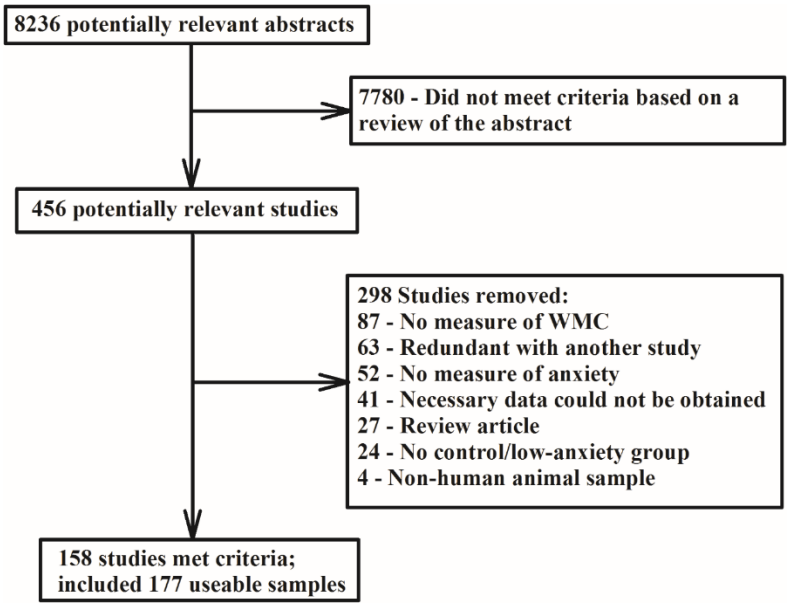
Supplemental Materials

Meta-Analytic Methods

Literature Search. Published articles were obtained via a search of PubMed, PsycInfo, Scopus, EMBASE and Google Scholar databases. Additionally, unpublished theses and dissertations were obtained via a search of ProQuest. I crossed the following words – anxiety, arousal, GAD, OCD, PTSD, phobia, panic, social anxiety, and worry – with working memory, working memory capacity, capacity, span, memory span, complex span, simple span, OSPAN, RSPAN, digit span, and N-Back. The references for all empirical and review articles were systematically searched for additional articles.

The following criteria were used to include/exclude studies from the meta-analysis: 1) The study was published (or in press) in the English language by October 2015. 2) The study included either participants who received an anxiety diagnosis using the DSM or ICD classification systems or participants who completed a self-report measure of anxiety. 3) The study included either a simple or complex memory span task (e.g. digit span or OSPAN, respectively) or dynamic task (e.g. N-Back). And 4) the study reported enough information to allow for the computation of an effect size for the association between anxiety symptoms and behavioral performance in a WMC task. If it appeared that relevant data were collected but were not explicitly reported, the corresponding author was emailed a request for the necessary data. The study-selection process is depicted in Supplementary Figure 1. These criteria resulted in 177 samples which included N = 22,061 individuals.

Supplementary Figure 1.



Supplementary Figure 1. A flowchart of the literature review process for the primary meta-analysis.

Coding Decisions.

1) When participants were divided into more than 2 groups based on degree of self-reported anxiety (e.g. low, medium and high), the two most extreme groups were selected for inclusion.
2) When a study reported on multiple WMC tasks that were equivalent based on the criteria listed in the main text (e.g. two simple span tasks involving verbally-rehearsable material), I computed an effect size that reflected the average of the effect sizes for those measures.
3) When multiple time-points were collected (e.g. pre- and post-treatment), only the first time-point was included in the analysis. Similarly, if a study involved a comparison between WMC recorded in a baseline condition and WMC recorded during a manipulation, only the baseline condition was included in this analysis.

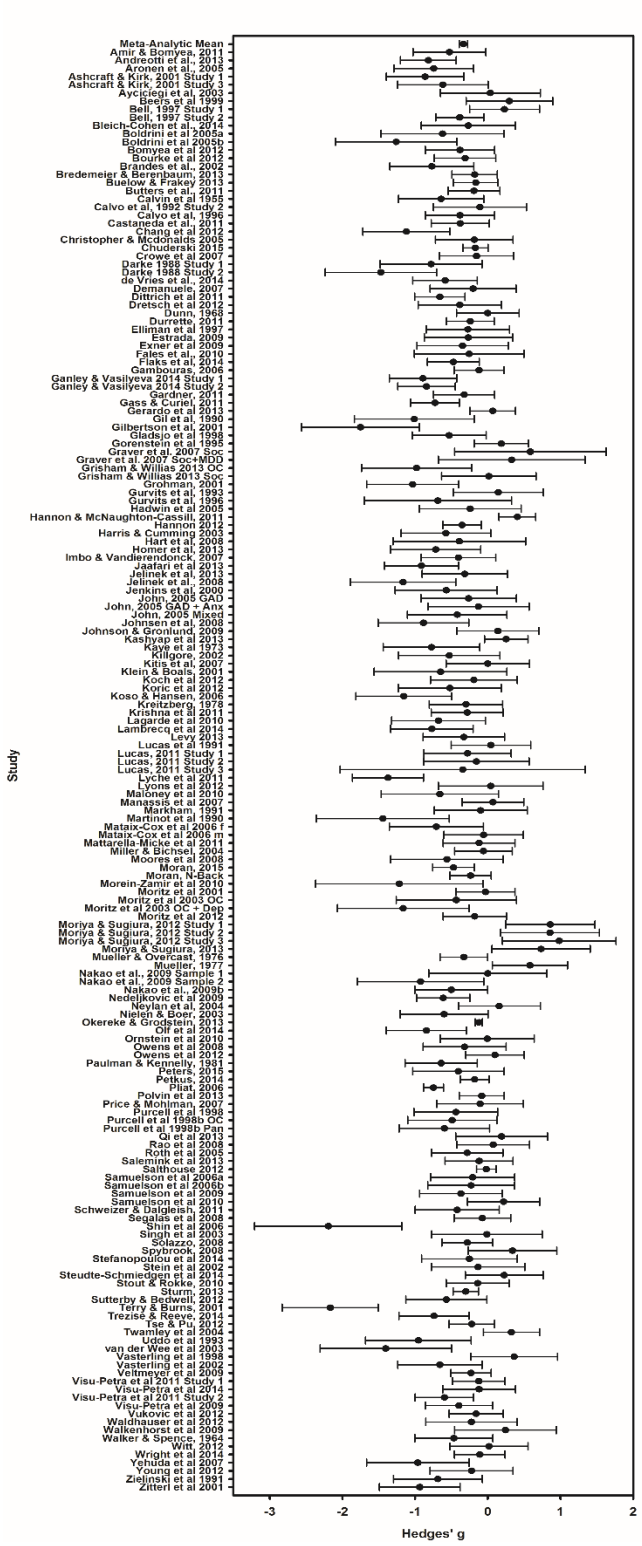
4) When studies involved a comparison of WMC for affectively-neutral material with WMC for affective material, only the neutral condition was included.
5) When two or more anxiety groups were compared with a single control group, the sample size for the control group was divided by the number of anxiety groups to avoid artificially inflating the sample size (see Bar-Haim et al., 2007 for a similar procedure).
6) For studies in which multiple types of anxiety were measured in a single sample of participants (e.g. trait-anxiety and worry), I included the type of anxiety for which there were fewer samples available (e.g. worry was less common than trait-anxiety; therefore, worry was selected over trait-anxiety when they were presented in the same sample) in order to increase the power to detect effects involving type of anxiety. Similarly, when multiple, theoretically non-equivalent, types of WMC were collected (e.g. simple and complex span) I included the WMC measure that was less common.
7) Given that N-Back results were inconsistently reported (some report the results of all levels of n separately; others report only a subset of n and still other report only the main effect of anxiety status across all levels of n), the results presented in meta-analysis reflect the average effect size for all levels of n presented in a given study.
8) When an effect was reported as non-significant and the information required to compute an effect size could not be obtained, I estimated an effect size assuming $p = .5$ in order to ensure non-significant results were not ignored (Cooper & Hedges, 1994).

An independent rater coded the primary articles included in the meta-analysis in order to establish inter-rater reliability. Across all variables, kappas ranged between .71 and 1.00 (*M* =

.92). Disagreements were resolved by discussion. The final set of studies and coding decisions reflected the consensus reached by the raters.

Meta-Analytic Computations. The focal effect size for this analysis was Hedges' g (Hedges, 1981). Hedges' *g* is computed from the more commonly used Cohen's d (the difference between two means standardized by their pooled standard deviations) and can be interpreted in much the same way. Hedges (1981) noted the tendency for Cohen's *d* to inflate the population effect size when estimated with small samples. To correct for this inflation, Hedges (1981) recommended that *d* values be multiplied by a correction factor approximately equal to $1 - \frac{3}{4*\mathrm{df}-1}$. Effect sizes were coded such that a negative *g* value indicates that anxious individuals show reduced working memory capacities whereas a positive value indicates that anxious individuals show greater capacities. All computations were conducted using Comprehensive Meta-Analysis Software (v.2; Borenstien et al., 2005). Given that 1) this review covers multiple types of anxiety/populations, 2) this review covers multiple types of WMC tasks (i.e. simple/complex and phonological/spatial) and 3) the studies reported rather heterogeneous effect sizes ($Q(176) = 536.89$, $p < .001$, $I^2 = 67.22$), this meta-analysis was conducted within the framework of a random-effects model. The meta-analytic mean and individual effects are presented in Supplementary Figure 2.

Supplementary Figure 2.



Supplementary Figure 2. A forest plot depicting individual effects and 95% CIs for the primary meta-analysis.

Supplementary Results for the Meta-Analysis.

There were 20 effect sizes included in the meta-analysis that could be considered outliers (≥ 2.5 SDs from the mean). This section presents the results with those effects removed. Consistent with the main results, the pooled effect size was moderate and significant (g = -.31, k = 157, 95% CI: -.36; -.26, $p = 10^{-30}$).

Moderator analyses are presented in Supplementary Table 1. These results closely resemble those presented in the manuscript. The main exception is that the effect size for non-spatial visual tasks is now negative and significant. This is due to the fact that many of the studies that included non-spatial visual tasks reported large, positive effects.

Supplementary Table 1. Meta-Analytic Results for High-Anxiety vs. Low-Anxiety Comparisons

Moderator	g (95% CI)	k	p	Q[a] (df)	p[a]
Age				1.46(1)	.23
Adults	-.32 (-.38; -.27)	132	<.001	–	–
Children	-.23 (-.36; -.11)	25	<.001	–	–
Anxiety Dimension				1.15(1)	.28
Arousal	-.33 (-.44; -.21)	43	<.001	–	–
Worry	-.45 (-.65; -.25)	11	<.001	–	–
Trait/State				0.12(1)	.73
Trait	-.24 (-.36; -.13)	25	<.001	–	–
State	-.21 (-.38; -.03)	11	.02	–	–
Severity				4.84(1)	.03
Clinical	-.37 (-.45; -.30)	88	<.001	–	–
Non-Clinical	-.25 (-.33; -.18)	68	<.001	–	–
WM Span Type				0.50(2)	.78
Complex	-.35 (-.46; -.23)	28	<.001	–	–
Simple	-.30 (-.36; -.24)	112	<.001	–	–
Dynamic	-.29 (-.46; -.13)	17	<.001	–	–
WM Span Domain				3.42(2)	.18
Spatial	-.38 (-.47; -.29)	50	<.001	–	–
Phonological	-.27 (-.34; -.21)	98	<.001	–	–
Visual	-.29 (-.51; -.08)	7	.008	–	–
WM Span Type X Severity: Clinical				3.29(2)	.19
Complex	-.57 (-.82; -.32)	8	<.001	–	–
Simple	-.32 (-.41; -.24)	72	<.001	–	–
Dynamic	-.33 (-.58; -.08)	8	.01	–	–
WM Span Type X Severity: Non-Clinical				0.01(2)	.99
Complex	-.25 (-.40; -.14)	20	<.001	–	–
Simple	-.24 (-.36; -.18)	39	<.001	–	–
Dynamic	-.26 (-.48; -.05)	9	.02	–	–
Anxiety Dimension X Domain: Arousal				3.42(2)	.18
Spatial	-.39 (-.59; -.19)	13	<.001	–	–
Phonological	-.27 (-.40; -.13)	28	<.001	–	–
Visual	-.71 (-1.2; -.21)	2	.005	–	–
Anxiety Dimension X Domain: Worry				2.43(1)	.12
Spatial	-.68 (-1.02; -.33)	3	<.001	–	–
Phonological	-.34 (-.58; -.10)	7	.006	–	–
Manuscript Type				0.22(1)	.64
Peer-Reviewed	-.30 (-.35; -.25)	135	<.001	–	–
Thesis/Dissertation	-.33 (-.46; -.21)	22	<.001	–	–

[a] Q statistic and p value for comparison between subcategories of a moderator

N-Back Methods

Participants. 197 undergraduates (*M age* = 20.29, *SD* = 2.90) from a large Midwestern university completed a 3-Back version of the N-Back in exchange for partial course credit.

Stimuli and Procedure. Participants were presented with a sequence of letters selected from the following list of phonologically distinct letters: B, F, H, K, M, Q, R and X. Participants were required to indicate whether the current letter matched the letter presented three back as quickly and accurately as possible (for example, the second "B" in this sequence: B-Q-X-B-R). There were four lists of 48 trials. During each list, each letter appeared as non-targets (items that do not match the 3-back letter) five times and as a target once.

Each trial began with the presentation of a black fixation cross against a white background for 500 ms. This was followed by a letter stimulus presented in a standard, black font for 500 ms. Each letter was followed by a 2000 ms ISI during which participants were able to respond. Letter presentations were randomly varied between upper and lower case to prevent matching based on perceptual features alone.

Self-Reports. Following the N-Back, participants completed the Penn State Worry Questionnaire (PSWQ; Meyer, Miller, Metzger, & Borkovec, 1990) and the Anxious Arousal Subscale of the Mood and Anxiety Symptoms Questionnaires (MASQ; Watson & Clark, 1991). Individual items were averaged to yield scale scores.

References

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychological Bulletin, 133*, 1-24.

Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2005). Comprehensive Meta-analysis Version 2. Englewood, NJ: Biostat.

Cooper, H. & Hedges, L.V. (1994). *The handbook of research synthesis.* New York: Russell Sage Foundation.

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.

Meyer, T.J., Miller, M.L., Metzger, R.L., & Borkovec, T.D. (1990). Development and validation of the penn state worry questionnaire. *Behaviour Research and Therapy, 28*, 487-495.

Watson, D., & Clark, L.A. (1991). The Mood and Anxiety Symptom Questionnaire. Unpublished manuscript. University of Iowa, Iowa City, IA.