Psychological Assessment
2012, Vol. 24, No. 3, 640–646

1040-3590/11/$12.00 DOI: 10.1037/a0026510

The Factor Structure of the Autobiographical Memory Test in Recent Trauma Survivors

James W. Griffith
Northwestern University

Birgit Kleim
King's College London

Jennifer A. Sumner
Northwestern University

Anke Ehlers
King's College London and Oxford University

The objective of this study was to examine the psychometric properties of the Autobiographical Memory Test (AMT), which is widely used to measure overgeneral autobiographical memory in individuals with depression and a trauma history. Its factor structure and internal consistency have not been explored in a clinical sample. This study examined the psychometric properties of the AMT in a sample of recent trauma survivors ($N = 194$), who completed the AMT 2 weeks after a trauma. Participants were also assessed with structured clinical interviews for current acute stress disorder and current and past major depressive disorder. Confirmatory factor analysis and item response theory were used to analyze the AMT in the whole sample. The factor structure of the AMT was also compared for (a) individuals with and without lifetime major depressive disorder and (b) individuals with current (posttrauma) major depressive disorder and/or acute stress disorder versus those with neither disorder. In all of these analyses, the AMT with cues of positive and negative valence had a 1-factor structure, which replicates work in nonclinical samples. Based on analyses of the whole sample, scores from the AMT had a reliability estimate of .72, and standard error of measurement was lowest for people who scored low on memory specificity. In conclusion, the AMT measures 1 factor of memory specificity in a clinical sample and can yield reliable scores for memory specificity. More psychometric studies of the AMT are needed to replicate these results with similar and other clinical populations.

Keywords: psychometrics, Autobiographical Memory Test, overgeneral autobiographical memory, acute stress disorder, major depressive disorder

Since Williams and Broadbent's (1986) seminal study on autobiographical memory in suicide attempters, there has been great interest in *overgeneral autobiographical memory* (OGM), which is defined as a difficulty in retrieving specific autobiographical memories. When participants are presented with a cue word and asked to provide a specific memory in response, OGM is manifest when the participant provides a summary of events (e.g., "Whenever I go cycling") or memories for events that lasted for more than 1 day (e.g., "My first year of college"), rather than retrieving a specific personal event that occurred at a particular time and place and lasted for less than 1 day (e.g., "My first day of kindergarten"). The intensive study of OGM has been motivated by evidence of its relationship with depression, trauma-related psychopathology (for reviews, see Moore & Zoellner, 2007; Williams et al., 2007), and difficulties with social problem solving (e.g., Goddard, Dritschel, & Burton, 1996). Furthermore, OGM predicts increases in depressive symptoms after stress (e.g., Van Minnen, Wessel, Verhaak, & Smeenk, 2005) and has been observed in individuals in remission from depression (e.g., Mackinger, Pachinger, Leibetseder, & Fartacek, 2000). OGM also predicts a worse course of depression (Hermans, Vandromme, et al., 2008; Sumner, Griffith, & Mineka, 2010; Sumner, Griffith, Mineka, et al., 2010) and the onset of major depressive disorder (MDD) and posttraumatic stress disorder (PTSD) after trauma (Kleim & Ehlers, 2008).

There are some basic methodological issues that hamper interpretation of the extant research on OGM, and potentially limit the impact of future studies. Griffith et al. (in press) detailed several measurement issues in OGM research, such as a lack of cross-study consistency in the measurement of OGM and the lack of information on psychometric properties of certain OGM measurement tools. In particular, more research is needed on the psychometric properties of the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986), a cuing methodology that is widely used to measure OGM.

This article was published Online First December 12, 2011.

James W. Griffith, Department of Medical Social Sciences, Northwestern University; Birgit Kleim, Institute of Psychiatry, King's College London, London, England; Jennifer A. Sumner, Department of Psychology, Northwestern University; Anke Ehlers, Institute of Psychiatry, King's College London, London, England, and Oxford University, Oxford, England.

We gratefully acknowledge the support to Anke Ehlers of the Psychiatry Research Trust and the Wellcome Trust (Grant 069777). This article derives, in part, from discussions at the Seventh Special Interest Meeting on Autobiographical Memory and Psychopathology, October 1–3, 2009, Seminaris Campus Hotel, Berlin, Germany.

Correspondence concerning this article should be addressed to James W. Griffith, Department of Medical Social Sciences, Northwestern University, Feinberg School of Medicine, 625 N. Michigan Ave., 27th Floor, Chicago, IL 60611. E-mail: j-griffith@northwestern.edu or jameswgriffith@gmail.com

Some psychometric properties of the AMT have been examined in past research. For example, Raes, Williams, and Hermans (2009) examined the test–retest reliability of alternate forms of the AMT in six samples, with retest intervals ranging from 1 to 5 months. Five of the samples were college undergraduates, and one sample was composed of depressed inpatients. Raes et al. (2009) found that Spearman rank correlations between baseline and follow-up scores ranged from .53 to .68. In addition, Griffith et al. (in press) examined the internal consistency of the AMT in an undergraduate sample using the latent-variable method of Raykov, Dimitrov, and Asparuhov (2010) and found a reliability coefficient of .79 (95% CI = [.74, .84]).

Studies that have used the AMT have included a number of different cue words, and there are questions that remain regarding how different cue-word characteristics influence performance on the AMT. Because most studies use positive and negative emotion words as cues on the AMT, valence is one cue-word characteristic that is relevant to much of the research on OGM. A meta-analytic review of factors that impact AMT performance did not find significant differential effects of cue valence on memory specificity (van Vreeswijk & de Wilde, 2004), but few factor analytic studies have explored the role of valence. One exception is the study by Griffith et al. (2009): Using data from three studies that used positive and negative cue words on the AMT, they found that a one-factor model provided the best fit to the AMT data, which suggests that OGM is a generalized style that influences memories that are elicited by both positive and negative cues. This generalized tendency to retrieve nonspecific memories may develop as a way to avoid negative emotions associated with specific memories (Hermans, de Decker et al., 2008; Hermans, Defranc, Raes, Williams, & Eelen, 2005; Raes, Hermans, de Decker, Eelen, & Williams, 2003). One limitation of the work of Griffith et al. (2009, in press) is the reliance on nonclinical samples of participants (e.g., students who were not selected on the basis of depression), who generally perform better on the AMT than participants drawn from clinical samples. Thus, it is unknown whether a one-factor structure would be found in a clinical sample, or whether a different factor structure would emerge.

The current study aims to address this gap in the literature. In a sample of recent trauma survivors, a proportion of whom met diagnostic criteria for acute stress disorder (ASD) and/or MDD, we examined the factor structure of the AMT, as well as the reliability of AMT scores. Given previous results in nonclinical samples, we hypothesized that a one-factor model would provide a good fit, whereas a more complex model with factors for positive and negative word valence would not provide a better fit.

Method

Participants

Participants were assault survivors who were treated for their injuries at a hospital emergency department between July 2003 and December 2004. To be eligible for the study, participants had to understand and speak English fluently enough to complete interviews and questionnaires. A priori exclusion criteria were current psychosis, substance dependence, or not being able to remember the event (e.g., because of a head injury). Inclusion and exclusion criteria were assessed in a telephone screening interview before the research session was scheduled. Those with minor head injuries who were nevertheless able to remember the assault in detail were not excluded. The sample comprised 194 participants (70% male). Regarding ethnicity, 57% of the participants were White; 34% were Black; 9% self-identified as "other"; and 1% were Indian, Pakistani, or Bangladeshi. Ages ranged from 18 to 64 years (M = 34.3, SD = 10.8). Data on years of education were available for 171 participants (M = 13.9 years, SD = 4.8). Participants were mainly physically assaulted, with an average of two assailants involved (M = 2.4, SD = 2.8), of whom approximately 50% had a weapon. The average assault duration was 8 min (M = 7.8, SD = 8.7). Participants attended the research session where the AMT was conducted at approximately 2 weeks post-assault (M = 17 days since assault, SD = 8).

Autobiographical Memory Test (AMT)

The AMT (Williams & Broadbent, 1986) was administered individually by the second author following the procedures outlined in Williams's (2003) manual. Participants saw 12 cue words (six each of positive and negative words; see Table 1), which occurred in random order on a computer screen. The participants' task was to retrieve and briefly describe a specific personal memory in response to each word. Participants were given 30 s to begin responding after the cue word was presented. A trial was coded as a *specific* memory if the memory was of a personal event that occurred at a particular time and place within the course of 1 day. Nonspecific responses included *extended* memories, which lasted for more than 1 day (e.g., "My weekend in Zürich"); *categoric* memories, which are thematic summaries of events (e.g., "Whenever I go for a bike ride"); *nonmemories* (a.k.a. *semantic associates*; e.g., "I am a happy person"); and *omissions*, in which the participant was unable to furnish a response (e.g., "I don't know").

Structured Clinical Interviews

Current and past (i.e., prior to the assault) MDD, as well as current ASD, were diagnosed using the Structured Clinical Interview for *DSM–IV* (SCID; First, Spitzer, Gibbon, & Williams, 2002), and the Acute Stress Disorder Scale (ASDS, interview version; Bryant & Harvey, 2000), respectively. Clinical interviews were conducted at the end of the research session, which occurred after participants had already completed the AMT. A trained clinical psychologist conducted all interviews under the supervision of an expert clinical psychologist. Interview diagnoses showed good inter-rater reliability (κ = .97 for ASD, κ = 1.00 for MDD) in a randomly selected subset of 56 interviews.

Analytic Strategy

We used confirmatory factory analysis (CFA) to determine whether AMT performance represents a single dimension of autobiographical memory specificity or whether it is characterized by a two-dimensional structure reflecting cue valence. We also examined the AMT using item response theory (IRT), a psychometric method based on the notion that an individual's performance on a test item can be predicted by latent traits (Hambleton, Swaminathan, & Rogers, 1991). In IRT, the probability of responding to an item in a particular way (e.g., providing a specific memory) is

Table 1
*Item Response Theory Statistics for a One-Factor Model Across the Sample of Individuals With (*n = *90) and Without (*n = *104) a Lifetime History of MDD*

Cue word	Item slope	Item threshold	Omission	Nonmemory	Categoric	Extended	Specific
Guilty	.61	−.24	4%	22%	6%	8%	60%
MDD			2%	20%	8%	9%	61%
Non-MDD			6%	23%	5%	8%	58%
Cheer	.59	−.57	4%	14%	7%	4%	72%
MDD			2%	13%	11%	4%	69%
Non-MDD			5%	14%	4%	3%	74%
Relieved	.52	−.87	1%	13%	3%	3%	81%
MDD			0%	16%	2%	4%	78%
Non-MDD			1%	11%	4%	1%	83%
Awful	.50	−.75	3%	9%	5%	5%	77%
MDD			1%	10%	6%	7%	77%
Non-MDD			5%	9%	5%	3%	78%
Worse	.50	−.34	3%	22%	4%	7%	63%
MDD			2%	22%	1%	8%	67%
Non-MDD			4%	22%	7%	7%	61%
Lively	.49	−.22	3%	21%	7%	10%	59%
MDD			2%	18%	4%	11%	64%
Non-MDD			4%	24%	10%	9%	54%
Pleased	.45	−.43	3%	19%	5%	6%	67%
MDD			3%	16%	4%	8%	69%
Non-MDD			2%	22%	6%	5%	65%
Hopeless	.44	−.24	6%	21%	6%	7%	59%
MDD			0%	22%	8%	8%	62%
Non-MDD			12%	20%	4%	7%	57%
Glorious	.41	−.27	9%	14%	6%	10%	61%
MDD			7%	15%	7%	12%	60%
Non-MDD			11%	13%	6%	9%	62%
Peaceful	.33	.13	3%	21%	20%	11%	45%
MDD			2%	19%	18%	11%	49%
Non-MDD			4%	23%	22%	11%	41%
Grave	.32	−.60	4%	16%	2%	5%	72%
MDD			3%	17%	3%	8%	69%
Non-MDD			4%	16%	1%	3%	76%
Ugly	.28	−.09	6%	32%	4%	5%	54%
MDD			6%	24%	5%	8%	58%
Non-MDD			7%	38%	3%	2%	50%

Note. MDD = major depressive disorder. All item slopes are statistically significantly different from zero ($p \leq .011$). Percentages are from nonmissing Autobiographical Memory Test responses only. Percentages may not sum to 100% because of rounding errors.

a function of a latent trait. IRT and CFA can be unified in a single mathematical framework (Muthén & Muthén, 1998–2010), so we simultaneously examined the factor structure of the AMT, as well as IRT parameters for individual items. We also calculated the internal consistency reliability of scores from the AMT, which has not been examined in a clinical sample.

To evaluate model fit in CFA, we used a comparative fit index (CFI; Bentler, 1990; Hu & Bentler, 1999) and a root-mean-square error of approximation (RMSEA; Browne & Cudeck, 1993). We used cutoffs as a guide to model interpretation by seeking models with CFI $\geq$ 0.95 and RMSEA $<$ 0.06, as suggested by Hu and Bentler (1999), but we did not reject models outright if certain fit indices were slightly outside these cutoffs (for discussion, see Marsh, Hau, & Wen, 2004). We fit each model using a robust (mean- and variance-adjusted) method of weighted least squares (WLSMV) estimation.

To reduce the number of thresholds to be estimated, we treated each response on the AMT as a dichotomy: specific versus nonspecific. Omissions were included in the analyses as nonspecific responses. Thus, two parameters were estimated for each cue word: one item-slope and one threshold. A slope parameter describes the degree of relationship between an item and a latent trait; it also captures the ability of an item to discriminate between people who are high and low on the trait being studied (memory specificity in this case). We sought standardized item slopes of .30 or larger. Threshold parameters are placed along a standardized continuum in our analyses. Threshold parameters help to determine the point at which a person is more than 50% likely to respond with a response category above the threshold (Embretson & Reise, 2000, pp. 98–99). Threshold parameters convey information about how difficult it is for individuals to generate a specific response to a cue word. In IRT, the slope and threshold parameters jointly determine the likelihood of a particular response to an item, given a person's level of ability on the trait being studied.

Results

Prior to analyses, we examined frequency distributions for each cue word. Table 1 presents the percentage of response types among nonmissing data for the total sample, and separately for those with

and without a lifetime history of MDD. Specific memories were the modal response for each cue word.

One- Versus Two-Factor Models

We attempted to fit a two-factor model with factors for positive and negative words. For each of the cue words, we estimated an item-slope and a threshold parameter. When we attempted to fit a two-factor model to the whole sample ($N = 194$), we obtained good fit indices overall, CFI = .93, RMSEA = .03, WLSMV $\chi^2(53) = 63.80, p = .15$.[1] The correlation between the positive and negative word factors was .98 (95% CI = [.71, 1.00], $p < .001$). The very large factor correlation suggested that a one-factor model should provide a good fit. The one-factor model, which is more parsimonious, had similar fit indices, CFI = .94, RMSEA = .03, WLSMV $\chi^2(54) = 63.64$, $p = .17$, and did not significantly differ from the two-factor model, difference $\chi^2(1) = 1.02$, $p = .31$. Based on these results, we retained the one-factor model as the final one. Standardized item slopes ranged from .28 to .61 and were all statistically significantly different from zero ($p < = .011$). Standardized thresholds ranged from –.87 to .13 (see Table 1).

Comparison of Participants With and Without a Lifetime History of MDD

To examine whether the psychometric properties differed as a function of depression, we used multiple group CFA with lifetime MDD status as the grouping variable. Different CFA/IRT parameters were estimated separately for the groups with and without a lifetime history of MDD (*n*s = 90 and 104, respectively). Lifetime history of MDD was defined as having a current or past diagnosis, whereas the no-history group did not have a current or past diagnosis. Among the 90 participants with a lifetime history, 37 had a current diagnosis of MDD. Based on our CFA findings from the whole sample, we fit a one-factor model in each group. Because the group sample sizes were small, it was necessary to impose further simplifying assumptions to identify the model. Across both groups, we constrained all the items to have the same item slope. For each group, a separate threshold was estimated for each word. The factor means of the two groups were both set to zero. This model allowed the thresholds across the two groups to be compared on the same scale. The model fit well across groups, CFI = .96, RMSEA = .03, WLSMV $\chi^2(119) = 125.93$, $p = .31$. We compared this model with a fully metric invariant model in which all the parameters were identical across the groups. The metric invariant model was not significantly different, so we accepted it as the final model, difference $\chi^2(12) = 8.32$, $p = .76$; CFI = .99, RMSEA = .01, WLSMV $\chi^2(131) = 133.48$, $p = .42$. Table 1 shows that the groups were comparable on their distributions of response types for each cue word.

Comparison of Participants With and Without a Current Diagnosis of ASD and/or MDD

To address the role of current trauma-related psychopathology, we compared participants with a current (posttrauma) diagnosis of ASD and/or MDD at the time of the AMT with participants with neither current diagnosis. Further subdivision into psychopathology groups was not possible because of small sample sizes. Using multiple group CFA, we solved for different CFA/IRT parameters in each group ($n = 56$ for ASD and/or MDD; $n = 138$ for neither diagnosis). There were 19 participants with current ASD only, 23 participants with current MDD, and 14 participants with comorbid MDD and ASD. We constrained all items to have the same item slope in both groups, but a separate threshold was estimated for each word, and thresholds were free to vary across groups. As before, the factor means of the groups were set to zero, allowing the thresholds to be compared on the same scale.

The multiple group CFA fit well, CFI = .97, RMSEA = .02, WLSMV $\chi^2(119) = 122.90$, $p = .38$. We compared this model with a fully metric invariant model in which parameters were identical across groups. The metric invariant model was not significantly different, difference $\chi^2(12) = 18.88$, $p = .09$; CFI = .92, RMSEA = .03, WLSMV $\chi^2(131) = 141.81$, $p = .24$.[2]

Internal Consistency Reliability of AMT Scores

Reliability is the proportion of observed variance that is due to true score variance. Because Cronbach's alpha underestimates reliability for binary data (Raykov et al., 2010), we used the method of Raykov et al. (2010) for estimating reliability for test scores composed of dichotomous items. Reliability for scores from the AMT was .72 (95% CI = [.67, .77]), based on dichotomous (specific vs. not specific) scoring of AMT responses from the whole sample. We also calculated coefficient alpha using a scoring algorithm suggested at an AMT consensus meeting. Specific, extended, categoric, and nonmemories were coded 4–1, respectively (see Raes, Hermans, Williams, & Eelen, 2007, footnote 2). Alpha was .67, which was slightly lower than the reliability coefficient yielded by the Raykov et al. (2010) method, even though more response types were included in the calculations.

Summary of IRT Results

Given that the one-factor model fit well, we present detailed IRT statistics for the one-factor model across the entire sample (see Table 1). With the exception of *ugly*, all of the words had an item slope of .30 or greater. Most of the thresholds were below zero, with the exception of *peaceful*. Figure 1 shows the test information function, which plots test *information*, a way to express measurement precision, as a function of memory specificity. Test information is the sum of information conveyed by each individual test item. It is related to standard error of measurement (SEM) by the equation information = $1/SEM^2$. SEM is also depicted in Figure 1 on a standardized scale ($M = 0$, $SD = 1$). SEM has its lowest value (SEM = 0.65) at a memory specificity score of –.83.

Discussion

We examined the psychometric properties of the AMT in a clinical sample of trauma survivors, as well as in subgroups of (a)

[1] It should be noted that the WLSMV chi-square and the accompanying degrees of freedom cannot be interpreted in the usual ways. Therefore, only the statistical significance of the chi-square is interpretable.

[2] IRT parameters, as well as descriptive statistics for participants with ASD and/or MDD versus participants with neither disorder, are available from the authors upon request.

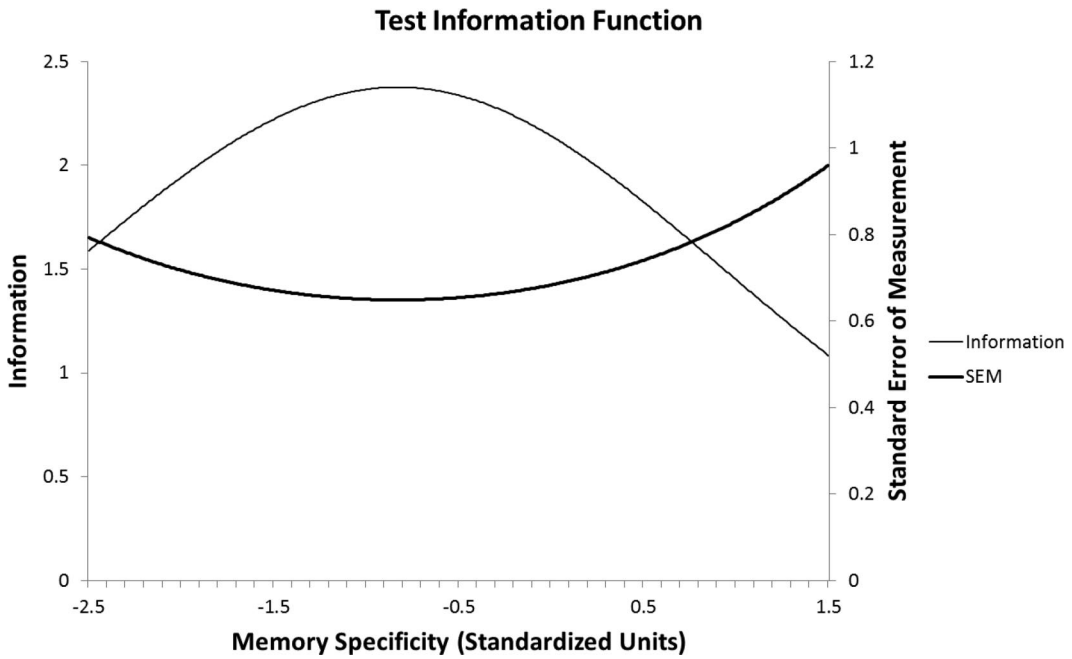

Figure 1. Test information function and standard error of measurement (SEM). Autobiographical Memory Test specificity and SEM are measured on a standardized scale. *Information*, on the ordinate, is the precision of the test at a particular point, measured in a squared metric.

participants with and without a lifetime history of MDD and (b) participants with current (posttrauma) diagnoses of MDD and/or ASD versus those with neither diagnosis. It is important to note that our sample comprised individuals who had all experienced a recent trauma. Some studies of trauma survivors include participants with remote histories of trauma, or a mixture of participants with recent or remote trauma history. The results of this study are consistent with those of Griffith et al. (2009) in nonclinical samples: A one-factor model fit the AMT well, suggesting that this test measures a unidimensional construct of autobiographical memory specificity. Inspection of factor loadings suggested that the words had adequate loadings on memory specificity, although *ugly* had a standardized item slope that was slightly below .30. Moreover, item thresholds tended to be below the mean, which suggests that even people with below average memory specificity will be more likely than not to retrieve specific memories for most cue words; in the current study, only *peaceful* had a threshold above zero in the overall sample.

SEM is a common metric of test imprecision and can be used to construct confidence intervals around individual test scores. As shown in Figure 1, the AMT had the lowest SEM for individuals who scored .83 standard deviations below the mean on memory specificity. In other words, the precision of the test is greatest for people who score low on memory specificity relative to other participants. Nevertheless, it is noteworthy that SEM was below 1.0 for the range of OGM studied, which suggests that the AMT is a precise test in this clinical sample. As pointed out by Cronbach and Shavelson (2004), approximately two thirds of observed scores will fall within one SEM of a person's true score. However, the precision of the test is lower for participants who are high on memory specificity, relative to low-specificity participants. Reliability of the AMT scores appears adequate (.72), although the 95% CI of [.67, .77] suggests that low reliability might be considered for studies with null findings. Interestingly, the reliability of AMT scores was slightly lower when using a 1–4 rating scale for the specificity of each memory. This may be because the rating scale assumed an interval scale for response to each cue word, which is perhaps a tenuous assumption for the AMT.

Implications for Future Studies

The AMT appears to be unidimensional and does not appear to differentiate memories that are elicited from positive and negative words in terms of specificity, even in individuals with a history of MDD and/or ASD. Thus, when measuring OGM, an overall specificity score should be considered. However, this study is the first to examine this question in a clinical sample, so these findings await replication and extension to samples with other forms of psychopathology. We had too few participants to examine those with current MDD, as opposed to lifetime MDD, as a separate group. It is possible that current mood state interacts with cue valence and that a different factor structure might emerge in people suffering from current MDD. An IRT/CFA analysis of the AMT in a large sample of people with current MDD would be informative to explore this possibility. Future studies might also employ the bifactor approach (Holzinger & Swineford, 1937) to examine whether group factors for valence can be identified in addition to a general factor of memory specificity. This approach would help to identify valence-specific effects that are independent of an overall memory specificity factor. This approach was not possible in this sample because of its small size.

In addition to further examination of the dimensionality of OGM, it is important for future research to link these dimensions to their causal mechanisms. Williams (2006) reviewed the literature and proposed three mechanism of OGM: capture and rumination, functional avoidance, and decreased executive control. Research has begun to examine these mechanisms in trauma survivors. For example, Dalgleish and his colleagues have examined the role of decreased executive control as a cause of OGM (Dalgleish et al., 2007), but affect regulation, as opposed to decreased executive function, accounted for OGM in trauma survivors (Dalgleish, Rolfe, Golden, Dunn, & Barnard, 2008).

The results of IRT analyses, including those in this study, may be useful for planning future studies and for creating an AMT for clinical use. For example, investigators should note that the AMT may be prone to ceiling effects because of the low threshold for generating a specific memory. In this study, SEM was higher for high scorers on the AMT, relative to low scorers. This finding should be considered when using the AMT with high-functioning samples, because these samples might best be assessed with cues that have higher thresholds (i.e., cues that are more difficult in the sense of being able to elicit a specific memory). IRT results might also be inspected to determine whether certain cues should be included. For example, in this study, *peaceful*, *grave*, and *ugly* all had standardized item slopes smaller than .4, suggesting a weak relationship with the underlying dimension of memory specificity. Of course, the magnitude of an individual IRT parameter is subject to sampling error like any statistic, which underscores the need for future studies.

Limitations

Although our sample contained individuals with a history of MDD and/or ASD, we did not examine the effect of comorbidity. We also did not examine other disorders besides MDD and ASD. Even with these disorders, it was not possible to separate them into pure groups because of small sample sizes. PTSD, for example, is known to be associated with OGM, but we did not examine PTSD because lifetime history of PTSD prior to the AMT was not assessed. Further investigation of the psychometric properties of the AMT in samples of individuals with trauma-related psychopathology is still needed. Such research would help determine whether the AMT might be useful for differential diagnosis.

The results of this study may not generalize well to other samples. For example, our sample had more men than women. It is possible that gender differences might be observed in the factor structure of the AMT across genders and across racial and ethnic groups. To conduct such studies, large subsamples would be needed for each group. This was not available for the current study, so this remains an important issue for future research. A larger sample than the current one would have enabled other analyses, such as examining each response type on the AMT as an ordered category, rather than simply a dichotomy of specific versus nonspecific. More categories per item require more threshold parameters to be estimated, but fine-grained analyses that distinguish between various memory types of the AMT (specific, extended, categoric, extended, etc.) will help to identify specific response patterns that are related to various forms of psychopathology.

As with other psychometric studies of the AMT, the results are limited to the particular set of cue words that were used. It is unknown whether similar results would be obtained with different words. Heretofore, only positive and negative emotion words have been examined in psychometric studies of the AMT, although some studies that have used the AMT have included emotionally neutral words (e.g., Henderson, Hargreaves, Gregory, & Williams, 2002). Investigators should be encouraged to examine psychometric issues in their own data, especially given the varying methodologies used to measure OGM. Although some psychometric analyses require large samples, they can aid in the interpretation of research findings, can further inform investigators about the psychometric properties of the AMT, and can suggest refinements to it.

Summary and Conclusions

There have been no IRT or factor analyses of the AMT in a sample that grouped individuals according to lifetime MDD status or current MDD and/or ASD diagnoses. Consistent with prior work in nonclinical samples, we found that a parsimonious one-factor structure fit the data well. This study supports the notion that OGM is a unidimensional construct, but replication is needed. We hope that more precise measurement of OGM will yield further insights into its role in psychopathology and suggest clinical applications of assessing OGM.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. doi:10.1037/0033-2909.107.2.238

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural models* (pp. 136–162). Newbury Park, CA: Sage.

Bryant, R. A., & Harvey, A. G. (2000). *Acute stress disorder: A handbook of theory, assessment, and treatment.* Washington, DC: American Psychological Association. doi:10.1037/10346-000

Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64,* 391–418. doi:10.1177/0013164404266386

Dalgleish, T., Rolfe, J., Golden, A.-M., Dunn, B. D., & Barnard, P. J. (2008). Reduced autobiographical memory specificity and posttraumatic stress: Exploring the contributions of impaired executive control and affect regulation. *Journal of Abnormal Psychology, 117,* 236–241. doi:10.1037/0021-843X.117.1.236

Dalgleish, T., Williams, J. M. G., Golden, A.-M. J., Perkins, N., Barrett, L. F., Barnard, P. J., . . . Watkins, E. (2007). Reduced specificity of autobiographical memory and depression: The role of executive control. *Journal of Experimental Psychology: General, 136,* 23–42. doi:10.1037/0096-3445.136.1.23

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM–IV–TR Axis I Disorders* (Research Version, Non-patient ed.). New York, NY: Biometrics Research, New York State Psychiatric Institute.

Goddard, L., Dritschel, B., & Burton, A. (1996). Role of autobiographical memory in social problem solving and depression. *Journal of Abnormal Psychology, 105,* 609–616. doi:10.1037/0021-843X.105.4.609

Griffith, J. W., Sumner, J. A., Debeer, E., Raes, F., Hermans, D., Mineka, S., . . . Craske, M. G. (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory, 17,* 609–623. doi:10.1080/09658210902939348

Griffith, J. W., Sumner, J. A., Raes, F., Barnhofer, T., Debeer, E., &

Hermans, D. (in press). The relationship between autobiographical memory and psychopathology: Psychometric and methodological issues. *Journal of Behavior Therapy and Experimental Psychiatry.*

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Henderson, D., Hargreaves, I., Gregory, S., & Williams, J. M. G. (2002). Autobiographical memory and emotion in a non-clinical sample of women with and without a reported history of childhood sexual abuse. *British Journal of Clinical Psychology, 41,* 129–142. doi:10.1348/014466502163921

Hermans, D., de Decker, A., De Peuter, S., Raes, F., Eelen, P., & Williams, J. M. G. (2008). Autobiographical memory specificity and affect regulation: Coping with a negative life event. *Depression and Anxiety, 25,* 787–792. doi:10.1002/da.20326

Hermans, D., Defranc, A., Raes, F., Williams, J. M. G., & Eelen, P. (2005). Reduced autobiographical memory specificity as an avoidant coping style. *British Journal of Clinical Psychology, 44,* 583–589. doi:10.1348/014466505X53461

Hermans, D., Vandromme, H., Debeer, E., Raes, F., Demyttenaere, K., Brunfaut, E., & Williams, J. M. G. (2008). Overgeneral autobiographical memory predicts diagnostic status in depression. *Behaviour Research and Therapy, 46,* 668–677. doi:10.1016/j.brat.2008.01.018

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2,* 41–54. doi:10.1007/BF02287965

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Kleim, B., & Ehlers, A. (2008). Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. *Journal of Consulting and Clinical Psychology, 76,* 231–242. doi:10.1037/0022-006X.76.2.231

Mackinger, H. F., Pachinger, M. M., Leibetseder, M. M., & Fartacek, R. R. (2000). Autobiographical memories in women remitted from major depression. *Journal of Abnormal Psychology, 109,* 331–334. doi:10.1037/0021-843X.109.2.331

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341. doi:10.1207/s15328007sem1103_2

Moore, S. A., & Zoellner, L. A. (2007). Overgeneral autobiographical memory and traumatic events: An evaluative review. *Psychological Bulletin, 133,* 419–437. doi:10.1037/0033-2909.133.3.419

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Raes, F., Hermans, D., de Decker, A., Eelen, P., & Williams, J. M. G. (2003). Autobiographical memory specificity and affect regulation: An experimental approach. *Emotion, 3,* 201–206. doi:10.1037/1528-3542.3.2.201

Raes, F., Hermans, D., Williams, J. M. G., & Eelen, P. (2007). A sentence completion procedure as an alternative to the Autobiographical Memory Test for assessing overgeneral memory in non-clinical populations. *Memory, 15,* 495–507. doi:10.1080/09658210701390982

Raes, F., Williams, J. M. G., & Hermans, D. (2009). On the test–retest reliability of the autobiographical memory test. In L. B. Palcroft & M. V. Lopez (Eds.), *Personality assessment: New research* (pp. 391–397). New York, NY: Nova Science.

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling, 17,* 265–279. doi:10.1080/10705511003659417

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour Research and Therapy, 48,* 614–625. doi:10.1016/j.brat.2010.03.013

Sumner, J. A., Griffith, J. W., Mineka, S., Rekart, K. N., Zinbarg, R. E., & Craske, M. G. (2010). Overgeneral autobiographical memory and chronic interpersonal stress as predictors of the course of depression in adolescents. *Cognition & Emotion.* Advance online publication. doi:10.1080/02699931003741566

Van Minnen, A., Wessel, I., Verhaak, C., & Smeenk, J. (2005). The relationship between autobiographical memory specificity and depressed mood following a stressful life event: A prospective study. *British Journal of Clinical Psychology, 44,* 405–415. doi:10.1348/014466505X29648

van Vreeswijk, M. F., & de Wilde, E. J. (2004). Autobiographical memory specificity, psychopathology, depressed mood, and the use of the Autobiographical Memory Test: A meta-analysis. *Behaviour Research and Therapy, 42,* 731–743. doi:10.1016/S0005-7967(03)00194-3

Williams, J. M. G. (2003). *Autobiographical Memory Test scoring manual.* Unpublished manuscript, Oxford University.

Williams, J. M. G. (2006). Capture and rumination, functional avoidance, and executive control (CaRFAX): Three processes that underlie overgeneral memory. *Cognition & Emotion, 20,* 548–568. doi:10.1080/02699930500450465

Williams, J. M. G., Barnhofer, T., Crane, C., Hermans, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin, 133,* 122–148. doi:10.1037/0033-2909.133.1.122

Williams, J. M. G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology, 95,* 144–149. doi:10.1037/0021-843X.95.2.144

Received February 8, 2011
Revision received August 9, 2011
Accepted August 25, 2011 ■