

SHORT PAPER

Development and validation of the Eyewitness Metamemory Scale

Renan Benigno Saraiva^{1,2}  | Inger Mathilde van Boeijen³ | Lorraine Hope¹  |
Robert Horselenberg² | Melanie Sauerland³  | Peter J. van Koppen^{2,4} 

¹Department of Psychology, University of Portsmouth, Portsmouth, UK

²Department of Criminal Law and Criminology, Maastricht University, Maastricht, The Netherlands

³Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands

⁴Department of Criminal Law and Criminology, VU University Amsterdam, Amsterdam, The Netherlands

Correspondence

Renan Benigno Saraiva, Department of Psychology, Faculty of Science, University of Portsmouth, Portsmouth, UK.
Email: renan.saraiva@port.ac.uk

Funding information

Erasmus Mundus Joint Doctorate Program The House of Legal Psychology (EMJD-LP), Grant/Award Number: 2013-0036

Summary

Metamemory can be defined as the knowledge about one's memory capabilities and about strategies that can aid memory. In this paper, we describe the development and validation of the Eyewitness Metamemory Scale (EMS), tailored specifically for use in face memory and eyewitness identification settings. Participants ($N = 800$) completed the EMS and other measures on general metamemory. Results from exploratory and confirmatory factor analysis revealed good factorial validity, internal consistency, and content validity. The EMS items emerged into three distinct factors: memory contentment, memory discontentment, and memory strategies. The EMS is a brief and easily administrable questionnaire that might be used to assess self-ratings of face recognition capacity and use of strategies to encode faces.

KEYWORDS

eyewitness identification, factor analysis, face recognition, metamemory, scale development

1 | INTRODUCTION

Metamemory can be defined as the knowledge about one's memory capabilities and strategies that can aid memory (Shimamura, 2008). This construct has been the subject of a substantial amount of research, sparked by developmental studies investigating how the ability to evaluate one's memory processes and mnemonic strategies improved learning during early childhood (Cavanaugh & Perlmutter, 1982). Metamemory research has since expanded to a variety of domains such as cognitive neuropsychology, educational psychology, and cognitive psychology, motivating the development of diverse self-report measures on memory monitoring and control (Pannu & Kaszniak, 2005). However, current psychometric instruments for assessment of metamemory typically focus on broad memory domains (e.g., episodic memory or semantic memory), and there appears to be

an absence of self-assessment instruments of memory capacity for faces and person recognition. In this paper, we present the development and initial validation evidence for a metamemory assessment scale tailored specifically to face memory and eyewitness identification settings.

Metamemory research is essential for a comprehensive understanding of how people use and perceive their own memory, providing a theoretical framework that can generate testable hypotheses. For example, in research examining feeling-of-knowing judgements, participants decide whether they have studied some new information sufficiently for future recall. If the subjective memory confidence experienced indicates they have not sufficiently learned the material, they may employ mnemonic strategies or engage in further study to better learn the material (Koriat, 1993). Other important branches of metamemory research include investigations on the relationship

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Applied Cognitive Psychology Published by John Wiley & Sons Ltd

between metacognitive judgement and memory performance (Kelemen, 2000), use of memory strategies (Guerrero Sastoque et al., 2019), regulation of retrieval (Goldsmith, Pansky, & Koriat, 2014), and how metamemory changes across the lifespan (Ghetti, Lyons, Lazzarin, & Cornoldi, 2008).

Interest in assessing different aspects of metamemory has stimulated the development of various self-report measures that differ in content and item format. The content may include different aspects of metamemory so that respondents are asked to indicate the frequency of forgetting, the vividness of remembering, contentment with one's memory, and perceived changes or decay in their capabilities. The item format can also vary so that some instruments focus on the relative frequency of memory issues in relation to others or in relation to one's own performance across a specified period. For example, the Metamemory in Adulthood Questionnaire (Dixon, Hultsch, & Hertzog, 1988) assesses individual's knowledge of general memory processes and tasks, frequency of memory strategy use, self-rated memory ability, perceptions of memory stability over time, anxiety regarding memory, memory and achievement motivation, and locus of control in memory abilities. The Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002) was developed to assess separate dimensions of memory ratings that are applicable to clinical assessment and intervention. This instrument includes scales of contentment regarding one's memory, self-appraisal of one's memory capabilities, and reported frequency of memory strategy use. Another example is the Squire Subjective Memory Questionnaire (SSMQ; van Bergen, Brands, & Jelicic, 2010; Squire, Wetzel, & Slater, 1979), assessing how one's memory trust has developed over time.

Despite the existence of several self-report memory questionnaires, there seems to be an absence of instruments that focus specifically on self-rated memory capacity for faces and person recognition. Most of the current measures have a strong focus on clinical assessments or interventions and typically include items concerning self-evaluation of general memory ability or items concerning semantic or episodic memory issues. One notable exception is the newly developed Stirling Face Recognition Scale (SFRS; Bobak, Mileva, & Hancock, 2019). The SFRS was developed to assess face recognition ability, ranging from developmental prosopagnosia (i.e., a neurological disorder characterized by the inability to recognize faces) to superrecognition. It has two components, face processing and face memory, which correlated moderately with objective face matching tests (correlations between $r = .28$ and $r = .34$). However, this instrument has not yet been subjected to factor analysis, and the reliability of each SFRS component is unknown. Furthermore, The SFRS does not include items related to other person identification elements that may be relevant in eyewitness settings.

Self-report instruments specifically developed to measure face recognition ability and person identification would have important implications for research and practice. One important issue in the criminal justice system, for instance, is to distinguish accurate from inaccurate eyewitness identifications. Evidence obtained from witnesses of crimes can be very influential in court decisions, but inaccurate witness identifications can impair investigations and in more severe cases

contribute to miscarriages of justice. Some postdictors of eyewitness identification accuracy have been identified, such as early statements of confidence (Brewer & Wells, 2006), decision time (Sporer, 1993), and decision process (i.e., absolute vs. relative judgements, Dunning & Stern, 1994). However, under certain circumstances, the predictive value of those factors is undermined, for example, when eyewitnesses are exposed to biased lineups (Charman, Wells, & Joy, 2011) or receive feedback after an identification is made (Semmler, Brewer, & Wells, 2004). This limitation highlights the importance of investigating new factors that may be used to estimate eyewitness accuracy that are less undermined by external factors. One such potential estimator is self-efficacy in face recognition, which has shown to be predictive of eyewitness accuracy performance (Olsson & Juslin, 1999; Perfect, 2004). However, previous studies on this issue have used single items of unknown reliability and validity, limiting conclusions regarding the relation between self-efficacy and objective memory performance. A reliable and valid metamemory scale tailored specifically to eyewitness settings would improve the inferences in studies investigating the relation between self-ratings of memory ability and objective memory accuracy.

Another important theoretical implication of an eyewitness metamemory scale is that it would help elucidate the relation between self-ratings of memory ability and expressions of confidence. Koriat (1993) has proposed that expressions of memory confidence are partly based on the encoding experience (i.e., characteristics of the stimuli) and on internal cues or beliefs about memory capacity (i.e., "am I good at recognizing this type of stimuli?"). However, general theories of memory confidence have not yet been thoroughly examined in eyewitness contexts. In forensic settings, for example, eyewitness confidence judgements are commonly used for assessing the likelihood that the eyewitness memory is accurate (Wixted & Wells, 2017). The ability to accurately evaluate one's own memory performance is a critical feature of metamemory function, but laboratory manipulations have shown that eyewitness confidence can be inflated by factors such as postidentification feedback (Douglass & Steblay, 2006) and repeated recall (Odinot & Wolters, 2006). It has been suggested that confidence expressed by witnesses is also influenced by internal cues (Leippe & Eisenstadt, 2014), but the extent to which memory accuracy and confidence for faces is related to self-perceived recognition skill is relatively unknown. In one of the few studies on the matter, Olsson and Juslin (1999) found that people who claim to be good face recognizers show slightly higher accuracy and better confidence-accuracy calibration in eyewitness identifications, but that study is limited by the use of single items of unknown reliability and validity. The absence of valid measures of eyewitness face recognition ability impairs the advancement of this theoretical line of research. With such a measure, it would be possible to better examine the relation between beliefs of memory capacity and expressions of confidence in forensic relevant contexts.

Despite the benefits of self-report tools, it can be argued that memory accuracy could be better estimated by objective tests of memory performance. In fact, it has been proposed that tests of face recognition performance are informative estimators of proclivity to

choose and identification accuracy (Baldassari, Kantner, & Lindsay, 2019; Russ, Sauerland, Lee, & Bindemann, 2018). However, in practical terms, objective tests of face recognition are more difficult to implement in applied and research settings. That is because commonly used tests of face recognition or face match ability are computerized and include many repeated trials (e.g., Dowsett & Burton, 2015; Russell, Duchaine, & Nakayama, 2009). Ideally, both objective memory tests and self-ratings of memory performance could be deployed to estimate eyewitness identification accuracy, but such approach may not always be possible due to time and resources constraints. In this scenario, brief self-ratings of memory ability may be a feasible alternative to provide estimates of accuracy in practical settings and in empirical studies, although the relation between eyewitness self-ratings of memory capacity and objective performance has yet to be elucidated (Olsson & Juslin, 1999).

In sum, a metamemory instrument tailored specifically to eyewitness settings would be of considerable value in several lines of research and has the potential to aid end-users in forensic contexts. Obtaining valid measures of metamemory for eyewitness identification is essential in research investigating the relation between self-efficacy, objective accuracy, and expressions of confidence. Depending on the results and development of this line of research, self-ratings of memory ability may also be employed to distinguish accurate from inaccurate identifications or to identify individuals with superior face recognition abilities (Russell et al., 2009). In this article, we present the development steps and initial evidence of the psychometric validity of the Eyewitness Metamemory Scale (EMS), a self-report memory instrument tailored specifically to face recognition and eyewitness identification settings. For the purposes of this study, we aimed to develop the instrument and test its factorial structure, while also testing for its convergent and discriminant validity through associations with other metamemory measures.

2 | MATERIALS AND METHODS

2.1 | Participants and procedure

A total of 1,347 participants proceeded past the informed consent page, although 143 cases were removed for failure to complete the metamemory measures. Several exclusion criteria were adopted to ensure the quality of the data: (a) 38 cases were removed for taking more than 90 min to complete the experiment (without outliers, the study took in average 30 min to be completed); (b) 145 cases were removed for completing the experiment in under 15 min (i.e., an impossible time to attentively complete the study); (c) 78 cases were removed for not passing all of the attention checks; and (d) 137 cases were removed due to suspicious bot activity (i.e., Prims & Motyl, 2018). The final sample ($N = 800$) comprised 62% female participants and had a mean age of $M = 29.83$, ranging from 18 to 72 years ($SD = 11.89$). The sample was from Amazon Mechanical Turk (48%), university students attending U.K. and Dutch institutions (32%), and participants recruited through social media (20%). Participants

recruited via Amazon Mechanical Turk received U\$0.50, students received course credits, and participants recruited via social media were entered a prize draw for the prize of two £50 Amazon vouchers.

In an online survey presented via Qualtrics, participants first completed the EMS, followed by other general metamemory scales. The EMS was always shown first, whereas the other metamemory scales were presented in a random order.¹ Demographic information including gender, age, and level of education was also obtained, and on completion of all tasks, participants were debriefed and thanked for their participation.

2.2 | Materials and instruments

2.2.1 | Eyewitness Metamemory Scale

Two qualitative approaches were adopted to develop an initial pool of items for the EMS. First, we closely examined the items of other metamemory measures and, where possible, based our item development on these items. Then a semistructured interview was conducted with a group of legal psychologists and graduate students working in this field of research ($N = 14$) to obtain additional information regarding memory self-assessment in eyewitness contexts. The initial pool of items consisted of 35 items, including eyewitness specific items and items concerning facial recognition adapted from various metamemory questionnaires. All items were rated on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). We did not establish specific hypotheses concerning the factorial structure that would emerge from these items but rather used an exploratory approach to establish its factorial structure.

2.2.2 | General metamemory instruments

In addition to the EMS, participants also completed the MMQ (Troyer & Rich, 2002) and the SSMQ (van Bergen, Horselenberg, Merckelbach, Jelicic, & Beckers, 2010; Squire et al., 1979). The MMQ has three subscales: contentment, ability, and strategy. All items are measured on a 5-point Likert scale. The contentment scale has 18 items (e.g., "I am generally pleased with my memory ability"; $\alpha = .92$) rated from 1 (*strongly disagree*) to 5 (*strongly agree*), with higher scores indicating higher memory contentment. The ability scale has 20 items related to experiences with common memory errors over the past 2 weeks (e.g., "how often do you forget an appointment?"; $\alpha = .92$) from 1 (*all the time*) to 5 (*never*), with higher scores indicating better self-reported ability. The strategy scale has 19 items concerning the use of memory strategies during the past 2 weeks (e.g., "how often do you use a timer or alarm to remind you when to do something?"; $\alpha = .88$). The items are assessed on a

¹Participants then took part in an eyewitness paradigm consisting of a mock crime video and two identification tasks with confidence judgements. These data were obtained as part of a larger research project aiming to investigate the relation between metamemory measures and eyewitness memory performance. Due to space and focus, we only report on those measures that are relevant to the development of the EMS.

scale ranging from 1 (*never*) to 5 (*all the time*), with higher scores indicating greater use of memory strategies. The MMQ has shown good high test-retest reliability and high internal consistency in the original study by Troyer and Rich (2002) and in adaptations to different countries (e.g., Fort, Adoul, Holl, Kaddour, & Gana, 2004; van der Werf & Vos, 2011). The SSMQ consists of 18 items related to memory trust (e.g., “my ability to recall things when I really try is”; $\alpha = .94$). Participants rated the items on a 9-point scale ranging from -4 (*worse than ever*) to 4 (*better than ever before*). This instrument has shown good psychometric properties in different studies and has been correlated in a meaningful way with age, cognitive failures, and susceptibility to misinformation (van Bergen, Brands, et al., 2010; van Bergen, Horselenberg, et al., 2010). The MMQ and SSMQ differ mainly in response format. Although both instruments tap into self-rated memory ability, the MMQ focuses on present ability (i.e., “I am generally pleased with my memory ability”), whereas the SSMQ focuses on memory development over time (“my memory ability is better than ever before”). Those instruments were selected to test convergent and divergent validity of the EMS due to their good psychometric properties and high content validity in assessing metamemory traits such as self-ratings of memory capacity and memory trust. However, no specific hypotheses were established a priori concerning the specific relation between each of the SSMQ and MMQ factors with the factors obtained for the EMS, given that the factorial structure of the EMS was unknown prior to our analysis. Therefore, the convergent and divergent analysis in this study were exploratory, and it was generally expected that factors in the EMS would relate meaningfully with factors from the MMQ and SSMQ given the similarities between those instruments in assessing metamemory.

2.2.3 | Attention checks

Three attention checks were included within the metamemory assessment, in which participants were asked to select a specific response for that item such as “for this question, please select the option 4 (*better than ever before*).” The attention checks were included as an exclusion criterion (see Section 2.1).

3 | RESULTS

To examine the validity of the EMS factorial solutions, a within-sample replication strategy was adopted, and the total sample was randomly split in half (Osborne & Fitzpatrick, 2012). The first half was treated as a training dataset for obtaining an initial factorial solution via exploratory factor analysis (EFA). The second half was treated as a test dataset for examining the fit of the initial solutions obtained in the training dataset via confirmatory factor analysis (CFA). All analyses were performed in the statistical software package R (2019). The dataset and data analysis script can be found in Data S1 to S3.

3.1 | Exploratory factor analysis

Prior to the analysis, one item was removed because of a semantic error in the survey. A correlation matrix of the remaining 34 items was screened to identify items that were poorly correlated with the others, or items that were highly correlated and generating multicollinearity issues. Eight items were excluded for showing weak item-total correlations ($r < .30$). Two other items were excluded for presenting high correlations ($r > .65$) and redundant content in relation to other items.

Diagnostic tests were performed on the remaining 24 items to examine the assumptions for EFA. Data gathering for the metamemory measures were performed in an online setting with forced responses, so no missing responses were present. Graphical inspection and significant Shapiro Wilk tests for all the items indicated significant univariate nonnormality, with skewness ranging from -0.87 to $+0.89$ and kurtosis ranging from -1.01 to $+0.32$. This observation was supported by the statistically significant Mardia's test, indicating that the assumption of multivariate normality was violated. Therefore, a weighted least squares extraction method for EFA was used, which provides standard errors and tests of model fit that are robust to the nonnormality of the data. The items showed good factorability (Kaiser–Meyer–Olkin test = 0.89 and significant Bartlett's test) and did not present multicollinearity or singularity issues (determinant >0.00001).

Parallel analysis and scree plots were used as factor retention criteria and suggested the presence of four factors. A four-factor solution was extracted using oblimin rotation to allow for correlations between the factors. This solution revealed four distinguishable factors, but one factor had only four emerging items that appeared to be related to memory development over time. These items presented high cross-loadings with two of the other factors in the solution, indicating that a four-factor solution might not be robust. We proceeded with the extraction of a three-factor solution using oblimin rotation. Examining the pattern matrix, we decided to exclude one item from the first factor for high cross-loadings and a content that was dissonant with the other items (i.e., “People are generally good at remembering unfamiliar faces”). The same three-factor extraction was then repeated on the remaining 23 items (the pattern matrix for this solution is presented in Table 1). Items had high loadings on their respective factor, with no cross loadings higher than $.30$. We termed the three factors memory contentment (10 items explaining 19% of the total scale variance), memory discontentment (eight items explaining 15% of the variance), and memory strategies (five items explaining 10% of the variance). The memory contentment factor combined items related to positive self-perception of memory ability, including keywords such as “satisfied,” “confident,” and “better.” The memory discontentment factor combined items related to negative self-perception of memory ability, including keywords such as “trouble” and “worse.” The memory strategies factor combined items related to the use of memory strategies in the context of person identification and could be defined as the extent to which an individual adopts strategies to better recognize someone in the future. Reliability of the factors was examined using omega coefficients instead of alpha, given that assumptions for alpha are rarely met in psychometric research

TABLE 1 Item-total correlations, communalities and pattern matrix for the Eyewitness Metamemory Scale

Scale items	ITC	h^2	Factor loadings		
			1	2	3
Items Relating to Memory Contentment					
1. My ability to remember faces is much better than other people's ability to remember faces.	.33	.67	.84	–	–
2. I am confident with my ability to remember faces in a stressful situation.	.43	.57	.74	–	–
3. Compared to other people, I think I would be a much better eyewitness.	.40	.60	.73	–	–
4. If I saw someone commit a crime, I am certain that I would remember his/her face.	.45	.55	.70	–	–
5. If I witnessed a robbery, I would be able to recognize the perpetrator a month later.	.54	.46	.65	–	–
6. As I age, I find my ability to remember faces is getting better.	.58	.42	.65	–	–
7. I am generally satisfied with my ability to remember the faces of people I have only met once.	.50	.50	.61	–.24	–
8. My ability to correctly remember where and when I saw a particular face has improved over time.	.62	.38	.57	–	–
9. I can give a detailed description of a friend's face to a stranger.	.74	.26	.37	–.10	.20
10. I recognize relatively unknown actors, if I have seen them in another movie/television show.	.80	.20	.34	–.16	.11
Items Relating to Memory Discontentment					
11. Sometimes I have trouble recognizing a person that I know relatively well.	.44	.56	.26	.77	–.11
12. My ability to remember faces is much worse than other people's ability to remember faces.	.35	.65	–.19	.75	–
13. It often happens that a person who seems familiar starts a conversation with me, but I have no idea who the person is.	.54	.46	.11	.68	–
14. Whenever I meet an important person, I am worried that I will not be able to recognize him/her a week later.	.53	.47	–	.67	–
15. Compared to other people, I think I would be a much worse eyewitness.	.49	.51	–.25	.62	–
16. My ability to correctly remember where and when I saw a particular face has deteriorated over time.	.62	.38	–.10	.57	.12
17. When I see a person that looks familiar, I often do not know where I have seen that person before.	.66	.34	–.11	.55	–
18. As I age, I find my ability to remember faces is getting worse.	.65	.35	–.12	.54	–
Items Relating to Memory Strategies					
19. Compared to other people, I more often use a strategy (e.g., focus on specific facial features such as eyes) to remember a person's face.	.32	.68	–	–	.76
20. Compared to ten years ago, I more often use a strategy (e.g., focus on specific facial features such as eyes) to remember a person's face.	.36	.64	–	.14	.75
21. In order to remember a perpetrator's face, I would definitely use a strategy (e.g., focus on specific facial features such as eyes) to remember the perpetrator's face.	.56	.44	–	–.10	.70
22. I often focus on specific facial features such as nose and eyes when I am paying attention to a face that I have to remember.	.53	.47	–	–.10	.67
23. I often create a visual image in my mind of a face that I want to remember.	.73	.27	.18	–	.41
Eigenvalues			4.36	3.55	2.41
Percentage of variance explained			.19	.15	.10

Note. Factor loadings higher than >.40 are presented in bold.

Abbreviation: ITC, item-total correlations.

(Dunn, Baguley, & Brunnsden, 2014). Good reliability scores were found for the memory contentment ($\omega = .88$, 95% CI [.86, .90]), memory discontentment ($\omega = .86$, 95% CI [.83, .88]), and memory strategies factors ($\omega = .82$, 95% CI [.78, .85]). A separate factorial structure with two factors was also extracted for comparison purposes.

3.2 | Confirmatory factor analysis

The purpose of the subsequent analysis was to confirm the factor structure for the 23-item EMS on a separate subset of our sample.

The results from the EFA indicated that a three-factor solution was the most appropriate to describe the EMS. A two-factor structure was also submitted for analysis as a plausible competing model for comparing fit indices. This two-factor solution was fitted to further examine whether the contentment and discontentment factors in the three-factor solution emerged due to phrasing method rather than to the constructs the factors represent (Podsakoff, MacKenzie, & Podsakoff, 2012). Confirmatory factor analyses were conducted to test both models. Goodness of fit was evaluated using the robust root mean square error of approximation (RMSEA) and its 90% confidence

interval, robust comparative fit index (CFI), robust Tucker–Lewis index (TLI), and expected cross-validation index. These fit indices provide different types of information (i.e., absolute fit, fit adjusting for model parsimony, and fit relative to a null model), and when combined, they provide a reliable and conservative evaluation of model fit (Schreiber, Nora, Stage, Barlow, & King, 2006). The chi-square test is reported, but not relied upon to evaluate model fit due to its oversensitivity to sample size and the fact that it tests for perfect fit. The evaluation of the models was based on (a) conventional criteria for good model fit (RMSEA < .08, CFI > .90, TLI > .90, smallest expected cross-validation index) and (b) the interpretability of the solution (i.e., the comprehensibility of the factors on a conceptual level).

Diagnostic tests were performed on the 23 items to examine the assumptions for CFA and indicated that the assumption of multivariate normality was violated. Therefore, we estimated parameters in CFA using a maximum likelihood estimation with robust standard errors, which provides tests of model fit that are robust to the nonnormality of the data (Finney & DiStefano, 2013). Figure 1 presents the model specification and goodness of fit indices for the three-factor model and two-factor model. The model fit indices suggested that the three-factor solution had a better fit compared with the two-factor solution. However, the three-factor solution did not fit the data particularly well (e.g., RMSEA > 0.08, CFI < 0.90, and TLI < 0.90). In an exploratory approach, we revised the three-factor model by evaluating its modification indices, adopting only theoretically sound modifications to avoid overspecification of the model. Following this approach, we included two new correlations between errors

of Items 6 and 8 and Items 16 and 18. These modifications were based on the content of the items, which seem to be closely related to memory development over time (e.g., “As I age, I find my ability to remember faces is getting better”). The revised model resulted in an acceptable fit to data (see Figure 1).

3.3 | Convergent and discriminant validity

Convergent and discriminant validity was examined by correlating the factors of the EMS with factors from other established metamemory measures (i.e., MMQ and SSMQ). This analysis was conducted on the complete dataset ($N = 800$) instead of the training or testing dataset, given that those subsets were used only to perform independent EFA and CFA analyses. The three-factor solution demonstrated better validity in both the EFA and CFA, so the three factors emerging from this solution (i.e., EMS–Contentment, EMS–Discontentment, and EMS–Strategies) were used in the correlation analyses. Table 2 presents the correlation matrix of the EMS factors and other metamemory measures. A small negative correlation was found between EMS–Contentment and EMS–Discontentment ($r = -.29$). A moderate positive correlation was found between the EMS–Contentment and EMS–Strategies factors ($r = .44$), but no relation was found between EMS–Discontentment and EMS–Strategies ($r = .04$). The EMS–Contentment and EMS–Discontentment factors presented a small to moderate relation with most of the other metamemory measures, whereas the EMS–Strategies factor presented

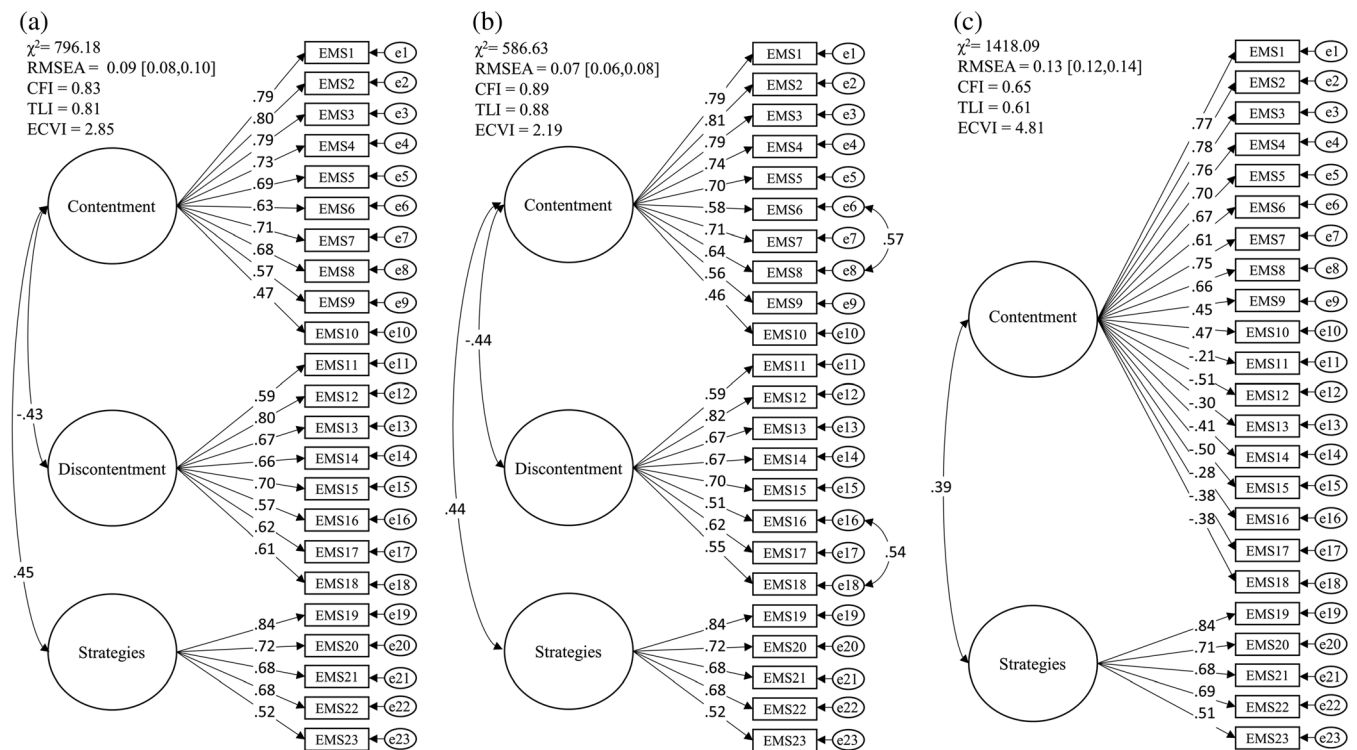


FIGURE 1 Model specification and model fit indices for the three-factor solution (a), revised three-factor solution (b), and two-factor solution (c) of the Eyewitness Metamemory Scale (EMS). CFI, comparative fit index; ECVI, expected cross-validation index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation

TABLE 2 Means, standard deviations, and correlations with confidence intervals of the EMS factors and other metamemory measures

Variable	M	SD	1	2	3	4	5	6
1. EMS-Contentment	4.22	1.02						
2. EMS-Discontentment	3.53	1.09	-.29** [-.35, -.23]					
3. EMS-Strategies	4.51	1.10	.45** [.39, .50]	.04 [-.03, .11]				
4. MMQ-Contentment	3.62	0.69	.26** [.19, .32]	-.54** [-.59, -.49]	.07 [-.00, .14]			
5. MMQ-Ability	3.57	0.62	.31** [.24, .37]	-.34** [-.40, -.27]	.19** [.12, .25]	.53** [.48, .58]		
6. MMQ-Strategy	2.86	0.64	.09** [.02, .16]	.23** [.16, .30]	.20** [.13, .26]	-.29** [-.36, -.23]	-.40** [-.46, -.34]	
7. SSMQ	5.91	1.21	.62** [.57, .66]	-.20** [-.26, -.13]	.38** [.32, .44]	.45** [.39, .50]	.42** [.36, .47]	.07* [.01, .14]

Note. Values in square brackets indicate the 95% confidence interval for each correlation.

Abbreviation: EMS, Eyewitness Metamemory Scale; MMQ, Multifactorial Memory Questionnaire; SSMQ, Squire Subjective Memory Questionnaire.

* $p < .05$. ** $p < .01$.

a positive relation with MMQ-Ability, MMQ-Strategy, and SSMQ-memory trust.

4 | DISCUSSION

In this article, we present a new self-report metamemory questionnaire developed specifically for face recognition and eyewitness identification settings. Overall, the analyses revealed good evidence of factorial validity, internal consistency, and both convergent and discriminant validity. EFA yielded three meaningful factors, each of which is associated with high loadings by the items on one factor but not on the other. This pattern of loadings provided an initial factorial validity that was also confirmed in a subset sample, corroborating the instrument division into three scales representing different dimensions of eyewitness metamemory (i.e., memory contentment, memory discontentment, and memory strategies).

Convergent validity was demonstrated by small to large correlations between the EMS factors and other scales on multidimensional metamemory questionnaires. EMS-Contentment was positively related to self-perceived contentment and ability for general memory capacity (MMQ) and had a large correlation with memory trust (SSMQ; $r = .62$). The EMS-Discontentment factor was negatively related to self-perceived contentment and ability for general memory capacity. Interestingly, some divergent pattern of results can be observed between EMS-Contentment and EMS-Discontentment. EMS-Contentment and EMS-Discontentment were not strongly related, and a model aggregating both factors in a single memory contentment factor presented poor fit to the data in this study. Furthermore, eyewitness memory contentment was positively related to the use of strategies for person identification ($r = .45$), but this relation was not observed for eyewitness memory discontentment ($r = .04$). It may be the case that individuals with higher contentment with their own memories seek additional strategies to maintain performance or that adopting strategies to better recognize someone result in higher satisfaction with one's memory capacity (Meinhardt, Persike, & Meinhardt, 2014). These findings indicate that, at least in part,

contentment and discontentment with one's own capacity for face and person recognition may represent independent constructs, rather than opposite ends of the same spectrum. The EMS-Strategies factor had only small to moderate correlations with self-rated contentment and ability for general memory capacity and memory trust. This divergent correlation pattern seems related to the fact that EMS-Strategies items focus specifically on the use of memory strategies to encode and remember faces, which appears to be somewhat independent from the use of strategies and self-appraisal for general memory.

Regarding discriminant validity, part of the nonshared variance between EMS and the other scales may be due to differences in content and item format. The EMS focuses specifically on memory for faces and person identification, whereas the other measures have a broader scope of items related to different memory domains. Contemporary memory models often consider that memory is composed of relatively independent domains (Repovs & Baddeley, 2006), and there is some evidence that people have distinct self-perceived capacity for different memory domains (Tonković & Vranić, 2011). In terms of item format, most items in the EMS are responded in relation to present contentment with memory, whereas in the SSMQ, for example, items are responded in relation to memory development over time (e.g., "better than ever before"). The SSMQ memory development focus may be especially appropriate in clinical contexts, where changes in memory perception can indicate the advancement of medical conditions (Mitchell, 2008).

Due to space and focus, we report in the current paper the development and evidence for factorial, convergent, and divergent validity of the EMS. In two other studies, it was observed that some of the EMS factors are related to eyewitness identification performance (Saraiva, Van Boeijen, Hope, Sauerland, et al., 2019; Saraiva, Van Boeijen, Hope, Horselenberg, & Van Koppen, 2019). In Saraiva, Van Boeijen, Hope, Sauerland, et al. (2019), it was found that lineup choosers (i.e., witnesses who select someone from a lineup) with higher EMS-Discontentment were more likely to be inaccurate, in both biased ($OR = 0.57$, $p < .001$) and unbiased lineups ($OR = 0.56$, $p < .001$). In Saraiva, Van Boeijen, Hope, Horselenberg, and Van Koppen (2019), it was found that for each unit increase in EMS-

Contentment score, the odds of making a correct identification increased by a factor of 1.41 ($p < .001$), and the odds of making a false identification decreased by a factor of 0.79 ($p = .002$). In both studies, it was also observed that EMS–Contentment and EMS–discontentment were significant predictors of identification confidence, suggesting that expressions of confidence are partially influenced by self-ratings of face recognition ability. Taken together, these studies provide initial evidence for the content validity and predictive validity of the EMS.

The EMS fills an important gap in the literature on face recognition and eyewitness testimony that might benefit a variety of research subjects. Eyewitnesses in criminal cases, for example, can provide unique evidence that may help solve investigations, hint to primary suspects, or potentially identify a perpetrator (Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006). However, eyewitness memory is malleable and susceptible to contamination, which may impair investigations or in more severe cases contribute to wrongful convictions. Discriminating accurate from inaccurate eyewitnesses is a challenging issue, but some postdictors of eyewitness identification accuracy have been identified, such as decision time during identifications (Sauer, Brewer, & Wells, 2008), self-reported decision-making process (Sauerland & Sporer, 2007), and early statements of confidence (Wixted & Wells, 2017). Metamemory judgements and individual differences in face recognition capacity may also relate to eyewitness performance, but this hypothesis has been relatively unexplored. Some studies have suggested that people have only moderate insight into their face recognition and face perception abilities (Bobak et al., 2019), but expressions of confidence may have a stronger relation to self-perceived memory ability. This is of importance because confidence statements are often used to discriminate accurate from inaccurate witnesses, but little is known about whether confidence statements are affected by individual differences (Leippe & Eisenstadt, 2014). Research adopting self-report instruments of face recognition capacity such as the EMS could help clarify the relationship between past experiences with memory and confidence judgements.

Research on prosopagnosia and superrecognizers could also benefit from the use of self-rated measures of face memory capacity. From a theoretical perspective, it is not clear whether prosopagnosia and superior face recognition represent opposite ends of the same continuum of face memory ability (Bobak et al., 2019). Comparing self-reported scores of face recognition capacity on patients with prosopagnosia and super recognizers could help clarify whether objective memory capacity has a linear relation with subjective memory experience. From an applied perspective, superrecognizers are considered as particularly valuable to national security agencies and border control due to their extraordinary ability to match and recognize faces from video footage or lineups (Bobak, Dowsett, & Bate, 2016). Valid face memory capacity self-report questionnaires could be used as screening tools among many participants prior to other behavioural testing, helping identifying individuals with remarkable face memory skill.

The EMS is a brief, easily administrable metamemory questionnaire focusing on face recognition and eyewitness contexts. It has been developed on a large and relatively heterogeneous sample and

showed good psychometric properties, although future amplification of its validity is desired. Self-report assessments of memory add a unique element to the assessment of memory performance that cannot be obtained in objective memory testing alone. Self-report tools allow the measurement of overarching memory issues and experiences rather than artificial laboratory-based memory problems, providing insights on an individual's memory functioning. Consequently, such tools have an important role in research and theory development regarding how memory performance relates to one's theory and one's previous experiences with memory.

ACKNOWLEDGMENTS

This research was supported by a fellowship award from the Erasmus Mundus Joint Doctorate Program The House of Legal Psychology (EMJD-LP) with Framework Partnership Agreement (FPA) 2013-0036 and Specific Grant Agreement (SGA) 2013-0036.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this article.

AUTHOR CONTRIBUTION

R. B. S. and I. M. V. B. worked on the study conception and design, acquisition of data, and analysis. R. B. S. worked on interpretation of data and drafting the manuscript. L. H., R. H., M. S., and P. V. K. made substantial contributions to the design and interpretation of data and critically revised the manuscript.

ORCID

Renan Benigno Saraiva  <https://orcid.org/0000-0001-9461-7594>

Lorraine Hope  <https://orcid.org/0000-0002-0012-8165>

Melanie Sauerland  <https://orcid.org/0000-0002-7401-1365>

Peter J. van Koppen  <http://orcid.org/0000-0002-3127-7649>

REFERENCES

- Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: toward a Lineup Skills Test. *Cognitive Research: Principles and Implications*, 4, 2. <https://doi.org/10.1186/s41235-018-0150-3>
- Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common sense: Comparing jurors, judges and law enforcement to eyewitness experts. *Applied Cognitive Psychology*, 20, 115–129. <https://doi.org/10.1002/acp.1171>
- Bobak, A. K., Dowsett, A. J., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS One*, 11, e0148148. <https://doi.org/10.1371/journal.pone.0148148>
- Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, 72, 872–881. <https://doi.org/10.1177/1747021818776145>
- Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity,

- and target-absent base rates. *Journal of Experimental Psychology: Applied*, 12(1), 11–30. <https://doi.org/10.1037/1076-898X.12.1.11>
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, 53, 11–28. <https://doi.org/10.2307/1129635>
- Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior*, 35, 479–500. <https://doi.org/10.1007/s10979-010-9261-1>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Dixon, R. A., Hulstsch, D. F., & Hertzog, C. (1988). The Metamemory in Adulthood (MIA) questionnaire. *Psychopharmacology Bulletin*, 24, 671–688.
- Douglass, A. B., & Steblay, N. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology*, 20, 859–869. <https://doi.org/10.1002/acp.1237>
- Dowsett, A. J., & Burton, A. M. (2015). Unfamiliar face matching: Pairs outperform individuals and provide a route to training. *British Journal of Psychology*, 106, 433–445. <https://doi.org/10.1111/bjop.12103>
- Dunn, T. J., Baguley, T., & Brunson, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399–412. <https://doi.org/10.1111/bjop.12046>
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, 67, 818–835.
- Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modelling: A second course* (pp. 439–492). Charlotte, NC: Information Age Publishing.
- Fort, I., Adoul, L., Holl, D., Kaddour, J., & Gana, K. (2004). Psychometric properties of the French version of the Multifactorial Memory Questionnaire for adults and the elderly. *Canadian Journal on Aging*, 23, 347–357. <https://doi.org/10.1353/cja.2005.0020>
- Ghetti, S., Lyons, K. E., Lazzarin, F., & Cornoldi, C. (2008). The development of metamemory monitoring during retrieval: The case of memory strength and memory absence. *Journal of Experimental Child Psychology*, 99, 157–181. <https://doi.org/10.1016/j.jecp.2007.11.001>
- Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In T. J. Perfect, & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 481–500). Sage: London. <https://doi.org/10.4135/9781446294703.n27>
- Guerrero Sastoque, L., Bouazzaoui, B., Burger, L., Froger, C., Isingrini, M., & Tacannat, L. (2019). Optimizing memory strategy use in young and older adults: The role of metamemory and internal strategy use. *Acta Psychologica*, 192, 73–86. <https://doi.org/10.1016/j.actpsy.2018.11.002>
- Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, 92, 800–810. <https://doi.org/10.1037/0022-0663.92.4.800>
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609–639. <https://doi.org/10.1037/0033-295X.100.4.609>
- Leippe, M. R., & Eisenstadt, D. (2014). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology: Memory for people* (pp. 377–425). New York: Psychology Press.
- Meinhardt, B., Persike, M., & Meinhardt, G. (2014). Holistic processing and reliance on global viewing strategies in older adults' face perception. *Acta Psychologica*, 151, 155–163. <https://doi.org/10.1016/j.actpsy.2014.06.001>
- Mitchell, A. J. (2008). The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: A meta-analysis. *International Journal of Geriatric Psychiatry*, 23, 1191–1202. <https://doi.org/10.1002/gps.2053>
- Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology*, 20, 973–985. <https://doi.org/10.1002/acp.1263>
- Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *The Journal of Applied Psychology*, 84, 42–49. <https://doi.org/10.1037/0021-9010.84.1.42>
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research & Evaluation*, 17, 1–8. ISSN: 1531-7714
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, 15, 105–130. <https://doi.org/10.1007/s11065-005-7091-6>
- Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, 18, 157–168. <https://doi.org/10.1002/acp.952>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Prims, J., & Motyl, M. (2018). A tool for detecting low quality data in internet research. GitHub: <https://github.com/SICLab/detecting-bots>
- Repos, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–21. <https://doi.org/10.1016/j.neuroscience.2005.12.061>
- Russ, A. J., Sauerland, M., Lee, C. E., & Bindemann, M. (2018). Individual differences in eyewitness accuracy across multiple lineups of faces. *Cognitive Research: Principles and Implications*, 3, 30. <https://doi.org/10.1186/s41235-018-0121-8>
- Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, 16, 252–257. <https://doi.org/10.3758/PBR.16.2.252>
- Saraiva, R. B., Van Boeijen, I. M., Hope, L., Horselenberg, R., & Van Koppen, P. (2019, July 5). Using general and eyewitness-specific metamemory assessments to estimate performance in multiple identifications. <https://doi.org/10.31219/osf.io/pnhm3>
- Saraiva, R. B., Van Boeijen, I. M., Hope, L., Sauerland, M., Horselenberg, R., & Van Koppen, P. (2019, July 5). Eyewitness metamemory predicts identification performance in biased and unbiased lineups. <https://doi.org/10.31219/osf.io/tay75>
- Sauer, J. D., Brewer, N., & Wells, G. L. (2008). Is there a magical time boundary for diagnosing eyewitness identification accuracy in sequential line-ups? *Legal and Criminological Psychology*, 13, 123–135. <https://doi.org/10.1348/135532506X159203>
- Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of

- identification accuracy. *Psychology, Crime & Law*, 13, 611–625. <https://doi.org/10.1080/10683160701264561s>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modelling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99, 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology*, 89, 334–346. <https://doi.org/10.1037/0021-9010.89.2.334>
- Shimamura, A. P. (2008). A neurocognitive approach to metacognitive monitoring and control. In J. Dunlosky, & R. A. Bjork (Eds.), *Handbook of Metamemory and Memory* (pp. 373–390). Psychology Press: New York.
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78, 22–33. <https://doi.org/10.1037/0021-9010.78.1.22>
- Squire, L. R., Wetzel, C. D., & Slater, P. C. (1979). Memory complaint after electroconvulsive therapy: Assessment with a new self-rating instrument. *Biological Psychiatry*, 14, 791–801.
- Tonković, M., & Vranić, A. (2011). Self-evaluation of memory systems: Development of the questionnaire. *Aging & Mental Health*, 15, 830–837. <https://doi.org/10.1080/13607863.2011.569483>
- Troyer, A. K., & Rich, J. B. (2002). Psychometric properties of a new metamemory questionnaire for older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 57, 19–27. <https://doi.org/10.1093/geronb/57.1.P19>
- van Bergen, S., Brands, I., & Jelicic, M. (2010). Assessing trait memory distrust: Psychometric properties of the Squire Subjective Memory Questionnaire. *Legal and Criminological Psychology*, 15, 373–384. <https://doi.org/10.1348/135532509X471960>
- van Bergen, S., Horselenberg, R., Merckelbach, H., Jelicic, M., & Beckers, R. (2010). Memory distrust and acceptance of misinformation. *Applied Cognitive Psychology*, 24, 885–896. <https://doi.org/10.1002/acp.1595>
- van der Werf, S. P., & Vos, S. H. (2011). Memory worries and self-reported daily forgetfulness: A psychometric evaluation of the Dutch translation of the multifactorial memory questionnaire. *The Clinical Neuropsychologist*, 25, 244–268. <https://doi.org/10.1080/13854046.2010.543290>
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18, 10–65. <https://doi.org/10.1177/152910061668696>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Saraiva RB, van Boeijen IM, Hope L, Horselenberg R, Sauerland M, van Koppen PJ. Development and validation of the Eyewitness Metamemory Scale. *Appl Cognit Psychol*. 2019;33:964–973. <https://doi.org/10.1002/acp.3588>