive functions are attributable to the fact that some tasks are poor representations of their purported function. A similar argument can be applied at the level of latent variables. Work by Miyake and colleagues (Friedman et al., 2008, 2011; Miyake & Friedman, 2012) suggests that inhibition, shifting, and updating are not equally related to “common executive functioning” (i.e., the variance shared between inhibition, shifting, and updating). Specifically, whereas loadings for shifting and updating tend to be between .6 and .7, the loading for inhibition tends to be 1. That is, inhibition, at least as it was measured in these samples, is indistinguishable from the higher order, more domain-general ability (however, also see Friedman et al., 2006; Shipstead et al., 2014). Thus, anxiety may be more reliably related to inhibition than other executive functions because inhibition, as it is commonly measured, is difficult or impossible to distinguish from higher order abilities.

Regardless of the specific model, the literature supports a fairly domain-general account of the association between anxiety and WMC. Anxiety is most strongly related to attentionally demanding tasks and is also related to both attentional performance in the absence of a storage component and storage in the absence of an obvious attentional component. However, there are still a number of outstanding questions. As noted above, ample evidence supports a link between anxiety and inhibitory control. However, it is currently unclear whether this an example of a specific deficit in inhibition or the result of deficits in higher order fluid cognition.

In addition to being related to measures of simple and complex span, the current review found that anxiety—both experimentally induced and measured as an individual difference—was related to dynamic span measures of WMC (e.g., the N-Back). As noted in the introduction, although most anxiety researchers assume that the N-Back and other measures of WMC are measuring the same underlying WM system (e.g., Robinson et al., 2013; Shackman et al., 2006; Vytal et al., 2012, 2013), mounting evidence suggests that this is not the case. A recent meta-analysis found low correlations between the N-Back and measures of simple and complex span (Redick & Lindsey, 2013). Similarly, Kane et al. (2007) demonstrated that the N-Back and OSPAN accounted for independent variance in general fluid intelligence. Results such as these imply that these tasks are not simply interchangeable and may be measures of separate underlying processes.

\(^1\) This discrepancy is likely to be attributable to statistical power. Nearly all dual-task studies were conducted in undergraduates and the meta-analysis revealed that the effect size for subclinical samples was \(g = -0.215\) (see Table 1). Achieving a reasonable level of statistical power (\(\gtrsim .80\)) would require more than 500 participants.
In attempting to explain the relationship between anxiety and the N-Back, there would seem to be several possibilities. First, while complex and dynamic spans are only weakly related, they do share some variance and it is currently unclear exactly what this variance represents. The current review cannot rule out the possibility that anxiety is associated with the variance that is shared between complex and dynamic spans. This possibility would be consistent with the model depicted in Figure 1. That is, that anxiety is associated with broad deficits in fluid cognition which manifest in the performance of a number of different tasks, including the N-Back. Second, Shipstead and colleagues (2014) analyzed candidate mechanisms underlying individual differences in WMC. They reported that, whereas complex span tasks related to performance on measures of attentional control, what they called “running span tasks” were related to overall storage capacity (e.g., Cowan, 2005). Importantly, running span tasks are superficially quite similar to the N-Back insofar as they both require attending to serially presented memoranda, maintaining the most recent n items and discarding items that are no longer relevant. This finding may imply that anxiety is associated with a limited capacity store that is even more restrictive than the typical 3- to 5-item limit (e.g., Cowan, 2001, 2005; Luck & Vogel, 1997; Rouder et al., 2011). This possibility is quite intriguing given that it implies more fundamental capacity limits beyond the attentional limitations typically associated with anxiety (e.g., Eysenck et al., 1992, 2007). However, it seems unlikely that a short-term anxiety induction (Robinson et al., 2013; Shackman et al., 2006; Vytal et al., 2012, 2013) would alter some fundamental capacity of the working memory system. Rather, it is more likely that anxiety-related processes interfered with active maintenance in some way.

The third explanation regards the differing task demands of complex/simple spans and the N-Back. Complex/simple tasks require participants to serially recall items with no external cues and thus require explicit recollection of an item in the context in which it was encountered. The N-Back, on the other hand, involves recognizing recently presented items and possibly discriminating them from recent nontargets. Thus, in addition to tapping recollection, the N-Back also involves familiarity-based processes that might obscure the relationship between the N-Back and complex spans (Oberauer, 2005; Szmalec, Verbruggen, Vandierendonck, & Kemps, 2011; also see Yonelinas, 2002 for a review). Indeed, this difference in retrieval demands appears to play a large role in psychometrically dissociating these tasks. In particular, Shelton and colleagues (Shelton et al., 2007, 2009) required participants to complete a modified N-Back that involved free recall of a subset of the most recent items. Items were serially presented as in the N-Back; however, when the list was complete, participants were required to recall the item that appeared 0-, 1-, 2-, or 3-back. Participants did not know which item they were required to recall until after the list was complete. Although the correlation between the standard recognition N-Back and complex span tasks tends to be a modest .20 (Redick & Lindsey, 2013), Shelton and colleagues found correlations as large as about .50.

What does this mean for anxiety? It is possible that anxiety is differentially associated with recollection and familiarity-based processes. For example, tagging an item as familiar is thought to be a relatively automatic and undemanding process whereas explicit recollection is more effortful (e.g., Yonelinas, 2002) and a great deal of research suggests that anxious individuals are more likely to be reliant on less demanding strategies (see above; also, Eysenck et al., 2007). Alternately, it is, perhaps, unsurprising that anxious individuals tend to believe that their environment is less predictable and amenable to control relative to low-anxious individuals (e.g., APA, 2013; Archer, 1979; Judge et al., 2002; Watson, 1967; note that much of the experimental work involving the N-Back placed participants in a situation in which they lacked control over their environment—i.e., by delivering randomly timed electric shocks—e.g., Shackman et al., 2006). Some work suggests that when individuals lack control over their environment they are more likely to falsely recognize patterns and rate novel stimuli as familiar in a presumed attempt to gain control over uncertain environments (e.g., Shermer, 2011; Simonov et al., 1977; Whitson

Figure 1. A potential model describing the relationship between anxiety and cognition. Anxiety is assumed to impair fluid cognition—that is, domain-general abilities that are common to the lower order abilities. Black, unbroken arrows are meant to depict significant causal paths. Gray, broken arrows are meant to depict nonsignificant paths.
Both of these possibilities suggest that anxious individuals are falsely determining N-Back nontargets to be familiar and predict that anxious individuals’ impaired performance on the N-Back results from an increase in false alarms specifically. However, this is difficult to test as most studies have reported either overall accuracy or sensitivity (i.e., d’) thereby obscuring the distinction between hits and false alarms.

To examine the relationship between anxiety and false alarms, I collected N-Back and anxiety scores from 197 undergraduates. Participants completed a letter version of the 3-Back as well as the Penn State Worry Questionnaire and Anxious Arousal Subscale of the Mood and Anxiety Symptoms Questionnaire (see Supplemental Materials for a detailed description of the methods). As can be seen in Table 3, both the Anxious Arousal subscale and the PSWQ predicted the false alarm rate but not the hit rate. To further examine these relationships, a regression was conducted in which both self-report measures were used to predict the false alarm rate. The overall regression was significant, F(2, 194) = 9.82, p < .001. Interestingly, the Anxious Arousal subscale remained a significant predictor in this regression (β = .28, t = 3.77, p < .001) but the PSWQ did not (β = .07, t = 0.87, p = .38).

Finally, response bias, C, was computed using Snodgrass and Corwin (1988)’s formula:

$$C = 5\left(\frac{LN\left(\frac{1 - FA(1 - H)}{FA(H)}\right)}{LN}\right)$$

where LN is the natural logarithm, H is the proportion of hits, and FA is proportion of false alarms. H and FA values equal to 0 or 1 were adjusted by .01. Positive values of C indicate a conservative “no” bias whereas negative values indicate a liberal “yes” bias. Higher Anxious Arousal scores predicted a “yes” bias, r = −.24, p < .001. This finding suggests that anxiously aroused individuals tend to set a response criterion such that items are more likely to be tagged as familiar even if they are nontargets (e.g., MacMillan & Creelman, 2005).

It should, however, be noted that Hansen and colleagues (2009) found that participants in the “safe” condition were more likely to commit false alarms in a threat-of-shock design. It is unclear why results of Hansen et al. (2009) differ so dramatically from those presented above (also see Shin et al., 2006). Perhaps it resulted from the cognitive load used in that study (Hansen used a 2-Back whereas the current study used a 3-Back). Alternately, the current study utilized a lower target to nontarget ratio (see Supplemental Materials) than Hansen et al. (2009). Varying the proportion of targets present in a task may differently affect anxious and non-anxious individuals’ response criteria. In any case, these results suggest that the relationship between anxiety and the N-Back is likely to be somewhat nuanced. Fully understanding it will require a detailed understanding of the effects of different types of anxiety (e.g., worry/arousal), different operationalizations of anxiety (e.g., self-report/induction) and N-Back task parameters.

To thoroughly test whether anxiety increases reliance on familiarity processes, performance on lure sequences would need to be examined (e.g., Yonelinas, 2002). In the N-Back, lure sequences are sequences wherein an item matches a recent but irrelevant item (e.g., L-F-X-L in a 2-Back task) and are likely to be flagged as familiar due to their recent appearance (e.g., Szmaciel, et al., 2011). Unfortunately, few studies included in this review presented data for lure sequences or even indicated whether lure sequences were included in the task. Fales et al. (2008) did examine lure trials but found no effects. However, as noted above, it is unclear whether this design successfully induced lasting effect. Overall, it seems that anxiety is robustly and causally (see below) related to performance on the N-Back. But fully understanding this relationship is going to require both a more detailed account of how the N-Back relates to other measures of working memory as well as more thorough descriptions of anxious individuals’ performance.

Perhaps the most surprising finding of the meta-analysis was that social anxiety appears to be either unrelated to WMC or related to increased visual WMC in East Asian populations (e.g., Moriya & Sugiuira, 2012, 2013). This is quite unexpected insofar as no existing theory makes specific predictions for visual WMC, for cultural differences, or for social anxiety specifically. Moriya and Sugiuira (2012) speculate that socially anxious individuals work to maintain a large number of visual features so as to keep track of audience member responses in social situations. However, this study only examined students attending East Asia universities. It is true that social phobia is among the most culturally sensitive anxiety disorders insofar as it manifests as a fear of negative evaluation in Western cultures and as a fear of offending others in East Asian cultures (e.g., Chang, 1997; Okazaki et al., 2002; Okazaki, 1997). However, in both cases, socially anxious individuals must keep track of others’ reactions. Additionally, Western and East Asian students do not appear to differ in the amount or type of information that is used when evaluating faces (e.g., Caldara et al., 2010). At present, it seems likely that this finding can be attributed to measurement issues. For example, all studies demonstrating this enhanced visual WMC used the same simple span measure (i.e., change detection) as well as the same measure of social anxiety. One cannot rule out the possibility that these findings reflect more task-specific factors rather than a true association between social anxiety and WMC for visual features. Alternately, it is possible that the psychometric properties of the social anxiety measure are not invariant across cultures—that is, it may not be measuring the same construct in both cultures (Chen, 2008). This possibility is especially likely given cultural differences in social anxiety. In any case, given that this finding persists across several samples and is not predicted by any existing theory, it would seem to warrant future attention.

### Limitations and Future Directions for Correlational Studies

The meta-analysis, along with studies of dual-tasking, revealed a robust relationship between measures of anxiety and WMC. However, the most common method of demonstrating a relation-
ship between anxiety and WMC consisted of presenting scores on a single measure of WMC as a function of a single measure of anxiety. Although this has undoubtedly contributed to our understanding, the aim of this line of inquiry is not just to show that test scores are related. Rather, the aim is to understand the relationship between the emotional state and the cognitive abilities that underlie these test scores. One of the most prominent threats to understanding the anxiety/WMC relationship at the level of abstract constructs has been the tacit assumption that scores on a single test are process-pure indicators of that construct.

It is generally understood that scores on a given measure can be partitioned into at least two sources of variance (e.g., Kim & Mueller, 1978; Shadish et al., 2001). The first is “true score” variance which represents variance attributed to the construct under investigation (e.g., anxiety). Statistically, true score variance is instantiated as “valid” variance or variance that is shared with other measures of the same construct. The second source of variance is error variance which can be further partitioned into two sources. The first of these sources is random error that arises during measurement; importantly, this random error is unrelated to other sources of variance and tends to average to zero in large samples. Random error can be statistically estimated as the proportion of unreliable variance. Finally, the second source of error variance is systematic error. Systematic error refers to some construct that is being reliably measured but differs from the characteristic of interest (e.g., scores on a measure of anxiety may be influenced by variance related to depression). Such variance is statistically reliable but is not necessarily shared with other measures of the same construct.

For example, the reliability coefficient of the State–Trait Anxiety Inventory (STAI) tends to be approximately .96 (e.g., Nitschke et al., 2001). With respect to validity, correlations between the STAI and other anxiety measures tend to range between .35 and .70—1 will assume .70, but any reasonable validity estimate will yield the same conclusions. Such combinations of high reliability coefficients and relatively low validity coefficients lead to interpretive difficulties that are often unappreciated.

Based on those estimates, one can infer that about 81% of the variance of STAI scores can be considered reliable or “true score” variance whereas the remaining 19% can be attributed to measurement error. Approximately half (49%) of the total variance can be considered valid. Thus, approximately 32% of the variance in the STAI is reliable but not valid. That is, it reliably measures something other than anxiety as defined by the STAI. What, then, does this mean for the study of anxiety and working memory? If one were to conduct a study using the STAI and a single measure of working memory, it would be impossible to unambiguously attribute a significant correlation to trait-anxiety, per se. Rather, any combination of the valid variance and the unaccounted for reliable variance could be driving this relationship.

This has indeed been the case. Nearly all correlational studies presented bivariate associations between single measures of working memory and anxiety. Although a few studies did present interrelationships between several constructs (e.g., measures of trait anxiety, worry and working memory), this is not the same as a true multimeasure assessment. In these studies, multiple measures were recorded to demonstrate a mediational effect or to show some degree of specificity (e.g., that a measure of WMC predicted worry but not state anxiety); however, each construct was still represented by a single measure. To show a relationship at the construct level, multiple measures of each construct must be included and aggregated. Thus, although it seems fairly clear that measures of anxiety and WMC do interrelate, future research will benefit from showing that this relationship holds at the construct level.

This leads to a second, related limitation. One of the primary dimensions along which theories differ is the extent to which they predict domain-specific versus domain-general effects. That is, theories differ both with respect to the roles they posit for the worry and arousal dimensions of anxiety as well as with respect to the WM domain(s) most strongly related to anxiety. In general, the notions that worry impacts phonological storage and arousal impacts spatial storage have found some support in this review; what is missing is strong evidence for the predicted degree of specificity. That is, demonstrating that one dimension of anxiety predicts performance in a WM task is not the same thing as demonstrating that this dimension of anxiety uniquely predicts performance in that task. A good example of this comes from the work of Owens et al. (2014). In this study, worry and WMC were found to mediate the association between trait anxiety and academic performance in adolescents. However, this study did not include a measure of arousal thus allowing for the possibility that this finding is not specific to worry (while they did include a measure of depression, this was analyzed separately and not controlled). Indeed, work by Putwain, Connors, and Symes (2010) has found that both worry and arousal symptoms independently predict academic performance in adolescents. Similarly, although the discussion thus far has focused on measures of anxiety, the same could be said for both simple and complex measures of span. For example, phonological and spatial storage tend to be highly related and both are strongly correlated with domain-general WMC (i.e., complex spans; Kane et al., 2004).

The facts that (a) all measures of anxiety appear to predict task performance and (b) and all of the WMC measures included in this review appear to be related to anxiety lead to an interpretative difficulty. Namely, although it is clear from the meta-analysis and review that measures of anxiety predict measures of WMC, it is unclear what these correlations represent. For example, the meta-analysis found that worry predicts WMC for both phonological and spatial material. It may be that worry involves both specific phonological and specific spatial deficits, that worry predicts only the shared variance (i.e., more domain-general variance), or that worry predicts the shared variance as well as specific phonological deficits once the shared variance has been accounted for. This latter possibility would be broadly consistent with theorizing on worry (Eysenck et al., 1992; Sarason, 1988). Though it is less frequently studied, the same could be said of arousal-based measures. For example, it may be that arousal predicts the domain-general variance was well as specific spatial deficits once the shared variance has been accounted for (e.g., Robinson et al., 2013; Vytal et al., 2012, 2013).

2 For the sake of simplicity, I will focus on internal consistency. But the same conclusions will be reached if other types of reliability are considered.

3 Working memory measures tend to show reliability > .7 and tend to correlate with other WMC measures between .3 and .75 (e.g., Kane et al., 2004). Thus, similar arguments could be adduced for measures of WMC.
Figure 2 presents a potential model that would be consistent with existing research. Given that worry and arousal seem to predict WMC domains broadly, the shared variance between worry and arousal is assumed to predict domain-general processes. Variance specific to worry and arousal, on the other hand, predict processes specific to phonological and spatial storage, respectively (the WM component of the model was based on Kane et al., 2004). Given the importance of this issue for existing theories, there is a clear need to examine the interrelationships between worry, arousal, phonological storage, spatial storage, and domain-general processes in the same participants. Such detailed studies would allow researchers to examine the structure of the interrelationships between anxiety and cognition and adjudicate between competing models in ways that previous research has not. It should be noted that this limitation is also true of constructs, such as depression, which are closely related to anxiety. Although some studies did examine both anxiety and depression (e.g., Stout & Rokke, 2010; Owens et al., 2014), most studies analyzed these measures separately. Given that depression is also related to executive functioning (e.g., Snyder, 2013), this means the current review cannot rule out the possibility that the anxiety/WMC association is attributable to shared variance with depression.

Finally, the relationship between trait and state anxiety has received very little attention in the literature. The current review suggests that state anxiety, whether induced or measured via self-report, is detrimental to complex cognition. Recent work by Grillon and colleagues (Grillon, Robinson, Mathur, & Ernst, 2015; Robinson, Krimsky, & Grillon, 2013), on the other hand, has found that the threat-of-shock improves performance on response inhibition tasks. It is true that these studies did not measure WMC directly; however, as noted above, work by Friedman and Miyake (Friedman et al., 2008, 2011) has suggested that, once the variance common to all of the executive functions—thought to reflect domain general attention (McCabe et al., 2010)—is accounted for, there is no inhibition-specific variance. That is, inhibition appears to be a highly domain-general process.

Such findings might be best understood in terms of the differences between trait and state anxiety. Indeed, researchers have long posited that state anxiety is nonlinearly related to task performance and that it interacts with trait anxiety to affect performance whereas trait anxiety is often thought to uniformly impair cognitive task performance (Easterbrook, 1959; Eysenck, 1979; Eysenck et al., 1992, 2007; Yerkes & Dodson, 1908). For example, trait anxiety may interfere with WMC by predisposing one to worry whereas the effects of state anxiety may be mediated through an arousal mechanism which is curvilinear and/or dependent on the level of trait anxiety. However, this possibility is difficult to evaluate in the current review. The majority of studies in this review have not explicitly evaluated the trait-by-state anxiety interaction and, among those that have, the results have been quite mixed. Hodges and Durham (1972) found no trait × ego-threat interaction on digit span performance whereas Walker and Spence (1964) found a significant interaction such that trait anxiety predicted digit span performance only in the control condition. More recently, Sorg and Whitney (1992) found that a stress induction impaired complex span performance only for trait-anxious individuals suggesting that the stress induction impaired performance only for those already predisposed to anxiety (also see Shin et al., 2006). Similarly, Salthouse (2012) found a curvilinear relationship between trait anxiety and several measures of cognitive ability (also see Walkenhorst & Crowe, 2009). However, trait and state anxiety are often highly correlated and this relationship may have been attributed to individual differences in state anxiety. Overall, the current review was able to find very little evidence bearing on the relationship between state and trait anxiety. Given the historical and theoretical importance of the distinct and inter-

Figure 2. A potential model describing the relationship between anxiety and WMC. In this model, processes specific to worry are assumed to impair phonological storage whereas processes specific to arousal are assumed to impair spatial storage. Processes that are shared between worry and arousal are assumed to impair domain-general abilities. Singled-headed arrows depict assumed causal pathways whereas doubled-headed arrows depict correlations between constructs.
active roles of trait and state anxiety (e.g., Easterbrook, 1959; Eysenck, 1979; Eysenck et al., 1992, 2007; Yerkes & Dodson, 1908), there would seem to be a clear need for more detailed study in this area.

The correlational studies reviewed here provide clear evidence for an association between anxiety symptoms and WMC. That is, a meta-analysis of 177 effect sizes, as well as a narrative review of dual-tasking studies, indicate that anxiety is related to poorer performance on measures of WMC. Additional evidence suggests that WMC is capable of mediating the relationship between anxiety and important real-world outcomes, such as academic performance. However, the review also identified a number of limitations to be addressed by future research: (a) there is a clear need to conduct multimeasure assessments that are capable of decomposing variance associated with measures of anxiety and WMC into theoretically meaningful partitions and to examine their interrelations (e.g., worry, arousal and depression for anxiety and domain-general executive attention and modality-specific storage for WMC); (b) determine the nature of the association between anxiety and the N-Back task; (c) explore the independent and interactive effects played trait and state anxiety and evaluate the possibility of curvilinear effects; and (d) determine what role, if any, culture plays in the anxiety/WMC relationship.

Experimental Studies and the Causal Relationship

There currently exist competing hypotheses regarding the nature of anxiety-related deficits in cognitive performance. Many modern theories propose that processes related to anxiety interfere with cognition in some way—for example, by demanding attention, consuming phonological resources, or by competing with storage-related processes for neural representation (Eysenck et al., 1992, 2007; Robinson et al., 2013; Shackman et al., 2006). Such frameworks are generally agnostic with respect to the etiology of anxious symptoms but suggest that, once present, these symptoms produce a number of downstream effects on measures of cognitive ability. Other, more clinically focused theories (e.g., Mathews & MacLeod, 2005; Ouimet et al., 2009) have tended to shift the causal burden to cognitive biases and deficits. In these theories, preexisting individual differences in cognitive abilities may act as a risk factor which, when paired with subsequent stress, predispose an individual for later anxiety symptoms. Thus, in these theories, WMC is one of many causal factors contributing to the etiology of anxiety.

Unfortunately, the current review can only offer a partial resolution to this question. The strongest evidence bearing on this question comes from threat-of-shock studies which have shown, fairly unequivocally, that performance on at least one measure of WMC can be impaired by an anxiety induction. What makes these studies such strong demonstrations is their focus on internal validity. First, these studies were some of the few that experimentally manipulated anxiety thereby helping to rule out preexisting individual differences. Second, these studies used a well-validated and widely studied method of inducing anxiety (e.g., Davis et al., 2010; Schmitz & Grillon, 2012). Third, they provided both self-reported and physiological evidence that the induction indeed modulated participants’ anxiety and Shackman et al. (2006) provided evidence that this increase in anxiety measures mediated the threat-of-shock effect.

The review presented above indicates that ego-threat manipulations are also capable of affecting performance on WM measures. Early studies, in which participants were faced with either actual failure or the threat-of-failure, found robust effects of the ego-threat manipulation on digit span performance. More socially focused manipulations, such as the Trier Social Stress Task, appears to affect performance in the phonological 2- and 3-Back (Schoofs et al., 2008) and to some extent in the Sternberg task (Oei et al., 2006).

These findings are quite promising insofar as they suggest that at least two manipulations are capable of affecting performance on a number of WM tasks (i.e., N-Back, digit span, and Sternberg task). However, these results should be interpreted in the light of a number of caveats. First, these studies have generally not shed light on the underlying mechanisms. Threat-of-shock paradigms robustly increase physiological symptoms of anxiety (Grillon, 2008), increase vigilance (Cornwell et al., 2007), and activate the amygdala (Phelps et al., 2001), which are symptoms that most closely map onto the arousal dimension of anxiety (e.g., Nitschke et al., 2001; Heller, Nitschke, Etienna, & Miller, 1997; Heller, Nitschke, & Lindsay, 1997). However, it cannot be said with certainty that the threat-of-shock manipulation did not induce worry. Similarly, ego-threat studies place participants in a stressful, evaluative situation which is often associated with arousal symptoms. For example, Beilock and colleagues (2007) found that the effects of stereotype threat—a manipulation that is conceptually very similar to ego-threat—had its largest effects on tasks involving phonological material. These studies would be aided by including even a self-reported measure of worry to examine variance that is specific to worry. Additionally, the literature would benefit from studies that attempt to experimentally induce or reduce worry. A number of dual-tasking studies did attempt to do this; however, it is unclear whether asking people to engage in worry in a laboratory setting mimics the process of worry in real life. Expressive writing may offer one avenue forward—although there are likely to be many. Expressive writing has shown some promise with respect to temporarily alleviating worry and its effects on WMC (e.g., Ramirez & Beilock, 2011; also see Klein & Boals, 2001a). Thus, although there does appear to be clear evidence that certain anxiety manipulations are capable of affecting tests scores, there is still much to be learned about the underlying mechanisms.

The second caveat is that anxiety inductions and measures of WMC were often confounded. For example, all studies using the standard ego-threat manipulation measured WMC with a simple span task. Similarly, nearly all of the threat-of-shock studies inferred working memory deficits from versions of the N-Back. Given that simple spans, complex spans, and the N-Back appear to index processes that are at least partially separable (e.g., the OSPAN; Kane et al., 2004, 2007; Redick & Lindsey, 2013), it is unclear whether these effects would generalized to other WM measures. The results of Pyke and Agnew (1963) suggest that it is unlikely that the threat-of-shock design only affects processes specific to the N-Back; however generalizing these results will require the use of multiple measures of WMC.

Finally, there seems to be a WM load × WM modality × affect interaction that inconsistently appears across studies. Threat-of-shock studies have generally found that anxiety impaired performance on all levels of load for a spatial N-Back but only for low
levels of load for a phonological N-Back (however, see Kalisch et al., 2006). Conversely, Schoofs et al. (2008) found that the Trier Social Stress Task affected both phonological 2- and 3-Back performance. Using the Sternberg task, studies have found that anxiety inductions impair performance at low (Porcelli et al., 2008; Exp. 1; however see Porcelli et al., 2008; Exp. 2), moderate load (Duncko et al., 2009) and high load (Oei et al., 2006).

Taking an individual differences approach, Stefanopoulou et al. (2014) failed to find the expected group × load interaction in a phonological N-Back in patients with generalized anxiety disorder. Klein and Boals (2001b) reported that correlations between a measure of state-anxiety and OSPAN scores were fairly reliable across set sizes. Similarly, Schweizer and Dalgleish (2011) presented reading span scores as a function of set size for a sample of PTSD patients. Although the interaction term was not presented, and inspection of their Table 1 suggests that the interaction was likely nonsignificant for affectively neutral material. Lastly, this review presented novel data indicating that self-reported arousal predicts poorer performance on a phonological 3-Back.

One could rebut this limitation by proposing (a) that some of these manipulations were of questionable effectiveness, (b) that symptoms were not provoked in Stefanopoulou et al. (2014); Klein and Boals (2001b) and Schweizer and Dalgleish (2011) and it is therefore unclear whether participants were actually symptomatic during the task, or (c) it may be that these different versions of the N-Back were not well-matched or that the variable requirements of the N-Back, OSPAN and Sternberg tasks make it impossible to directly compare performance. Such rebuttals would certainly be fair. However, these points do suggest that at the very least there are subtle differences between self-report measures, manipulations and WM measures that have not been accounted for. There would seem to be a clear need to use well-characterized measures and manipulations as well as to understand the boundary conditions on the effects of anxiety.

With respect to models that consider low WMC to be a risk factor for anxiety (e.g., Quinet et al., 2009; also see Posner & Rothbart, 2000), the current review found mixed evidence relevant to the causal status of WMC. There appear to be two primary factors responsible for limiting the conclusions that can be drawn regarding WMC as a risk factor. First, although anxiety can be readily manipulated in a laboratory setting, it is unclear whether interventions are capable of manipulating WMC (e.g., Redick et al., 2013; Shipstead et al., 2010). This situation severely limits the causal inferences that can be drawn for WMC and anxiety. A few studies found that adaptive working memory training programs failed to reduce anxiety symptoms (Åkerlund et al., 2013; Roughan & Hadwin, 2011). On the other hand, Sari and colleagues (in press) demonstrated that the degree to which anxious individuals engaged in the training task predicted decreased anxiety following training. It is currently unclear why these findings diverge. It may be that participants in Åkerlund et al. (2013) and Roughan and Hadwin (2011) were less engaged in the training task than in Sari et al. (in press) or that some methods of WMC training (e.g., CogMed vs. dual N-Back) might be more effective than others. Alternately, training may be more effective for some populations than others. For example, Sari et al. (in press) included participants preselected for trait anxiety. Similarly, Studer-Luuethi et al. (2012) and Jaeggi et al. (2014) have shown that individual differences in conscientiousness, mindset and neuroticism moderate the effectiveness of WMC training (however, see Thompson et al., [2013] for a failure to replicate some of these findings). In any case, Sari et al. (in press) provides some of the most direct evidence that reduced WMC can act as a risk factor for anxiety; however, given some of the inconsistencies in the literature, as well as the debate regarding WMC training, replicating these findings and determining under what circumstances WMC training affects anxiety seem to warrant further work.

Second, the majority of the studies in the current review have focused on adult populations. In the absence of novel, prolonged stressors (e.g., combat; Macklin et al., 1998), adults are somewhat unlikely to show large increases in anxiety or rates of diagnosis over a short period of time (Bredemeier & Berenbaum [2013] proposed that their effect resulted from upcoming midterm exams). In general, adult-like anxiety, as well as the stress responses which predispose one to anxiety, begin to appear during pubertal development (e.g., Gunnar & Vazquez, 2006; Reardon et al., 2009; Sumter et al., 2010; also see APA, 2013). With respect to child studies, the present meta-analysis found that effect sizes were statistically equivalent for adults and children. However, there were relatively few studies conducted in child populations and many of the child studies included somewhat wide age ranges (e.g., 6–13 years old). This means that the current meta-analysis lacked the sensitivity to detect the developmental course of anxiety-related impairments in WMC.

Conducting longitudinal/prospective studies would seem to represent a fruitful avenue for moving forward. The small number of prospective correlational studies also hint at the possibility that lower WMC may act as a risk factor for later anxiety. Specifically, Macklin et al. (1998) and Bredemeier and Berenbaum (2013) have shown that lower general intelligence and N-Back scores, respectively, predict later anxiety. However, as noted earlier, the inferences that can be drawn from these studies is limited by several factors including (a) the use of intelligence scores rather than WMC scores, (b) confounding factors such as combat severity, (c) designs that were not fully crossed, and (d) the fact that these studies tended to be conducted in adults. Both the risk factors for developing anxiety as well as the downstream effects of anxiety are likely going to be most easily detected in longitudinal studies which follow children and adolescents through puberty. Moving forward, then, it seems that there is much to be gained from conducting lagged cross-correlational studies in children. If WMC is indeed a risk factor for later anxiety, then measures of WMC recorded prior to puberty would be expected to predict post-pubertal anxiety; additionally, measures of anxiety recorded prior to puberty should not predict post-pubertal measures of WMC. This design has the additional advantage of being able to evaluate more sophisticated models.

For example, it is possible that the relationship between anxiety and WMC is more cyclical than many theories propose. Perhaps individual differences in WMC increase the likelihood of later anxiety which, in turn, interferes with working memory (as depicted in Figure 3; see Van Bockstaele et al., 2014 for a similar proposal regarding attentional biases). This possibility would be broadly consistent with the work of Hirsch and Mathews’ (2012) proposal that the control of worrisome thoughts is reliant on domain-general attentional control abilities. Importantly, worry itself is assumed to impair attentional control abilities in this model. Thus, in this model, worry and WMC are capable of...
bidirectionally affecting each other. Hallion and colleagues (2014) have provided some initial support of this model. In this study, 2-Back performance suffered when participants attempted to control worries relative to neutral thoughts suggesting that the control of worry is reliant on available WMC (also see Gorlin & Teachman, 2015).

Although there has not been as much experimental work as there has been correlational work, a few conclusions regarding the causal status of anxiety and WMC seem warranted. Most importantly, work using the threat-of-shock method has provided fairly strong evidence that an anxiety induction is capable of impairing performance on at least one measure of WMC—the N-Back. Additionally, ego-threat studies, while suffering from some limitations, appear capable of impairing performance on simple span measures. However, the review also identified a number of limitations to be addressed by future research: (a) experimental studies have often paired a particular anxiety induction with a particular WMC task rather than examining the generality of their findings, (b) lagged longitudinal studies can help establish the temporal precedence of one construct over the other, (c) determine the conditions under which cognitive load affects the anxiety/WMC association, and (d) further explore the effects of WMC-training of anxiety symptoms using well-validated training methods.

Conclusions

Over the last 60 years, a great deal of research has been conducted with the intent of examining the relationship between anxiety and WM. It seems clear form this review that (a) measures of anxiety predict poorer performance on measures of WMC, (b) impaired WMC mediates the association between anxiety and some real-world outcomes such as academic performance and treatment response, and (c) anxiety inductions adversely affect performance on WM measures at least under certain circumstances. However, there are still basic questions regarding the WM processes most affected by anxiety. Future work using multimethod assessments and well-characterized experimental manipulations will help to answer these questions.

References

References marked with an asterisk were included in the primary meta-analysis.


tions of worry, somatic anxiety, and depression on emotional experience. 
http://dx.doi.org/10.1016/0005-7916(88)90006-7


at school. Neurobiology of Learning and Memory, 83, 33–42. 
http://dx.doi.org/10.1016/j.nlm.2004.06.010


Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, 
http://dx.doi.org/10.3758/s13423-014-0699-x

"Aycicegi, A., Dinn, W. M., Harris, C. L., & Erkmen, H. (2003). Neuro-
psychological function in obsessive-compulsive disorder: Effects of comorbid conditions on task performance. European Psychiatry, 18, 


Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., 
and nonanxious children: A meta-analytic study. Psychological Bulletin, 133, 
1–24. http://dx.doi.org/10.1037/0033-2909.133.1.1

Barlow, D. H., Chorpita, B. F., & Turovsky, J. (1996). Fear, panic, anxiety, 
and disorders of emotion. In D. A. Hope (Ed.), Nebraska Symposium on 
Motivation, 1995: Perspectives on anxiety, panic, and fear. Current 
time and research in motivation (Vol. 43, pp. 231–238). Lincoln, NE: 
University of Nebraska Press.

making and the orbitofrontal cortex. Cerebral Cortex, 10, 295–307. 
http://dx.doi.org/10.1093/cercor/10.3.295

Mahoney (Ed.), Cognitive and constructive psychotherapies: Theory, 
research and practice (pp. 29–40). New York, NY: Springer.

anxiety: Automatic and strategic processes. Behaviour Research and 
Therapy, 35, 49–58. http://dx.doi.org/10.1016/S0005-7967(96)00069-1

Beers, S. R., Rosenberg, D. R., Dick, E. L., Williams, T., O’Hearn, K. M., 
Birmaher, B., & Ryan, C. M. (1999). Neuropsychological study of frontal lobe function in psychotrópic-naive children with obsessive-
compulsive disorder. The American Journal of Psychiatry, 156, 777– 
779.

and working memory: Mechanisms, alleviation, and spillover. Journal 

Doctoral Dissertation, The University of Utah.

Thomas.

Berggren, N., & Derakshan, N. (2013). Attentional control deficits in trait 
anxiety: Why do you see them and why don’t. Biological Psychology, 
92, 440–446. http://dx.doi.org/10.1016/j.biopsycho.2012.03.007

Bleich-Cohen, M., Tzur, T., Weizman, R., Faragian, S., Weizman, A., 
2013.02.004

Boldrini, M., Del Pace, L., Placidi, G. P. A., Keilp, J., Ellis, S. P., Signori, 
S., ... Cappa, S. F. (2005). Selective cognitive deficits in obsessive-

cognitive control and posttraumatic stress symptoms. Journal of Behav 


0067-5967(83)90121-3

Bourke, C., Porter, R. J., Carter, J. D., McIntosh, V. V., Jordan, J., Bell, 
C. ... Joyce, P. R. (2012). Comparison of neuropsychological func 
tioning and emotional processing in major depression and social anxiety 
disorder subjects, and matched healthy controls. Australian and Ne 

Brandes, D., Ben-Schachar, G., Gilboa, A., Bonne, O., Freedman, S., & 
Shalev, A. Y. (2002). PTSD symptoms and cognitive performance in 
recent trauma survivors. Psychiatry Research, 110, 231–238. http://dx .doi.org/10.1016/S0165-1781(02)00125-7

Bredemeier, K., & Berenbaum, H. (2013). Cross-sectional and longitudi 

Brewin, C. R., & Smart, L. (2005). Working memory capacity and sup 
pression of intrusive thoughts. Journal of Behavior Therapy and Exper 
006.

1080/02699938808410922

WAIS-IV arithmetic performance in undergraduates. Archives of Clini 

Butters, M. A., Bhalla, R. K., Andrecsu, C., Wetherell, J. L., Mantella, 
functioning following treatment for late-life generalised anxiety disor 

ONE, 5, e9708. http://dx.doi.org/10.1371/journal.pone.0009708


ANXIETY AND WORKING MEMORY


ANXIETY AND WORKING MEMORY


Johnson, D. R., & Granel, S. D. (2009). Individuals lower in working memory capacity are particularly vulnerable to anxiety’s disruptive effect on performance. Anxiety, Stress, & Coping, 22, 201–213. http://dx.doi.org/10.1080/10615800802291277


ANXIETY AND WORKING MEMORY

31


convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, 104, 3–14. [dx.doi.org/10.1037/0021-843X.104.1.3](http://dx.doi.org/10.1037/0021-843X.104.1.3)


Received February 9, 2015
Revision received February 1, 2016
Accepted February 4, 2016