



Published in final edited form as:

Memory. 2009 August ; 17(6): 609–623. doi:10.1080/09658210902939348.

An item response theory/confirmatory factor analysis of the Autobiographical Memory Test

James W. Griffith,

Northwestern University, Evanston, IL, USA

Jennifer A. Sumner,

Northwestern University, Evanston, IL, USA

Elise Debeer,

University of Leuven, Belgium

Filip Raes,

University of Leuven, Belgium

Dirk Hermans,

University of Leuven, Belgium

Susan Mineka,

Northwestern University, Evanston, IL, USA

Richard E. Zinbarg, and

Northwestern University and The Family Institute at Northwestern University, Evanston, IL, USA

Michelle G. Craske

University of California, Los Angeles, CA, USA

Abstract

The Autobiographical Memory Test (AMT) is used to assess the degree of specificity of autobiographical memory. The AMT usually contains cue words of both positive and negative valence, but it is unclear whether these valences form separate factors or not. Accordingly, confirmatory factor analysis assessed whether the AMT measures one overall factor, or whether different cue types are related to different factors. Results were consistent across three datasets ($N=333$, $N=405$, and $N=336$). A one-factor model fitted each dataset well, which suggests that responses to positive and negative cues are related to the one construct. In addition, item response theory analyses showed that the AMT is most precise for people who score low on memory specificity. Implications for using the AMT with high-functioning samples are discussed.

Keywords

Autobiographical memory; Overgeneral memory; Item response theory; Factor analysis; Emotional valence

Autobiographical memory refers to a network of personal knowledge that includes both specific (episodic) and more generic (conceptual) self-related information (Conway &

Pleydell-Pearce, 2000; Raes, Hermans, Williams, & Eelen, 2007). Autobiographical memory is widely recognised as an integral component of human functioning, and has been the subject of numerous studies investigating its relationship to different aspects of human experience. Research has shown that overgenerality in autobiographical memory is associated with some forms of psychopathology, with the majority of studies focused on the relationship between overgeneral memory (OGM) and depression (Williams et al., 2007).

When asked to recall specific autobiographical memories, individuals with depression are more likely to report overgeneral memories and less likely to recall specific memories than non-depressed controls. A specific memory is typically defined as a particular event that lasted less than a day. Overgeneral memories, in contrast, refer to more generic material, such as a whole class of events (categorical memories) or extended periods of time (extended memories). Some investigators have found OGM in individuals whose depression has remitted (Mackinger, Pachinger, Leibetseder, & Fartacek, 2000b; Williams & Dritschel, 1988), which suggests that OGM may be a trait marker for depression. OGM also seems to predict the course of depression (e.g., Brittlebank, Scott, Williams, & Ferrier, 1993; Mackinger, Loschin, & Leibetseder, 2000a).

Although OGM is a replicable phenomenon in individuals with clinical levels of depression, findings are less consistent among the smaller number of studies conducted with non-clinical samples. Although some studies have found that individuals with sub-clinical levels of depression (dysphoria) are less specific in their memory than non-dysphoric respondents (e.g., Goddard, Dritschel, & Burton, 1997; Ramponi, Barnard, & Nimmo-Smith, 2004; Rekart, Mineka, & Zinbarg, 2006), other studies with non-clinical participants have failed to find this phenomenon. For example, Raes, Pousset, and Hermans (2004) found no significant differences between the number of specific or overgeneral memories generated by high- versus low-dysphoric students. In fact, greater memory specificity has sometimes been found to be related to *higher* levels of depressive symptoms in non-clinical participants (Debeer, Hermans, & Raes, 2005). In response to these inconsistencies in the literature, Raes et al. (2007) proposed that the standard method of assessing OGM (the Autobiographical Memory Test, described below) might not be sufficiently sensitive for use with non-clinical samples.

THE AUTOBIOGRAPHICAL MEMORY TEST

Most studies of OGM have used some version of the Autobiographical Memory Test (AMT; Williams & Broadbent, 1986), in which participants are asked to produce a specific memory in response to a presented cue word within a given time limit (e.g., 30 or 60 seconds). The cue words vary in emotional valence, and most studies include positive (e.g., *happy*) and negative (e.g., *sad*) words (van Vreeswijk & de Wilde, 2004). Individuals' responses are scored as specific or non-specific, with some studies distinguishing between different forms of non-specific memories, such as categorical or extended memories. Despite the widespread use of the AMT, few studies have examined its psychometric properties. Our goal was to investigate the psychometric properties and factor structure of the AMT.

Many investigators examine valence effects when analysing AMT performance by comparing the number of specific and/or overgeneral memories recalled to positive versus negative cue words. The results of valence analyses have been inconsistent (Williams et al., 2007). Some studies have found that depressed individuals generate fewer specific and more overgeneral memories to positive cue words, compared to negative cue words (e.g., Park, Goodyer, & Teasdale, 2002). Others studies have detected the opposite pattern (e.g., Mackinger et al., 2000b). From their meta-analysis, van Vreeswijk and de Wilde (2004) concluded that, compared to non-depressed controls, individuals with depression tend to be

less specific and more overgeneral in their memories recalled to both positive and negative cue words. Within studies, effect sizes for positive and negative cue words were highly intercorrelated for both specific and overgeneral memories (van Vreeswijk & de Wilde, 2004).

Given that responses to positive and negative cues appear to correlate, it is unclear what knowledge is gained from separately examining responses to positive versus negative words. Manipulation of valence would seem to imply that positive and negative cue words elicit distinct types of responses, but this idea has not been thoroughly tested. More research is needed on certain psychometric properties of the AMT, such as the factor structure of the positive and negative words that comprise the test.

CONFIRMATORY FACTOR ANALYSIS AND ITEM RESPONSE THEORY

To investigate psychometric properties of the AMT we used confirmatory factor analysis (CFA) to determine whether AMT performance is characterised by a factor structure that reflects cue word characteristics. Given van Vreeswijk and de Wilde's (2004) findings, we hypothesised that a one-factor model of the AMT would be superior to a two-factor model that incorporated two different valences (positive versus negative). The one-factor model specified a general measure of autobiographical memory specificity, with responses to all cue words loading on a single factor. In contrast, the two-factor model had latent variables that corresponded to positive versus negative valence. To investigate the generalisability of findings we analysed three independent samples: one sample of American high school students from the Northwestern/UCLA Youth Emotion Project (YEP) and two samples of Belgian college students from the University of Leuven. We also examined whether word meaning, beyond emotional valence, affects responding to the AMT. To further examine word meaning we used the YEP sample to examine how the relevance of cue words to anxiety and depression affects memory specificity.

Another goal of this study was to examine the AMT in the context of 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). The relationship between performance on an item and a latent trait is described by a mathematical function, which is known as an item characteristic curve. In IRT, the probability of responding to an item in a particular way (e.g., providing a specific memory) is a function of the level of the latent trait. IRT and CFA can be unified in a single mathematical framework (Muthén & Muthén, 1998–2007), so we examined both the factor structure of the AMT and IRT parameters for individual items.

IRT is an item-oriented, rather than a test-oriented, approach to psychometric analysis. An IRT analysis of the AMT can provide information about how difficult it is to generate a specific memory in response to a particular cue word, as well as how well a cue word discriminates between different levels of memory specificity ability. Relatively little work has examined how cue word characteristics are related to AMT performance (but see Williams, Healy, & Ellis, 1999, for a discussion of the effects of cue word imageability). Thus an IRT analysis might offer suggestions for how to improve word selection. Because the AMT may be insufficiently sensitive to assess OGM in non-clinical samples (Raes et al., 2007), findings from an IRT analysis might also inform ways to modify the AMT so that it is more appropriate for use with non-clinical respondents.

IRT is also well suited as an analytic tool for understanding measures of autobiographical memory specificity because the underlying statistical model is in line with theories of autobiographical memory. According to Conway and Pleydell-Pearce (2000) there is a

continuous hierarchy in autobiographical memory, with representations ranging from broad themes in the life story (e.g., work, relationships), to lifetime periods (e.g., when I lived in London), to general events (e.g., psychology classes during freshman year), and to individual episodic memories (e.g., my father's 56th birthday). This continuum corresponds to different response categories on the AMT that are ordered from least-to-most specific (semantic associate, categoric memory, extended memory, specific memory; see Williams et al., 2007, p. 139). Specialised IRT models, such as Samejima's (1997) graded response model, correspond exactly to this notion that a latent trait (memory specificity) is related to ordered, categorical responses (specific, categoric, etc.). Thus, Conway and Pleydell-Pearce's theoretical conceptualisation of autobiographical memory can be effectively analysed with an IRT model.

GOALS OF THE CURRENT STUDY

The current study had two goals. First we sought to examine the structure of AMT items, and compared a one-factor model to multifactor models that incorporated information about cue word valence and relevance to anxiety versus depression. Second, we used IRT to examine characteristics of individual items on the AMT.

METHOD

Overview

To test our main hypotheses about the factor structure of the AMT, we conducted confirmatory factor analyses on three separate samples: (1) A sample of American high school students from the YEP study, (2) a sample of Belgian undergraduates from *Katholieke Universiteit Leuven* (the University of Leuven), and (3) a second sample of undergraduates from the University of Leuven who were administered one of two different versions of the AMT (between-participants).

YEP high school sample

Participants—The AMT data used in this sample derives from a large, two-site study of risk for emotional disorders. Participants in the YEP were sampled on the basis of neuroticism scores, which is thought to be a risk factor for anxiety and depression. High-scorers on neuroticism were over-sampled (for further details about the sample, see Griffith et al., 2009; Zinbarg et al., 2009). A subset of the overall sample ($N=627$) was randomly allocated to complete the AMT and other cognitive tasks. The participants ($N=333$) ranged in age from 16 to 18 years ($M=17.1$, $SD=0.4$), and were predominantly female (69%).

Materials and procedure—Experimenters administered the AMT to participants individually. Each stimulus word was printed in a booklet and presented one at a time. The 16 test words were divided into four types: positive words relevant to depression, negative words relevant to depression, positive words relevant to anxiety, and negative words relevant to anxiety (see Table 1 for the list of words). The words were administered in a fixed order that alternated among the four types. Words were culled from previous AMT studies (e.g., Park et al., 2002) and were equated in terms of length and usage frequency (Kučera & Francis, 1967). AMT interviewers were trained undergraduates, graduate students, and Bachelors-level research assistants.

The AMT interviewer administered practice items until the participant provided two specific memories without requiring feedback. The experimenter then presented the test words. Participants described their memories verbally within a 30-second time window. Interviewers classified each trial as a specific, extended, or categoric memory, semantic

associate, or omission/failed trial. No feedback was given on test trials. For inter-rater reliability, raters trained in the AMT listened to recordings of a subset of interviews. Kappa was .78 for both within-site ($n=73$) reliability, and for cross-site reliability ($n=37$).

A trial was coded as specific if the memory was of an event that occurred at a particular time and place within the course of one day. An example of a specific memory from the sample for the word *calm* is “I felt calm on the evening of the Fourth of July as the workday was winding down and I was getting ready to have a party that night.” A memory was coded as *extended* if it was of an event that occurred at a specific time and place, but lasted more than 1 day (e.g., “I felt calm over the summer”). A memory was coded as *categoric* if it described a summary or class of past events without indicating a specific time and place (e.g., “I feel calm whenever I watch a movie”). A trial was coded as a *semantic associate* if the person provided a verbal response within the allotted time, but the response was derived from general semantic knowledge rather than a memory (e.g., “I might feel calm when travelling to places that I haven’t been to before”). Finally, a trial was coded as an *omission/failed trial* if the participant could not produce a response within the allotted 30 seconds.

University of Leuven Dataset #1

Participants—Participants were all first-year psychology students who participated in return for course credit. The sample consisted of 405 participants (81% female). The mean age of the sample was 18.4 years ($SD=1.4$; $range=17-36$).

Materials and procedure—An experimenter administered a written version of the AMT to a group of participants (see Raes, Hermans, de Decker, Eelen, & Williams, 2003). Participants wrote their responses in a booklet that contained 12 cue words; each word was printed on a separate page. The first two cue words were practice items, and the test words consisted of five positive and five negative emotion words, strictly alternating with valence (see Table 2 for the list of words).

The experimenter read the cue words aloud, and instructed all participants to write down one specific memory for each cue word. Participants were asked to provide a memory that was at least 1 week old. All participants were given 60 seconds to respond, and were instructed to turn to the following cue word when this time had elapsed. Experimenters coded each response as specific or non-specific, and non-specific responses were further coded as categoric memories, extended memories, semantic associates, omissions, same events (referring to an event already mentioned), or incorrect specifics (referring to an event of the past week).¹ Using this scoring procedure, previous studies obtained good reliability (Raes et al., 2003, 2004), with inter-rater agreement ranging from 92% to 99% ($kappa=.83$ to $.96$).

University of Leuven Dataset #2

Participants—Participants were all first-year psychology students who participated in return for course credit. The sample consisted of 336 participants (79% female), with a mean age of 18.1 years ($SD=1.3$; $range=17-30$). All participants were different people from those participating in the previous study. Further details about the participants can be obtained from Debeer, Hermans, and Raes (2009).

Materials and procedure—An experimenter administered the same written version of the AMT as for the University of Leuven Dataset #1 to groups of participants. Each group

¹We classified some trials as specific/extended if a memory was most likely specific but could be interpreted as an extended event. An example of such a memory would be “When my grandfather died.” We adopted a stringent approach to scoring a memory as overgeneral by coding these as specific memories.

received one of two different sets of instructions. Approximately half of the participants were given the traditional AMT instructions, in which individuals were directed to write down one unique specific memory for each cue word (*Traditional Instructions Group*; $n=162$). The other half received a minimal set of instructions, in which participants were asked to generate unique memories without explicitly stating that the memories should be specific (*Minimal Instructions Group*; $n=174$).

Experimenters coded each trial as a specific, extended, or categoric memory, semantic associate, or omission/failed trial (as in the previous two datasets), and participants also rated their responses as a means to check the experimenters' coding (see also Raes et al., 2007, p. 500). Experimenters consulted the self-reported code when clarification was necessary.

RESULTS

Descriptive statistics

Tables 1–3 show the percentages of each response category in the YEP sample, University of Leuven Dataset #1, and University of Leuven Dataset #2. Most participants provided specific memories in response to each item when given the traditional AMT instructions. As shown in Table 3, participants who received the traditional AMT instructions retrieved more specific memories than those who received the minimal instructions, $F(1, 334)=146.67$, $p=.00$, Cohen's $d=1.3$ (Cohen, 1988). For more details about the effects of different AMT instructions, see Debeer et al. (2009).

Confirmatory factor analyses

In each factor analysis described below, the variance of each factor was set to 1.0 so that the loading of each item could be freely estimated. Each item was treated as an ordered categorical indicator in Mplus, version 5.0 (Muthén & Muthén, 1998–2007). We used probit regression assumptions to assess the relationship between factors and indicators. Further details about the analyses are presented below in the *Item response theory analyses* section. Because our indicators were ordinal categories, a mean- and variance-adjusted weighted least squares chi-square was used as the estimator (for a discussion of the WLSMV estimator, see Muthén & Muthén, 1998–2007). We used a *comparative fit index* (CFI; Bentler, 1988; Hu & Bentler, 1999) and a *root mean square error of approximation* (RMSEA; Browne & Cudeck, 1993) to assess the fit of each model. We used cut-offs as a guide to model interpretation by seeking models with $CFI > .95$ and $RMSEA < .06$, as suggested by Hu and Bentler, as well as models with nonsignificant χ^2 statistics. However, we did not necessarily reject models outright if certain fit indices were slightly outside these cut-offs (for a discussion of the use of cut-offs for fit indices, see Marsh, Hau, & Wen, 2004).

YEP high school sample—In this dataset we examined four models. To investigate our main hypothesis that positive and negative cue words both related to one memory specificity factor, we compared a one-factor model to a two-factor oblique model in which the eight negative cue words loaded onto one factor and the eight positive cue words loaded onto another factor. We also compared a one-factor model to a two-factor oblique model with depression- and anxiety-word factors. Finally, we compared the one-factor model to a four-factor oblique model in which one factor was created for each cue word type: positive-depression, negative-depression, positive-anxiety, and negative-anxiety. The four items within that cue word type were used as indicators.

One participant had one missing item due to experimenter error, but was included in the analyses because Mplus provides maximum likelihood estimates for parameters when data are missing at random. As shown in Table 1, some items elicited semantic associates infrequently. Thus we collapsed semantic associates and omissions/failed trials into one category, which represented failure to retrieve a memory. The four categories in the analyses were specific, extended, categoric, and failed.

The one-factor model provided an excellent fit to the observed data as indicated by $CFI=.98$, $RMSEA=.02$, and a non-significant $\chi^2(56, N=333)=63.30, p=.23$. Standardised factor loadings ranged from .43 to .69. We attempted to fit two two-factor models (one with positive-and negative-word factors, and the other with depression- and anxiety-word factors) and a four-factor model, but they could not be empirically identified. Specifically, the estimated correlations between the latent variables exceeded 1.0 in all the models.² We accepted the one-factor model as the final model.

University of Leuven Dataset #1—The base rate of semantic associates was also low in this data set (see Table 2), so we combined these responses with omissions/failed trials. In addition, the base rate of extended memories was low, with a mean percentage of 1.0% across participants. Therefore categoric and extended memories were combined into one category.³

The one-factor model again provided an excellent fit to the data as indicated by $CFI= 1.0$, $RMSEA=.00$, and a non-significant $\chi^2(25, N=405)=23.48, p=.55$. Standardised factor loadings ranged from .35 to .68. A two-factor model with negative- and positive-word factors provided an almost identical fit, $CFI=1.0, RMSEA=.00$, and this model did not fit better than the one-factor model, $\chi^2(1)=1.20, p=.27$. The correlation between the two factors was .91. Because it was parsimonious and fitted well, the one-factor model was accepted as the final model.

University of Leuven Dataset #2—This dataset also contained cue words of positive and negative valence, so we compared one- and two-factor models. As with the other two datasets, semantic associate responses were collapsed with omissions/failed trials. Unclassifiable responses were recoded into missing data.⁴ A one-factor model provided an excellent fit to the data as indicated by $CFI=.97, RMSEA=.03$, and a non-significant $\chi^2(28, N=336)=37.92, p>.09$. Standardised factor loadings ranged from .41 to .60. A two-factor model had almost identical fit indices, $CFI=.96$ and $RMSEA=.03$, and this model did not fit better than the one-factor model, $\chi^2(1)31.05, p=.31$. In this model the correlation between the negative-word factor and the positive-word factor was .93. We accepted the one-factor model as the final model because it provided as good a fit to the data as the two-factor model.

Because participants received two different forms of instructions, we attempted to fit separate models using multiple-group confirmatory factor analysis. Because we expected the item parameters to differ across the two AMT instruction groups, we tested a model with configural invariance, which allows item parameters to vary across groups. The within-

²We attempted to overcome this convergence problem by specifying in Mplus that the correlation between the factors could be no larger than .71, which would mean that the two factors shared no more than 50% of their variance. In other words, this parameter was still free to vary, but had a ceiling of .71. Although the two-factor models had one more free parameter than the one-factor model, both of the two-factor models had *worse* fit indices. Other attempts were made to re-specify the two- and four-factor models. However, we were unable to fit a sensible model that was superior to the one-factor model. Further details are available from the first author.

³One item, *verrast (surprised)*, elicited very low rates of extended and categoric memories. We dichotomised this into specific versus non-specific.

⁴As shown in Table 3, the rate of extended memories was low for three items: *verrast (surprised)*, *lomp (stupid)*, and *brutaal (bold)*. For these items, we combined categoric and extended memories into one category.

group samples were relatively small for CFA ($n=162$ for the traditional instructions, $n=174$ for minimal instructions), so the models were simplified to reduce the number of parameters to be estimated. Each item on the AMT was treated as a dichotomy: specific versus non-specific. The one-factor model fitted well across groups according to an $RMSEA < .04$ and a non-significant $\chi^2(50, N=336)=61.03, p=.14$, but the CFI of .89 was below recommended cut-offs for finding good-fitting models.

In the configural invariant model the factor loadings for bang (*scared*) and prettig (*pleasurable*) were below .3 for the minimal instructions group. Because the minimal instructions condition was an exploratory method, we dropped these two items to investigate whether a better fit could be obtained. After we dropped these items, the one-factor configural invariant model had an excellent fit as indicated by a CFI of 1.0, an $RMSEA$ of .00, and a non-significant $\chi^2(32, N=336)=30.62, p=.54$. Standardised factor loadings ranged from .38 to .62 in the original instructions group and from .30 to .57 in the minimal instructions group.

The configural invariant model was compared to a metric invariant model, in which all parameters were constrained to be equal across groups. As expected, the metric invariant model fitted poorly, and had a significantly worse fit than the configural invariant model, $\chi^2(12, N=336)=120.59, p=.00$. In summary, the multiple-group CFA indicated that the instructions used in the AMT affect the psychometric properties of the test. More detail about how the instructions affected the psychometric properties of the AMT is presented below in the *Item Response Theory Analyses* section.

Item response theory analyses

IRT parameters were derived from a normal-ogive graded response model (Samejima, 1997). Specifically, each observed item was related to the latent trait of *autobiographical memory specificity* by a probit regression analysis. These analyses yield two types of parameters: item-slopes (sometimes referred to as discrimination in IRT parlance) and thresholds (sometimes referred to as difficulty). An item-slope relates the item to a latent trait, and it captures the ability of an item to discriminate between people who are high and low on the latent trait being studied. Assuming equal thresholds, participants who retrieve specific memories in response to highly discriminating items are higher on memory specificity than participants who retrieve specific memories in response to items with lower discrimination parameters.

The second type of parameter is a threshold. Thresholds are placed along a standardised continuum in our analyses, and the number of thresholds is equal to the number of categories minus 1. If a person's level of memory specificity exceeds a particular threshold, then he or she would be more than 50% likely to respond to that item with one of the response categories above the threshold (Embretson & Reise, 2000, pp. 98–99). Thus, the threshold conveys information about how difficult it is for individuals to generate a particular response to a cue word.

YEP high school dataset—As shown in Table 4, all of the items on the AMT had substantive item-slopes of .43 or larger. Responses to each item were classified into one of four categories. Therefore, three thresholds were modelled for each item. All of the thresholds in Table 4 were lower than zero. The fact that Threshold 3 (the threshold between specific and extended memories) was below zero for all of the items indicated that, on average, each item was more likely than not to elicit specific memories, even in people with below-average memory specificity ability.

We also examined a test information function, which plots information and standard error of measurement as a function of examinee ability (see Figure 1). Information, in the IRT sense, refers to the degree to which an item or test contributes to the estimation of ability. This plot showed that the AMT was most informative (i.e., was most precise) for individuals with a trait value of -1.7 (standardized units) on autobiographical memory specificity. Test information functions are derived by examining the summed contribution of each item (for further details, see Embretson & Reise, 2000; Hambleton et al., 1991). These functions are model based, and therefore assume that the underlying statistical model applies to the data in question. As shown by the CFA, our underlying statistical model fit these data very well.

To illustrate information yielded by IRT analyses, Figure 2 and 3 present item characteristic curves for two AMT cues: *energetic* and *lonely*. Table 4 indicates that *energetic* had the largest item-slope (.69) of all of the cues in this AMT. This can be seen in Figure 2 by the sharp increase in the probability of retrieving a specific memory as autobiographical memory specificity increases. In Figure 3, the relationship between the latent autobiographical memory specificity trait and the probability of retrieving a specific memory was less steep than for *energetic*, which is consistent with the lower item-slope for *lonely* (.49).

The thresholds presented in Table 4 are related to the position of the curves along the latent dimension of autobiographical memory specificity. The two items presented in Figure 2 and 3 had similar thresholds: $-.89$ and $-.86$, respectively. Thus, the two curves are positioned at approximately the same location. An item that elicited specific memories more easily, such as *happy*, would have a lower threshold, and therefore a curve shifted to the left on autobiographical memory specificity (graph not shown).

University of Leuven Dataset #1—Table 5 shows IRT parameters. Inspection of the test information function (see Figure 4) indicated that this AMT was maximally precise for individuals whose true score was two standard deviations below the mean on autobiographical memory specificity. Item-slopes were all .35 or greater, and the values for the threshold distinguishing overgeneral from specific responses (Threshold 2) were all negative, which suggests that each item was more likely than not to elicit specific memories among people with average ability. In fact, for each item except *moed* (*courage*) and *gerust* (*calm/at ease*), Threshold 2 was less than -1.0 . For most items, a specific memory was more likely than not, even for people scoring one standard deviation below the mean on autobiographical memory specificity.

University of Leuven Dataset #2—Because of low base rates for individual types of non-specific memories, we only modelled thresholds that distinguished specific from non-specific responses in this dataset. As shown in Table 6, the thresholds in the traditional instructions group were all negative, and these low thresholds were consistent with the findings from the YEP data and Leuven Dataset #1. The thresholds for retrieving a specific memory were higher for items presented with the minimal instructions than with the traditional instructions. For example, *gerust* (*calm/at ease*) had a threshold of $-.47$, which indicates that participants scoring higher than $-.47$ on the autobiographical memory specificity trait would be more likely than not to generate a specific memory when presented with the traditional instructions. However, for the minimal instructions, the threshold for *gerust* was .45. Therefore, participants would need to be approximately half a standard deviation above the mean on memory specificity to be more likely than not to retrieve a specific memory in response to this word when given the minimal instructions.

Differences between the two versions of the AMT are also apparent in Figure 5, which shows the test information function and standard error of measurement for the two versions

of the AMT. The peak of the information curve for the minimal instructions AMT was farther to the right on the autobiographical memory specificity trait than the peak for the AMT with the traditional instructions. This pattern indicates that the minimal instructions version of the AMT was more precise with individuals who score higher on autobiographical memory specificity. The traditional instructions AMT was maximally precise for participants 1.5 standard deviations below the mean on autobiographical memory specificity, whereas the minimal instructions AMT was maximally precise for participants scoring near the mean (maximum information was provided at a location of -0.1 on memory specificity). Figure 5 suggests that standard error of measurement begins to increase dramatically for people more than three standard deviations above the mean for the traditional instructions AMT and increases to a lesser extent for people more than three standard deviations below the mean for the minimal instructions AMT.

DISCUSSION

As predicted, we found that a single trait of autobiographical memory specificity seems to account for responses on the AMT. Across three samples, a one-factor model was characterised by a good fit, and this fit did not differ significantly from the fits of models that took cue word characteristics into account. Because all of the items in the AMT were indicative of a single construct, the best measure of autobiographical memory specificity is one that incorporates information from all of the items. Investigators should therefore be cautioned that interpretation of so-called valence effects might not be due to the emotional associations of words, but rather due to idiosyncratic psychometric properties of items.

In addition, IRT analyses demonstrated that when the traditional instructions are used, the AMT measures autobiographical memory specificity most precisely for individuals who are low on this trait. When minimal instructions were given, the AMT elicited more overgeneral memories and was most precise for people of average ability on the latent trait of memory specificity. Thus the AMT, as it is traditionally implemented, may have limited utility for discriminating individuals across a wide range of autobiographical memory specificity ability.

As described below, our IRT analyses show that item characteristics can vary within a cue word type, even if the words have been equated in the usual ways (e.g., matching on length and usage frequency). This finding is not surprising, and it has implications for examining valence effects. Researchers who find a main effect of valence may need to consider the possibility that their findings are due to idiosyncrasies in a particular pool of words. If an easy (in an IRT sense) set of negative cue words were compared to a difficult set of positive cue words, then one might find a main effect of word type. Consequently, it would be tempting to conclude that the negative content of the cue words elicited more specific memories. However, the possibility that the findings were due to differential cue word characteristics, rather than to the emotional content of the word, could not be ruled out.

To make a strong case for valence effects, researchers would need to examine a large comprehensive corpus of words in a large sample of participants. IRT could be employed to examine characteristics of individual cues. Different sets of words (e.g., positive and negative words) could then be equated on item-slopes and thresholds, which subsequently could provide a fair test of valence effects. Valence effects are inconsistent in the literature, so analyses that rule out alternative explanations, such as psychometric differences across positive and negative word sets, would be a valuable addition to the OGM literature.

For each item in the three datasets using the traditional AMT instructions, the threshold that separated specific and non-specific memories was below zero on the latent trait of

autobiographical memory specificity. Thus, each cue word was more likely than not to elicit a specific memory among people of average ability. This is also true for people of below average ability, to a point, because the highest threshold obtained in the samples that received the traditional instructions was $-.33$ (see Tables 4–6.). Thus individuals who are low on autobiographical memory specificity may perform well on an AMT that uses the traditional instructions. Indeed, Raes et al. (2007) noted the low frequency of overgeneral memories retrieved by non-clinical samples with the traditional AMT. Consequently, the traditional AMT may be less apt to find differences among samples with better memory functioning. Our IRT results confirm this observation, and offer suggestions for future research with the AMT.

The results of IRT analyses would be helpful in piloting versions of the AMT for use with various samples. As one example, the threshold for specific memories in response to *happy* in the YEP dataset was low at a value of -1.40 on the latent trait (see Table 4). Therefore this item may be too easy (in an IRT sense) to help distinguish between individuals who are high and low on autobiographical memory specificity. In contrast, *calm* had a higher threshold of $-.60$. Although it is below zero, it may be a more useful item in AMT research because it should elicit more diverse responses. Examination of item characteristics curves, such as in Figure 2 and 3, can help to select items that are likely to elicit specific response types that may be of special interest in depression research, such as categoric memories.

The IRT analyses also have implications for how instructions are worded. As seen in Table 6, the threshold that separates specific and non-specific memories is larger for the minimal instructions group compared to the traditional instructions group. Thus, the minimal instructions version of the AMT may be more useful in samples expected to have average memory functioning, such as college students, whereas the traditional instructions version may be useful in samples with poor memory functioning, such as severely depressed participants. This difference helps elucidate why the AMT is consistently related to depression in clinical samples, but less so in high functioning samples (e.g., Debeer et al., 2005; Raes et al., 2004). However, because the minimal instructions AMT does not explicitly require participants to generate a specific memory, it is possible that some respondents may exhibit a more overgeneral style of reporting memories on this test, even though they would be able to provide a specific memory if prompted to do so. Thus, the two versions of the AMT may be measuring different constructs. Within-participants research using both versions of the AMT would be useful in answering this question.

Limitations

A low rate of non-specific memories was obtained in our three samples of students. Like other investigators (e.g., Brittlebank et al., 1993; Scott, Williams, Brittlebank, & Ferrier, 1995), we were forced to collapse certain response categories in order to analyse our data. A second limitation is that our word sets were limited. Although we failed to find differences across cue word types, we cannot rule out the possibility that our findings are specific to the small corpus of words that we examined. It is possible that other sets of positive, negative, or other words would yield different results. A third limitation of this study is that depression was not examined as it relates to the structure of the AMT. The YEP dataset contained diagnostic and self-report measures of depression (not used in the current paper), but there were far too few cases of major depressive disorder ($n=58$) to factor analyse them as a separate group. IRT analyses yield statistics that are relevant to the properties of a test, but it is unknown whether IRT analyses would yield different results in a clinical sample. Our results may not generalise to clinical samples, but IRT analyses using clinical samples would be a rich direction for future research. A different factor structure for positive and negative AMT cue words might be found in a clinical sample. Finally, our three datasets

contained adolescents and young adults. Future studies should extend this research to adult samples.

Future directions

Some investigators have examined the effects of cue word characteristics other than valence on autobiographical memory specificity. For example, Dalgleish et al. (2003) hypothesised that the effects of cue words on the AMT are due more to word meaning than to valence. Recent studies have begun to investigate this idea. At least three studies have shown that cues that are relevant to an individual's concerns and self-concept are more likely to elicit non-specific memories than cues without such self-relevance (Barnhofer, Crane, Spinhoven, & Williams, 2007; Crane, Barnhofer, & Williams, 2007; Spinhoven, Bockting, Kremers, Schene, & Williams, 2007). Thus, characteristics of AMT cues may interact with individual difference variables (e.g., rumination) or idiosyncratic concerns. Future studies should examine whether certain cues perform differently in certain subgroups. In the IRT literature, these analyses are referred to as analyses of *differential item functioning* (Embretson & Reise, 2000). These analyses may yield interesting clinical insights, but such studies would require a range of cues and adequate variation in the traits being studied.

In addition to research on personal relevance, Williams et al. (1999) showed that highly imageable cue words elicit specific memories more often than less imageable cue words. Although studies of mean differences are informative, IRT allows for the examination of specific words within a cue type. Thus IRT can help to explore the effects of individual cue words, and to plan autobiographical memory studies by selecting discriminating cue words with thresholds that are appropriate to the population being studied.

Another possibility for future research would be to conduct a large-scale investigation of cue words to find cues that are maximally informative for different populations. Different sets of cue words could be identified that are discriminating for individuals in the range of the autobiographical memory specificity trait that is most relevant for a particular population, such as college students or patients with major depression. Such a project could establish a common word set to be used in studies with similar individuals, thus facilitating comparison across different studies.

Summary and conclusions

To our knowledge, we are the first investigators to use CFA and IRT to examine the psychometric properties of the AMT. Our findings indicate that a one-factor model of autobiographical memory specificity provides a good conceptualisation of AMT performance, at least in non-clinical samples. This result is consistent with the finding that responses to positive and negative cue words are highly intercorrelated (van Vreeswijk & de Wilde, 2004). Additionally, a one-factor model of the AMT is congruent with the notion that overgenerality develops as an overall response style over time, as proposed by the functional avoidance hypothesis (Williams et al., 2007).

The findings of the current study also demonstrate that the AMT is not maximally informative for individuals at all levels of autobiographical memory specificity. Rather, certain characteristics of the AMT, such as particular cue words and the nature of the instructions, may influence the types of responses obtained. These results offer a number of suggestions for ways to modify the AMT to obtain the most relevant outcomes for a particular study. Future AMT studies may be well served by examining IRT parameters for individual cues.

Acknowledgments

We would like to thank the National Institutes of Health for supporting our research (Grant# R01MH065652 to Drs Zinbarg and Mineka, and R01MH065651 to Dr Craske), and we acknowledge the assistance of countless undergraduate and graduate students for their help in collecting and processing our data. We would also like to thank an anonymous reviewer for his or her helpful comments.

References

- Barnhofer T, Crane C, Spinhoven P, Williams JMG. Failure to retrieve specific memories in previously depressed individuals: Random errors or content-related? *Behaviour Research and Therapy*. 2007; 45:1859–1869. [PubMed: 17399680]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1988; 107:238–246. [PubMed: 2320703]
- 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]
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural models*. Newbury Park, CA: Sage Publications; 1993.
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1988.
- Conway MA, Pleydell-Pearce CW. The construction of autobiographical memories in the self-memory system. *Psychological Review*. 2000; 107:261–288. [PubMed: 10789197]
- Crane C, Barnhofer T, Williams JMG. Cue relevance affects autobiographical memory specificity in individuals with a history of major depression. *Memory*. 2007; 15:312–323. [PubMed: 17454667]
- Dalgleish T, Tchanturia K, Serpell L, Hems S, Yiend J, de Silva P, et al. Self-reported parental abuse relates to autobiographical memory style in patients with eating disorders. *Emotion*. 2003; 3:211–222. [PubMed: 14498792]
- Debeer, E.; Hermans, D.; Raes, F. Unpublished raw data. 2005. Autobiographical memory specificity and depression in first-year psychology students.
- Debeer, E.; Hermans, D.; Raes, F. Associations between components of rumination and autobiographical memory specificity as measured by a Minimal Instructions Autobiographical Memory Test. 2009. Manuscript submitted for publication
- Embretson, SE.; Reise, SP. *Item response theory for psychologists*. London: Lawrence Erlbaum Associates Ltd; 2000.
- 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]
- Griffith, JW.; Zinbarg, RE.; Craske, MG.; Mineka, S.; Rose, RD.; Waters, AM., et al. Neuroticism as a common dimension in the internalizing disorders. 2009. Manuscript submitted for publication
- Hambleton, RK.; Swaminathan, H.; Rogers, HJ. *Fundamentals of item response theory*. New-bury Park, CA: Sage Publications; 1991.
- Hu L, Bentler PM. Cut-off criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Kučera, H.; Francis, N. *Computational analysis of present-day American English*. Providence, RI: Brown University Press; 1967.
- Mackinger HF, Loschin GG, Leibetseder MM. Prediction of postnatal affective changes by autobiographical memories. *European Psychologist*. 2000a; 5:52–61.
- Mackinger HF, Pachinger MM, Leibetseder MM, Fartacek RR. Autobiographical memories in women remitted from major depression. *Journal of Abnormal Psychology*. 2000b; 109:331–334. [PubMed: 10895571]
- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cut-off values for fit indices and dangers in overgeneralising Hu and Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 5. Los Angeles: Muthén & Muthén; 1998–2007.

- Park RJ, Goodyer IM, Teasdale JD. Categorical overgeneral autobiographical memory in adolescents with major depressive disorder. *Psychological Medicine*. 2002; 32:267–276. [PubMed: 11866322]
- Raes F, Hermans D, de Decker A, Eelen P, Williams JMG. Autobiographical memory and affect regulation: An experimental approach. *Emotion*. 2003; 3:201–206. [PubMed: 12899419]
- 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. Unpublished manuscript. 2004. Correlates of autobiographical memory specificity in a non-clinical student population.
- Ramponi C, Barnard PJ, Nimmo-Smith I. Recollection deficits in dysphoric mood: An effect of schematic models and executive mode? *Memory*. 2004; 12:655–670. [PubMed: 15615322]
- Rekart KN, Mineka S, Zinbarg RE. Autobiographical memory in dysphoric and non-dysphoric college students using a computerised version of the AMT. *Cognition and Emotion*. 2006; 20:506–515.
- Samejima, F. Graded response model. In: van der Linden, WJ.; Hambleton, RK., editors. *Handbook of modern item response theory*. London: Springer; 1997. p. 85-100.
- Scott J, Williams JMG, Brittlebank A, Ferrier IN. The relationship between premorbid neuroticism, cognitive dysfunction and persistence of depression: A 1-year follow-up. *Journal of Affective Disorders*. 1995; 33:167–172. [PubMed: 7790668]
- Spinhoven P, Bockting CLH, Kremers IP, Schene AH, Williams JMG. The endorsement of dysfunctional attitudes is associated with an impaired retrieval of specific autobiographical memories in response to matching cues. *Memory*. 2007; 15:324–338. [PubMed: 17454668]
- 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]
- 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]
- Williams JMG, Dritschel BH. Emotional disturbance and the specificity of autobiographical memory. *Cognition & Emotion*. 1988; 2:221–234.
- Williams JMG, Healy HG, Ellis NC. The effect of imageability and predictability of cues in autobiographical memory. *The Quarterly Journal of Experimental Psychology*. 1999; 52A:555–579. [PubMed: 10504900]
- Zinbarg, RE.; Mineka, S.; Craske, MG.; Griffith, JW.; Sutton, J.; Rose, RD., et al. The Northwestern-UCLA Youth Emotion Project: Associations of cognitive vulnerabilities, neuroticism and gender with past diagnoses of emotional disorders in adolescents. 2009. Manuscript submitted for publication

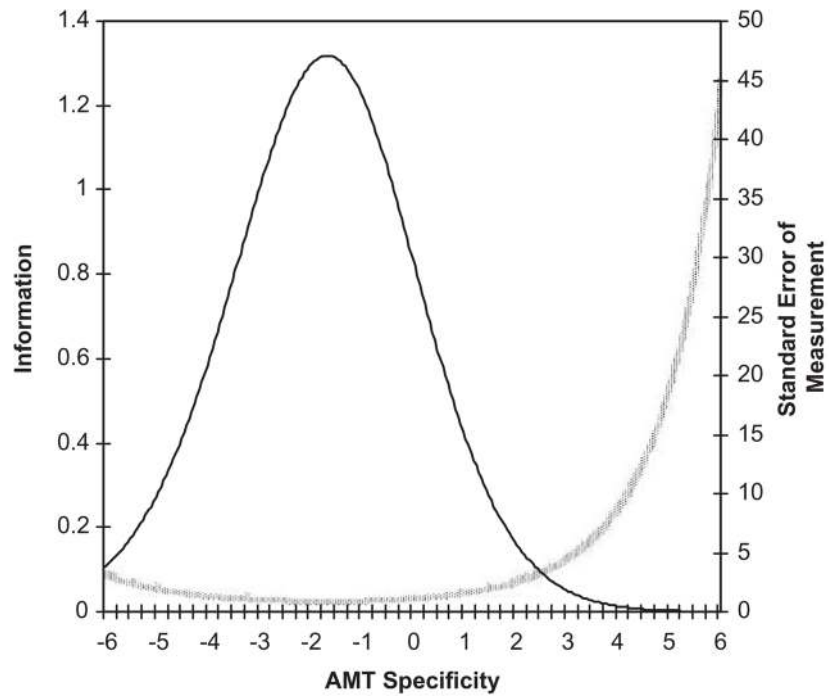


Figure 1. Test information function and standard error of measurement (SEM) for the AMT in the Youth Emotion Project. AMT specificity and SEM are measured on a standardised scale. Information, on the ordinate, is the precision of the test at a particular point, measured in a squared metric.

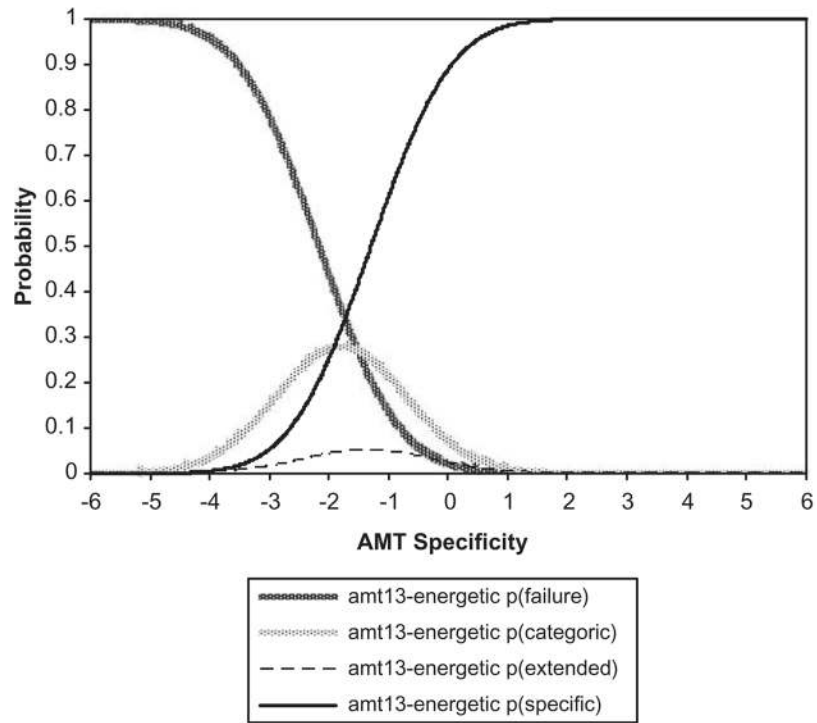


Figure 2. Item characteristic curve for the AMT cue energetic. AMT specificity is on a standardised scale. The ordinate is the probability of responding in a given category as a function of true score on memory specificity.

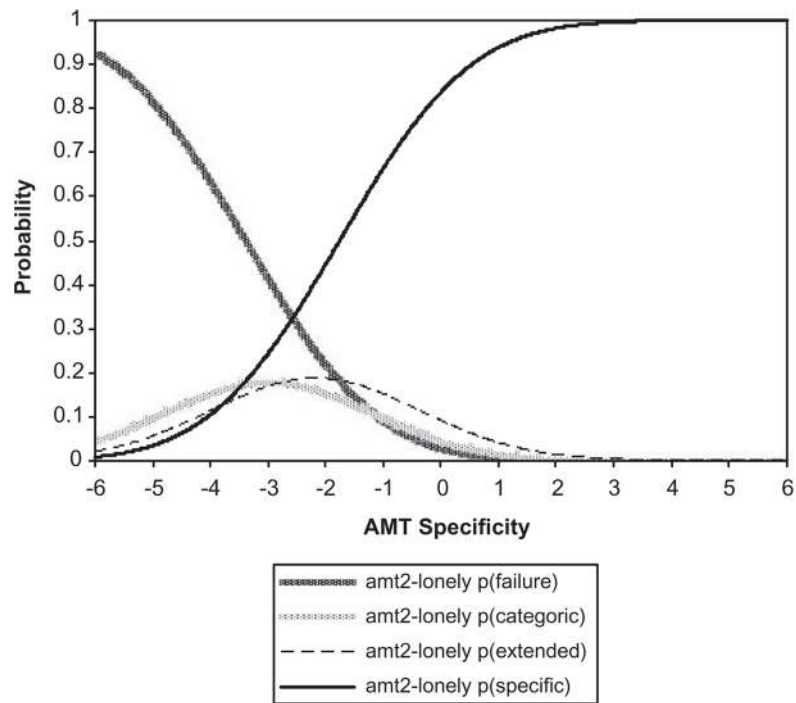


Figure 3. Item characteristic curve for the AMT cue lonely. AMT specificity is on a standardised scale. The ordinate is the probability of responding in a given category as a function of true score on memory specificity.

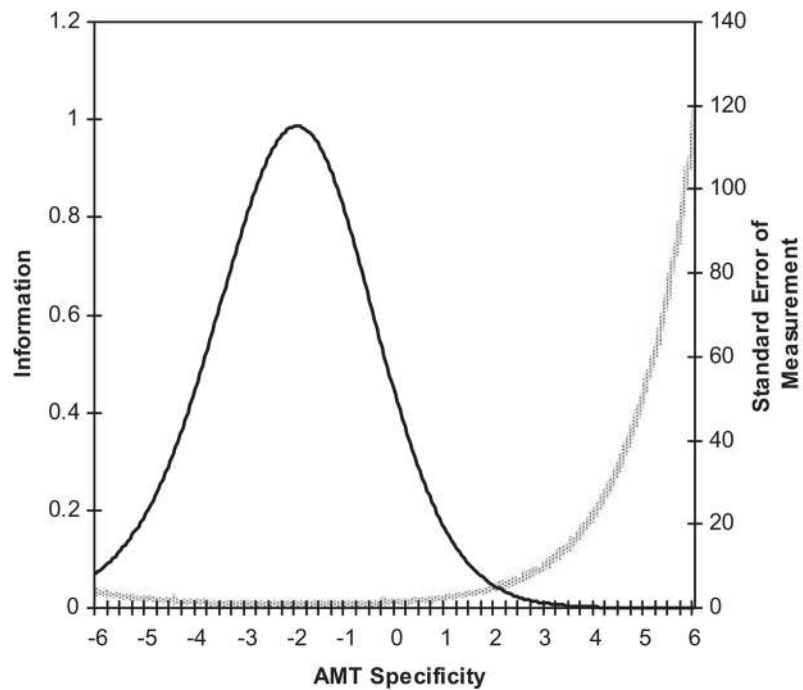


Figure 4. Test information function and standard error of measurement (SEM) for the AMT in Leuven Dataset #1. AMT specificity and SEM are measured on a standardised scale. Information, on the ordinate, is the precision of the test at a particular point, measured in a squared metric.

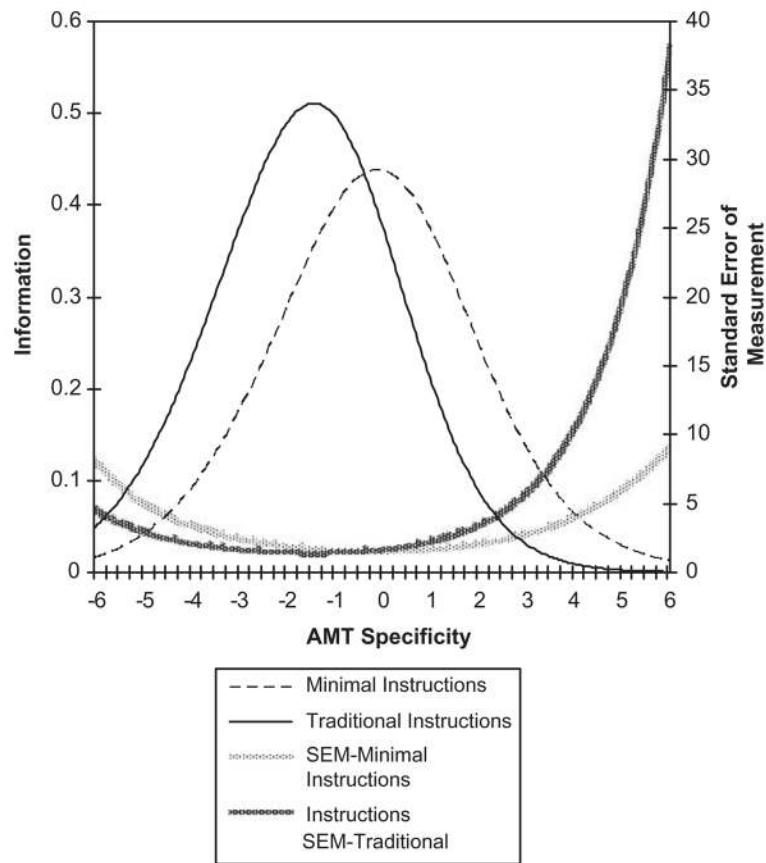


Figure 5.

Test information function and standard error of measurement for the traditional instructions and minimal instructions AMTs. SEM=Standard error of measurement. AMT specificity and SEM are measured on a standardised scale. Information, on the ordinate, is the precision of the test at a particular point, measured in a squared metric.

TABLE 1

Percentage of response type for AMT items: YEP sample

Word type	Response type			
	Omission/failed	Semantic associate	Categoric	Specific
Positive Depression-Relevant				
Energetic	6.3%	0.3%	9.6%	81.4%
Happy	1.5%	1.5%	2.4%	91.9%
Loved	9.9%	1.5%	8.7%	76.0%
Successful	5.4%	0.9%	3.6%	88.3%
Positive Anxiety-Relevant				
Brave	23.7%	1.2%	3.9%	69.4%
Calm	10.5%	2.1%	11.1%	72.7%
Comfortable	6.3%	3.9%	10.2%	76.9%
Safe	18.0%	0.3%	14.7%	63.1%
Negative Depression-Relevant				
Failure	6.9%	1.2%	3.9%	85.0%
Hopeless	18.3%	1.8%	4.2%	66.1%
Lonely	3.9%	0.9%	5.4%	80.5%
Sad	10.8%	0.9%	3.9%	82.0%
Negative Anxiety-Relevant				
Afraid	6.0%	1.8%	3.6%	86.5%
Angry ^a	7.8%	0.6%	6.0%	83.7%
Tense	7.5%	2.7%	7.8%	78.1%
Worried	10.8%	3.3%	7.2%	72.7%

N=333.

^a One participant did not complete this item due to experimenter error.

TABLE 2

Percentage of response type for AMT items: University of Leuven Dataset #1

	Omission/failed	Semantic associate	Categorie	Extended	Specific
<i>Positive words</i>					
Vertrouwen (Confidence/trust)	6.4%	1.2%	0.2%	1.2%	90.9%
Prettig (Pleasurable)	10.4%	0.0%	1.5%	2.2%	85.9%
Moed (Courage)	9.9%	2.7%	2.0%	2.5%	83.0%
Genust (Calm/at ease)	16.3%	0.5%	4.2%	1.7%	77.3%
Verrast (Surprised)	6.4%	0.7%	0.2%	0.0%	92.6%
<i>Negative words</i>					
Bang (Scared)	4.4%	1.0%	1.7%	0.7%	92.1%
Boos (Angry)	8.1%	0.7%	3.2%	0.0%	87.9%
Droevig (Sad)	5.9%	1.0%	2.0%	1.5%	89.6%
Brutaal (Bold)	9.1%	1.0%	2.7%	0.2%	86.9%
Lomp (Stupid)	10.6%	0.7%	1.7%	0.0%	86.9%

N=405.

TABLE 3

Percentage of response type for AMT items: University of Leuven Dataset #2

	Unclassifiable	Omission/failed	Categoric	Extended	Specific
<i>Original instructions</i>					
<i>Positive words</i>					
Vertrouwen (Confidence/trust)	3.1%	12.3%	2.5%	3.7%	78.4%
Prettig (Pleasurable)	2.5%	14.2%	4.3%	3.7%	75.3%
Moed (Courage)	3.7%	22.8%	4.3%	4.3%	64.8%
Gerust (Calm/at ease)	1.2%	17.3%	11.1%	3.1%	67.3%
Verrast (Surprised)	1.2%	7.4%	1.2%	0.0%	90.1%
<i>Negative words</i>					
Bang (Scared)	1.2%	7.4%	4.3%	2.5%	84.6%
Boos (Angry)	1.9%	11.7%	8.6%	5.6%	72.2%
Droevig (Sad)	0.0%	9.9%	3.1%	10.5%	76.5%
Brutaal (Bold)	2.5%	15.4%	7.4%	1.2%	73.5%
Lomp (Stupid)	2.5%	9.3%	3.1%	1.2%	84.0%
<i>Minimal instructions</i>					
<i>Positive words</i>					
Vertrouwen (Confidence/trust)	3.4%	19.5%	14.9%	14.9%	47.1%
Prettig (Pleasurable)	0.6%	7.5%	30.5%	21.3%	40.2%
Moed (Courage)	1.7%	21.3%	19.5%	14.9%	42.5%
Gerust (Calm/at ease)	1.1%	17.8%	37.9%	10.9%	32.2%
Verrast (Surprised)	0.6%	8.6%	3.4%	0.6%	86.8%
<i>Negative words</i>					
Bang (Scared)	0.6%	9.2%	20.1%	12.6%	57.5%
Boos (Angry)	1.1%	12.1%	26.4%	12.6%	47.7%
Droevig (Sad)	0.6%	7.5%	19.0%	21.3%	51.7%
Brutaal (Bold)	0.0%	16.7%	28.7%	8.0%	46.6%
Lomp (Stupid)	1.1%	16.1%	27.0%	1.1%	54.6%

N=336. The omission/failed category includes non-memory responses, such as semantic associates.

TABLE 4

Item response parameters: YEP sample

Word	Item-slope	Threshold 1(Failed Categorical)	Threshold 2(Categorical Extended)	Threshold 3(Extended Specific)
Energetic	.69	-1.51	-0.99	-0.89
Angry	.58	-1.38	-1.06	-0.98
Loved	.57	-1.21	-0.84	-0.71
Successful	.53	-1.53	-1.29	-1.19
Hopeless	.53	-0.84	-0.70	-0.41
Safe	.53	-0.90	-0.44	-0.33
Sad	.51	-1.19	-1.01	-0.92
Brave	.51	-0.68	-0.56	-0.51
Failure	.50	-1.40	-1.17	-1.04
Afraid	.49	-1.42	-1.21	-1.10
Lonely	.49	-1.66	-1.27	-0.86
Calm	.48	-1.15	-0.72	-0.60
Happy	.47	-1.88	-1.61	-1.40
Worried	.47	-1.08	-0.80	-0.60
Tense	.55	-1.27	-0.92	-0.78
Comfortable	.43	-1.27	-0.83	-0.74

The first category represents responses that were non-memories: omissions, failed trials, and semantic associates.

TABLE 5

Item response theory parameters: University of Leuven Dataset #1

Word	Item-slope	Threshold 1(Failed Overgeneral)	Threshold 2(Overgeneral Specific)
Prettig (Pleasurable)	.68	-1.26	-1.08
Moed (Courage)	.62	-1.15	-0.95
Lomp (Stupid)	.63	-1.21	-1.12
Bang (Scared)	.59	-1.60	-1.41
Boos (Angry)	.59	-1.35	-1.17
Vertrouwen (Confidence/trust)	.58	-1.43	-1.33
Verrast (Surprised)	.54	<i>a</i>	-1.45
Gerust (Calm/at ease)	.51	-0.96	-0.75
Brutaal (Bold)	.50	-1.28	-1.12
Droevig (Sad)	.35	-1.48	-1.26

Failed trials also included semantic associates. This category included all trials that were not unique memories that were at least 1 week old. Overgeneral memories included extended and categoric memories.

^aFor this word there were no extended memories and only one categoric memory. Thus, we dichotomised this item into specific versus non-specific. The only threshold was the border between specific and non-specific memories.

TABLE 6

Item response theory parameters: University of Leuven Dataset #2

Word	Traditional instructions		Minimal instructions	
	Item-slope	Threshold(Non-specific Specific)	Item-slope	Threshold(Non-specific Specific)
Gerust (Calm/at ease)	.62	-0.47	.43	0.45
Droevig (Sad)	.52	-0.72	.49	-0.05
Lomp (Stupid)	.48	-1.08	.50	-0.13
Boos (Angry)	.48	-0.63	.50	0.04
Verrast (Surprised)	.47	-1.36	.50	-1.14
Brutaal (Bold)	.47	-0.68	.33	0.09
Vertrouwen (Confidence/trust)	.43	-0.87	.57	0.03
Moed (Courage)	.38	-0.45	.30	0.17