



Published in final edited form as:

Behav Res Ther. 2010 July ; 48(7): 614–625. doi:10.1016/j.brat.2010.03.013.

Overgeneral Autobiographical Memory as a Predictor of the Course of Depression: A Meta-Analysis

Jennifer A. Sumner^{a,*}, James W. Griffith^{a,1}, and Susan Mineka^a

^aNorthwestern University, Department of Psychology, 2029 Sheridan Road, Suite 102 Evanston, IL 60208, USA

Abstract

Overgeneral autobiographical memory (OGM) is a robust phenomenon in depression, but the extent to which OGM predicts the course of depression is not well-established. This meta-analysis synthesized data from 15 studies to examine the degree to which OGM 1) correlates with depressive symptoms at follow-up, and 2) predicts depressive symptoms at follow-up over and above initial depressive symptoms. Although the effects are small, specific and categoric/overgeneral memories generated during the Autobiographical Memory Test significantly predicted the course of depression. Fewer specific memories and more categoric/overgeneral memories were associated with higher follow-up depressive symptoms, and predicted higher follow-up symptoms over and above initial symptoms. Potential moderators were also examined. The age and clinical depression status of participants, as well as the length of follow-up between the two depressive symptom assessments, significantly moderated the predictive relationship between OGM and the course of depression. The predictive relationship between specific memories and follow-up depressive symptoms became greater with increasing age and a shorter length of follow-up, and the predictive relationship was stronger for participants with clinical depression diagnoses than for nonclinical participants. These findings highlight OGM as a predictor of the course of depression, and future studies should investigate the mechanisms underlying this relationship.

Keywords

Autobiographical memory specificity; Overgeneral autobiographical memory; Course of depression; Meta-analysis

Overgeneral Autobiographical Memory as a Predictor of the Course of Depression: A Meta-Analysis

Over the past 20 years, a large body of research has accumulated on the overgeneral autobiographical memory (OGM) phenomenon in depression. First described by Williams and Broadbent (1986) in their study of suicidal patients, OGM refers to the finding that, when asked to come up with a specific memory in response to a cue word, some individuals are less specific

© 2010 Elsevier Ltd. All rights reserved.

*Correspondence concerning this article should be addressed to Jennifer A. Sumner, Department of Psychology, Northwestern University, 2029 Sheridan Road, Suite 102, Evanston, IL 60208. Jennifer.Sumner@gmail.com, Tel.: 1 847 491 3843; fax: 1 847 491 7859.

¹Present address: Department of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, Chicago, IL.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and/or more overgeneral in their memory retrieval than others. In particular, much research has shown that individuals with depression are characterized by higher levels of OGM than nondepressed controls (Williams et al., 2007). Moreover, OGM has been proposed as a risk factor for the onset and course of depression. This phenomenon appears to be relatively specific to depression rather than being characteristic of psychopathology in general (although it has also been associated with traumatic experiences and trauma-related disorders, such as PTSD and acute stress disorder; Moore & Zoellner, 2007; Williams et al., 2007).

In the majority of studies on OGM, the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986) has been used to assess the specificity of autobiographical memory. On the AMT, individuals are presented with cue words of different valences, and are asked to produce a specific memory related to the cue word within a given time limit (e.g., 30 seconds). A specific memory is defined as a memory for an event that occurred at a particular time and place and lasted less than one day (e.g., “my high school graduation;” Williams et al., 2007). In contrast, overgeneral memories include both categoric memories that refer to a class of generic events (e.g., “parties with my friends”) and extended memories that refer to an event lasting more than one day (e.g., “when I was on vacation last month”). Researchers have most frequently analyzed specific memories and/or categoric memories, although some have presented results for overgeneral memories more broadly (e.g., categoric and extended memories). In this review, and in most other work on this topic, OGM refers to either the retrieval of fewer specific memories and/or more categoric/overgeneral memories.

The OGM phenomenon appears to be a robust and replicable phenomenon among individuals with clinical depression as evidenced by recent meta-analytic and literature reviews (van Vreeswijk & de Wilde, 2004; Williams et al., 2007). However, in contrast to studies of individuals with clinical depression, there is less consistency in the findings of studies with nonclinical samples.¹ Some studies have found that dysphoric individuals are less specific in their memory than nondysphoric respondents (e.g., Goddard, Dritschel, & Burton, 1997), but other studies have failed to detect this phenomenon (e.g., Raes, Pousset, & Hermans, 2004). However, this pattern of results may be due to insensitivity of the AMT as a measure of OGM in nonclinical samples rather than to an absence of the phenomenon in these groups. For example, our item response theory analyses of AMT performance suggested that the AMT may be insufficiently sensitive to measure OGM in nonclinical samples (Griffith et al., 2009). Furthermore, Raes, Hermans, Williams, and Eelen (2007) used an alternative sentence-completion methodology that does not explicitly prompt respondents to retrieve specific memories. They found that overgeneral responding on this measure was indeed associated with increased levels of depressive symptoms in nonclinical samples, even when traditional AMT performance was not.

OGM has been posited as a trait-like characteristic that may serve as a vulnerability factor for depression (Williams et al., 2007). For example, OGM has been associated with later increases in depressive symptoms in nonclinical samples. In one study, higher levels of OGM (relative to lower levels) predicted higher levels of depressive symptoms after a failed in vitro fertilization attempt (van Minnen, Wessel, Verhaak, & Smeenk, 2005). Higher levels of OGM are also sometimes (but not always) observed in individuals in remission from depression (e.g., Mackinger, Pachinger, Leibetseder, & Fartacek, 2000; Wessel, Meeren, Peeters, Arntz, & Merckelbach, 2001), thereby suggesting that OGM is not merely a correlate of depressed mood.

¹In this review, nonclinical samples refer to samples of individuals who were not selected on the basis of the level of depressive symptoms or a depression diagnosis. Examples include pregnant women who were recruited from the community and college students in an Introductory Psychology course. Individuals in these samples may be described as dysphoric (those with elevated scores on a measure of depressive symptoms) or nondysphoric. However, in nonclinical studies, the presence of clinical depression is generally not assessed with a diagnostic interview.

As a vulnerability factor, OGM should also predict the course of depression (Williams et al., 2007), but the results of empirical studies are inconsistent. Some studies, for example, have shown that AMT performance predicts depressive symptoms at follow-up approximately 7 months after an initial assessment (e.g., Brittlebank, Scott, Ferrier, & Williams, 1993; Dalgleish, Spinks, Yiend, & Kuyken, 2001; Raes et al., 2006). Specifically, clinically depressed individuals who retrieve more categoric and/or fewer specific memories have higher levels of depressive symptoms at follow-up, even after covarying baseline symptoms. In addition, Hermans et al. (2008) showed that patients with major depressive disorder (MDD) who retrieved fewer specific memories and more categoric memories upon hospital admission were more likely than patients not characterized by OGM at the initial assessment to still meet criteria for MDD 3–4 weeks later. Together, these studies suggest that OGM predicts the maintenance of depression (i.e., compared to lower levels of OGM, higher levels of OGM are associated with less of a decrease in depressive symptoms over time). However, Brewin, Reynolds, and Tata (1999) failed to detect a significant predictive relationship between OGM and depressive symptoms at 6-month follow-up as measured by the Beck Depression Inventory (BDI) in a sample of patients with MDD.

The aim of the current review was to perform a systematic quantitative analysis of the extent to which OGM predicts the course of depression. This was examined in two ways. First, we analyzed correlations between OGM and the level of depressive symptoms at a follow-up assessment. Second, we examined standardized regression (β) coefficients for OGM predicting depressive symptoms at follow-up in order to examine the predictive power of OGM over and above initial depressive symptom levels. All studies included in this second set of analyses incorporated only 1) a measure of OGM and a baseline measure of depressive symptoms as predictors, and 2) a follow-up measure of depressive symptoms as an outcome variable in their regression models.

Consideration of Moderator Variables

Several variables could moderate the relationship between OGM and the course of depression. Thus, we had a secondary goal of examining two classes of potential moderators: characteristics of the sample and characteristics of the study design. Given that there is little research on this issue, many of these analyses represent preliminary exploratory investigations, although the potential moderator variables we chose to examine are relevant from a theoretical standpoint.

Characteristics of the sample—As described above, OGM is a replicable phenomenon among individuals with clinical depression, but it is less consistently detected in nonclinical samples. Thus, we anticipated that the clinical depression status of participants (i.e., patients with clinical depression diagnoses versus nonclinical participants) would be an important moderator. Specifically, we hypothesized that the predictive relationship between AMT performance and the course of depression might be greater for samples of individuals with a clinical diagnosis of depression than for nonclinical samples, as suggested by Raes et al. (2007).

We also examined the age of participants as a moderator. Research shows that aging is associated with declines in executive functioning (e.g., Salthouse, Atkinson, & Berish, 2003), which in turn may contribute to difficulties in retrieving specific memories (e.g., Dalgleish et al., 2007). Consequently, older participants may retrieve fewer specific memories on the AMT than younger participants. Indeed, Ros, Latorre, and Serrano (2009) recently found that older adults generated fewer specific memories and more categoric memories on the AMT than did younger adults, and deficits in working memory were associated with lower levels of memory specificity. Given this more pronounced OGM phenomenon in older (compared to younger) adults, we were interested in exploring the preliminary hypothesis that the predictive

relationship between OGM and the course of depression might also be greater for older than younger participants.

Characteristics of study design—This class of potential moderator variables included: a) the valence of the AMT cue word, b) the measure of depressive symptoms used, and c) the length of follow-up between the two assessments of depressive symptoms.

Cue word valence was chosen as a potential moderator variable because it is often taken into account in studies of OGM, such that memories retrieved in response to positive and negative cue words are analyzed separately. However, findings with respect to valence effects have been highly inconsistent. For example, some studies have found that depressed individuals generate fewer specific and/or more overgeneral memories to positive than negative cue words (e.g., Park, Goodyer, & Teasdale, 2002), whereas others have detected the opposite pattern (e.g., Mackinger, Pachinger et al., 2000). The prediction of depressive symptoms over time based on memories to cue words of different valence has also been inconsistent across studies (e.g., a significant predictive relationship with positive cues, Brittlebank et al., 1993, versus a significant predictive relationship with negative cues, Raes et al., 2006).

However, when 14 studies of the OGM phenomenon were meta-analyzed, van Vreeswijk and de Wilde (2004) found that effect sizes for positive and negative cue words on the AMT were highly inter-correlated. Additionally, in a study of three independent samples, confirmatory factor analyses of the structure of responses on the AMT found that a one-factor model provided a good fit, and did not differ significantly from models that included factors for cue word valence (Griffith et al., 2009). However, neither of these studies used meta-analytic techniques to examine whether cue word valence might moderate the predictive relationship between OGM and the course of depression, as we do here.

We also hypothesized that the measure of depressive symptoms employed might moderate the predictive relationship between OGM and the course of depression. As mentioned above, Brewin et al. (1999) failed to detect a significant predictive relationship between OGM and depression using the BDI as the measure of depressive symptoms, and several subsequent studies using the BDI also obtained nonsignificant results (e.g., Dalgleish et al., 2001; Hermans et al., 2008). After finding discrepant results when using the BDI and Hamilton Rating Scale for Depression (HRSD) to measure change in depressive symptoms over time, Dalgleish et al. (2001) noted that these measures differ in both their mode of administration (self- versus clinician-rated) and content (the BDI emphasizes cognitive symptoms of depression, whereas the HRSD focuses on somatic-vegetative symptoms of depression). They suggested that either of these differences might explain the differential results. However, aside from their study, comparisons of the relationship between OGM and change in both the BDI and HRSD have not been conducted within a single investigation. Additionally, other research examining OGM and the somatic and cognitive-affective symptoms of depression has not found consistent evidence of differential relationships between OGM and these two types of symptoms (e.g., Mackinger & Svaldi, 2004; Roberts, Carlos, & Kashdan, 2006). Therefore, a more systematic investigation of this issue would be useful to better understand the relationship between OGM and the course of depression. We focused on the rating method for depressive symptoms (self-rated versus clinician-rated) because this information was available for all of the measures employed in the studies included in this meta-analysis.

Finally, we also examined whether the predictive relationship between OGM and the course of depression might vary as a function of the length of follow-up between the two assessments of depressive symptoms. There is substantial variability in the length of follow-up in the extant literature, with follow-up periods ranging from less than one month (e.g., 25.9 days; Mackinger et al., 2004) to four years (Bryant, Sutherland, & Guthrie, 2007). Thus, we were interested in

conducting a systematic investigation of how the length of follow-up might be associated with the relationship between OGM and the course of depression.

Summary

This meta-analytic review had several aims. First, the main goal was to examine the overall relationship between OGM and the course of depression. This was done by examining whether OGM: 1) is correlated with a measure of depressive symptoms at follow-up, and 2) predicts depressive symptoms at follow-up over and above initial levels of depressive symptoms. Second, several characteristics of the sample and study design were examined as potential moderators of the predictive relationship between OGM and the course of depression.

Design of the Meta-Analysis

Assessment of OGM

In the current review, two operational definitions of OGM were used: the number of specific memories and the number of categoric/overgeneral memories. In Conway and Pleydell-Pearce's (2000) hierarchy of autobiographical memory, different levels reflect more general knowledge (such as that at the level of lifetime periods or general events) and more event-specific knowledge. They propose that during the strategic retrieval of a specific memory, individuals move from the more general levels of the hierarchy to the level of event-specific knowledge. Thus, both more overgeneral and fewer specific memories retrieved during the AMT are thought to be indices of the OGM phenomenon. Some studies present results for overgeneral memories whereas others present results for different subtypes of overgeneral memories, such as categoric or extended memories (Williams et al., 2007). OGM measures based on overgeneral and categoric memories (the most frequently reported single index of overgeneral memories) were treated equivalently in this meta-analysis.

Even though some researchers consider specific and overgeneral memories to be at opposite ends of a single dimension, unfortunately the different handling of omissions on the AMT by different researchers precludes the combination of specific and overgeneral responses. Specifically, some studies have considered omissions to be overgeneral responses, whereas others have viewed them simply as a failure to respond (rather than a failure to retrieve a specific memory). Studies that employ the first approach could have an inflated estimate of overgeneral memories with respect to studies that adopt the latter approach, thereby seriously interfering with the comparability of results. Furthermore, information on how omissions are handled is reported only infrequently. Thus, specific and categoric/overgeneral memories were examined separately in the current review as in the meta-analysis by van Vreeswijk and de Wilde (2004).

Assessment of Depression

Studies using several measures of depression were included in the meta-analysis (as described below). All measures assessed levels of depressive symptoms, although some were clinician-rated [e.g., HRSD, Montgomery-Asberg Depression Rating Scale (MADRS)] and others were self-rated (e.g., BDI).

Method

Sample of Studies

A computerized search of electronic databases was conducted using the following key word terms: *autobiographical memory specificity*, *overgeneral autobiographical memory*, and *autobiographical memory and depression*. These terms were entered into the following databases from the beginning point of each database through December 2008: PsycINFO,

PubMed, Social Sciences Abstracts, and ProQuest Dissertations and Theses. Web of Science was also searched for references citing the seminal article by Brittlebank et al. (1993) on OGM as a predictor of the course of depression. In addition, the reference list of a recent review article on OGM in emotional disorders by Williams et al. (2007) was reviewed for possible studies to include. Finally, we sent a request for relevant unpublished data to an established email network for researchers with an interest in the OGM phenomenon.

Inclusion and Exclusion Criteria

Studies were included in the meta-analysis only if they: 1) used the AMT as the measure of OGM; 2) presented results for specific memories and/or categoric memories and/or overgeneral memories; 3) measured depressive symptoms at a minimum of two time-points; 4) presented the correlation between OGM and depressive symptoms at follow-up and/or the standardized regression coefficient for OGM predicting follow-up depressive symptoms with an initial level of depressive symptoms as a covariate; and 5) were published in English.

The search of the PsycINFO, PubMed, Social Sciences Abstracts, and ProQuest Dissertations and Theses databases provided 2653 citations, and 164 citations were obtained from the Web of Science search for references citing Brittlebank et al. (1993). An additional nine studies were identified by checking the references of Williams et al. (2007), and we obtained one unpublished data set from the email network of researchers interested in the OGM phenomenon. After adjusting for duplicates, 2210 citations remained. Of these, 2182 were discarded because it was clear that they did not meet the inclusion criteria after reviewing the abstracts. The remaining 28 citations were assessed for eligibility in more detail through manuscript review and application of the inclusion criteria. We excluded four studies because they used dependent variables that were incommensurate with the inclusion criteria (Hermans et al., 2007; Park, Goodyer, & Teasdale, 2005; Sidley, Calam, Wells, Hughes, & Whitaker, 1999; Spinhoven et al., 2006), two studies because they lacked two assessments of depressive symptoms (Raes et al., 2005; Sampson, Kinderman, Watts, & Sembi, 2003), two studies because of overlap with other study samples (Gibbs, 2004; Mackinger & Svaldi, 2004), and two studies because they provided insufficient information to compute effect sizes (Sutherland & Bryant, 2007; van Minnen et al., 2005). Finally, we excluded three studies that reported regression models with additional predictors besides an initial measure of depressive symptoms and a measure of OGM, and did not report any correlations between OGM and follow-up depressive symptoms (Bryant et al., 2007; Hermans et al., 2008; Svaldi & Mackinger, 2003). As a result, all of the studies in the analyses of standardized regression coefficients only included two predictors of follow-up depressive symptoms: 1) a baseline measure of depressive symptoms, and 2) a measure of OGM. We were able to include two studies (Mackinger et al., 2004; Mackinger, Loschin, & Leibetseder, 2000) in the meta-analysis even though they did not present the relevant regression models in the original publications because we obtained the data from Herbert Mackinger and conducted the analyses ourselves. In sum, a total of 15 studies were included in the meta-analysis (see Figure 1 for a flow diagram of study selection).

Variables Coded From Each Study

A standard coding sheet was completed for each study. We coded the following general information: 1) date of publication; 2) publication type (journal article, thesis or doctoral dissertation, unpublished report); 3) the mean age of participants; 4) the proportion of females in the sample; 5) the diagnostic status of participants [clinical with some psychiatric diagnosis, nonclinical, mixed sample (i.e., both individuals with and without a psychiatric diagnosis)]; 6) presence of a diagnosed clinical depressive disorder in participants (yes, no); and 7) total sample size (used in the original analyses).

We also coded characteristics of the AMT procedure including: 1) number of cue words; 2) type of cue words (positive and negative words, positive, negative, and neutral words, other); 3) method of cue word presentation (orally, visually, both orally and visually); 4) duration of the AMT response time limit (in seconds); 5) type(s) of indices of OGM used in the analyses (specific, categoric, overgeneral); and 6) statistical measure(s) of OGM used in the analyses (number of memories, proportion of memories, percentage of memories).

The following information about the measure of depressive symptoms was coded as well: 1) depressive symptom measure used [BDI, HRSD, MADRS, Edinburgh Postnatal Depression Scale (EPDS), Self-Rating Depression Scale (SDS), Structured Clinical Interview for DSM-IV (SCID) symptom severity scale, Hospital Anxiety and Depression Scale (HADS), Center for Epidemiologic Studies Depression Scale (CES-D)]; 2) rating method for depressive symptoms (self-rated, clinician-rated); and 3) average length of time between the baseline and follow-up assessments (in days). Some studies presented results for more than one measure of depressive symptoms, so we separately coded the presence versus absence of each of the different measures.

The first two authors (J. A. Sumner and J. W. Griffith) each coded all studies. In order to determine intercoder reliability, kappas were computed for categorical variables and intraclass correlation coefficients were computed for continuous variables. Kappa coefficients ranged from .56 to 1.00, with a mean of .92 and a median of 1.00. Intercoder reliability was lowest for rating the statistical measure of OGM used in the analyses. Upon further inspection of the articles, it was noted that there was some inconsistency in how the OGM measures were described in the articles (e.g., in some studies, both proportions and percentages were referred to in different sections of the articles, without specifying which variables were used in the analyses). Intraclass correlation coefficients ranged from .98 to 1.00, with a mean and median of 1.00. Discrepancies were resolved by discussion and re-examination of the studies.

Computation and Analysis of Effect Sizes

Two effect size indices were used in this meta-analysis: Fisher's Z_r -transform and the standardized regression (β) coefficient. As is common in meta-analysis, Fisher's Z_r -transform was used in calculations rather than the product-moment correlation because the latter is characterized by some undesirable statistical properties, such as a problematic standard error formulation (Lipsey & Wilson, 2001).² Because the standard error is used to compute weights when conducting meta-analytic computations, it is thus recommended that Fisher's Z_r -transform be used as the effect size metric (Rosenthal, 1994). Standard errors were computed, and each effect size estimate was weighted by the inverse of its variance in order to give more influence to studies with more reliable estimates. If information for calculating the standard error of the standardized regression coefficient was not reported, then we contacted authors to request relevant statistics.

Given the evidence for a one-factor conceptualization of OGM (Griffith et al., 2009), an overall OGM measure collapsing across cue valence was used in this meta-analysis. Some studies presented effect size estimates based on responses to all cue words in addition to separate valence-specific effect size estimates, some studies only presented separate valence-specific effect size estimates, and one study presented relevant effect size estimates only for responses to cue words of a single valence (e.g., responses to positive, but not negative, cue words). When overall OGM effect sizes were not reported, the separate valence-specific effect size estimates were averaged to compute an overall measure. Because Dalgleish et al. (2001) only reported

²Correlations can be easily converted to the Fisher's Z_r -transform using the formula $0.5\ln[(1+r)/(1-r)]$, where r is the product-moment correlation. The standard error of Fisher's Z_r -transform is equal to $1/\sqrt{N-3}$, where N is the total sample size.

effect size estimates for responses to positive cue words, the reported effect size estimate was averaged with an effect size of zero to calculate a conservative overall effect size estimate. We selected a value of zero as the effect size estimate for responses to negative cue words in an attempt to adjust for the bias of only reporting the results that were statistically significant for a single cue word valence.

Fisher's Z_r -transform estimates and standard errors were computed based on reports of product-moment correlations and sample size. The standard errors for the standardized regression coefficients were based on standard errors, t tests, R^2 change values, and p values. Overall, 71 effect sizes ($N = 918$; 37 Fisher's Z_r -transform estimates and 34 standardized regression coefficients) were calculated. Several studies yielded more than one relevant effect size for a particular OGM index. For example, two studies provided effect size estimates for more than one measure of depressive symptoms (Dalgleish et al., 2001; Peeters, Wessel, Merckelbach, & Boon-Vermeeren, 2002). However, using more than one effect size per sample violates the independence assumptions of meta-analysis (Lipsey & Wilson, 2001). Thus, we averaged the multiple effect size estimates to create a data set that included only one effect size per sample per OGM index. In addition, the studies by Brittlebank et al. (1993), Hipwell, Reynolds, and Crick (2004), and Peeters et al. (2002) presented effect size estimates for multiple follow-up time points. In these cases, we used the effect size estimate for the follow-up time point that was closest to the mean of the follow-up periods across the other studies in this review ($M = 5.3$ months).

Two studies presented separate effect size estimates for self- and clinician-rated measures of depressive symptoms (Dalgleish et al., 2001; Peeters et al., 2002). The mean effect size estimate for each sample was used in all analyses except for the moderator analysis comparing self- and clinician-rated measures. In this moderator analysis, the effect size estimates for each depressive symptom measure were both entered; this still allowed only one effect size estimate from each sample in each class of the moderator variable. Although the effect sizes in the classes of the moderator variable are not independent, it has been shown that analyses of differences in groups of effect sizes that might be dependent tend to yield conservative results (Hedges, 2007). In addition, the study by Kleim and Ehlers (2008) presented data for the overall sample and for the two subsamples of individuals with and without a history of MDD. The effect size estimate based on the overall sample was used for all analyses except for the moderator analysis examining clinical depression status of participants. For this moderator analysis, the two subsamples were treated as separate studies, and we used both effect size estimates.

Effect sizes were analyzed with Comprehensive Meta-Analysis, version 2.2.046 (Borenstein, Hedges, Higgins, & Rothstein, 2007). Four separate analyses were conducted: an analysis for correlations based on specific memories, an analysis for correlations based on categoric/overgeneral memories, an analysis for standardized regression coefficients based on specific memories, and an analysis for standardized regression coefficients based on categoric/overgeneral memories. We assumed a random effects model for these analyses given that none of the studies included in this meta-analysis was an exact replication of another. This model includes a random component in addition to sampling error to account for variation in effect size estimates (Lipsey & Wilson, 2001). We examined the level of between-study heterogeneity empirically with the Q statistic and I^2 index. A nonsignificant Q statistic suggests that studies only differ as a result of sampling error at the subject level, whereas a significant Q statistic suggests that factors in addition to sampling error are needed to account for the variation in effect size estimates (Lipsey & Wilson, 2001). Because the Q statistic is based on statistical significance, its power is dependent on the number of studies included in the meta-analysis. Unlike the Q statistic, the I^2 index is not based on statistical significance but rather represents the percentage of the total variability in a set of effect sizes that is due to true heterogeneity

between studies (Higgins & Thompson, 2002). Conventions are such that I^2 indices of 25%, 50%, and 75% reflect low, medium, and high heterogeneity, respectively.

To address concerns of publication bias, we used Duval and Tweedie's (2000) trim and fill method. We did not compare effect sizes as a function of publication status (published versus unpublished) because we were only able to obtain two unpublished data sets for this meta-analysis. In the trim and fill method, studies that might be missing from the sample of studies included in the research synthesis as a result of publication bias are estimated using a funnel plot. These graphs plot effect size estimates on the x-axis and the inverse of their standard error on the y-axis. The plots generally resemble a funnel, with the estimates with the smallest standard errors found at the top of the funnel, and the less precise estimates found at the base of the funnel. Typically, there is asymmetry in the plot as a function of publication bias, such that nonsignificant effect size estimates or those that might be the reverse of the hypothesized finding are missing; this asymmetry is generally present at the bottom of the funnel plot. The trim and fill method estimates these missing effect size estimates and includes them in the analysis to produce an effect size estimate that is corrected for publication bias. The larger the difference between the trim and fill effect size estimate and the original uncorrected effect size estimate, the more the effect size estimate based on only reported studies may be influenced by publication bias. Although other methods have been developed to detect publication bias [e.g., Begg and Mazumdar's (1994) rank correlation; Egger's regression intercept (Egger, Smith, Schneider, & Minder, 1997)], these are often characterized by low power. However, Duval and Tweedie's trim and fill method has been found to be an appropriate index for detecting and correcting for publication bias in meta-analyses with smaller sample sizes (L.Hedges, personal communication, June 3, 2008).

Results

Characteristics of the Studies

The studies in this review were published between 1993 and 2008 (median year of publication = 2004), and consisted mainly of published data (13 journal articles, 1 unpublished dissertation, and 1 unpublished report). Most studies used samples from a clinical population with a psychiatric diagnosis ($n = 9$) or from a nonclinical population (e.g., college students, pregnant women from the community; $n = 5$). One additional study included assault survivors with and without psychiatric diagnoses. Samples generally included more female than male participants (mean proportion of females = .69). The mean age of participants was 36.1 years ($range = 18.4$ to 54.0 years). Additional characteristics of the studies included in the correlational and standardized regression coefficient analyses are presented in Table 1 and Table 2, respectively.

Relationship between OGM and Depressive Symptoms at Follow-Up

The results for studies that reported product-moment correlations between OGM and a measure of depressive symptoms at follow-up are presented in Table 3. Although analyses were conducted on Fisher's Z_r -transform estimates, correlations are presented for ease of interpretation. All analyses were performed with an overall measure of OGM collapsing across cue word valence.

The weighted mean effect size averaged across the four studies reporting correlations between specific memories and depressive symptoms at follow-up was $-.10$, and the weighted mean effect size averaged across the eight studies reporting correlations between categoric/overgeneral memories and depressive symptoms at follow-up was $.13$. The fewer specific memories and the more categoric/overgeneral memories that individuals retrieved at baseline, the higher the level of depressive symptoms at follow-up. The effect size for categoric/overgeneral memories was statistically significant; however, the effect size for specific

memories did not reach conventional levels of statistical significance ($p = .11$). Both represent effect sizes of a small magnitude based on Cohen's (1988) conventions. These findings suggest that OGM accounts for approximately 1–2% of the variance in depressive symptoms at follow-up.

Even though the Q statistic for the analysis for specific memories was nonsignificant, the I^2 index of 31% suggests that there is a low to medium degree of heterogeneity in the effect sizes. In contrast, the nonsignificant Q statistic and I^2 index of 0% for the analysis for categoric/overgeneral memories suggests that the heterogeneity among these effect sizes from different studies is consistent with what would have been expected due to subject-level sampling error alone.

OGM as a Predictor of Depressive Symptoms Over and Above Initial Symptom Levels

The results for studies that reported the standardized regression coefficient for OGM in predicting the level of depressive symptoms at follow-up over and above the initial level of depressive symptoms are also presented in Table 3. Again, all analyses were performed with an overall measure of OGM collapsing across cue word valence.

The weighted mean effect size averaged across the 6 studies reporting a standardized regression coefficient for specific memories was $-.17$, and the weighted mean effect size averaged across the 6 studies reporting a standardized regression coefficient for categoric/overgeneral memories was $.11$. Both effect sizes differed significantly from zero. The weighted mean effect size for the studies reporting specific memories indicates that for every one standard deviation unit increase in the measure of specific memories, depressive symptoms at follow-up are lower by $.17$ standard deviation units. The weighted mean effect size for the studies reporting categoric/overgeneral memories indicates that for every one standard deviation unit increase in the measure of categoric/overgeneral memories, depressive symptoms at follow-up are higher by $.11$ standard deviation units. Thus, the predictive power of OGM appears to be small, albeit statistically significant. The value of the Q statistic for the analysis for specific memories was significant, and the I^2 index of 57% reflects a medium degree of heterogeneity among these effect sizes. As with the correlational analyses, the Q statistic for the analysis for categoric/overgeneral memories was nonsignificant and the I^2 index was 0%. These statistics again suggest that variability in effect sizes can be explained by sampling error.

Results of Moderator Analyses

Moderator analyses were only conducted if there were at least three effect size estimates for a class of a potential moderator variable. It would be mathematically possible to perform the analyses with a single effect size estimate in a class of a moderator, but we preferred to conduct moderator analyses that reflected a synthesis of research. We used an analogue to ANOVA to analyze categorical moderator variables and meta-regression to analyze continuous moderator variables (Lipsey & Wilson, 2001). For these analyses, we employed a mixed effects model that posits that variation in effect size parameters is due to the moderator variable, subject-level sampling error, and an additional random component. Results of the categorical moderator variable analyses are presented in Table 4. Given that there were only four studies in the correlational data set for specific memories, no categorical moderator variables were tested for this set of effect sizes.

Characteristics of the sample—We examined the clinical depression status of participants (participants with a clinical depression diagnosis versus nonclinical participants) as a moderator, as well as the age of participants. The *clinical depression status of participants* was found to be a significant moderator of the predictive relationship between specific memories and the course of depression for standardized regression coefficients [$Q_B(1) = 8.60, p = .003$,

see Table 4]. Even though fewer specific memories predicted higher levels of depressive symptoms at follow-up for both clinical and nonclinical samples (although only at a trend level for nonclinical samples), there was a significantly stronger relationship between this measure of OGM and follow-up depressive symptoms for studies with participants with clinical depression diagnoses ($\beta = -.37, p < .0001$) than for studies with nonclinical participants ($\beta = -.10, p = .10$). However, the clinical status of participants was not found to be a significant moderator of the correlational relationship between categoric/overgeneral memories and the course of depression [$Q_B(1) = .21, p = .64$].

Using meta-regression, *age of participants* did not significantly predict mean effect size for either of the two categoric/overgeneral variables. Specifically, neither the regression model for correlations for categoric/overgeneral memories [$Q_R(1) = .18, p = .67$], nor the regression model for standardized regression coefficients for categoric/overgeneral memories [$Q_R(1) = .12, p = .72$] was significant. However, the regression model for correlations for specific memories approached the conventional level of significance [$Q_R(1) = 3.60, p = .06$]. As the mean age of participants became older, the correlation between specific memories and depressive symptoms at follow-up became larger and more negative ($b = -.01, p = .06$). Furthermore, the regression model for standardized regression coefficients for specific memories was significant [$Q_R(1) = 7.24, p = .01$]. Like with the regression for correlations for specific memories, as the mean age of participants became older, the standardized regression coefficient for specific memories became larger and more negative ($b = -.01, p = .01$).

Characteristics of the study design—This class of potential moderators included the valence of the AMT cue word (positive versus negative), the rating method for depressive symptoms (self- versus clinician-rated), and the length of follow-up between the two assessments of depressive symptoms (in days). For the correlation effect size measure for categoric/overgeneral memories, *cue word valence* did not emerge as a significant moderator [$Q_B(1) = 0, p = .99$]. Valence was also not a significant moderator of the predictive relationship between OGM and the course of depression for the two standardized regression coefficient effect size measures [$Q_B(1) = .13, p = .72$, for standardized regression coefficients for specific memories; $Q_B(1) = .45, p = .50$, for standardized regression coefficients for categoric/overgeneral memories].

The *rating method for depressive symptoms* approached the conventional level of statistical significance as a moderator of the predictive relationship between specific memories and the course of depression for standardized regression coefficients [$Q_B(1) = 2.96, p = .09$]. There was a trend for a larger predictive relationship for clinician-rated measures ($\beta = -.28, p = .01$) than for self-rated measures ($\beta = -.06, p = .39$). However, the rating method for depressive symptoms was not found to be a significant moderator for the standardized regression coefficients for categoric/overgeneral memories [$Q_B(1) = .01, p = .92$] or for the correlations for categoric/overgeneral memories [$Q_B(1) = 1.74, p = .19$].

Using meta-regression, the *length of follow-up* (measured in days) was a significant predictor of the mean effect size for standardized regression coefficients for specific memories [$Q_R(1) = 9.56, p = .002$]. As the length of follow-up increased, the relationship between specific memories and follow-up depressive symptoms became smaller and less negative ($b = .0009, p = .002$). However, none of the other regression models was significant [$Q_R(1) = 2.54, p = .11$, for correlations for specific memories; $Q_R(1) = 2.40, p = .12$, for correlations for categoric/overgeneral memories; $Q_R(1) = .44, p = .51$, for standardized regression coefficients for categoric/overgeneral memories].

Consideration of Publication Bias

To address concerns of publication bias, we used Duval and Tweedie's (2000) trim and fill method. No studies were trimmed for the standardized regression coefficients for categorical/overgeneral memories or for correlations for specific memories (both effect size estimates were identical). However, two studies were trimmed for the standardized regression coefficients for specific memories, and three studies were trimmed for the correlations for categoric/overgeneral memories. For both of these OGM measures, the trim and fill effect size estimates and the effect sizes calculated for this meta-analysis differed by .08. These results suggest that publication bias is minimal for the standardized regression coefficient analyses for categorical/overgeneral memories and the correlational analyses for specific memories, although there might be slight concern for the correlational analyses for categoric/overgeneral memories and standardized regression coefficient analyses for specific memories.

Discussion

Overview of the Results

Although the OGM phenomenon in depression is clearly well-established (van Vreeswijk & de Wilde, 2004; Williams et al., 2007), inconsistencies in the literature have left the predictive relationship between OGM and change in depressive symptoms over time as a matter of debate. Thus, the primary goal of this meta-analysis was to investigate the degree to which OGM predicts the course of depression in order to better elucidate the reliability and magnitude of this predictive relationship.

We examined the relationship between OGM and the course of depression in two ways: the degree to which OGM 1) is correlated with a measure of depressive symptoms at follow-up, and 2) predicts depressive symptoms at follow-up over and above initial levels of depressive symptoms. The results of these analyses revealed an overall small relationship between OGM and the course of depression ($r = -.10$ and $\beta = -.17$ for the analyses for specific memories; $r = .13$ and $\beta = .11$ for the analyses for categoric/overgeneral memories). In particular, the higher the levels of OGM (i.e., fewer specific memories and more categoric/overgeneral memories) at baseline, the higher the level of depressive symptoms at follow-up. Furthermore, the predictive relationship between OGM and depressive symptoms at follow-up is not accounted for simply by the level of depressive symptoms at the initial assessment. The mean effect sizes were significant for all analyses except for the analysis of correlations for specific memories. However, this analysis was based on the smallest number of studies (four), and the finding was nevertheless in the expected direction.

Interpretation of Moderator Analyses

A secondary goal of this meta-analysis was to examine whether characteristics of the sample and the study design moderated the predictive relationship between OGM and the course of depression. Although the proposed moderator variables are theoretically relevant, these analyses represent preliminary exploratory investigations given the lack of research on moderators of this relationship. Overall, the majority of the variables was not found to significantly moderate the relationship between OGM and the course of depression with several exceptions. For example, there were three significant moderator variables for standardized regression coefficients for specific memories: the clinical depression status of participants, the age of participants, and the length of follow-up between the two assessments of depressive symptoms. In addition, the test of age of participants approached conventional levels of statistical significance for correlations for specific memories. There was also a trend for the test of rating method for depressive symptoms to approach statistical significance for standardized regression coefficients for specific memories.

The clinical depression status moderator analyses suggested that the magnitude of the standardized regression coefficient for the predictive relationship between specific memories and depressive symptoms at follow-up was significantly larger for samples with clinical diagnoses of depression ($\beta = -.37$) than for nonclinical samples ($\beta = -.10$). There are several possible interpretations of this finding. One possibility is that this is a statistical artifact, owing to greater variability of depressive symptoms at follow-up in clinical samples, compared to nonclinical samples. Alternatively, this finding might also result from the AMT being an insufficiently sensitive measure of OGM in nonclinical samples, such as depressed college students who are relatively high-functioning, as argued by Raes et al. (2007). Yet another possibility is that there is a nonlinear effect of OGM on the course of depression, such that the predictive effect of OGM is greater among individuals with higher levels of depression. OGM may create a kind of vicious cycle in depression, such that as individuals become more depressed, they may be less able to problem-solve the negative consequences of their depressive symptoms, thereby leading to longer bouts of depression (Teasdale, 1983). This speculation could be tested in future research by longitudinally examining OGM, depression, and problem-solving ability. Furthermore, it should be noted that in this moderator analysis, a clinical depression diagnosis refers to current depression status at baseline. The one exception is the study by Kleim and Ehlers (2008), which presents information on whether or not participants had a history of MDD. We focused primarily on current depression because of the small number of published prospective studies on OGM as a predictor of the recurrence of depression (although see Spinhoven et al., 2006 and Sumner et al., in press). However, given that there is some (albeit inconsistent) evidence of OGM in individuals in remission from depression (e.g., Mackinger, Pachinger et al., 2000), it would be of interest to examine the extent to which OGM predicts recurrence as more studies accumulate.

Age also emerged as a significant moderator of the predictive relationship between OGM and the course of depression. The results of the meta-regression examining the relationship between age and effect size suggested that as age increases, the inverse relationship between specific memories and depressive symptoms at follow-up is larger (this was significant for standardized regression coefficients and approached conventional significance levels for correlations). The greater predictive relationship between OGM and depressive symptoms at follow-up for older individuals might be the result of a more pronounced OGM phenomenon at older ages, as reported by Ros et al. (2009). For example, younger individuals might exhibit a ceiling effect on the AMT such that most are able to come up with specific memories. Thus, the relationship between specific memories and depressive symptoms at follow-up might be attenuated as a result. It should be noted that the mean ages of participants in the studies in these analyses ($M = 33.5$ years) were younger than the mean age of participants in the study by Ros et al. (2009; 66 years), but there was some overlap in the age ranges. Ros et al. had participants aged 57–80 years and our analyses had some samples with age ranges that extended into the late 50s and mid 60s.

In addition, we examined the degree to which the significant clinical depression status moderator finding for specific memories might be due to an age difference between the clinical and nonclinical samples by examining whether samples with clinical diagnoses of depression were significantly older than nonclinical samples. Even though it was not significant, there was a trend for the mean age for the studies with samples with clinical diagnoses of depression ($M = 39.3$ years) to be greater than the mean for the studies with nonclinical samples ($M = 27.6$ years), $t(4) = -2.25, p = .09$. As more studies accumulate, it would be of interest to re-examine this finding in order to better understand how clinical status and age are associated with the predictive relationship between OGM and the course of depression.

The length of follow-up between the two assessments of depressive symptoms was also found to predict the mean effect size for standardized regression coefficients for specific memories.

As the length of follow-up increased, the relationship between specific memories and follow-up depressive symptoms became smaller and less negative. Thus, the relationship between these two variables seems to be greater the closer in time that they are both assessed. This might reflect the fact that as the follow-up interval increases, there is more time for other depressogenic factors to accumulate. However, as with all of the moderator results, this finding warrants replication.

The rating method for depressive symptoms approached conventional levels of statistical significance as a moderator of the predictive relationship between specific memories and the course of depression for standardized regression coefficients. There was a trend for a larger predictive relationship for clinician-rated measures ($\beta = -.28$) than for self-rated measures ($\beta = -.06$). This pattern of results is consistent with the findings of Dalgleish et al. (2001), in which OGM predicted change in depressive symptoms as measured with the HRSD, but not with the BDI. Dalgleish et al. (2001) noted that the BDI and HRSD also differ in content, with the BDI emphasizing cognitive symptoms of depression and the HRSD focusing on somatic-vegetative symptoms of depression. Unfortunately, information on the degree to which the various depressive symptom measures used in the studies included in this meta-analysis assess cognitive versus somatic-vegetative symptoms of depression was not readily available. Additional research using multi-method approaches to measuring depression, as well as measures that distinguish between the cognitive and somatic-vegetative symptoms of depression, is needed to better understand this issue. Furthermore, meta-analytic reviews suggest that the HRSD may be more sensitive to change in depressive symptoms than the BDI, such that the magnitude of clinical improvement for the HRSD is often greater than that for the BDI (e.g., Edwards et al., 1984). Even though the HRSD and BDI are each only one of a few clinician- and self-rated measures, respectively, that were included in this meta-analysis, this potential explanation for the observed moderator trend should be examined in the future as well.

One limitation of the moderator analysis of the rating method for depressive symptoms is that clinical depression status and rating method were confounded. In the set of studies reporting standardized regression coefficients for specific memories, clinician-rated measures were used more often with participants with clinical diagnoses of depression than with nonclinical participants, and only clinician-rated measures were used for participants with clinical depression diagnoses. Thus, a potential alternative explanation for the stronger predictive relationship between OGM and the course of depression for clinician-rated measures than for self-rated measures is that variation in the clinical depression status of participants accounted for these differences. This potential explanation should be examined in future studies.

Moderating effects emerged for only the analyses of effect sizes for specific, and not categoric/overgeneral, memories. Discrepant findings for these different operationalizations of AMT performance have been observed in prior research (e.g., Raes et al., 2006; Stokes, Dritschel, & Bekerian, 2004), and it is possible that such results are due to a restriction of range in the number of categoric memories retrieved. In general, there tends to be a greater range in the number of specific memories generated than in the number of categoric memories generated [e.g., range for specific = 1–10 versus 0–4 for categoric for a nonclinical sample (Hipwell et al., 2004); range for specific = 3–18 versus 0–7 for categoric for a clinical sample (Hermans et al., 2008)]. In addition, discrepant findings for the moderator analyses for correlations and standardized regression coefficients likely reflect the fact that the standardized regression coefficient (but not correlation) effect sizes are influenced by the initial symptom level covariate in the regression models.

Limitations

To our knowledge, this meta-analysis is the first to provide a systematic review of the research examining the predictive relationship between OGM and the course of depression. However, a number of limitations need to be acknowledged. First, we mostly reviewed published studies. Although we attempted to obtain unpublished data, and the results from the trim and fill method suggested that publication bias did not appear to be a major concern for the majority of the analyses, the small number of unpublished studies included in the analysis is still a limitation.

Second, there are limitations with using the standardized regression coefficient as a measure of effect size. The interpretation of weighted mean effect sizes for the standardized regression coefficient can be complicated if none of the studies are exact replications of one another. Nevertheless, because all of the regression models across studies included measures of the same constructs as predictors (i.e., OGM and initial depressive symptoms), they can thus estimate the degree to which OGM predicts depressive symptoms at follow-up over and above initial symptom levels. Additionally, other research syntheses using standardized regression coefficients as a measure of effect size have computed the median effect size to analyze effect magnitude (e.g., Greenwald, Hedges, & Laine, 1994). When we employed this approach, the results were similar to those obtained with the formal meta-analytic procedures (median $\beta = -.18$ for specific memories and $.13$ for categoric/overgeneral memories). Thus, despite the limitations associated with the standardized regression coefficient effect size measure, the analyses reported in this review permit a more sophisticated and informative examination of the extant literature than what would be obtained with a qualitative review or count of statistically significant results (Lipsey & Wilson, 2001).

Conclusions and Future Directions

The results of this systematic quantitative review suggest that there is a small but reliable relationship between OGM and the course of depression. In terms of the clinical relevance of this relationship, the effects are small. This might lead some to question whether interventions aimed at increasing autobiographical memory specificity are warranted. However, given that only approximately half of all patients respond to any one intervention for depression and that only approximately one third of patients go on to meet criteria for full remission (Hollon, Thase, & Markowitz, 2002), developing and testing techniques for increasing autobiographical memory specificity may still be useful, perhaps in conjunction with existing cognitive therapy techniques. One such intervention that appears promising is MEMory Specificity Training (MEST; Raes, Williams, & Hermans, 2009), a four-week group-based intervention aimed at increasing the specificity of memory retrieval. Preliminary findings suggest that MEST is successful at increasing autobiographical memory specificity, and increases in memory specificity were associated with decreases in rumination and with increases in problem-solving effectiveness. Even though MEST was designed as a stand-alone intervention, future research might examine incorporating its techniques into existing treatment protocols.

Future studies should also examine the mechanisms that may account for the predictive relationship between OGM and the course of depression. Williams et al. (2007) recently proposed a theoretical model of three mechanisms that may underlie OGM: capture and rumination, functional avoidance, and impaired executive control. Capture and rumination refers to when ruminative processes are activated during retrieval and disrupt the retrieval process, functional avoidance refers to when the retrieval of specific memories is avoided as a means of affect regulation, and impaired executive control refers to when deficits in executive resources limit the ability to conduct a successful retrieval search. As reviewed by Williams et al. (2007), research suggests that each of these mechanisms may contribute to OGM to some extent. Furthermore, there is some evidence that these mechanisms may have effects on aspects of psychological functioning, such as the ability to adequately resolve one's problems. For

example, Raes et al. (2005) found that rumination was negatively correlated with problem-solving effectiveness, and OGM was found to mediate this relationship. It would be of interest to investigate whether the mechanisms proposed by Williams et al. (2007), or their associated consequences (e.g., impaired problem-solving), may explain the relationship between OGM and change in depressive symptoms over time. For example, as a result of deficits in executive control, depressed individuals might be less specific in their personal memory recollection and thus be less likely to use specific details from past experiences to generate successful solutions to their problems. The stress associated with the continuation of such problems could then potentially contribute to a longer episode of depression. Longitudinal research designs with measures of these proposed mechanisms and OGM, in addition to multiple assessments of depression, will be critical for understanding the factors that contribute to the predictive relationship between OGM and the course of depression.

As more studies accumulate, the tests for moderators examined in this meta-analysis should also be reconsidered. In this meta-analysis, we required there to be at least three studies in a category of a moderator variable in order to conduct a moderator analysis. Even though an analysis of three studies reflects a synthesis of extant research, they are small sample sizes and are more susceptible to influence by extreme values. The moderator analyses in this review thus represent a preliminary investigation, and they warrant replication. Furthermore, with a larger sample of studies, the power to detect such effects will likely be greater. If additional significant moderators were to emerge in a larger sample of studies, then this might inform the design of future research examining the predictive relationship between OGM and the course of depression. Future work might also investigate additional moderators, such as the educational level of participants. There was insufficient information to examine whether education might moderate the predictive relationship between OGM and the course of depression in this meta-analysis; only 7 of the 15 studies reported this information, and it was often reported inconsistently across studies. Consistent reporting of study information, such as participant and AMT characteristics, across investigations will facilitate future meta-analyses.

Despite some inconsistencies among individual studies investigating the predictive relationship between OGM and the course of depression, the results of this meta-analysis suggest that such a relationship does indeed exist. We hope that this review will encourage researchers to continue to study OGM as a predictor of the course of depression, and that it will help to guide such future studies.

Acknowledgments

We would like to thank Larry Hedges for his help and guidance with this meta-analysis. We also extend thanks to the following authors who sent us requested data and statistical information: Anke Ehlers, Elise Debeer, Alison Hipwell, Birgit Kleim, Ismay Kremers, Max Leibetseder, Herbert Mackinger, Stephanie Rude, and Philip Spinhoven. We would also like to thank the National Institutes of Health for supporting our research (Grant# R01MH065652 to Dr. Mineka).

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088–1101. [PubMed: 7786990]
- Borenstein, M.; Hedges, L.; Higgins, J.; Rothstein, H. *Comprehensive Meta Analysis* (Version 2.2.046) [Computer software]. Englewood, NJ: Biostat; 2007.
- Brewin CR, Reynolds M, Tata P. Autobiographical memory processes and the course of depression. *Journal of Abnormal Psychology* 1999;108:511–517. [PubMed: 10466275]

- Brewin CR, Watson M, McCarthy S, Hyman P, Dayson D. Memory processes and the course of anxiety and depression in cancer patients. *Psychological Medicine* 1998;28:219–224. [PubMed: 9483699]
- Brittlebank AD, Scott J, Williams JMG, Ferrier IN. Autobiographical memory in depression: State or trait marker? *British Journal of Psychiatry* 1993;162:118–121. [PubMed: 8425125]
- Bryant RA, Sutherland K, Guthrie RM. Impaired specific autobiographical memory as a risk factor for posttraumatic stress after trauma. *Journal of Abnormal Psychology* 2007;116:837–841. [PubMed: 18020729]
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2nd ed.. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- Conway MA, Pleydell-Pearce CW. The construction of autobiographical memories in the self-memory system. *Psychological Review* 2000;107:261–288. [PubMed: 10789197]
- Dalgleish T, Spinks H, Yiend J, Kuyken W. Autobiographical memory style in seasonal affective disorder and its relationship to future symptom remission. *Journal of Abnormal Psychology* 2001;110:335–340. [PubMed: 11358027]
- Dalgleish T, Williams JMG, Golden AJ, Perkins N, Barrett LF, Barnard PJ, et al. Reduced specificity of autobiographical memory and depression: The role of executive control. *Journal of Experimental Psychology: General* 2007;136:23–42. [PubMed: 17324083]
- Debeer E, Raes F, Hermans D. [Overgeneral autobiographical memory as a predictor of depressive symptoms and clinical depression in a college sample]. 2008 Unpublished raw data.
- Duval SJ, Tweedie RL. A non-parametric “trim and fill” method of assessing publication bias in meta-analysis. *Journal of the American Statistical Association* 2000;95:89–98.
- Edwards BC, Lambert MJ, Moran PW, McCully T, Smith KC, Ellingson AG. A meta-analytic comparison of the Beck Depression Inventory and the Hamilton Rating Scale for Depression as measures of treatment outcome. *British Journal of Clinical Psychology* 1984;23:93–99. [PubMed: 6722384]
- Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997;315:629–634. [PubMed: 9310563]
- Gibbs, BR. *Dissertation Abstracts International*. Vol. 64. 2004. Overgeneral cognitive style: The impact of physical and emotional adjustment to life stress; p. 4075 (Doctoral dissertation, University of Texas at Austin, 2002)
- Gibbs BR, Rude SS. Overgeneral autobiographical memory as depression vulnerability. *Cognitive Therapy and Research* 2004;28:511–526.
- Goddard L, Dritschel B, Burton A. Social problem solving and autobiographical memory in non-clinical depression. *British Journal of Clinical Psychology* 1997;36:449–451. [PubMed: 9309360]
- Greenwald R, Hedges LV, Laine RD. When reinventing the wheel is not necessary: A case study in the use of meta-analysis in education finance. *Journal of Education Finance* 1994;20:1–20.
- Griffith JW, Sumner JA, Debeer E, Raes F, Hermans D, Mineka S, et al. An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory* 2009;17:609–623. [PubMed: 19585345]
- Hedges, LV. Meta-analysis. In: Rao, CR.; Sinharay, S., editors. *The handbook of statistics*. Vol. 26. Amsterdam: Elsevier; 2007. p. 919-953.
- Hermans D, de Decker A, de Peuter S, Raes F, Eelen P, Williams JMG. Autobiographical memory specificity and affect regulation: Coping with a negative life event. *Depression and Anxiety* 2007;25:787–792. [PubMed: 17393378]
- Hermans D, Vandromme H, Debeer E, Raes F, Demyttenaere K, Brunfaut E, et al. Overgeneral autobiographical memory predicts diagnostic status in depression. *Behaviour Research and Therapy* 2008;46:668–677. [PubMed: 18342835]
- Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 2002;21:1539–1558. [PubMed: 12111919]
- Hipwell AE, Reynolds S, Crick EP. Cognitive vulnerability to postnatal depressive symptomatology. *Journal of Reproductive and Infant Psychology* 2004;22:211–227.
- Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychological Science in the Public Interest* 2002;3:39–77.

- Johnson, RJ. Dissertation Abstracts International. Vol. 66. 2006. A short-term longitudinal study of the relations among depression, stressful life events, and autobiographical memory; p. 6276 (Doctoral dissertation, University of Kansas, 2005)
- Kleim B, Ehlers A. Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. *Journal of Consulting and Clinical Psychology* 2008;76:231–242. [PubMed: 18377120]
- Kremers IP, Spinhoven P, Van der Does AJW, Van Dyck R. Autobiographical memory in depressed and nondepressed patients with borderline personality disorder after long-term psychotherapy. *Cognition and Emotion* 2006;20:448–465.
- Lipsey, MW.; Wilson, DB. Practical meta-analysis. Thousand Oaks, CA: Sage Publications; 2001.
- Mackinger HF, Leibetseder MF, Kunz-Dorfer AA, Fartacek RR, Whitworth AB, Feldinger FF. Autobiographical memory predicts the course of depression during detoxification therapy in alcohol dependent men. *Journal of Affective Disorders* 2004;78:61–65. [PubMed: 14672798]
- Mackinger HF, Loschin GG, Leibetseder MM. Prediction of postnatal affective changes by autobiographical memories. *European Psychologist* 2000;5:52–61.
- Mackinger HF, Pachinger MM, Leibetseder MM, Fartacek RR. Autobiographical memories in women remitted from major depression. *Journal of Abnormal Psychology* 2000;109:331–334. [PubMed: 10895571]
- Mackinger HF, Svaldi JJ. Autobiographical memory predicts cognitive but not somatic change in sleep apnea patients vulnerable for affective disorder. *Journal of Affective Disorders* 2004;81:17–22. [PubMed: 15183595]
- Moore SA, Zoellner LA. Overgeneral autobiographical memory and traumatic events: An evaluative review. *Psychological Bulletin* 2007;133:419–437. [PubMed: 17469985]
- Park RJ, Goodyer IM, Teasdale JD. Categorical overgeneral autobiographical memory in adolescents with major depressive disorder. *Psychological Medicine* 2002;32:267–276. [PubMed: 11866322]
- Park RJ, Goodyer IM, Teasdale JD. Self-devaluative dysphoric experience and the prediction of persistent first-episode major depressive disorder in adolescents. *Psychological Medicine* 2005;35:539–548. [PubMed: 15856724]
- Peeters F, Wessel I, Merckelbach H, Boon-Vermeeren M. Autobiographical memory specificity and the course of major depressive disorder. *Comprehensive Psychiatry* 2002;43:344–350. [PubMed: 12216009]
- Raes F, Hermans D, Williams JMG, Beyers W, Brunfaut E, Eelen P. Reduced autobiographical memory specificity and rumination in predicting the course of depression. *Journal of Abnormal Psychology* 2006;115:699–704. [PubMed: 17100527]
- Raes F, Hermans D, Williams JMG, Demyttenaere K, Sabbe B, Pieters G, et al. Reduced specificity of autobiographical memory: A mediator between rumination and ineffective social problem-solving in major depression? *Journal of Affective Disorders* 2005;87:331–335. [PubMed: 15979154]
- Raes F, Hermans D, Williams JMG, Eelen P. A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. *Memory* 2007;15:495–507. [PubMed: 17613793]
- Raes F, Pousset G, Hermans D. Correlates of autobiographical memory specificity in a non-clinical student population. 2004 Unpublished manuscript.
- Raes F, Sienaert P, Demyttenaere K, Peuskens J, Williams JMG, Hermans D. Overgeneral memory predicts the stability of short-term outcome of electroconvulsive therapy for depression. *Journal of ECT* 2008;24:81–83. [PubMed: 18379339]
- Raes F, Williams JMG, Hermans D. Reducing cognitive vulnerability to depression: A preliminary investigation of Memory Specificity Training (MEST) in inpatients with depressive symptomatology. *Journal of Behavior Therapy and Experimental Psychiatry* 2009;40:24–38. [PubMed: 18407245]
- Roberts JE, Carlos EL, Kashdan TB. Impact of depressive symptoms, self-esteem and neuroticism on trajectories of overgeneral autobiographical memory over repeated trials. *Cognition and Emotion* 2006;20:383–401.
- Ros L, Latorre JM, Serrano JP. Working memory capacity and overgeneral autobiographical memory in young and older adults. *Aging, Neuropsychology, and Cognition* 2010;17:89–107.

- Rosenthal, R. Parametric measures of effect size. In: Cooper, H.; Hedges, LV., editors. *The handbook of research synthesis*. New York: Russell Sage Foundation; 1994. p. 231-244.
- Salthouse TA, Atkinson TM, Berish DE. Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General* 2003;132:566–594. [PubMed: 14640849]
- Sampson MJ, Kinderman P, Watts S, Sembi S. Psychopathology and autobiographical memory in stroke and non-stroke hospitalized patients. *International Journal of Geriatric Psychiatry* 2003;18:23–32. [PubMed: 12497553]
- Sidley GL, Calam R, Wells A, Hughes T, Whitaker K. The prediction of parasuicide repetition in a high-risk group. *British Journal of Clinical Psychology* 1999;38:375–386. [PubMed: 10590825]
- Spinhoven P, Bockting CLH, Schene A, Koeter MWJ, Wekking EM, Williams JMG, et al. Autobiographical memory in the euthymic phase of recurrent depression. *Journal of Abnormal Psychology* 2006;115:590–600. [PubMed: 16866599]
- Stokes DJ, Dritschel BH, Bekerian DA. The effect of burn injury on adolescents' autobiographical memory. *Behaviour Research and Therapy* 2004;42:1357–1366. [PubMed: 15381443]
- Sumner JA, Griffith JW, Mineka S, Rekart KN, Zinbarg RE, Craske MG. Overgeneral autobiographical memory and chronic interpersonal stress as predictors of the course of depression in adolescents. *Cognition and Emotion*. (in press).
- Sutherland K, Bryant RA. Autobiographical memory in posttraumatic stress disorder before and after treatment. *Behavior Research and Therapy* 2007;45:2915–2923.
- Svaldi JJ, Mackinger HF. Obstructive sleep apnea syndrome: Autobiographical memory predicts the course of depressive affect after nCPAP therapy. *Scandinavian Journal of Psychology* 2003;44:31–37. [PubMed: 12603001]
- Teasdale JD. Negative thinking in depression: Cause, effect, or reciprocal relationship? *Advances in Behaviour Research and Therapy* 1983;5:3–25.
- van Minnen A, Wessel I, Verhaak C, Smeenk J. The relationship between autobiographical memory specificity and depressed mood following a stressful life event: A prospective study. *British Journal of Clinical Psychology* 2005;44:405–415. [PubMed: 16238885]
- van Vreeswijk MF, de Wilde EJ. Autobiographical memory specificity, psychopathology, depressed mood, and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy* 2004;42:731–743. [PubMed: 15081887]
- Wessel I, Meeren M, Peeters F, Arntz A, Merckelbach H. Correlates of autobiographical memory specificity: The role of depression, anxiety and childhood trauma. *Behavior Research and Therapy* 2001;39:409–421.
- Williams JMG, Barnhofer T, Crane C, Hermans D, Raes F, Watkins E, et al. Autobiographical memory specificity and emotional disorder. *Psychological Bulletin* 2007;133:122–148. [PubMed: 17201573]
- Williams JMG, Broadbent K. Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology* 1986;95:144–149. [PubMed: 3711438]

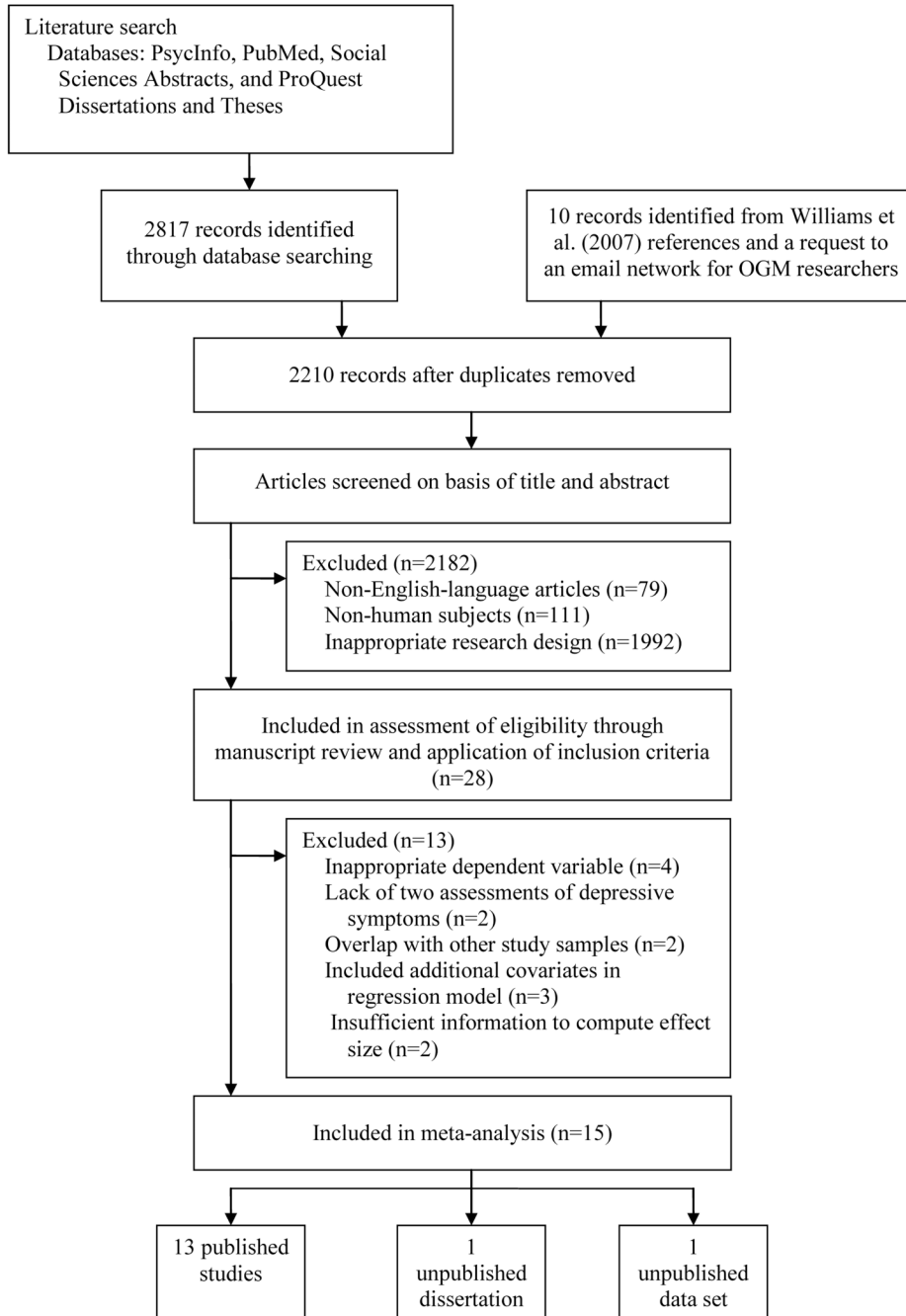


Figure 1. Flow diagram depicting the stages of study selection for inclusion in the meta-analysis.

Table 1

Characteristics of Studies Included in the Correlational Analyses

Author(s)	Year of Publication	Sample	Depression Measure	Length of Follow-Up	Memories Reported	Study's Conclusion	Product-Moment Correlation	Fisher's Z_r Transform (SE)
Brewin, Reynolds, & Tata	1999	Patients with MDD (N = 40)	BDI	6 months	Overgeneral	NS	.02	.02 (.16)
Brewin, Watson, McCarthy, Hyman, & Dayson	1998	Depressed cancer patients (N = 36)	HADS	6 months	Overgeneral	NS	.12	.12 (.17)
Brittlebank, Scott, Ferrier, & Williams	1993	Patients with MDD (N = 19)	HRSD	7 months	Overgeneral	S	.39	.42 (.25)
Dalgelish, Spinks, Yiend, & Kuyken	2004	Patients with SAD (N = 21)	BDI <i>HRSD</i>	6-8 months	Overgeneral	NS	.20 ^a	.21 (.24) ^a
Debeer, Raes, & Hermans	N/A ^b	College students (N = 112)	BDI	13 months	Overgeneral	S	.26	.26 (.24) ^a
Gibbs & Rude	2004	College students (N = 81)	BDI	4-6 weeks	Categorical Specific	S NS	.26 .05	.27 (.10) .05 (.10)
Hipwell, Reynolds, & Crick	2004	Primiparous women (N = 94)	EPDS	8.4 weeks	Categorical	NS	.02	.02 (.11)
Johnson	2006 ^b	Female college students (N = 88)	CES-D	6 months	Specific	NS	-.12	-.12 (.10)
Kleim & Ehlers	2008	Assault survivors (N = 187)	<i>SCID sx severity scale</i>	6 months	Overgeneral	S	.02	.02 (.11)
Peeters, Wessel, Merckelbach, & Boon-Vermeeren	2002	Patients with MDD (N = 23)	<i>MADRS</i> SDS	7 months	Specific	NS	-.20	-.20 (.07)
Raes, Sienaert, Demyttenaere, Peuskens, Williams, & Hermans	2008	Patients with MDD (N = 25)	<i>HRSD</i>	1 week	Specific Overgeneral	NS NS	-.08 -.11	-.09 (.22) -.11 (.22)
						NS	.24	.24 (.21)

Note. Sample: SAD = seasonal affective disorder. Sample sizes are those included in the analysis. Depression Measure: BDI = Beck Depression Inventory; HADS = Hospital Anxiety and Depression Scale; HRSD = Hamilton Rating Scale for Depression; EPDS = Edinburgh Postnatal Depression Scale; CES-D = Center for Epidemiologic Studies Depression Scale; SCID sx severity scale = Structured Clinical Interview for DSM-IV symptom severity scale; MADRS = Montgomery-Asberg Depression Rating Scale; SDS = Self-Rating Depression Scale. Italicized depression measures are clinician-rated, whereas non-italicized measures are self-rated. Study's Conclusion: NS = nonsignificant; S = significant. Fisher's Z_r -Transform (SE): Effect sizes are for the overall OGM measure (when separate results were presented for different cue valences, the Z_r -transform values for these statistics were averaged to obtain an overall estimate); SE = standard error.

^aConservative estimate based on reported effect size averaged with value of zero for responses to cue words of unreported valence.

^bUnpublished data.

Table 2
 Characteristics of Studies Included in the Standardized Regression (β) Coefficient Analyses

Author(s)	Year of Publication	Sample	Depression Measure	Length of Follow-Up	Memories Reported	Study's Conclusion	β Effect Size (SE)
Dalglish, Spinks, Yiend, & Kuyken	2004	Patients with SAD (N = 21)	BDI	6-8 months	Overgeneral	NS	.20 (.22) ^a
Debeer, Raes, & Hermans	N/A ^b	College students (N = 112)	BDI	13 months	Categoric	S	.21 (.18) ^a
Kleim & Ehlers	2008	Assault survivors (N = 187)	SCID xx severity scale	6 months	Specific	NS	.14 (.07)
Kremers, Spinhoven, Van der Does, & Van Dyck	2006	Patients with Borderline PD (N = 55)	BDI	15 months	Categoric	NS	-.03 (.09)
Mackinger, Leibetseder, Kunz-Dorfer, Fartacek, Whitworth, & Feldinger	2004	Alcohol dependent men with DD NOS (N = 65)	MADRS	25.9 days	Specific	S	-.14 (.07)
Mackinger, Loschin, & Leibetseder	2000	Primiparous women (N = 50)	EPDS	5 months	Categoric	NS	-.07 (.12)
Raes, Hermans, Williams, Beyers, Brunfaut, & Eelen	2006	Patients with MDD (N = 22)	HRSD	7 months	Overgeneral	NS	.02 (.12)
					Specific	S	-.42 (.10) ^c
					Categoric	NS	.12 (.12) ^c
					Categoric	NS	.20 (.14) ^c
					Specific	S	-.22 (.14) ^c
					Overgeneral	NS	-.10 (.30)
					Specific	S	.38 (.28)

Note. Sample: SAD = seasonal affective disorder. Sample sizes are those included in the analysis. Depression Measure: BDI = Beck Depression Inventory; HRSD = Hamilton Rating Scale for Depression; EPDS = Edinburgh Postnatal Depression Scale; SCID xx severity scale = Structured Clinical Interview for DSM-IV symptom severity scale; MADRS = Montgomery-Asberg Depression Rating Scale. Italicized depression measures are clinician-rated, whereas non-italicized measures are self-rated. Study's Conclusion: NS = nonsignificant; S = significant. β Effect Size (SE): Effect sizes are for the overall OGM measure; SE = standard error.

^a Conservative estimate based on reported effect size averaged with value of zero for responses to cue words of unreported valence.

^b Unpublished data.

^c Not reported in published article but calculated based on the data provided by the first author.

Table 3

Summary of Overall Effect Sizes for the Predictive Relationship between Specific and Categorical/Overgeneral Memories and Depressive Symptoms at Follow-up

Variable	Correlational Data Set		Standardized Regression Coefficient Data Set	
	Specific	Categoric/Overgeneral	Specific	Categoric/Overgeneral
Number of studies	4	8	6	6
Weighted overall effect size(ES) ^a	-.10 ⁺	.13 ^{**}	-.17 [*]	.11 [*]
95% CI for ES	-.22, .02	.04, .23	-.31, -.03	.02, .20
Homogeneity (<i>Q</i>)	4.38	6.34	11.54 [*]	3.58
I ² index	31%	0%	57%	0%

Note. CI = confidence interval. Nonsignificance of the *Q* statistic indicates a failure to reject the hypothesis of homogeneity. The I² index represents the percentage of the total variability in effect size estimates that is due to true heterogeneity between studies. Effect sizes are based on a random effects model.

^a Effect sizes are weighted by the reciprocal of the variance.

⁺ $p < .11$,

^{*} $p < .05$,

^{**} $p < .01$

Table 4

Results of Analyses for Categorical Moderators

Moderator Variable	Between-classes Effect (Q_B)	n	Mean Weighted Effect Size	95% CI	Homogeneity Within Each Class (Q_W)	I^2 Index
Correlations for Categorical/Overgeneral Memories						
Clinical Status			.21			
Clinical Depression		5	.17 ⁺	-.01, .33	2.17	0%
Nonclinical		3	.11	-.06, .27	3.97	50%
Cue Word Valence	0					
Positive		4	.22	-.13, .52	9.55*	69%
Negative		3	.22*	.02, .41	.80	0%
Depressive Symptoms Rating Method	1.74					
Self-Rated		6	.11*	.01, .21	4.47	0%
Clinician-Rated		3	.29*	.04, .51	.34	0%
Standardized Regression Coefficients for Specific Memories						
Clinical Status			8.60**			
Clinical Depression		3	-.37***	-.51, -.23	.45	0%
Nonclinical		3	-.10 ⁺	-.21, -.02	1.44	0%
Cue Word Valence	.13					
Positive		3	-.19	-.44, .06	4.44	55%
Negative		3	-.27	-.65, .10	7.49*	73%
Depressive Symptoms Rating Method	2.96 ⁺					
Self-Rated		3	-.06	-.18, .07	1.92	0%
Clinician-Rated		3	-.28*	-.51, -.06	5.33 ⁺	62%
Standardized Regression Coefficients for Categorical/Overgeneral Memories						
Cue Word Valence			.45			
Positive		5	.11	-.10, .31	8.78 ⁺	55%

Moderator Variable	Between-classes Effect (Q_B)	n	Mean Weighted Effect Size	95% CI	Homogeneity Within Each Class (Q_W)	I^2 Index
Negative		4	-.002	-.24, .24	7.96*	62%
Depressive Symptoms Rating Method	.01					
Self-Rated		4	.11*	.01, .21	3.05	2%
Clinician-Rated		3	.12	-.06, .30	.75	0%

Note. CI = confidence interval. Significance for Q indicates a rejection of the homogeneity hypothesis. The I^2 index represents the percentage of the total variability in effect size estimates that is due to true heterogeneity between studies. Effect size based on a mixed effects model.

+ $p < .11$,

* $p < .05$,

** $p < .01$,

**** $p < .0001$