Article

Contrasting the Efficacy of the Content Analysis Categories of the Global Evaluation System to Discern Between Experienced Memories and Video Observed Memories

Jéssica Sanmarco [ID], Adriana Selaya [ID], Verónica Marcos [ID], & Álvaro Montes [ID]

Unidad de Psicología Forense, Universidad de Santiago de Compostela, Spain

ABSTRACT

Background/Aim: Witness credibility is commonly the keystone to support criminal cases, mainly in crimes against persons in the private sphere. As for the evaluation of the witness credibility, in Spain a content analysis technique, the Global Evaluation System (SEG), was scientifically and judicially validated. A critical concern with these techniques is their ability for the correct classification of memories of video viewed events as non-experienced memories. An experimental study was designed to test the capacity of the content analysis categorical system of the SEG to discriminate between memories of experienced events (autobiographical memories) and memories of a video observed event (non-experienced event). **Method:** Two trained interviewers interviewed (cognitive interview) to 103 participants about an autobiographical event and a video observed event (the order was counterbalanced). The interviews were analysed with the content analysis categorical system of the SEG by two trained coders which were concordant in their encodings. **Results:** The results exhibited that the statements of viewed events were not sufficient and valid evidence in 60% of cases. Moreover, the results showed that the content categories were significantly productive in experienced memories and discriminated between experienced and non-experienced memories. **Conclusions:** Strict forensic decision models, i.e., guaranteeing the presumption of innocence principle, were derived from the results.

Contrastando la Eficacia de las Categorías de Análisis de Contenido del Sistema de Evaluación Global Para Discernir Entre Memorias Experimentadas y Observadas en Video

RESUMEN

Antecedentes/Objetivo: La credibilidad de los testigos suele ser la piedra angular para apoyar los casos penales, principalmente en delitos contra personas en la esfera privada. En cuanto a la evaluación de la credibilidad del testigo, en España una técnica de análisis de contenido, el Sistema de Evaluación Global (SEG), fue científica y judicialmente válida. Una preocupación crítica con estas técnicas es su capacidad para la correcta clasificación de los recuerdos de eventos vistos en video como recuerdos no experimentados. Se diseñó un estudio experimental para probar la capacidad del sistema categórico de análisis de contenido del SEG para discriminar entre recuerdos de eventos experimentados (recuerdos autobiográficos) y recuerdos de un evento observado en video (evento no experimentado). **Método:** Dos entrevistadores entrenados entrevistaron (entrevista cognitiva) a 103 participantes sobre un evento autobiográfico y un evento observado en video (el orden fue contrarrestado). Las entrevistas fueron analizadas con el sistema categórico de análisis de contenido del SEG por dos codificadores capacitados que fueron concordantes en sus codificaciones. **Resultados:** Los resultados mostraron que las declaraciones de eventos vistos no fueron evidencia suficiente y válida en el 60% de los casos. Además, los resultados mostraron que las categorías de contenido fueron significativamente productivas en recuerdos experimentados y discriminaron entre recuerdos experimentados y no experimentados. **Conclusiones:** De los resultados se derivaron modelos forenses de decisión estricta, es decir, que garantizan el cumplimiento estricto del principio de presunción de inocencia.

Introduction

Since time immemorial, the determination of witness credibility has been the cornerstone on which the judicial judgment making pivots. Initially, the assessment of witness credibility consisted of detecting the lie by testing the defendant's testimony based primarily on the association between lying and physical, physiological, and behavioural evidence (Porter & ten Brinke, 2010). These techniques were not only ineffective in detecting lies (Oberlader et al., 2016; Vrij et al., 2010), but also classified the true testimony as false and ended up declaring themselves judicially invalid (e.g., the defendant has no obligation to testify against itself, reversal of the burden of proof; Arce, 2017). In a second procedural moment, the focus was shifted to the assignment of the witness-complainant credibility, given that the burden of proof corresponds to the prosecution. Generally, the witness-complainant testimony is not sufficient evidence for the conviction of the accused, so evidence is required that gives it the value of proof. These confirmatory proves do not usually occur in crimes committed against people in the private sphere (e.g., sexual abuse, intimate partner violence), so forensic proves are central to credibility assessment of the testimony (Novo & Seijo, 2010). For this purpose, the analysis of the declaration content based on criteria has been the proof with the greatest scientific support and judicial acceptance (Arce, 2017; Novo & Seijo, 2010; Steller & Böhm, 2006). The forensic technique par excellence based on content analysis, the Criteria Based Content Analysis (CBCA; Steller & Köhnken, 1989), has enjoyed scientific support (Amado et al., 2015, 2016; Oberlader et al., 2016) and also of judicial acceptance. However, it does not meet the Daubert Criteria (Daubert vs. Merrel Dow Pharmaceuticals, 1993) to be considered scientific evidence, as well as the legal and the doctrine of precedent criteria to be an admissible burden proof with sufficient probative value (Arce, 2017).

Based on the validity of the content analysis model for discrimination between memories of experienced events and fabricated memories and the limitations of the CBCA, Arce and Fariña (2005, 2006, 2009) created and validated (meets the Daubert criteria, and the scientific, legal, and doctrine of precedent criteria applicable to forensic evidence) the Global Evaluation System (SEG for its initials in Spanish) for different types of criminal victimization and populations (Arce & Fariña, 2007, 2012, 2013, 2014). The study of the testimony, in this forensic technique, is structured in two phases: analysis of the validity of the statement and analysis of the content of the statement. The analysis of the validity of the statements is carried out prior to the analysis of the content of the testimony to determine if the evidence (statement) is sufficient and valid. The sufficiency is measured in whether the statement exceeds the witness's memory capacity (learning control) and includes a complete narrative event of the facts. The validity of the proof (if any of these criteria is not met, the evidence is invalid) derives from the internal and external consistency of the statement, the consistency between statements, the persistence of the statements, and the consistency with scientific and natural laws. The results of the study of the sufficiency and validity of the evidence have shown that between 80 and 90% of the fabricated memories of a complaint are insufficient and/or invalid evidence (Arce, 2017; Arce et al., 2013; Redondo, 2021).

As for the content analysis of statements, it is executed with a methodic categorial system (that is, reliable and valid) that combines criteria of reality (SRA/CBCA; Steller & Köhnken, 1989; Undeutsch, 1989), memory attributes (RM; Gancedo et al., 2021; Johnson & Raye, 1981), and content categories of self-experienced event memories (Arce & Fariña, 2013, 2014), resulting in a system of 24 categories applicable to all types of cases and populations (see Annex I). The presence of the criteria is interpreted in line with the validation of a self-experienced event memory, while their absence does not imply a fabricated memory. For the development of the categorical system, it was followed the guidelines for the construction of a methodic system: mutual exclusion (control of duplication of measures), homogeneity (internal consistency), objectivity (empirical and exhaustive definition of the categories and the unit of analysis), adequacy or relevance (valid for the classification of memories of self-experienced events), completeness (concretion of new categories through a procedure by successive approximations), fidelity (creation of a coding manual for the training of coders to ensure faithful coding of categories) and productivity (categories are productively in self-experienced event memories) (Arce, 2017; Bardin, 1996). The resulting categorical system was tested (Arce et al., 2007, Arce et al., 2013; Redondo, 2021; Vilariño, 2010; Vilariño et al., 2011) observing that the SEG content categories discriminated significantly between self-experienced event memories and fabricated memories of an event.

This forensic technique specifies that the statement, as well as the repetition of the statement, must be obtained on sensory memories (as opposed to a (re)construction of the event). Therefore, this technique involves the use of the cognitive interview (Geiselman et al., 1984), or an adapted version if the witness has a handicap (Arce & Fariña, 2012). This interview technique reduces sources of external errors, memory contamination, and maximizes recall of objective information (Colwell et al., 2002). The repetition of the statement is made at least 7 days after the initial (forgetting curve). Therefore, an analysis of the consistency in the time of the testimony can be carried out. For this purpose, police or judicial statements are not valid because they are contaminated and the contents are not comparable (Memon et al., 2010). This consistency analysis is understood in terms of the centrality and peripherality of the contradictive (Undeutsch, 1967). Only consistency that refers to central elements of the event will be relevant, but not for peripheral elements (fabricated memories are more consistent – constructive hypotheses – than those of self-experienced events).

To study the effectiveness of these tools, field studies have been designed that take statements from real court cases to be tested (Craig et al., 1999; Lamb et al., 1997), thus having greater ecological validity. However, these are very minority, lack internal validity and are often tautological. In this type of design, the most critical condition to detect is the transposition of experienced events from one scenario to another. That is, the memory is of a certain event, but not of the judicial case investigated (for example, transforming a couple's argument into an act of inter-partner violence). Tested, the forensic technique of the SEG was effective in classifying these memories and in discriminating between these memories and that of the self-experienced event object of the judicial investigation (Arce et al., 2007, Arce et al., 2013; Redondo, 2021; Vilariño, 2010; Vilariño et al., 2011).

The quasi-experimental investigations in which witnesses are asked to assume the truth or lie role and provide a statement about a specific event (Colwell et al., 2002; Porter et al., 1999). The external

validity of these designs is deficient leading to results quantitatively and qualitatively different from studies with real cases (Arce, 2017). To overcome this limitation, Steller (1989) proposed three requirements for these studies (high-fidelity simulations; Fariña et al., 1994): 1) direct emotional and cognitive involvement of the person in the reported event, 2) loss of control over the situation, and 3) that the event constitutes an emotionally negative experience. These studies took as memories of lived events both memories of self-experienced events and memories of a video observed event, finding that the criteria of content of the statements are equally valid in the discrimination between both types of memories and fabricated memories of an event (Amado et al., 2015, Amado et al., 2016).

Based on the state of the literature, an investigation has been designed in order to study whether the criteria for content analysis of the SEG are productive (prevalence) in memories of self-experienced events, if they discriminate and with what effectiveness between memories of experienced events and memories of a video observed event, and define a strict decision criterion that allows, in line with the principle of presumption of innocence, the correct classification of all the video observed memories.

Method

Participants

A total of 103 people (63 women, 39 men, 1 non-binary) with a mean age of 36.74 years (*SD* = 14.59; range 19–73) participated in the study. Academically, 4.9% had completed primary education; 6.8% secondary; 22.3% of Baccalaureate; 20.4% professional training; 40.81% were university students, 2.9% were doctoral students, and 1.9% did not respond.

Procedure and Design

A repeated measures design was executed to contrast memories of experienced events (true condition) with memories of video observed events (simulated condition). In the true condition, participants were instructed to report an autobiographical experience memory of a negative nature directly lived event and that the situation and its outcome were out of their control (Steller, 1989). In the simulated condition, participants were asked to describe, as if they had experienced them, events viewed on video. The video, lasting 5 minutes and 48 seconds and from a security camera, contained a criminal action (a robbery at a jewellery store) that involved the direct victimization of two people. The order of obtaining statements about self-experienced and video observed events was counterbalanced by the standard rotation procedure (A-B, B-A). The statements were videotaped and transcribed for encoding. The interview technique used was the cognitive interview (Geiselmen et al., 1984) which was shown to be the most productive in obtaining information, while presenting the best balance between hits and errors in the information retrieved (Memon et al., 2010), and does not contaminate memory (Arce, 2017). From the cognitive interview, a modified version was taken that reduces the application to two techniques, free recall and context reinstatement, which provide results in the productivity of remembered information similar to the original and reduce the contamination on the memory due to the effect of the interview (Davis et al., 2005).

The participants were contacted through a staged online survey with restricted access, so that the distribution of the link was made through direct contact with the participants through social networks and by a non-probabilistic snowball sampling. Firstly, it was presented and explained the study, as well as the conditions of participation: voluntary and with informed consent. Secondly, participants completed a sociodemographic questionnaire (i.e., age, sex, gender, residence, education, occupation, criminal victimization). All those who agreed to participate, were no underage and had not been victims of a robbery with violence (exclusion criteria), viewed a real video of a robbery with violence and were asked for a contact detail (telephone and/or email) to be able to make subsequent appointments. Thirdly, each participant was contacted, the participation in the study was explained once again, and what their role in it would be. After approval, attendants signed informed consent to participate and to record the interviews. Fourthly, the first interview was obtained, and participants were convened for the second interview that was obtained between 7 and 10 days (effect of the forgetting curve). After the second interview, they were thanked for their participation and a debriefing was carried out, confirming that the testimony had no adverse effects.

A total of 404 protocols (interviews) were collected, 202 protocols of each condition (memory of an experienced event vs. memory of a video observed event), being 396 repeated measures (participants testified 2 under both conditions), and 8 of a single measure (experimental death in the second statement: 4 simulated statements and 4 real).

Measurement Instrument

The interviews (protocols) were subjected to a content analysis with the general categories of content (applicable to all types of memories i.e., crimes, events) of the Global Evaluation System (SEG; see list in the Annex) of Arce & Fariña (2005, 2006, 2013, 2014). Succinctly, they created a methodic, that is, reliable (α = .789, Arce et al., 2013; α = .856, Vilariño et al., 2011; α = .822; Vilariño, 2010) and valid (criterion validity: Amado et al., 2015, Amado et al., 2016; discriminant and productive validity: Vilariño, 2010; Vilariño et al., 2011), categorical system.

To these categories were added five additional ones that were described in the literature in adult population reports (Amado et al., 2016): the statement style (prodigality, the witness describes irrelevant aspects that were not required); showing signs of insecurity (uncertainty in the description of any point of the event); explanations of the lack of memory (explains the reasons why he/she/they is not able to give a detailed description); clichés (expressions or statements that delay the declaration); and repetitions (elements already described are repeated without additional details).

Encoder Training and Coding Fidelity

Two trained researchers were provided with a coding manual with the operational definition (when to code the category as present) and examples of each category, as well as the unit of analysis, served as evaluators. The statements were divided, 264 and 160 protocols, between coders, sharing 20 common statements (inter-coder agreement). After one week of the original coding, each coder analysed 10 statements (intra-coder agreement). Fidelity (concordance) was computed as inter- and intra-coder agreement.

Fidelity in coding was estimated with true kappa (Fariña et al., 2002) that corrects Cohen's kappa, verifying the exact correspondence in the encodings. The results (see Annex II) showed (Landis & Koch, 1977) an inter- and intra-coder concordance between substantial (= 0.61–0.80) and quasi-perfect (= 0.81–1.00), with an average fidelity almost perfect. In sum, the coders analysed the statements applying faithfully the content categories.

Training of Interviewers and Effect of the Interviewer on the Contents

The interviews were conducted by a researcher with a light forensic experience and a forensic expert with extensive experience in court cases who has systematically obtained valid and productive judicial statements in terms of elicitation of content categories and the correct classification of statements of memories of experienced events and discrimination of fabricated memories. In addition, this interviewer has been equally productive in eliciting content categories with other interviewers in other studies (Monteiro et al., 2018). The researcher interviewer was trained by forensic experts experienced in the technique of cognitive interview. The same interviewer obtained both statements from the witness.

The effect of the interviewer was studied in the evocation of categories for the interviewer factor. The results exposed an equal evocation of content categories in experienced memories, $t(100) = 0.43$, *ns*, and in video observed memories, $t(27.8) = 0.57$, *ns*, for both interviewers. In short, the interviewer did not play a distinctive role in the evocation of content categories. In addition, and since the forensic interviewer also obtained similar interviews to other interviewers, it can be concluded that another equally trained interviewer would obtain similar interviews in content categories.

Data Analysis

The study of the inter-declaration stability of content analysis categories was addressed by the true kappa ($\bar{k}$); the True Concordance Index ($\overline{TCI}$), which corrects the concordance by verifying the exact correspondence in the codifications and which is the statistic of use in applied research; and the True Forensic Concordance Index ($\overline{TFCI}$), used in forensic practice, which limits true agreement to the classification of the presence of content categories (Fariña et al., 2002). That is, in the forensic task it is only of interest when the category is coded (true positive or false positive), not providing information the non-coding of the category. That is why the other measures of stability are overestimated. As a decision criterion for an acceptable agreement, .80 was taken (Tversky, 1977), such that it reduces the variance of the error by 80% (LeBreton et al., 2003). The stability measured with Kappa was interpreted according to Landis & Koch's criteria (1977).

The study of the productivity of the SEG and additional categories was performed by contrasting the observed probability with a constant (trivial prevalence, .05; common prevalence, .50; and overall prevalence, .95) with Zeta, estimating the effect size with Cohen's *h* (*h* = 0.20 small size; *h* = 0.50 moderate size; *h* = 0.80 large size; *h* = 1.20 size more than large; Arce et al., 2015; Cohen, 1988).

Additionally, and as the forensic error type I (false positives) must be zero (principle of presumption of innocence), Arce (Fandiño et al., 2021; Redondo et al., 2019) created an estimate of the effect size

to appreciate this type of error and to create evidence-based decision models: the Forensic Discrimination Index ($ID_f = (a - b) / N$). This index evaluates the net classification capacity of the categories, interpreting an ID_f as significant if the lower limit of the 95% CI is > .50 (and not significant if the interval passes through .50, that is, the capacity is equal to chance, while if the upper limit is < .50, it will qualify worse than chance). The magnitude of the effect is small if $ID_f = .55$ ($d \approx 0.20$); moderate if $ID_f = .60$ ($d \approx 0.50$); large if $IDf = .65$ ($d \approx 0.80$); and more than large if $ID_f = .75$ ($d \approx 1.20$). In addition, Arce (Fandiño et al., 2021) also proposed, for the creation of forensic decision models, the Forensic Error (E_f), an estimate of the probability of type I error. Such that if the lower limit of the CI of 95% of $E_f \leq .05$, the probability of forensic error is trivial (negligible), while if it is > .05 the probability of forensic error is significant. In addition, if the 95% confidence interval of Ef includes zero (or the lower limit is practically zero), the error is negligible; if it includes or exceeds .50 it is common (i.e., the probability of error is common, $\geq 50\%$, the forensic error is common or usual); and if it includes or exceeds .95 it is general (i.e., The probability of error is normal, general, $\geq 95\%$, the forensic error is general).

The effect size of the comparison of means for related samples (d_{rm}) was obtained with the Lipsey's & Wilson's formula (2001) $d_{rm} = (M_{em} - M_{obm}/DT_{Diff}) * $,where M_{em} is the mean of the condition of experienced memories; M_{vom} is the mean of the condition of video observed memories; DT_{Diff} is the standard deviation of the mean difference; and *r* is the correlation between the measurement on experienced and non-experienced memories. On the other hand, the error of the statistical model was estimated with the Probability of an Inferiority Score (PIS; Gancedo et al., 2021).

Results

Study of the Stability of Content Categories

Table 1 summarizes stability in coding content categories by memory type in statistical (Cohen's kappa), professional (concordance index) and forensic (forensic concordance index) measures. Succinctly, the results exhibited, in general, an almost perfect stability (= 0.81 – 1.00) in the content categories both in the memories of an experienced event and in those of a video observed event. However, in the forensic measure, the quantity of details, concreteness and clarity of the account, contextual embedding and temporal space, description of interactions, reproduction of conversations, unusual details, and accounts of subjective mental state and affection categories exceeded the criterion of stability (.80; Tversky, 1977) in experienced memories, while not in non-experienced (do not reach the criterion of stability). Contrariwise, the logical structure, realism and reconstruction of the events, unexpected complications during the incident, superfluous details, originality of the expressions, accounts of subjective mental state and affection, and attribution of the mental state of the offender content categories are also stable in video observed memories.

Study of the Sufficiency and Validity of the Statement, and of the Productivity and Discrimination of the Content Categories

The results displayed (see Table 2) that the statement make up significantly more sufficient evidence (proof) and with a larger effect size in memories of experienced events. Particularly, the

Table 1

Study of the Stability of the Categories of Content Analysis of the SEG in Experienced and non-Experienced Memories

Variable	$\overline{k}$ [IC 95%]	$\overline{TCI}$	$\overline{TFCI}$	$\overline{k}$ [IC 95%]	$\overline{TCI}$	$\overline{TFCI}$
	Experienced Memories			**Observed Memories**		
SEG Content Categories						
1.Logical structure, realism, and reconstruction of events	1.00[.803, 1.20]	1.00	1.00	.799[.604, .994]	.929	.913
2.Unstructured production	.655[.470, .840]	.869	.581	.662[.477, .848]	.990	.500
3.Quantity of details	.942[.745, 1.14]	.980	.974	.846[.650, 1.04]	.949	.783
4.Concreteness and clarity of the account	.887[.691, 1.08]	.960	.949	.781[.587, .975]	.919	.714
5.Contextual embedding and temporal space	1.00 [.803, 1.20]	1.00	1.00	.805[.608, 1.00]	.939	.727
6. Description of interactions	.858 [.662, 1.05]	.930	.875	.584[.387, .781]	.818	.763
7.Reproduction of conversations	.792 [.506, .898]	.920	.908	.612[.416, .809]	.818	.746
8.Unexpected complications during the incident	.942 [.745, 1.14]	.980	.913	.942[.745, 1.14]	.990	.900
9.Perceptual information	.462 [.265, .659]	.737	.571	.292[.098, .485]	.919	.200
10.Cognitive operations	1.00[----, ----]	1.00	----	.619[.422, .816]	.818	.739
11.Unusual details	.861 [.665, 1.06]	.940	.914	.805[.609, 1.00]	.919	.758
12.Superfluous details	.799 [.603, .995]	.929	.913	.978[.781, 1.17]	.990	.972
13.Misunderstanding of details accurately reported	.711 [.515, 1.20]	.970	.571	1.00[----, ----]	.990	----
14.Originality of the expressions	1.00 [.803, 1.20]	1.00	1.00	.887 [.691, 1.08]	.960	.840
15.Related external associations	.555 [.359, .751]	.848	.483	1.00[----, ----]	.970	----
16.Accounts of subjective mental state and affection	.651 [.455, .847]	.970	.970	.682 [.485, .879]	.859	.808
17.Attribution of the mental state of the offender	.499 [.302, .695]	.808	.871	.518 [.322, .713]	.838	.814
18.Spontaneous corrections, specifications, and memory complements	.321 [.124, .518]	.780	.286	.270 [.079, .461]	.879	.200
19.Admission of lack of memory	.575 [.378, .771]	.788	.632	.465 [.268, .662]	.798	.429
20.Raise doubts about one's testimony	.606 [.409, .803]	.808	.627	.394 [.214, .575]	.758	.368
21.Self-deprecation	.574 [.378, .771]	.879	.478	1.00[----, ----]	.990	----
22.Pardoning the offender	.795 [.602, .988]	.990	.667	1.00[----, ----]	1.00	----
23.Characteristic details of the crime/action	1.00 [.803, 1.20]	1.00	1.00	1.00[----, ----]	1.00	----
24.Subtle clinical symptoms of the psychological harm	.731 [.537, .925]	.889	.676	1.00[----, ----]	.990	----
Additional Categories						
25.The Statement Style	.865 [.668, 1.06]	.949	.935	.914 [.717, 1.11]	.960	.897
26.Clichés	1.00[----, ----]	----	----	1.00[----, ----]	----	----
27.Repetitions	1.00[----, ----]	----	----	1.00[----, ----]	----	----
28.Showing signs of insecurity	1.00[----, ----]	----	----	1.00[----, ----]	----	----
29.Explanations of the lack of memory	.313 [.116, .509]	.960	.200	.884 [.688, 1.08]	.990	.800

Note. N = 99; $\overline{k}$ = True Kappa; $\overline{TCI}$ = True Concordance Index; $\overline{TFCI}$ = True Forensic Concordance Index; ----= could not be calculated as record value 0.

Table 2

Study of the Sufficiency and Validity of the Proof

Variable	$f(p)_{em}$	$f(p)_{vom}$	p_{diff}[IC 95%]	Z	h
Sufficient Evidence	102(.990)	63(.612)	.388[.321, .455]	7.71***	1.34
Internal Consistency	103(1.00)	103(1.00)	.000[----, ----]	----	----
External Consistency	103(1.00)	61(.592)	.408[.341, .475]	8.43***	1.39
Previous/Successive Consistent Statement	103(1.00)	82(.796)	.204[.149, .259]	5.14***	0.94
Persistence in statements	103(1.00)	96(.932)	.068[.034, .102]	2.74**	0.53
Scientifics/Nature Laws Consistency	103(1.00)	103(1.00)	.000[----, ----]	----	----
Valid Evidence	103(1.00)	58(.563)	.437[.369, .505]	8.94***	1.44
Valid and Sufficient Evidence	102(.990)	38(.369)	.631[.565, .697]	12.79***	1.84

Note. $f(p)_{em}$= frequency (observed proportion) in the memories of an experienced event; $f(p)_{vom}$ = frequency (observed proportion) in the memories of a video observed event; p_{diff} = probability difference (advantages).

effect on experienced events statements is 99%, while about 40% (38.8%) of the statements of video observed events are not sufficient evidence. Likewise, the validity criteria external consistency, consistency with previous/successive statement, and persistence of statements (see Table 2) were significantly more associated with memories of an experienced event with moderate effect size in persistence of statements (advantage ratio: 6.8%); with a large effect in consistency with previous/successive statement (advantage ratio: 20.4%); and with a more than large effect in external consistency (advantage ratio: 40.8%). In line with these results, the validity evidence is also significantly associated with memories of an experienced event with a more than large effect size (advantage ratio: 43.7%); and with valid and sufficient evidence with a more than large effect size (advantage ratio: 63.1%). In sum, statements from memories of experienced events are valid and sufficient evidence in a systematic way, while the memory of a video observed event is more than common (> .50), .631 [.538, .724], that they are not valid and sufficient evidence.

In relation to the productivity of the content categories of the SEG, the results showed that the cognitive operations, misunderstanding of details accurately reported and pardoning the offender categories were not productive (the confidence interval for probability includes. 05) in memories of experienced events; while in the memories of video observed events were not productive the unstructured production, perceptual information, misunderstanding of details accurately reported, related external associations, self-disapproval, pardoning the offender, characteristic details of the crime/action and subtle clinical symptoms of the psychological harm categories (see Table 3). Likewise and cumulatively, the results warned of a production significantly, $t(102) = 15.50$, $p < .001$, higher of content criteria in memories of an experienced event ($M = 13.36$) than in memories of a video observed event ($M = 7.17$), with an effect size between moderate and large, $d = 0.63$. In short, the SEG content categories were more productive in memories of experienced events. However, the probability that the memories of experienced events contained fewer criteria than the mean of the memories of video observed viewed (statistical model error) was 26.4% (PIS = .264).

The results of the association between the content categories of the SEG and the type of memory (see Table 3) showed a significantly higher prevalence of content categories in memories of experienced events being:

a) The effect more than large ($h > 1.20$) in the quantity of details, concreteness and clarity of the account, and unusual details categories with a non-significant (95% CI of the ID_f includes .50) net classification capacity of memories of experienced events, and with a significant forensic error (the lower limit of $E_f > .05$; in fact, the error of this measure is practically only forensic); in the contextual embedding and temporal space, superfluous details and originality of the expressions categories with a large (the CI of ID_f includes .65) net classification capacity of experienced event memories, and with a significant (the lower limit of $E_f > .05$; in fact, the error of this measurement is practically only forensic) forensic error; in the perceptual information, characteristic details of the crime/action, and in the subtle clinical symptoms of the psychological harm category with a non-significant (the 95% CI of the ID_f is less than .50) net classification capacity

of memories of experienced events, but with a negligible forensic error (i.e., the lower limit of the 95% CI of the E_f is practically 0); and in the accounts of subjective mental state and affection category with a more than large (the 95% CI of the ID_f includes .75) net classification capacity of memories of experienced events, but with a more than common forensic error (the lower limit of the 95% CI of E_f includes .65).

b) The effect large ($h > 0.80$) in the logical structure, realism and reconstruction of events, and reproduction of conversations categories with a more than large (the 95% CI of the ID_f includes .75) net classification capacity of memories of experienced events, but with a more than common forensic error (the lower limit of the CI of 95% of the E_f includes .65); in the unstructured production, related external associations and self-disapproval categories with a non-significant (the 95% CI of the ID_f does not pass .50) net classification capacity of the memories of experienced events, but with a negligible forensic error (i.e., the lower limit of the CI of 95% of E_f is practically 0); and in the admission of lack of memory category with a non-significant (the 95% CI of the ID_f does not pass .50) net classification capacity of experienced memories, and with a significant forensic error (the lower limit of $E_f > .05$; in fact, the error of this measurement is practically only forensic).

c) The effect moderate ($h > 0.50$) in the unexpected complications during the incident and misunderstanding of details accurately reported categories with a non-significant (the 95% CI of the ID_f does not pass through .50) net classification capacity of memories of experienced events, while the forensic error was negligible (i.e., the lower limit of the CI of 95% of E_f is practically 0).

In relation to the additional categories (see Table 3), the statement style is significantly more associated with statements of memories of experienced events, with a more than large effect size ($h > 1.20$), but the net classification capacity of the memories is not significant (the CI of 95% of the ID_f includes .50), and the forensic error is significant (the lower limit of $E_f > .05$). Hence, it is not appropriate to add these additional categories or combine them with the SEG content analysis categories.

Empirical Model for Forensic Decision

In line with literature reviews, content categories may discern with a significant differential efficacy between experienced and fabricated memories i.e., some content categories discriminate memories significantly better than others, being the most powerful complications, details, plausibility, details of the offence and sensory information (Amado et al., 2015, 2016; Gancedo et al., 2021; Vrij, Deeb et al., 2021; Vrij, Mann et al., 2021; Vrij, Palena et al., 2021). Having this in mind, the forensic efficacy of 3 empirical models for the definition of a strict decision criterion was contrasted (in the forensic evaluation the probability of false positive, type I error, must be zero to comply with the principle of presumption of innocence): a complete categorial model with all the content categories of the SEG (general model, $k = 24$); a model of categories that discriminate significantly between experienced and non-experienced memories (significant model, $k = 19$); and a model with robust discriminative capacity categories (robust model), that

Table 3

Study of the Difference in Probabilities of Appearance of SEG Criteria and Type of Memory (Experienced vs. non-Experienced Event)

Variable	$f(p)M_{exp}$ [IC 95%]	$f(p)M_{vo}$ [IC 95%]	Z	h	p_{diff}[IC 95%]	ID_f[IC 95%] E_f[IC 95%]
SEG Content Categories						
1.	102(.990) [.993, 1.005]	82(.796) [.718, .874]	4.74***	0.91	.194[.142, .250]	.893[.851, .935] .796[.741, .851]
2.	33(.320) [.230, .410]	3(.029) [-.003, .061]	5.96***	1.14	.291[.229, .353]	.175[.227, .123] .029[.006, .052]
3.	79(.767) [.685, .849]	22(.214) [.135, .293]	9.53***	1.68	.553[.485, .621]	.490[.422, .558] .214[.158, .270]
4.	76(.738) [.653, .823]	27(.262) [.177, .347]	7.77***	1.52	.476[.408, .544]	.500[.432, .568] .262[.202, .322]
5.	102(.990) [.993, 1.005]	19(.184) [.109, .259]	20.45***	2.23	.806[.752, .860]	.587[.520, .654] .184[.131, .237]
6.	58(.563) [.467, .659]	70(.680) [.590, .770]	-1.74	-0.70	-.117[-.161, -.073]	.621[.555, .687] .680[.616, .744]
7.	88(.854) [.786, .922]	66(.641) [.548, .734]	3.63***	0.96	.213[.157, .269]	.748[.689, .807] .641[.575, .707]
8.	22(.214) [.135, .293]	11(.107) [.047, .167]	2.11*	0.67	.107[.065, .149]	.160[.110, .210] .107[.065, .149]
9.	60(.583) [.488, .678]	7(.068) [.019, .117]	9.44***	1.60	.515[.447, .583]	.325[.261, .389] .068[.034, .102]
10.	0(.000) ----	63(.612) [.518, .706]	-12.75***	-1.80	-.612[-.679, -.545]	.306[.243, .369] .612[.545, .679]
11.	71(.689) [.600, .778]	29(.282) [.195, .369]	6.4***	1.38	.407[.340, .474]	.485[.417, .553] .282[.221, .343]
12.	81(.786) [.707, .865]	39(.379) [.285, .473]	6.5***	1.38	.407[.340, .474]	.583[.516, .650] .379[.313, .445]
13.	6(.058) [.013, .103]	0(.000) ----	2.52*	0.49	.058[.026, .090]	.029[.006, .052] .000
14.	99(.961) [.924, .998]	26(.252) [.168, .336]	15.14***	2.00	.709[.647, .771]	.607[.540, .674] .252[.193, .311]
15.	21(.204) [.126, .282]	3(.029) [-.003, .061]	4.07***	0.86	.175[.123, .227]	.117[.073, .161] .029[.006, .052]
16.	98(.951) [.909, .993]	69(.670) [.579, .761]	5.51***	1.12	.281[.220, .342]	.823[.771, .875] [.670[.606, .734]
17.	76(.738) [.653, .823]	84(.816) [.741, .891]	-1.35	-0.57	-.078[-.115, -.041]	.777[.720, .834] .816[.763, .869]
18.	18(.175) [.102, .248]	11(.107) [.047, .167]	1.41	0.53	.068[.034, .102]	.141[.093, .189] .107[.065, .149]
19.	47(.456) [.360, .552]	26(.252) [.168, .336]	3.13**	0.94	.204[.149, .259]	.354[.289, .419] .252[.193, .311]
20.	44(.427) [.331, .522]	34(.330) [.239, .421]	1.44	0.63	.097[.057, .137]	.379[.313, .445] .330[.266, .394]
21.	19(.184) [.109, .259]	1(.010) [-.009, .029]	4.41***	0.86	.174[.122, .226]	.097[.057, .137] .010[-.004, .024]
22.	2(.019) [-.007, .045]	0(.000) ----	1.41	0.28	.019[.000, .038]	.010[-.004, .024] .000
23.	62(.602) [.507, .696]	0(.000) ----	12.48***	1.78	.602[.535, .669]	.301[.238, .364] .000
24.	34(.330) [.239, .421]	1(.010) [-.009, .029]	6.76***	1.20	.320[.256, .384]	.170[.119, .221] .000
Additional Categories						
25.	75(.728) [.642, .814]	40(.388) [.294, .482]	5.23***	1.25	.340[.275, .405]	.558[.490, .626] .388[.321, .455]
26.	0(.000) ----	0(.000) ----	----	----	----	----
27.	0(.000) ----	0(.000) ----	----	----	----	----
28.	0(.000) ----	0(.000) ----	----	----	----	----
29.	3(.029) [-.003, .061]	5(.049) [.007, .091]	-0.74	0.28	-.020[-.039, .001]	.039[.013, .065] .049[.020, .078]

Note. M_{exp} = memory of experienced events; M_{vo} = memory of the video observed event p_{diff}[IC 95%]= differential probability[95% confidence interval]; $f(p)_{exp}$= frequency (observed proportion); ID_f[IC 95%]= forensic discrimination index[95% confidence interval]; E_f[IC 95%]= forensic error [95% confidence interval]; Z .= zeta value for the difference between the observed probabilities of each type of memory; h= Cohen's h effect size.

is, categories with a moderate, large, or more than large effect size *(h),* with a discrimination index greater than .65 or that are not recorded in manufactured memories (robust model, *k* = 14).

The results (see Table 4) showed as a strict forensic criterion (Type I error = 0) the registration (true positives) of more than 12 criteria for the classification of the memory of events experienced in the general model, more than 10 in the significant model and more than 6 in the robust model. With these criteria, the estimated forensic error (false negatives – type II error) would be 40.6%, .406[.310, .502], 26.7%, .267[.181, .353], and 18.8%, .188[.112, .264] for the general, significant, and robust model, respectively.

Discussion

The results of this study should be taken with caution because of the limitations in its generalization. First, the design used assumes that the autobiographical accounts were authentically true by resorting to external sources, but there is not absolutely certainty that what was reported actually happened (Ruby & Brigham, 1998). Second, both the memories of an experienced event and of a video observed event are not equivalent to memories fabricated or experienced in a judicial context, which have direct effects on the results (Fariña et al., 1994). In fact, content analysis categories perform significantly better in field studies than in experimental simulation designs (Amado et al., 2015, 2016; Arce, 2017), with

direct effects in the decision models (Fariña et al., 1994). Third, the collaboration of the participants implied a great demand for time and involvement, so their performance could be diminished (Fariña et al., 2014; Rogers, 2018). Fourth, the codification criterion differs between the scientific and forensic context, being the most liberal criterion in the former, and more conservative in the second (principle of presumption of innocence), which implies a possible over-codification of categories that would not appear in a situation of forensic evaluation. Although coders (researchers) were trained and advertised about this contingency, data may be biased towards an overestimation of content categories. Fifth, the content categories of the SEG applied were of the general version (applicable to all types of events), being not used the prescribed in SEG specific content categories for robbery cases (unknown; only available for intimate partner violence, mobbing, and bullying cases). Thus, the effects may really be greater than observed.

The results of the study of the stability of content categories in revealed a lack of consistency (stability) in the forensic task (category registration; TFCI), in many of the content categories in both experienced and video observed memories. Not so in the scientific evaluation (agreement in the registration and non-registration of the category), so that artificially increases the rate of agreement. Therefore, the measure of scientific stability in this context is not valid for the derivation of conclusions for forensic practice. Regarding forensic stability, special caution should

Table 4

Empirical Models of Forensic Decision

Nº of Content Criteria	General Model		Significant Model		Robust Model	
	Memory$_{exp}$ f(%)	Memory$_{vo}$ f(%)	Memory$_{exp}$ f(%)	Memory$_{vo}$ f(%)	Memory$_{exp}$ f(%)	Memory$_{vo}$ f(%)
1						1(2.6)
2				1(2.6)		7(21.1)
3		1(2.6)	1(1.0)	6(18.4)	1(1.0)	15(60.5)
4	1(1.0)	1(5.3)		6(34.2)		7(78.9)
5	1(2.0)	4(15.8)	1(2.0)	5(47.4)	5(5.9)	6(94.7)
6		8(36.8)	4(5.9)	5(60.5)	**13(18.8)**	**2(100)**
7	5(6.9)	6(52.6)	5(10.9)	4(71.1)	24(42.6)	
8	5(11.9)	4(63.2)	8(18.8)	7(89.5)	20(62.4)	
9	8(19.8)	5(76.3)	4(22.8)		20(82.2)	
10	6(25.7)	4(86.8)	**4(26.7)**	**4(100)**	12(94.1)	
11	9(34.7)	3(94.7)	13(39.6)		4(98.0)	
12	**6(40.6)**	**2(100)**	20(59.4)			
13	5(45.5)		15(74.3)			
14	20(65.3)		14(88.1)			
15	15(80.2)		7(95.00)			
16	11(91.1)		4(99.0)			
17	5(96.0)					
18	3(99.0)		1(100)			
19	1(100)					
20						

Note. Memory$_{exp}$ = memories of experienced events; Memory$_{vo}$ = memories of video observed events; f(%) = frequency(accumulated percentage); cut-off point for the forensic decision criterion in bold; *N* of memories of experienced events = 101; *N* of memories of video observed events (restricted to valid and sufficient evidence) = 38; General Model: includes all the SEG criteria (*k* = 24); Model of significant effects: includes SEG criteria (*k* = 19) with significant effects; Robust Model: includes all criteria with an effect size (*h*) more than large, large or moderate and categories that were not registered in non-experienced memories (forensic error = 0) (*k* = 14).

be exercised in the categories of logical structure, realism and reconstruction of the events, unexpected complications during the incident, superfluous details, originality of the expressions, accounts of subjective mental state and affection, and attribution of the mental state of the offender, since these categories were stable in video observed memories too. These results are in line with the so-called hypothesis of repetition versus reconstruction (Granhag & Strömwall, 1999) which argues that the accounts of honest witnesses are reconstructed based on the memory of the events. On the other hand, dishonest witnesses, motivated by the persistence of the story, repeat the information they gave in the first statement. This assumes that the accounts of the honest witness will vary over time, not being the same in each repetition. On the contrary, the accounts of the dishonest witness will be practically the same over time, suffering very little variability (Arce & Fariña, 2009).

In relation to the sufficiency and validity of memories of experienced events and non-experienced events to proceed a content analysis, findings revealed that one-third of the statements of video observed events were classified as insufficient evidence, and around 50% were classified as invalid evidence. Combined both measures, approximately 60% of the statements of video observed events were classified as insufficient and/or invalid legal evidence. Consequently, the analysis of the sufficiency and validity of the statement classifies robustly the memories of a video observed events.

The productivity of the content categories (prevalence) was higher in memories of experienced than in non-experienced events, that is, the content criteria are characteristic of memories of an experienced event (i.e, memory characteristics identify the memories of experienced events). Likewise, and in line with the Undeutsch hypothesis, the total of categories registered discriminated between the types of memories, showing a significantly greater presence of categories in memories of experienced events. Specifically, in 19 of 24 content categories of the SEG, a significantly higher prevalence was recorded in memories of self-experienced events: logical structure, realism and reconstruction of the events, unstructured production; quantity of details; concreteness and clarity of the account; contextual embedding and temporal space; reproduction of conversations; unexpected complications during the incident; perceptual information; cognitive operations; unusual details; superfluous details; misunderstanding of details accurately reported; originality of the expressions; related external associations; accounts of subjective mental state and affection; admission of lack of memory; self-disapproval; characteristic details of the crime/action; and subtle clinical symptoms of the psychological harm. In conclusion, SEG content categories are valid for discerning between memories of experienced and non-experienced events.

Finally, and in line with the principle of presumption of innocence which implies that all innocent persons must be protected against unfounded convictions, while it is sufficient that guilty persons are generally punished (e.g., Sentencia del Tribunal Supremo [Spanish Supreme Court Sentence] 213/2002 de 14 de febrero, 2002), three strict criteria for forensic decision have been defined (false positives = 0): general model (more than 12 of 24 criteria), significant model (more than 10 of 19 criteria with significant discriminative capacity), and robust model (more than 6 of 14 criteria with significant discriminative capacity, with a moderate, large, or more than large effect size (h), and with a discrimination

index greater than .65 or that are not registered in memories of non-experienced events). Any of the three criteria that were met would be sufficient to guarantee that, a priori, no memory of non-self-experienced events would be classified as a self-experienced memory. However, the robust model has the greatest capacity for the correct classification of memories of self-experienced events (true positives), so it will be the one used by default.

In conclusion, the results validate scientifically (this study) and judicially (providing a strict decision criterion that guarantees the principle of presumption of innocence) the SEG technique of content analysis to provide validity (valid evidence) to the testimony of the complainant against viewed memories.

Future lines of research should deepen this research design to know the productivity, the ability to discriminate between memories and the effectiveness of the SEG content criteria in the classification of the memories of a viewed event in other contexts of criminal victimization (testimony of the complainant).

References

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., Farina, 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

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content analysis of the witness statements: Evaluation of the scientific and judicial validity of the hypothesis and the forensic proof]. *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. http://www.papelesdelpsicologo.es/pdf/1247.pdf

Arce, R., & Fariña, F. (2006). Psicología del testimonio: Evaluación de la credibilidad y de la huella psíquica en el contexto penal. In Consejo General del Poder Judicial (Ed.), *Psicología del testimonio y prueba pericial* (pp. 39-103). Consejo General del Poder Judicial.

Arce, R., & Fariña, F. (2007). Propuesta de un protocolo válido y fiable para la evaluación psicológico-forense de la imputabilidad. In R. Arce, F. Fariña, E. Alfaro, C. Civera, & F. Tortosa (Eds.), *Psicología Jurídica. Evaluación e intervención* (pp. 59-65) Servicio de Publicaciones de la Diputació de València.

Arce, R., & Fariña, F. (2009). Evaluación psicológica forense de la credibilidad y daño psíquico en casos de violencia de género mediante el sistema de evaluación global. In F. Fariña, R. Arce & G. Buela-Casal (Eds.), *Violencia de Género. Tratado psicológico y legal* (pp. 147-168). Biblioteca Nueva.

Arce, R., & Fariña, F. (2012). La entrevista psicológico forense a niños, adultos y discapacitados. In S. Delgado, F. Bandrés, & A. Tejerina (Eds.), *Tratado de medicina legal y ciencias forenses: Vol. V. Pediatría legal y forense. Violencia. Víctimas* (pp. 795-817). Bosch.

Arce, R., & Fariña, F. (2013). Evaluación psicológica-forense en casos de violencia de género: El Sistema de Evaluación Global (SEG). In S. P. Colín, E. García-López, & L. A. Morales (Eds.), *Ecos de la violencia, voces de la reconstrucción* (pp. 95-120). Servicio de Publicaciones de la Universidad de Morelia.

Arce, R., & Fariña, F. (2014). Práctica de la prueba psicológica-forense: El Sistema de Evaluación Global (SEG). In C. Chan, C. Estrada, & F. J. Rodríguez (Eds.), *Aportaciones a la psicología jurídica y forense desde Iberoamérica* (pp. 47-61). Manual Moderno.

Arce, R., Fariña, F., & Vivero, A. (2007). Estudio exploratorio de la efectividad de las técnicas de análisis de contenido de las declaraciones en casos de violencia de género. In C. Guillén & R. Guil (Eds.), *Psicología social: Un encuentro de perspectivas* (pp.590-604). Asociación de Profesionales de la Psicología Social.

Arce, R., Fariña, F., Novo, M., & Vilariño, M. (2013). Contraste de la efectividad en la detección de declaraciones fabricadas de los criterios de validez del sistema de evaluación global en casos de violencia contra la mujer. In F. Expósito, I. Valor-Segura, M. Vilariño, & A. Palmer (Eds.), *Psicología jurídica aplicada a los problemas sociales. Colección Psicología y Ley, 11* (pp. 151-159). Sociedad Española de Psicología Jurídica y Forense. http://dx.doi.org/10.13140/RG.2.1.3957.1363

Arce, R., Fariña, F., Seijo, D., & Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment, 22*(6), 769-777. http://dx.doi.org/10.1177/1073191114558111

Bardin, L. (1996). *El análisis de contenido* [Content analysis] (2nd ed.). Akal.

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). LEA. https://doi.org/10.4324/9780203771587

Colwell, K., Hiscock, C. K., & Memon, A. (2002). Interviewing techniques and the assessment of statement credibility. *Applied Cognitive Psychology, 16*, 287-300. https://doi.org/10.1002/acp.788

Craig, R. A., Scheibe, R., Raskin, D. C., Kircher, J. C., & Dodd, D. H. (1999). Interviewer questions and content analysis of children's statements of sexual abuse. *Applied Developmental Science, 3*, 77- 85. https://doi.org/10.1207/s1532480xads0302_2

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*, 75-93. https://doi.org/10.1002/acp.1048

Fandiño, R. Basanta, J., Sanmarco, J., Arce, R., & Fariña, F. (2021). Evaluation of the executive functioning and psychological adjustment of child to parent offenders: Epidemiology and quantification of harm. *Frontiers in Psychology, 12*, 616855. https://doi.org/10.3389/fpsyg.2021.616855

Fariña, F., Arce, R., & Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema, 14*, 39-46. http://www.psicothema.com/pdf/684.pdf

Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad [Lineups: A comparision of high fidelity research and research in a real context]. *Psicothema, 6*(3), 395-402. http://www.psicothema.com/pdf/935.pdf

Fariña, F., Arce, R., Vilariño, M., & Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *Spanish Journal of Psychology, 17*, e32, 1-10. http://dx.doi.org/10.1017/sjp.2014.30

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

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S. J., Avetissian, I. V., & Prosk, A. L. (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science & Administration, 12*, 74-80.

Granhag, P. A., & Strömwall, L. A. (1999). Repeated interrogations - Stretching the deception detection paradigm. *Expert Evidence, 7*, 163-174. https://doi.org/10.1023/a:1008993326434

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85. https://doi.org/10.1037/0033-295X.88.1.67

Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M. (1997). Criterion-Based Content Analysis: A field validation study. *Child Abuse and Neglect, 21*, 255-264. https://doi.org/10.1016/S0145-2134(96)00170-6

Landis J., & Koch G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. http://doi.org/10.2307/2529310

LeBreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., &James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6*, 80-128. https://psycnet.apa.org/doi/10.1177/1094428102239427

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage.

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*(4), 340-372. https://doi.org/10.1037/a0020518

Monteiro, A., Vázquez, M. J., Seijo, D., & Arce, R. (2018). ¿Son los criterios de realidad válidos para clasificar y discernir entre memorias de hechos auto-experimentados y de eventos vistos en vídeo? [Are the reality criteria valid to classify and to discriminate between memories of self-experienced events and memories of video-observed events?]. *Revista Iberoamericana de Psicología y Salud, 9*(2), 149-160. https://doi.org/10.23923/j.rips.2018.02.020

Novo, M., & Seijo, D. (2010). Judicial judgement-making and legal criteria of testimonial credibility. *European Journal of Psychology Applied to Legal Context, 2*, 91-115. https://sepjf.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=26&Itemid=110&lang=es

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), 440-456. https://dx.doi.org/10.1037/lhb0000193

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology, 15*, 57-75. https://doi.org/10.1348/135532509X433151

Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: Implications for the recovered memory debate. *Law and Human Behavior, 23*, 517-537. https://doi.org/10.1023/A:1022344128649

Redondo, L. (2021). *Propuesta de una técnica para la evaluación forense en violencia sexual a mujeres* (Doctoral dissertation, Universidade de Vigo, Spain). https://www.investigo.biblioteca.uvigo.es/xmlui/handle/11093/2616

Redondo, L., Fariña, F., Seijo, D., Novo, M., & Arce, R. (2019). A meta-analytical review of the responses in the MMPI-2/MMPI-2-RF

clinical and restructured scales of parents in child custody dispute. *Anales de Psicología, 35*(1) 156-165. http://dx.doi.org/10.6018/analesps.35.1.338381

Rogers R. (2018). Researching response styles. In En R. Rogers y S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 592-614). The Guilford Press.

Ruby, C. L., & Brigham, J. C. (1998). Can criteria-based content analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior*, *22*, 369-388. https://doi.org/10.1023/A:1025766825429

Sentencia del Tribunal Supremo 213/2002 de 14 de febrero. (2002). *Boletín Oficial del Estado*, *61*, 8-15. http://supremo.vlex.es/vid/agresion-sexualabsolucion-fa-u-15055496

Steller, M. (1989). Recent developments in statement analysis. In J. C. Tuille (Ed.), *Credibility assessment* (pp. 135-154). Kluwer. https://doi.org/10.1007/978-94-015-7856-1_8

Steller, M., & Böhm, C. (2006). 50 años de jurisprudencia del Tribunal Supremo alemán sobre psicología del testimonio: Balance y perspectiva. In T. Fabian, C. Böhm, & J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53-77). Lit Verlag.

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. https://doi.org/10.1037/t27704-000

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352. https://doi.org/10.1037/0033-295X.84.4.327

Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Aussagen. In U. Undeutsch (Ed.), *Handbuch der Psychologie Vol. 11: Forensische Psychologie* (pp. 26-181). Hogrefe.

Undeutsch, U. (1989). The development of statement reality analysis. In J. Yuille (Ed.), *Credibility assessment* (pp. 101-119). Kluwer Academic Publishers.

Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género?* (Doctoral dissertation, Universidade de Santiago de Compostela, Spain). https://minerva.usc.es/xmlui/bitstream/handle/10347/2831/9788498874327_content.pdf;jsessionid=6DAFFF191FD73C24C5C9B84A87361C8A?sequence=1

Vilariño, M., Novo, M., & Seijo, D. (2011). Estudio de la eficacia de las categorías de realidad del testimonio del Sistema de Evaluación Global (SEG) en casos de violencia de género. *Revista Iberoamericana de Psicología y Salud, 2*, 1-26. http://suips.webs.uvigo.es/index.php?option=com_docman&task=doc_download&gid=22&Itemid=77

Vrij, A., Deeb, H., Leal, S., Granhag, P., & Fisher, R. P. (2021). Plausibility: A verbal cue to veracity worth examining? *European Journal of Psychology Applied to Legal Context, 13*(2), 47-53. https://doi.org/10.5093/ejpalc2021a4

Vrij, A., Mann, S., Leal, S., & Fisher, R. P. (2021). Combining verbal veracity assessment techniques to distinguish truth tellers from lie tellers. *European Journal of Psychology Applied to Legal Context, 13*(1), 9-19. https://doi.org/10.5093/ejpalc2021a2

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

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*, 89-121. https://doi.org/10.1177/1529100610390861

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

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

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

Annexs

Annex I
SEG Content Categories of a Statement

GENERAL CHARACTERISTICS OF THE ACCOUNT. The criteria included here refer to the whole statement.

1.- **Logical structure, realism, and reconstruction of events.** The sequence of the statement and its episodes are causally ordered –physical and intentional–, forming a necessary and sufficient causal chain (logical structure); the story line is complex and framed in a broader context, consonance is observed between the development of facts and their presumed consequences, as well as the memories have quality (reconstruction of history); and plot realism, probability, plausbility and credibility of the story (faced with farfetched, extraordinary and incredible) (realism). The three must be complied with: logical structure, realism, and reconstruction of the facts. The unit of analysis is the whole statement.

2.- **Memory unstructured production.** The testimony is presented in a non-chronological order, in a unified manner, but without being subject to rigid narrative or temporal patterns. The testimony is the result of sensory memories. It does not follow a script or outline. The unit of analysis is the whole statement.

3.- **Quantity of details.** The details presented in the events narrative are varied, responding to the description of the experience held in its memory nodes. The analysis unit is the whole statement.

4.- **Concreteness and clarity of the account.** The testimony is vivid, clearly, and precisely described, convincing and substantiated the facts narrated. The unit of analysis is the whole statement.

SPECIFIC CONTENTS. Specific parts of the testimony referring to the presence or strength of certain types of descriptions are evaluated.

5.- **Contextual embedding and temporal space.** The facts are located in space and time, reporting the location, duration, and developing actions in a connected way between different spaces or temporary situations. The analysis unit are specific parts of the statement.

6.- **Description of interactions.** Interactions are reported between the actors of the events, reflecting a reciprocity of themselves (at least three interacting acts, being action-reaction-action). The analysis unit are specific parts of the statement.

7.- **Reproductions of conversations.** Whole conversations or fragments are reproduced between the actors of the events, being also necessary reciprocity, whether the narrative is both direct and indirect. The analysis unit are specific parts of the statement.

8.- **Unexpected complications during the incident.** During the events an unexpected incident, complication or difficulty arises. It is reflected as an unexpected interruption. The analysis unit are specific parts of the statement.

9.- **Perceptual information.** Visual, auditory, tactile, or gustatory details are narrated during the facts' narration, being of greater weight the tactile, gustative, and olfactory (more robust in the discrimination of experienced memories to be less frequent). This information may come from memory nodes associated with the event. The analysis unit are specific parts of the statement.

10.- **Cognitive operations*.** Thoughts and thinking made by the interviewed person at the time of the event, referring to unsubstantiated inferences on the event, made a posteriori. The analysis unit are specific parts of the statement.

CONTENT SINGULARITIES. Included here are those characteristics of a statement that increase its concreteness or vividness.

11.- **Unusual details.** Narration of unusual details, but that are significant within the narrative discourse of the facts, which can be from details of people to objects or uncommon actions, but that are not of impossible appearance. The analysis unit are specific parts of the statement.

12.- **Superfluous details.** Incorporation in the narrative of non-central details, which merge in the story with those of the context, and which are not necessary to narrate the event, but which do enrich it. The analysis unit are specific parts of the statement.

13.- **Misunderstanding of details accurately reported.** Narration of details that the person does not understand, or that does not manage to give a logical meaning, but that has a meaning within the criminal context, as is the case with rituals in sexual assaults. The analysis unit are specific parts of the statement.

14.- **Originality of the expressions.** The story is characterized by being its own, with specific personal information, fruit of the original experience of the event, which moves away from clichés or stereotypical narratives. The unit of analysis is the witness' cognitions.

15.- **Related external associations.** The person narrates information suggested by the event, relating a part of the event to previous information that shares some similarity in his/her/their memory, such as a resemblance of the offender to someone, a scene from a movie, etc. The analysis unit are specific parts of the statement.

16.- **Accounts of subjective mental state and affection.** Description of a change in the cognition, emotion, feeling or attitude of the witness, in response to the events narrated, from the evaluation of the account, to how he/she/they felt, the level of emotional intensity that the events produced, or his/her/their perceptions during the event. The unit of analysis is the witness' cognitions.

17.- **Attribution of the mental state of the offender.** The motivation, behaviour, thinking, physiological state or affective reactions of the actor or actors of the events are narrated. They should be differentiated from the attributions since this criterion is based on indicators perceived at the time of the event. The unit of analysis is the witness' cognitions.

CONTENTS RELATING TO MOTIVATION AND META-MEMORY. These criteria reveal the witness's motivation for making the statement.

18.- **Spontaneous corrections, specifications, and memory complements.** During the narration corrections occur, spontaneously, on the information given that increase its richness, specify details, or nuance developed issues. Complements or specifications are made to. The unit of analysis is the witness' cognitions.

19.- **Admission of lack of memory.** The witness verbalizes not remember information about the event, or that it has no knowledge about certain parts of the event, well limited to a temporal space, or consequence of an event within it. The unit of analysis is the witness' cognitions.

20.- **Raise doubts about one's testimony.** Doubts are expressed regarding the accuracy of what is narrated, how incredible it may sound, whether it contains errors or is accurate, as well as the general accuracy of what it narrates, in relation to the recovery of what is experienced. The unit of analysis is the witness' cognitions.

21.- **Self-disapproval.** Refers to a critical attitude towards one's own behaviour and its victimization connection. The unit of analysis is the witness' cognitions.

22.- **Pardon to the offender.** The witness-victim issues verbalizations in which, in some way, forgives the offender for the damage caused or the aggressions that have inflicted. The unit of analysis is the witness' cognitions.

SPECIFIC ELEMENTS OF THE ASSAULT. Elements of testimony that are not related to the general vividness of the statement, but to the crime.

23.- **Characteristic details of the crime/action.** It narrates characteristics of the crime that contradict the usual, stereotyped beliefs, that of how this type of aggression is carried out. For example, the narration of sexual relations without consent without being aware that they involve a crime, due to the normalization of these actions. The analysis unit are specific parts of the statement.

PSYCHOLOGICAL HARM ARISING FROM THE ACTION AT TRIAL.

24.- **Subtle clinical symptoms of the psychological harm.** Recording of clinical symptoms manifested spontaneously within the memory of events: a) recurring dreams about the victimization events that cause discomfort; b) feeling that the victimization (traumatic fact) is happening; c) physiological responses to exposure to stimuli associated with the victimization; d) efforts to avoid thoughts, feelings or conversations about the victimization (trauma); e) efforts to avoid activities, places or people who remember the victimization (trauma); f) inability to remember an important aspect of the events of victimization (trauma); g) irritability or outbursts of rage; h) hypervigilance; and i) overstated startle responses. The analysis unit was the clinical symptom record.

Note. The registered memory is exclusively limited to the investigated event (e.g., the memory of the accounted previous and subsequent to the investigated event [e.g., rape] is not coded); *The cognitive operations category in the original model (theoretical model of Reality Monitoring) is typical of memories of internal origin, that is, fabricated by the witness. However, in the SEG it is characteristic of memories of experienced events (empirical model).

Annex II

Study of the Coding Fidelity of the SEG Content Analysis Categories and Additional Criteria

Variable	Intra 1	Intra 2	Inter	Promedio
SEG Content Categories				
1.Logical structure, realism, and reconstruction of events	1.00	.894	1.00	.965
2.Unstructured production	1.00	.886	.875	.920
3.Quantity of details	.872	1.00	.900	.924
4.Concreteness and clarity of the account	1.00	.898	.894	.931
5.Contextual embedding and temporal space	1.00	.643	.894	.846
6. Description of interactions	.827	.802	1.00	.876
7. Reproduction of conversations	.773	.894	.857	.841
8.Unexpected complications during the incident	1.00	1.00	.643	.881
9.Perceptual information	.894	.615	.857	.789
10.Cognitive operations	1.00	.615	.765	.793
11.Unusual details	1.00	1.00	.886	.962
12.Superfluous details	1.00	.886	.894	.927
13.Misunderstanding of details accurately reported	1.00	1.00	1.00	1.00
14.Originality of the expressions	1.00	.900	.898	.933
15.Related external associations	1.00	.737	.773	.837
16.Accounts of subjective mental state and affection	1.00	.625	1.00	.875
17.Attribution of the mental state of the offender	.773	.886	1.00	.886
18.Spontaneous corrections, specifications, and memory complements	.894	1.00	1.00	.965
19.Admission of lack of memory	1.00	.794	1.00	.931
20.Raise doubts about one's testimony	1.00	1.00	.900	.967
21.Self-deprecation	1.00	1.00	1.00	1.00
22.Pardoning the offender	1.00	1.00	1.00	1.00
23.Characteristic details of the crime/action	1.00	1.00	1.00	1.00
24.Subtle clinical symptoms of the psychological harm	1.00	1.00	1.00	1.00
Additional Categories				
25.The Statement Style	1.00	.886	1.00	.962
26.Clichés	1.00	1.00	1.00	1.00
27.Repetitions	1.00	1.00	1.00	1.00
28.Showing signs of insecurity	1.00	1.00	1.00	1.00
29.Explanations of the lack of memory	1.00	1.00	1.00	1.00

Note. Values in true kappa.