

Article

## In Search of an Empirical Definition of a Social Model for the Assessment of the Quality of Memory

Adriana Selaya<sup>1</sup> , Manuel Vilarinho<sup>1,2</sup>  & Ramón Arce<sup>1,2</sup> 

<sup>1</sup> Unidad de Psicología Forense, Universidad de Santiago de Compostela (Spain)

<sup>2</sup> Departamento de Ciencia Política y Sociología, Universidad de Santiago de Compostela (Spain)

### ARTICLE INFO

Received: 03/10/2023  
Accepted: 22/11/2023

#### Keywords:

Memory myths  
False beliefs about memory  
Eyewitness memories  
Psychology of testimony  
Judicial judgement making  
Judicial judgement makers

### ABSTRACT

**Background/Objectives:** The cornerstone on which the burden of proof rests in criminal cases is the credibility attributed to the complainant's testimony. Two models have been formulated for the assessment of the credibility of testimony: a social one, performed by the general population based on socially learned experience and knowledge, and a scientific one, supported by scientific evidence and performed by expert psychologists. A study was designed with the aim of developing an empirical model of how social judgements are formed based on an assessment of the quality of memory and to find out the quality of judgements based on this model. **Method:** A total of 560 ( $M_{age} = 36.84$  years; 61.6% female) lay people in Psychology from the general population participated in the study and responded to an instrument measuring myths/false beliefs and scientific knowledge about memory. **Results:** The results showed that the general population (lay people in Psychology of testimony) use myths and false beliefs about memory, together with scientific knowledge, for the assessment of memory quality. In addition, they provided support (exploratory factor analysis) for a three-factor model of memory quality assessment: trauma-related memories, veracity of testimony, and memory span. A confirmatory factor analysis validated the factor structure. **Conclusions:** These results have direct implications for the assessment of the quality (credibility) of memories (testimony) in the judicial context. Thus, in the evaluation of the quality of the testimony, lay persons in Psychology of testimony, including judicial judgement makers and jurors, base the formation of judgments on erroneous criteria (myths and false beliefs about the quality of memory) and on a model of judgment formation that is not scientifically endorsed.

## En Búsqueda de una Definición Empírica de un Modelo Social de Evaluación de la Calidad de la Memoria

### RESUMEN

**Antecedentes/Objetivos:** La piedra angular sobre la que se sustenta la carga de la prueba en los casos penales es la credibilidad atribuida al testimonio del denunciante. Se han formulado dos modelos para la evaluación de la credibilidad de un testimonio: uno social, ejecutado por la población general basado en la experiencia y conocimientos socialmente aprendidos, y otro científico, apoyado en evidencia científica que ejecutan los peritos psicólogos. Se diseñó un estudio con el objetivo de desarrollar un modelo empírico sobre cómo se conforman los juicios a nivel social basados en una evaluación de la calidad de la memoria y conocer la calidad de los juicios sustentados en dicho modelo. **Método:** Participaron en el estudio un total de 560 ( $M_{edad} = 36,84$  años; 61,6% mujeres) individuos legos en psicología del testimonio tomados de la población general que respondieron a un instrumento de medida de mitos/falsas creencias y conocimientos científicos sobre el funcionamiento de la memoria. **Resultados:** Los resultados mostraron que la población general (legos en psicología del testimonio) usa mitos y falsas creencias sobre la memoria, junto con conocimientos científicos, para la evaluación de la calidad de la memoria. Además, prestaron apoyo (análisis factorial exploratorio) a un modelo de evaluación de la calidad de la memoria compuesto de tres factores: memorias relacionadas con el trauma, veracidad del testimonio, y capacidad de la memoria para recordar. Un análisis factorial confirmatorio validó la estructura factorial. **Conclusiones:** Estos resultados tienen implicaciones directas sobre la evaluación de la calidad (credibilidad) de la memoria (testimonio) en el contexto judicial. Así, en la evaluación de la calidad del testimonio, las personas legas en psicología del testimonio, incluidos los operadores jurídicos y jurados, descansan la formación de juicios en criterios erróneos (mitos y falsas creencias sobre la calidad de la memoria) y en un modelo de formación de juicios no avalado científicamente.

#### Palabras clave:

Mitos sobre la memoria  
Falsas creencias sobre la memoria  
Memoria de testigos  
Psicología del testimonio  
Formación de juicios legales  
Operadores jurídicos

## Introduction

In judicial setting, witness statement is obtained through (judicial) cross-examination and evaluated according to doctrinal criteria to determine the credibility of a testimony or by a forensic interview assessing the quality of the memory (e.g., reality, internal or external origin, memory of a lived event) with scientific-based criteria (Arce, 2017). Worldwide, three doctrinal criteria to assess the credibility of a testimony have been defined by Supreme Courts: the absence of subjective disbelief (e.g., the absence of any interest in the sentence of the accused farther than the guilt of the accused), verisimilitude (circumstantial evidence [e.g., expert testimony about the credibility of the testimony] to support the witness statement), and persistence of the incrimination (internal, over time and external consistency of the witness statement). The absence of subjective disbelief criterion rarely is met for the claimant testimony in crimes committed in private sphere (Arce, 2017). There are no judicial rules to assess witness credibility, then jurors and judges rest judgement making on non-supported by scientific evidence criteria (e.g., nonverbal and paraverbal indicators; Arce et al., 2003), which implies a social evaluation of the testimony (Akehurst et al., 1996). A social widely used criterion to assess credibility is the confidence-accuracy relationship. Nevertheless, Berkowitz et al. (2022) advised about the scarcity of research on the relationship between confidence in memory and actual accuracy of recall, concluding that the diagnostic value of this relationship cannot yet be relied upon, at least solely, especially because the confidence in own accuracy increases over time. In addition, Clifasefi et al. (2007) pointed out some criteria by which an account is socially determined as credible, although they do not necessarily imply accuracy. These were: how confident an eyewitness seems; the amount of detail they provide; the consistency (giving the same detail over time); and the emotional intensity of the account. All these indicators assess witness credibility, not the statement credibility.

On the other hand, credibility may be conferred by judges'/ jurors' judgment making by expert testimony via the verisimilitude criterion. Whereas there are no rules to estimate witness credibility for judges and jurors, expert credibility assessment, based on the quality of the memory (Undeutsch, 1967), has scientific support (Amado et al., 2015, 2016; Oberlader et al., 2016; Vrij et al., 2021). Four forensic tools have been construct based on this hypothesis: The Statement Reality Analysis (SRA; Undeutsch, 1967), the Criteria Based Content Analysis (CBCA; Steller & Köhnken, 1989), The Reality Monitoring (RM; Johnson & Raye, 1981; Sporer & Küpper, 1995) and the Global Evaluation System (GES; Arce & Fariña, 2005). The validity and limitations of these techniques was examined (Arce, 2017; Arce et al., 2023; Gancedo et al., 2021; Volbert & Steller, 2014). The quality of the memory is also assessed by lay people in credibility judgement making. Although it is well established that memory is based on reconstructive processes rather than just storing experiences the way they occurred (Conway & Howe, 2022; Schacter, 1999, 2022), people generally belief that memory works similarly to a video camera (Lilienfeld et al., 2010). In relation to the criteria employed by people to assess memory quality, Otgaar et al., 2019 have suggested myths about memory and beliefs about repressed traumatic memories and recovered from the unconscious are media employed by people to assess memory quality (statement credibility).

Considering the state of the role of the evaluation of the memory quality in judgement making, a study was designed with general population (lay in Psychology of testimony) with the aim of developing a quality of memory based empirical model of credibility judgement making (social model), and the knowledge of the quality of judgment making sustained in the social model.

## Method

### Participants

The sample consisted of a total of 560 participants from the general population, with a mean age of 36.84 years ( $SD = 14.85$ , range 18-92), of which 345 were female (61.6%), mostly identifying as cisgender (5 made a self-identification as binary or transgender). As for the academic level attained, 28 participants reached primary studies, 21 secondary studies, 110 professional training, 85 bachelor's degree, and 112 master's degree/postgraduate studies (none with knowledge in Psychology of testimony).

### Measurement Instrument

A sensitive search for measures of myths or false beliefs about memory quality was run in Google Scholar; the scientific databases Web of Science and Scopus, the specialized database PsycInfo, previous surveys about myths or false beliefs about memory and a review of the bibliographic references of the selected papers. The initial search in databases was based on broad descriptors (myths AND memory, false beliefs AND memory) followed by narrow descriptors obtained from revised literature. Two coders with research and judicial experience in memory credibility assessment searched independently for myths/false beliefs about memory and the real scientific knowledge in selected papers. The agreement in the identification of myths and false beliefs and scientific knowledge (literature pairs myths/false beliefs with the real scientific knowledge) about memory was total. A pool of 55 myths/false beliefs and scientific knowledge was identified. Items were drafted for each of the 55. Then, items were assessed (Thurstone's procedure) by 6 experts in Memory and Legal Psychology (judges in Thurstone procedure) in the relevance (validity), independence (control of duplicates), pertinence (significant prevalence) and clarity of the wording. 11 items were eliminated. Then, and following the same procedure, 25 lay people persons in Psychology evaluated the comprehension of the item content (rewording or addition of a definition of known terms). The remaining 44 items were ordered at random. The response scale was of 5-point Likert-type: 1 = *strongly disagree*; 2 = *disagree*; 3 = *neither agree nor disagree*; 4 = *agree*; 5 = *strongly agree*.

### Procedure

A survey to the general population with a non-probabilistic accidental sampling was conducted (confidence level: 95%; margin of error  $\pm 4.0\%$ ). Measures were administered to participants on paper ( $n = 387$ ) or online ( $n = 173$ ). Participants signed informed consent. Data were processed and stored in accordance with the Spanish Data Protection Act (*Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales*, 2018).

**Table 1**  
Descriptive, One-Sample *t*-test (Test Value: 3, Neither Agree nor Disagree) and Item-Total Statistics of Myths and Scientific Knowledge About Memory

Variable	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>	<i>d</i> [IC 95%]	<i>CV</i>	Corrected item-total correlation	Cronbach's alpha if item is deleted
<b>Myths or false beliefs about memory</b>								
Item 5	2.43	1.29	-10.40	<.001	-0.44[-0.61, -0.27]	53.09	.087	
Item 6	2.22	1.19	-15.54	<.001	-0.66[-0.83, -0.49]	53.60	.062	
Item 7	3.09	1.16	1.78	.075	0.07[-0.1, 0.24]	37.54	.075	
Item 8	3.65	1.11	13.81	<.001	0.58a[0.41, 0.75]	30.41	.172	
Item 10	2.68	1.27	-5.94	<.001	-0.25[-0.42, -0.08]	47.39	.098	
Item 11	3.54	1.10	11.64	<.001	0.49a[0.32, 0.66]	31.07	.183	
Item 12	2.23	2.13	-16.67	<.001	-0.36[-0.53, -0.19]	95.52	.033	
Item 16	1.60	0.87	-38.19	<.001	-1.61[-1.80, -1.42]	54.38	.091	
Item 17	1.69	0.93	-33.33	<.001	-1.41[-1.60, -1.22]	55.03	.184	
Item 18	2.39	1.14	-12.62	<.001	-0.54[-0.71, -0.37]	47.70	.100	
<b>Item 23</b>	<b>3.91</b>	<b>1.10</b>	19.55	<.001	0.83a[0.66, 1.00]	<b>28.13</b>	<b>.292</b>	.815
<b>Item 24</b>	<b>4.21</b>	<b>0.79</b>	36.29	<.001	1.53a[1.34, 1.72]	<b>18.76</b>	<b>.399</b>	.814
<b>Item 25</b>	<b>4.39</b>	<b>0.77</b>	42.76	<.001	1.81a[1.61, 2.01]	<b>17.54</b>	<b>.436</b>	.813
<b>Item 26</b>	<b>3.86</b>	<b>0.92</b>	22.30	<.001	0.94a[0.77, 1.11]	<b>23.83</b>	<b>.328</b>	.817
<b>Item 27</b>	<b>4.10</b>	<b>0.83</b>	31.11	<.001	1.31a[1.13, 1.49]	<b>20.24</b>	<b>.398</b>	.816
Item 31	3.59	0.89	15.55	<.001	0.66a[0.49, 0.83]	24.79	.175	
Item 32	3.44	0.88	12.02	<.001	0.51a[0.34, 0.68]	25.58	.179	
Item 33	3.53	1.01	12.48	<.001	0.53a[0.36, 0.70]	28.61	.148	
<b>Item 34</b>	<b>3.24</b>	<b>1.18</b>	4.67	<.001	0.14a[-0.03, 0.31]	<b>36.42</b>	<b>.333</b>	.812
Item 36	3.56	1.09	12.22	<.001	0.52a[0.35, 0.69]	30.62	.159	
<b>Item 38</b>	<b>3.73</b>	<b>1.09</b>	15.90	<.001	0.67a[0.50, 0.84]	<b>29.22</b>	<b>.306</b>	.813
Item 39	3.13	1.18	2.58	.010	0.11a[-0.06, 0.28]	37.7	.118	
Item 40	3.27	1.11	5.72	<.001	0.24a[0.07, 0.41]	33.94	.131	
Item 41	2.56	1.00	-10.43	<.001	-0.44[-0.61, -0.27]	39.06	.015	
Item 42	2.96	1.11	-0.80	.425	-0.03[-0.2, 0.14]	37.50	.165	
Item 43	3.25	1.10	5.30	<.001	0.22a[0.05, 0.39]	33.85	.023	
<b>Scientific knowledge</b>								
Item 1	3.45	1.16	9.20	<.001	0.39[0.22, 0.56]	33.62	.178	
Item 2	3.28	1.20	5.45	<.001	0.23[0.06, 0.40]	36.59	.121	
Item 3	3.34	1.32	6.07	<.001	0.26[0.09, 0.43]	39.52	.152	
Item 4	3.66	1.00	15.59	<.001	0.66[0.49, 0.83]	27.32	.196	
Item 9	3.65	1.04	14.67	<.001	0.62[0.45, 0.79]	28.49	.101	
Item 13	3.64	1.08	14.10	<.001	0.60[0.43, 0.77]	29.67	.184	
<b>Item 14</b>	<b>3.62</b>	<b>1.18</b>	12.50	<.001	0.53[0.36, 0.70]	<b>32.60</b>	<b>.245</b>	.812
<b>Item 15</b>	<b>2.74</b>	<b>1.12</b>	-5.57	<.001	-0.24b[-0.41, -0.07]	<b>40.88</b>	<b>.201</b>	.822
<b>Item 19</b>	<b>3.46</b>	<b>1.26</b>	8.71	<.001	0.37[0.20, 0.54]	<b>36.42</b>	<b>.216</b>	.806
<b>Item 20</b>	<b>3.37</b>	<b>1.15</b>	7.60	<.001	0.32[0.15, 0.49]	<b>34.12</b>	<b>.298</b>	.802
<b>Item 21</b>	<b>2.74</b>	<b>1.22</b>	-5.15	<.001	-0.22b[-0.39, 0.05]	<b>44.53</b>	<b>.247</b>	.811
<b>Item 22</b>	<b>3.68</b>	<b>0.97</b>	16.53	<.001	0.70[0.53, 0.87]	<b>26.36</b>	<b>.253</b>	.818
Item 28	3.70	0.98	16.92	<.001	0.72[0.55, 0.89]	26.49	.194	
<b>Item 29</b>	<b>3.93</b>	<b>1.04</b>	21.10	<.001	0.89[0.72, 1.06]	<b>26.46</b>	<b>.330</b>	.827
<b>Item 30</b>	<b>4.06</b>	<b>0.94</b>	26.68	<.001	1.13[0.95, 1.31]	<b>23.15</b>	<b>.251</b>	.816
Item 35	3.92	0.95	22.95	<.001	0.97[0.79, 1.15]	24.23	.191	
Item 37	3.93	1.07	20.58	<.001	0.87[0.70, 1.04]	27.23	.133	
Item 44	4.00	1.02	23.27	<.001	0.98[0.80, 1.16]	25.50	.148	

Note. CV: Coefficient of variation; in bold, the selected items accepted myth, disagreeing to scientific of knowledge.

**Data Analysis**

Item homogeneity i.e., an estimation if items are measuring the same was measured with the item-total correlation.  $r < .200$  (item shared < 2% with the remaining items) was the criterion to eliminate. The empirical definition of a latent structure since it may not be assumed or designed by researchers, requires of the performance of an exploratory factor analysis. Principal component analysis with varimax rotation (factors' independence, orthogonal)

was the appropriate for our purposes: definition of an empirical model.

The resulting factor structure was validated with a confirmatory factor analysis using the statistical analysis software *R* version 4.2.2. For this purpose, there are numerous goodness-of-fit indices for a model, each assessing a certain aspect of goodness-of-fit and with their respective strengths and limitations. Because of this, it is recommended to provide a set of diverse indices that allows for a

more comprehensive interpretation of model fit but does not imply an arbitrary selection of reported indices.

It has been developed a classification of fit indices according to 3 levels of verification or test: the discrepancy hypothesis (sample or population), the model implication (relative or absolute) and the complexity of the fit (matched or mismatched; Sun, 2005; Tanaka, 1993). In addition, each index has certain characteristics that may have implications for interpreting the fit of a model, such as the availability of cut-off scores, sensitivity to sample size, sensitivity to model specification and sensitivity to estimation methods (Sun, 2005). Thus, a combination of indexes should be reported (Hu & Bentler, 1999; Jackson et al., 2009; Tanaka, 1993):  $\chi^2/df$ , TLI and CFI (incremental fit) RMSEA and SRMR (residual), and RMSEA, TLI and CFI (discriminant validity).

Criteria for an optimal fit have been proposed are  $\chi^2/df < 2-3$ ; RMSEA and SRMR  $< .05$ ; TLI and CFI  $> .95$ , meanwhile for a good fit  $\chi^2/df < 4$ ; RMSEA and SRMR between  $\leq .08$  and  $.10$  and TLI and CFI  $> .90$  (Anderson & Gerbing, 1984; Brooke et al., 1988; Browne & Cudeck, 1992; Cole, 1987; Hu & Bentler, 1999; Marsh et al., 1988).

### Results

Table 1 shows the descriptive statistics for each of the items included in the questionnaire. The coefficient of variation reports that the spread around mean scores is low for the selected items ( $< 45\%$ ). In relation to the validity of the criteria to estimate the quality of the memory, results advise about a significant acceptance (error) of 16 (61.5%) of the myths and false beliefs about memory quality (see Table 1) i.e., the probability of the acceptance of myths or false beliefs is 50% (constant = .50),  $Z = 1.17$ ,  $p = .242$ . Conversely, regarding to scientific-based knowledge criteria to assess memory quality, people normally (constant = .95) agree (hit, .889),  $Z = 1.19$ ,  $p = .234$ . Nevertheless, in 11.1% of the scientific criteria people disagree with the scientific knowledge.

The results of the item homogeneity analysis of the 44 myths/false beliefs and scientific knowledge about memory items reduced ( $r < .200$ ) the list to 15 items (7 acceptance of myths/false beliefs, 2 reporting disagreeing with scientific knowledge and 6 in line with scientific knowledge) as the remaining items do not measure the same construct (see Table 1). Thus, half of the measure (constant: .50) of the memory quality,  $Z = 0.77$ ,  $p = .441$ , is based on error inferences (.60).

Bartlett's test of sphericity was significant,  $\chi^2(105) = 2078.38$ ,  $p < .001$ , and the Kaiser-Meyer-Olkin index (KMO = .876) showed a good sample adequacy of the data for conducting an exploratory factor analysis. Subsequently, an exploratory factor analysis (principal component analysis with varimax rotation) was performed for the 15 items returning a factorial structure composed by three factors, accounting for 48.82% of the variance (see Table 2): trauma-related memories (29.95% of the variance), veracity of the memory i.e., criterion to evaluate the validity of the testimony (11.91% of the variance) and memory capacity to remember (6.96% of the variance). The internal consistency for the total scale was good,  $\alpha = .834[.813, .853]$ , and  $\alpha = .755[.723, .785]$ , for the trauma-related memories factor,  $\alpha = .761[.727, .792]$  for the veracity of the memory factor and (11.91% of the variance), and  $\alpha = .598[.540, .649]$ , for the memory capacity to remember factor.

**Table 2**  
Model of Eyewitness' Memory Social Assessment

Dimension	Measure	$\lambda$	$\delta$
Factor 1: Trauma-related memories	23. It is possible to repress traumatic experiences in the unconscious.	.614	.623
	24. It is possible to recover memories that were repressed with therapy.	.749	.439
	25. There are repressed memories from childhood that generate psychological problems in adulthood.	.758	.425
	26. The repression of traumatic memories is an unconscious process.	.596	.645
	27. It is possible that a person, at a given moment, remembers something in detail that he/she did not remember before.	.610	.628
	34. It is possible to repress in the unconscious a crime that has been committed.	.441	.806
Factor 2: Veracity of the memory	38. It is possible that a person has physical symptoms associated with a trauma, but no memory of this trauma.	.385	.852
	15. It is possible for a person to remember experiencing abuse when it has not actually happened.	.597	.644
	19. It is possible to confess to a crime that has not been committed.	.698	.513
	20. It is possible to confess to a crime that has not been committed while believing that it has been committed.	.767	.412
	21. It is possible to confess to a crime that has not been committed and to remember it in detail.	.811	.342
	Factor 3: Memory capacity to remember	14. It is possible to remember something that has not happened as if it had been lived.	.589
22. When a person remembers an event, he/she remembers fragments and the rest he/she "fills in".		.555	.692
30. Not thinking about a trauma for a long time is different from being unable to remember it.		.485	.765
37. Memory is not like a video camera because we do not remember events exactly as they happened.		.787	.381

Note.  $\lambda$ : factor loading;  $\delta$ : error variance, variance that is 'unique' to the variable and not shared with other variables.

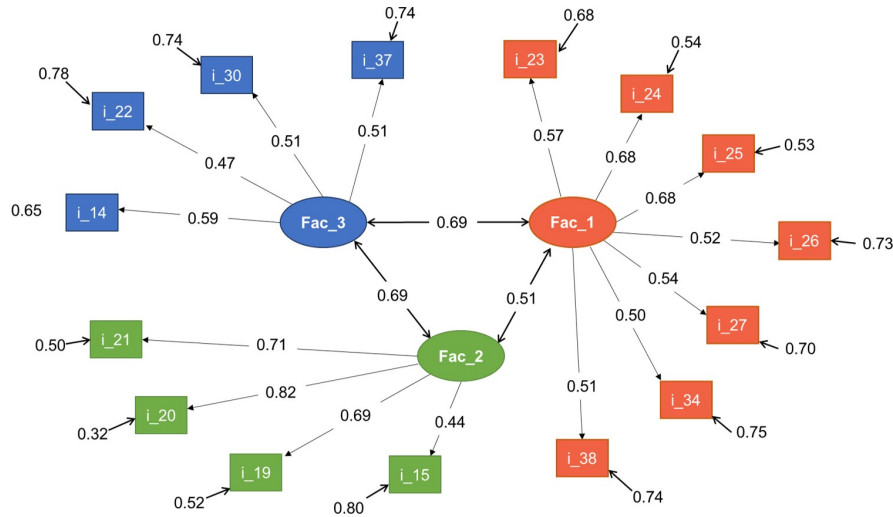
Construct validity of the social model was tested with a confirmatory factor analysis. The goodness-of-fit indices for the model (see Figure 1) were good:  $\chi^2(87) = 206.506$ ,  $p < .001$ ;  $\chi^2/df = 2.37$ ; CFI = .920; CFI(robust) = .924; TLI = .903; TLI(robust) = .908; RMSEA = .050 [90% CI: .042, .058]; SRMR = .056. Thus, social model about the estimation of the credibility of the memory of events is validated by data. In short, general population assesses the credibility of a testimony based on the trauma-related memories, veracity of the memory and memory capacity criteria.

### Discussion

The study was driven to two objectives: the empirical definition of a social model to assess the quality of the memory and the validity of the criteria employed by general population (lay in psychology of testimony) to assess the quality of the memory. The results support a 3-independent-factor empirical model



**Figure 1**  
Empirical Model of Social Evaluation Criteria for the Credibility of Testimony



(criteria) to assess the quality of the memories, comprising 15 items accounting for around 50% of the variance: trauma-related memories (factor/criterion 1), the veracity of the memory (factor/criterion 2) and the memory capacity to remember (factor/criterion 3). The criterion 1, explaining the bulk of the variance, serves to assess memory quality in relation to an experienced traumatic event (e.g., crime victimization). So, this is valid for claimant statement quality evaluation (the evaluation of the credibility of the claimant statement is ordinarily the key stone in judicial judgment making). Nevertheless, judgement inferences from this criterion rest all on the acceptance of myths/false beliefs about quality memory. In consequence, this criterion of quality of the memory (credibility of the statement in judicial setting) only introduces error in judgement making. The second criterion, veracity of the memory, close to the verisimilitude doctrinal criterion, measures scientific knowledge to estimate the veracity of the account. However, people mix correct and incorrect uses (denial of recovered memories, denial of false confessions). The third criterion, memory capacity to remember, is composed by scientific knowledge issues (items), assess the capacity to remember events of persons (witness). This criterion is correctly estimated (in line with scientific knowledge) by population.

In summary, general population (including law enforcement professionals) assess the quality of the memory (in judicial setting, testimony credibility) based on three criteria (i.e., trauma-related memories, veracity of the memory and memory capacity to remember), erroneously. These results strengthen the need of scientific based forensic psychological reports to assess witness statement credibility as non-expert evaluation is systematically based on errors sources in judgment making.

**Limitations and Future Literature**

It is important to stress the fact that the model presented is a social model derived from a scientific literature (criteria included in the scientific literature), i.e., it is unknown if other criteria are used by the general population in credibility judgement making,

in addition to those included in this model, could shape such judgements. Moreover, the statistical design of data analysis does not guarantee the invariance in results and model. In the same line, cultural effects may modulate the model. Additionally, it should be borne in mind that, although the fit indices of the presented model are good, they are not optimal. So, the margin error in modelling is not lesser. As a consequence, future literature should test the invariance of the model, the verification of the correct extension of the model to law enforcement professionals (e.g., judges, prosecutors, lawyers) and should be focused to verify if people use criteria beyond the reported in scientific literature.

**References**

Akehurst, L., Köhnken, G., Vrij, A., & Bull, R. (1996). Lay persons' and police officers' beliefs regarding deceptive behaviour. *Applied Cognitive Psychology*, 10(6), 461-471. [https://doi.org/10.1002/\(SICI\)1099-0720\(199612\)10:6<461::AID-ACP413>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1099-0720(199612)10:6<461::AID-ACP413>3.0.CO;2-2)

Amado, B. G., Arce, R., & Fariña, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *European Journal of Psychology Applied to Legal Context*, 7(1), 3-12. <https://doi.org/10.1016/j.ejpal.2014.11.002>

Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). Criteria-based content analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16(2), 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49, 155-173. <https://doi.org/10.1007/BF02294170>

Arce, R. (2017). Content analysis of the witness statements: Evaluation of the scientific and judicial validity of the hypothesis and the forensic proof [Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense]. *Acción Psicológica*, 14(2), 171-190. <https://doi.org/10.5944/ap.14.2.21347>

Arce, R., & Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio, la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG) [Psychological evidence in court on statement

- credibility, psychological injury and malingering: The Global Evaluation System (GES)]. *Papeles del Psicólogo*, 26, 59-77. <https://www.papelesdelpsicologo.es/pdf/1247.pdf>
- Arce, R., Fariña, F., & Seijo, D. (2003). Laypeople's criteria for the discrimination of reliable from non-reliable eyewitnesses. In M. Vanderhallen, G. Vervaeke, P. J. Van Koppen, & J. Goethals (Eds.). *Much ado about crime* (pp. 105-116). Uitgeverij Politeia NV.
- Arce, R., Selaya, A., Sanmarco, J., & Fariña, F. (2023). Implanting rich autobiographical false memories: Meta-analysis for forensic practice and judicial judgment making. *International Journal of Clinical and Health Psychology*, 23(4), 100386. <https://doi.org/10.1016/j.ijchp.2023.100386>
- Berkowitz, S. R., Garrett, B. L., Fenn, K. M., & Loftus, E. F. (2022). Eyewitness confidence may not be ready for the courts: A reply to Wixted et al. *Memory*, 30(1), 75-76. <https://doi.org/10.1080/09658211.2021.1952271>
- Brooke, P. P., Jr., Russell, D. W., & Price, J. L. (1988). Discriminant validation of measures of job satisfaction, job involvement, and organizational commitment. *Journal of Applied Psychology*, 73(2), 139-145. <https://psycnet.apa.org/doi/10.1037/0021-9010.73.2.139>
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230-258. <https://doi.org/10.1177/0049124192021002005>
- Clifasefi, S. L., Garry, M., & Loftus, E. (2007). Setting the record (or video camera) straight on memory: The video camera model of memory and other memory myths. In S. Della Sala (Ed.), *Tall tales about the mind & brain: Separating fact from fiction* (pp. 60-75). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198568773.003.0006>
- Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology*, 55, 584-594.
- Conway, M. A., & Howe, M. L. (2022). Memory construction: A brief and selective history. *Memory*, 30(1), 2-4. <https://doi.org/10.1080/09658211.2021.1964795>
- Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality monitoring: A meta-analytical review for forensic practice. *European Journal of Psychology Applied to Legal Context*, 13(2), 99-110. <https://doi.org/10.5093/ejpalc2021a10>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. <https://doi.org/10.1080/10705519909540118>
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23. <https://doi.org/10.1037/a0014694>
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88(1), 67-85. <https://doi.org/10.1037/0033-295X.88.1.67>
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y Garantía de los Derechos Digitales. (2018). *Boletín Oficial del Estado*, 294, 119788-119857. <https://www.boe.es/boe/dias/2018/12/06/pdfs/BOE-A-2018-16673.pdf>
- Lilienfeld, S. O., Lynn, S. J., Ruscio, J., & Beyerstein, B. L. (2010). *50 great myths of popular psychology: Shattering widespread misconceptions about human behavior*. Wiley-Blackwell.
- Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410. <https://psycnet.apa.org/doi/10.1037/0033-2909.103.3.391>
- Oberlader, V. A., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior*, 40(4), 440-457. <https://psycnet.apa.org/doi/10.1037/lhb0000193>
- Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus, E. F. (2019). The return of the repressed: The persistent and problematic claims of long-forgotten trauma. *Perspectives on Psychological Science*, 14(6), 1072-1095. <https://doi.org/10.1177/1745691619862306>
- Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, 54(3), 182-203. <https://doi.org/10.1037/0003-066X.54.3.182>
- Schacter, D. L. (2022). The seven sins of memory: An update. *Memory*, 30(1), 37-42. <https://doi.org/10.1080/09658211.2021.1873391>
- Sporer, S. L., & Küpper, B. (1995). Realitätsüberwachung und die Beurteilung des Wahrheitsgehaltes von Erzählungen: Eine experimentelle Studie [Reality monitoring and the judgment of the truthfulness of accounts: An experimental study]. *Zeitschrift für Sozialpsychologie*, 26(3), 173-193.
- Steller, M., & Köhnken, G. (1989). Criteria-Based Content Analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). Springer-Verlag.
- Sun, J. (2005). Assessing goodness of fit in Confirmatory Factor Analysis. *Measurement and Evaluation in Counseling and Development*, 37, 240-256. <https://doi.org/10.1080/07481756.2005.11909764>
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Sage.
- Undetsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen [Assessing the credibility of witness statements]. In U. Undetsch (Ed.). *Handbuch der psychologie, Vol. II: Forensische psychologie* (pp. 26-181). Verlag für Psychologie.
- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19(3), 207-220. <https://doi.org/10.1027/1016-9040/a000200>
- Vrij, A., Palena, N., Leal, S., & Caso, L. (2021). The relationship between complications, common knowledge details and self-handicapping strategies and veracity: A meta-analysis. *European Journal of Psychology Applied to Legal Context*, 13(2), 55-77. <https://doi.org/10.5093/ejpalc2021a7>

**Funding:** This research has been sponsored by a grant of the Spanish Ministry of Science and Innovation (PID2020-115881RB-I00) and by a grant of the Consellería de Cultura, Educación e Ordenación Universitaria, Xunta de Galicia (ED431B 2023/09).

**Data Availability:** The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Institutional Review Board Statement:** This study was approved by the Comité de Bioética of the Universidad de Santiago de Compostela (Code: USC-54/2022).