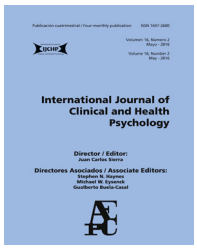
Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

International Journal of Clinical and Health Psychology (2016) **16,** 201–210

International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review

Bárbara G. Amado[a], **Ramón Arce**[a,*], **Francisca Fariña**[b], **Manuel Vilariño**[a]

[a] *Universidade de Santiago de Compostela, Spain*
[b] *Universidade de Vigo, Spain*

Received 24 November 2015; accepted 20 January 2016
Available online 16 March 2016

KEYWORDS
Criteria-Based Content Analysis;
Adults;
Statements;
Credibility;
Meta-analysis

Abstract *Background/Objective*: Criteria-Based Content Analysis (CBCA) is the tool most extensively used worldwide for evaluating the veracity of a testimony. CBCA, initially designed for evaluating the testimonies of victims of child sexual abuse, has been empirically validated. Moreover, CBCA has been generalized to adult populations and other contexts though this generalization has not been endorsed by the scientific literature. *Method*: Thus, a meta-analysis was performed to assess the Undeutsch Hypothesis and the CBCA checklist of criteria in discerning in adults between memories of self-experienced real-life events and fabricated or fictitious memories. *Results*: Though the results corroborated the Undeutsch Hypothesis, and CBCA as a valid technique, the results were not generalizable, and the *self-deprecation* and *pardoning the perpetrator* criteria failed to discriminate between both memories. The technique can be complemented with additional reality criteria. The study of moderators revealed discriminating efficacy was significantly higher in filed studies on sexual offences and intimate partner violence. *Conclusions*: The findings are discussed in terms of their implications as well as the limitations and conditions for applying these results to forensic settings.
© 2016 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

PALABRAS CLAVE
Criteria-Based Content Analysis;
adultos;

Criterios de realidad del Criteria-Based Content Analysis (CBCA) en adultos: una revisión meta-analítica

Resumen *Antecedentes/Objetivo*: El *Criteria-Based Content Analysis* (CBCA) constituye la herramienta mundialmente más utilizada para la evaluación de la credibilidad del testimonio.

* Corresponding author: Departamento de Psicoloxía Organizacional, Xurídico-Forense e Metodoloxía das Ciencias do Comportamento, Facultade de Psicoloxía, Campus Vida, s/n, 15782 Santiago de Compostela, España.
E-mail address: ramon.arce@usc.es (R. Arce).

http://dx.doi.org/10.1016/j.ijchp.2016.01.002
1697-2600/© 2016 Asociación Española de Psicología Conductual. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

declaraciones;
credibilidad;
meta-análisis

Originalmente fue creado para testimonios de menores víctimas de abuso sexual, gozando de amparo científico. Sin embargo, se ha generalizado su práctica a poblaciones de adultos y otros contextos sin un aval de la literatura para tal generalización. *Método*: Por ello, nos planteamos una revisión meta-analítica con el objetivo de contrastar la Hipótesis Undeutsch y los criterios de realidad del CBCA para conocer su potencial capacidad discriminativa entre memorias de eventos auto-experimentados y fabricados en adultos. *Resultados*: Los resultados confirman la hipótesis Undeutsch y validan el CBCA como técnica. No obstante, los resultados no son generalizables y los criterios *auto-desaprobación* y *perdón al autor del delito* no discriminan entre ambas memorias. Además, se encontró que la técnica puede ser complementada con criterios adicionales de realidad. El estudio de moderadores mostró que la eficacia discriminativa era significativamente superior en estudios de campo en casos de violencia sexual y de género. *Conclusiones*: Se discute la utilidad, así como las limitaciones y condiciones para la transferencia de estos resultados a la práctica forense.
© 2016 Asociación Española de Psicología Conductual. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The credibility of a testimony, primarily the victim's and in particular in relation to crimes committed in private (e.g., sexual offenses, domestic violence), is the key element determining legal judgements (Novo & Seijo, 2010), affecting an estimated 85% of cases worldwide (Hans & Vidmar, 1986). Though an array of tools for evaluating credibility have been designed and tested (Vrij, 2008), Criteria-Based Content Analysis [CBCA] (Steller & Köhnken, 1989) remains the technique of choice, enjoys wide acceptance among the scientific community (Amado, Arce, & Fariña, 2015), and is admissible as valid evidence in the law courts of in several countries (Steller & Böhm, 2006; Vrij, 2008). Though the technique was initially designed to be applied to the testimony of victims of child abuse sexual, its application has been extended to adults, witnesses, offenders, and other case types by Forensic Psychology Institutes in judicial proceedings (Arce & Fariña, 2012). The meta-analysis of Amado et al. (2015) found that the technique underpinning the Undeutsch Hypothesis (Undeutsch, 1967) that contends that memories of self-experienced events differ in content and quality to memories of fabricated or fictitious accounts, was equally valid in other contexts and age ranges up to the age of 18 years. Prior to the present review, empirical studies had already contrasted the validity of the Hypothesis in adult populations and in different contexts (Vrij, 2005, 2008). Moreover, as the Hypothesis was grounded on memory content, it had been theoretically advanced that the Hypothesis would be equally applicable to adults and contexts different to sexual abuse (Berliner & Conte, 1993).

CBCA consists of 19 reality criteria which are grouped into two factors: cognitive (criteria 1 to 13), and motivational (criteria 14 to 18). According to the original formulation, both factors are underpinned by the Undeutsch Hypothesis, but Raskin, Esplin, and Horowitz (1991) have underscored that only 14 conform to the aforementioned Hypothesis (14-criteria version).

CBCA has encompassed additional categories, some applicable to all contexts (Table 1) (Höfer, Köhnken, Hanewinkel, & Bruhn, 1993), and others for specific cases (Arce & Fariña, 2009; Juárez, Mateu, & Sala, 2007; Volbert & Steller, 2014), which may be combined with other techniques with diverse theoretical underpinnings such as memory attributes (Vrij, 2008).

Table 1 Additional criteria.

- Reporting style (is long-winded when interviewee described irrelevant aspects that were not asked).
- Display insecurities (uncertainty about the description of an item).
- Providing reasons lack memory (express reasons for not being able to give a detailed description).
- Clichés (expressions or utterances that introduce delays into the report).
- Repetitions (elements already described were repeated without additional details).

CBCA is extensively used in forensic practice as a tool for discriminating the memories of adults of self-experienced and fabricated events in different case types. However, due to the numerous inconsistencies in the literature (e.g., designs failing to meet the requirements for applying CBCA, conclusions of non-significant effects not substantiated by the data given the poor statistical power of the studies, $1\text{-}\beta<.80$), and the contradictory use of CBCA in adults, a meta-analysis was performed to assess the Undeutsch Hypothesis in an adult population; the discriminating efficacy of CBCA and additional reality criteria; and the effect of the context (case type), lie coaching effect, witness status, and the research paradigm.

Method

Literature search

An extensive scientific literature search was undertaken to identify empirical studies applying content analysis

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

to adult testimony in order to discriminate between self-experienced and fabricated statements, be they deliberately invented or implanted memories. The literature search consisted of a multimethod approach to meta-search engines (Google, Google Scholar, Yahoo); world leading scientific databases (PsycInfo, MedLine, Web of Science, Dissertation Abstracts International); academic social networks for the exchange of knowledge in the scientific community (i.e., Researchgate, Academia.edu); *ancestry approach* (crosschecking the bibliography of the selected studies); and contacting researchers to request unpublished studies mentioned in published studies. A list of descriptors was generated for successive approximations (i.e., the descriptors of the keywords in the selected articles were included): reality criteria, content analysis, verbal cues, verbal indicators, testimony, CBCA, Criteria Based Content Analysis, credibility, adult, statement, allegation, deception, detection, lie detection, truthful account, Statement Validity Assessment, SVA. These descriptors were used to formulate the search algorithms applied to the literature search.

Inclusion and exclusion criteria

Though reality criteria are mainly applied in judicial contexts to ensure a victim's testimony is admitted as valid evidence, a review of the literature reveals they have been also applied to both witnesses and offenders so both populations were included as the studies were numerically sufficient for performing a meta-analysis. The concept of adult in the judicial context is associated with being 18 years of age, and the vast majority of studies endorsed this legal age; notwithstanding, in a few studies the legal age was set at 17 years. Since this difference in age has no effect on the capacity to give testimony either on cognitive or legal grounds, the studies with 17-year-old adult populations were included. The inclusion criteria for primary studies were that the effect sizes of the reality criteria analysed for discriminating between truthful and fabricated statements were reported, or in their absence, the statistical data allowing for them to be computed, including studies with errors in data analysis that nonetheless enabled the effect sizes to be computed.

The exclusion criteria were data derived from a unit of analysis which was not the statement, or when two CBCA criteria were combined into one new criterion (failing the 'mutual' exclusion requirement for creating methodic categorical systems). As for the additional criteria, data that were not formulated as additional to CBCA or were specific to only one context were excluded. Likewise, the *duplicate publication of data* was eliminated, but not the piecemeal (independent data).

Finally, 39 primary studies fulfilling the inclusion and exclusion criteria were selected. Total CBCA score was calculated using 31 effect sizes, whereas as for the individual criteria, the effect sizes ranged from 5 for criteria 10 and 19, to 35 for criteria 3 and 8.

Procedure

The procedure observed the stages in meta-analysis by Botella and Gambara (2006). Having performed the literature search and selected the studies for the present meta-analysis, these were coded according to variables that have been found to have a moderating role i.e., previous studies (Fariña, Arce, & Real, 1994; Höfer et al., 1993; Raskin et al., 1991; Volbert & Steller, 2014; Vrij, 2005); previous meta-analysis with a child population (Amado et al., 2015); the research paradigm (field vs. experimental studies) under the US law of precedence (Daubert v. Merrell Dow Pharmaceuticals, 1993); compliance with the Daubert standard publication criterion (DSPC) i.e., peer-reviewed journals for evidence to be admitted as scientifically valid legal evidence; the lie coaching condition in reality criteria; and the version of the categorical system (full reality criteria vs. 14-criteria version). Having applied a procedure of successive approximations for the coding of the primary studies (Fariña, Arce, & Novo, 2002), the following moderators were detected: status of the declarant i.e., testimony target (victim, offender, or witness); event target (self-experienced events or video-observed events/witness), judicial context i.e., case type.

As some researchers had renamed the original criteria (Steller & Köhnken, 1989), a *Thurstone style* evaluation was used consisting of 10 judges who evaluated the degree of overlapping between the original and reformulated criterion. When the interval between Q1 and Q3 was within the region of criteria independence it was considered additional criteria, whereas when it was in the region of dependence with the original, it was considered original criteria.

The coding of the studies and moderators carried out by two independent researchers showed total coincidence (kappa = 1).

Data analysis

The effect sizes were taken directly from the primary studies when these were disclosed, or the effect size d was computed using the means, and standard deviations/standard error of the mean (Cohen's d when $N_1 = N_2$ and Glass's Δ when $N_1 \neq N_2$), the t value, or the F value. When the results were expressed as proportions the effect size δ (Hedges & Olkin, 1985) was equivalent to Cohen's d, whereas when they were expressed in 2X2 contingency tables, the phi obtained was transformed into Cohen's d.

The meta-analysis was performed in accordance with the procedure of Hunter and Schmidt (2015), the unit of analysis (n) was the number of statements, the effect sizes were weighted for sample size i.e., the number of statements (d_w), and effect sizes were corrected for criterion reliability (δ).

The differences between effect sizes were estimated using the difference between correlations (q statistic; Cohen, 1988), by transforming the effect sizes into correlations. In the study of moderators the average criteria for each moderator was computed.

In order to estimate the practical utility of the results of the meta-analysis in forensic settings, three recommended statistics were employed (Amado et al., 2015): U1, the Binomial Effect Size Display (BESD), and the Probability of Superiority (PS).

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Criterion reliability

Not all of the primary studies provided data on inter-rater reliability, or agreement for the reality criteria and for the total CBCA score. Moreover, the informed reliability coefficients varied among studies, and in some studies, several were reported, in which case those approximating the results obtained by Anson, Golding, and Gully (1993) and Horowitz et al. (1997) were taken. Owing to the lack of data on coding reliability in studies on specific criteria, average reliability was estimated for the criteria and for the total CBCA score, bearing in mind that reliability is different for the criteria than for the instrument (Horowitz et al., 1997). Reliability was estimated on the basis of reliability coefficients, since agreement indexes do not measure reliability. Thus, on the basis of 172 reliability coefficients of CBCA criteria in the primary studies, the average reliability for CBCA reality criteria was $r = .61$ ($EEM = .020$, 95%CI = 0.57, 0.65); and for the total CBCA score the Spearman-Brown prediction formula obtained an $r = .97$. Moreover, the average reliability for the proposed additional reality criteria was calculated using 7 reliability coefficients with an r of .74 ($EEM = 0.041$, 95%CI = 0.66, 0.82). The low average reliability observed was sometimes considered as a methodological weakness of the system. Nevertheless, this potential methodological deficiency is corrected for criteria unreliability in Hunter and Schmidt's (2015) meta-analytical procedure.

Results[1]

Study of outliers

An analysis and initial control of outliers was carried out in each of the reality criteria, and the total CBCA score and conditions. The criterion chosen was the ±3*IQR (extreme cases) of the simple size weighted mean effect size, given that the results of more conservative criteria such as ±1.5*IQR or ±2*SD*, eliminated more than 10% of the effect sizes, indicating they were more probably moderators than outliers (Tukey, 1960).

Global meta-analysis of reality criteria in adults

The results (Table 2) show a positive (between criteria presence and statement reality), and significant (when the confidence interval had no zero, indicating the effect size was significant) mean true effect size (δ) for the CBCA reality criteria, with the exception of 'self-deprecation' and 'pardoning the perpetrator' criteria, and the total CBCA score. Nevertheless, these results are not generalizable (criteria 10 and 19 were affected by a second order sampling error, so the results were invalid for this estimate) to future samples (when the credibility interval had zero, indicating the effect size was not generalizable to 80% of other samples). For the additional criteria (Höfer et al., 1993), the meta-analysis revealed a positive, significant and generalizable mean true effect size for 'reporting style' and 'display of insecurities' criteria. The mean true effect size for the 'repetitions' criterion was *negative* and significant, confirming it was not reality criteria, but no generalizable. As for the 'providing reasons for lack of memory' and 'clichés' criteria, a non-significant mean true effect size was found. The criteria repetitions and clichés were related to fabricated events, that is, they were not reality criteria in themselves, so they were not included in successive analyses. The 75% rule and the credibility interval (Hunter & Schmidt, 2015) warranted the study de moderators.

Study of moderators

The study of moderators (criteria average as dependent variable; Table 3) showed a positive and significant mean true effect size, but not generalizable, in all of the moderators analysed. As for the magnitude of the effect sizes, excluding the witness condition with a medium effect size ($\delta > 0.50$), all were small ($0.20 > \delta < 0.50$). Arce and Fariña (2009) have suggested (and designed) the specifications of categorical systems based on bottom-up rather than 'top-down' procedures to ensure only categories that effectively discriminate between memories of experienced events and fabricated memories form part of the system. This maximizes the efficacy of the resulting categorical system by eliminating the noise produced by non-discriminating 'top-down' categories. Thus, the meta-analyses were repeated with the categories of content analysis with significant effect size i.e., the confidence interval for d did not contain zero. The results (Table 3) revealed a significant increase in the effect size of field studies, $q_c = .119$, $p < .05$ (one-tailed; a larger effect size was expected with significant criteria), thus the effect size was significantly larger with significant criteria. Moreover, for significant criteria, the results (not all of the reality criteria were generalizable) became generalizable (the credibility interval had no zero). As for the experimental studies on the remaining moderators, the results did not corroborate the Hypothesis as the reality categories had been initially or subsequently screened to eliminate the non-significant ones.

The meta-analytical technique does not take into account the theoretical foundations or the reliability of the studies included in the original theories, that is, all of the studies on categories of reality are included. Moreover, the experimental designs of studies on witnesses are not really on witnesses of self-experienced events, but on non-self-experienced events i.e., video-observed events (watched on video, not involving self-experienced events) that do not fulfil the original theory hypothesizing that reality criteria discern between memories of self-experienced real-life events and fabricated or fictitious memories. Only one of the studies on witnesses involved self-experienced events (Gödert, Gamer, Rill, & Vossel, 2005), and for the total reality criteria, were found to discriminate significantly real witness from real offenders giving false testimony, $d = 0.59$, $1\text{-}\beta = .78$, and from uninvolved participants, $d = 0.83$, $1\text{-}\beta = .96$. Nevertheless, reality criteria also discriminated between both memories of video-observed events and fabricated events. The only study (Lee, Klaver, & Hart, 2008) comparing memories of self-experienced events (truth condition) and video-observed events (lie condition)

[1] Additional results and resources at http://www.researchgate.net/profile/Ramon_Arce

Table 2 Results of global meta-analysis for individual reality criteria and total CBCA score, and additional criteria.

CBCA Criterion	*k*	*n*	d_w	SD_d	SD_{pre}	SD_{res}	δ	$SD_δ$	%Var	95% CI_d	80% $CV_δ$
1. Logical structure	30	2,265	0.48	0.6990	0.2503	0.6527	0.62	0.8493	13	0.40, 0.56	−0.46, 1.71
2. Unstructured production	27	1,987	0.53	0.9241	0.2570	0.8876	0.69	1.1551	8	0.45, 0.61	−0.79, 2.17
3. Quantity of details	35	2,714	0.55	0.8294	0.2529	0.7899	0.71	1.0279	9	0.47, 0.63	−0.60, 2.03
4. Contextual embedding	29	2,137	0.19	0.6169	0.2372	0.5868	0.24	0.7411	15	0.11, 0.27	−0.70, 1.19
5. Description of interactions	29	2,243	0.27	0.3742	0.2349	0.2912	0.36	0.3790	39	0.19, 0.35	−0.13, 0.84
6. Reproduction conversations	34	2,528	0.34	0.4990	0.1780	0.4662	0.44	0.6067	13	0.26, 0.42	−0.33, 1.22
7. Unexpected complications	29	1,956	0.25	0.3788	0.2498	0.2847	0.32	0.3705	43	0.17, 0.33	−0.15, 0.79
8. Unusual details	35	2,441	0.31	0.6532	0.2489	0.6039	0.41	0.7859	14	0.23, 0.39	−0.59, 1.42
9. Superfluous details	27	1,863	0.14	0.5676	0.2437	0.5126	0.18	0.6670	18	0.04, 0.24	−0.67, 1.04
10. Details misunderstood	5	376	0.22	0.1208	0.2357	0.0000	0.28	0.0000	100	0.02, 0.42	0.28
11. External associations	22	1,612	0.26	0.4781	0.2405	0.3268	0.34	0.5376	25	0.16, 0.36	−0.35, 1.02
12. Subjective mental state	28	2,170	0.18	0.4843	0.2312	0.4256	0.23	0.5538	23	0.10, 0.26	−0.47, 0.94
13. Perpetrator's mental state	31	2,232	0.09	0.6212	0.2376	0.5741	0.11	0.7470	15	0.01, 0.17	−0.84, 1.07
14. Spontaneous corrections	29	1,842	0.16	0.5276	0.2545	0.4622	0.20	0.6014	23	0.06, 0.26	−0.56, 0.97
15. Admitting lack of memory	34	2,305	0.25	0.3823	0.2494	0.2897	0.32	0.3770	42	0.17, 0.33	−0.16, 0.80
16. Doubts one's testimony	26	1,755	0.20	0.4521	0.2478	0.3781	0.26	0.4919	30	0.10, 0.30	−0.37, 0.89
17. Self-deprecation	13	948	0.04	0.4629	0.2354	0.3985	0.05	0.5186	26	−0.08, 0.16	−0.61, 0.71
18. Pardoning the perpetrator	8	680	−0.02	0.2796	0.2178	0.1753	−0.02	0.2281	61	−0.18, 0.14	−0.31, 0.27
Details characteristics offence	5	562	0.28	0.1894	0.1966	0.0000	0.36	0.0000	100	0.12, 0.44	0.36
TOTAL CBCA SCORE	31	2,124	0.55	0.6759	0.2475	0.6290	0.56	0.6386	13	0.47, 0.63	−0.25, 1.37
Average (original criteria)	46	3,223	0.25	0.5032	0.2368	0.4269	0.33	0.5614	32	0.17, 0.33	−0.39, 1.05
Additional Criteria											
Reporting style	3	357	0.41	0.2030	0.1874	0.0781	0.48	0.0909	85	0.20, 0.63	0.36, 0.59
Display insecurities	3	297	0.67	0.5540	0.2111	0.5122	0.78	0.5965	14	0.43, 0.90	0.01, 1.54
Providing reasons lack memory	4	447	0.15	0.2877	0.1902	0.2158	0.18	0.2514	44	−0.03 0.33	−0.14, 0.50
Clichés	3	267	−0.18	0.5145	0.2134	0.4682	−0.21	0.5452	17	−0.41, 0.05	−0.90, 0.49
Repetitions	4	417	−0.47	0.5851	0.2011	0.5494	−0.54	0.6399	12	−0.67, −0.27	−1.36, 0.27
Average (original + additional)	46	3,223	0.27	0.4821	0.2313	0.4053	0.34	0.5275	34	0.19, 0.35	−0.33, 1.01

Note. *k* = number of studies; *n* = total sample size; d_w = effect size weighted for sample size; SD_d = observed standard deviation of *d*; SD_{pre} = standard deviations of observed *d*-values corrected from all artifacts; SD_{res} = standard deviation of observed *d*-values after removal of variance due to all artifacts; δ = effect size corrected for criterion reliability; $SD_δ$ = standard deviation of δ; %Var = variance accounted for by artifactual errors; 95% CI_d = 95% confidence interval for *d*; 80% $CV_δ$ = 80% credibility interval for δ.

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Table 3 Results of the meta-analysis of moderators.

Moderator	*k*	*n*	d_w	SD_d	SD_{pre}	SD_{res}	δ	$SD_δ$	%Var	95% CI_d	80% $CV_δ$
CBCA significant criteria (17)	46	3,223	0.27	0.5187	0.2380	0.4433	0.36	0.5835	31	0.19, 0.35	−0.39, 1.11
14-criteria version	45	3,143	0.28	0.5567	0.2394	0.4906	0.36	0.6465	25	0.22, 0.34	−0.47, 1.19
Daubert standard publication criterion											
All criteria (22)	35	2,256	0.20	0.4575	0.2407	0.3733	0.26	0.4786	39	0.12, 0.28	−0.35, 0.87
Self-experienced events											
All criteria (22)	34	2,277	0.26	0.4647	0.2371	0.3879	0.33	0.5022	40	0.18, 0.34	−0.31, 0.97
Non self-experienced events (witness)											
All criteria (13)	11	625	0.39	0.5835	0.2707	0.5032	0.51	0.6548	65	0.23, 0.55	−0.33, 1.35
Offenders											
All criteria (21)	11	1,067	0.27	0.4662	0.2024	0.3743	0.35	0.4975	41	0.15, 0.39	−0.29, 0.99
Victims											
All criteria (18)	11	840	0.27	0.4781	0.2355	0.4012	0.35	0.5221	35	0.13, 0.41	−0.32, 1.02
Field studies											
All field studies (18)	6	422	0.34	0.4948	0.2385	0.4153	0.45	0.5404	35	0.14, 0.54	−0.24, 1.14
Significant criteria (10)[a]	6	422	0.53	0.4774	0.2458	0.3834	0.69	0.4989	42	0.33, 0.73	0.05, 1.33
Sexual and IPV field studies											
All criteria (17)[b]	5	263	0.67	0.3587	0.2871	0.1957	0.87	0.2459	72	0.41, 0.92	0.55, 1.18
Significant criteria (15)[c]	5	263	0.74	0.3654	0.2892	0.2134	0.96	0.2478	72	0.48, 0.99	0.64, 1.28
Experimental studies											
All criteria (22)	39	2,721	0.25	0.4497	0.2336	0.3934	0.32	0.4933	37	0.17, 0.33	−0.31, 0.95

Note.
[a] Significant criteria (CBCA criteria, as for additional criteria, studies were insufficient): 1-3, 5-8, 11, 12 and 19.
[b] Significant criteria (CBCA criteria): 1-9, 11-18.
[c] Significant criteria (CBCA criteria): 1-9, 11-12, 14-17.

found CBCA reality criteria, and the total CBCA score discriminated significantly between both memories in line with the Undeutsch Hypothesis.

The high observed variability in effect sizes in field studies, which was mostly due to one study alone, suggested differences in experimental design (the crime context in this study was found to be different to the other studies). As the effect of context has been hypothesized (Köhnken, 1996; Volbert & Steller, 2014), and found (Arce, Fariña, & Vilariño, 2010; Vilariño, Novo, & Seijo, 2011) to mediate the discriminating efficacy of reality categories, the meta-analysis was repeated in field studies on sexual offences and intimate partner violence (IPV) cases (crimes committed in the privacy of one's home according to the categorization of Arce & Fariña, 2005). The results showed a positive, significant and generalizable (not generalizable in all field studies) mean true effect size for studies under this condition. Moreover, the magnitude of the effect sizes were significantly larger in sexual offences and IPV cases than in all field studies in all the reality criteria (0.45 for all field studies vs. 0.87 for sexual offences and IPV cases), $q_c = .199$, $p < .01$ (one-tailed; a higher effect size was expected in specific criminal contexts), and in the significant criteria, $q_c = .168$, $p < .05$ (0.69 vs. 0.96). Likewise, reality criteria were significantly more efficacious, $q_c = .2622$, $p < .01$, in sexual offences and IPV cases than in all other types of cases (0.32 vs. 0.87).

Results (meta-analysis could not be performed because *k*s and *n*s were insufficient and research designs incomparable) for the comparison between statements of participants instructed to lie (lie coaching condition) with truthful statements were inconclusive[2] in relation to the effectiveness of reality criteria to discriminate between truthful and false statements.

Discussion

The following conclusions may be drawn from the results of this study. First, the results confirmed the Undeutsch Hypothesis, that is, reality criteria discriminated between memories of self-experienced and fabricated events [File Drawer Analysis (FDA): to bring down this hypothesis to a trivial effect (McNatt, 2000), .05, for the average of the CBCA criteria, it would be necessary 184 studies with null effect; Hunter & Schmidt, 2015. It is unlikely to happen]. Besides fulfilling the DSPC, this Hypothesis was also valid for memories of victims/claimants and offenders (for witness of self-experienced events further research is required); and robust in both experimental studies (high internal validity), and field studies (high external validity). Notwithstanding, the reality criteria also discriminated between memories of video-observed events i.e., non-self-experienced events,

[2] Conclusions in the primary studies about non-significant results are inconclusive as the statistical power, $1\text{-}\beta < .80$, is insufficient to conclude ($d = -0.44$, $1\text{-}\beta = .41$, Bogaard, Meijer, & Vrij, 2013; $d = 0.37$, $1\text{-}\beta = .26$, Vrij, Akehurst, Soukara, & Bull, 2002; $d = 0.11$, $1\text{-}\beta = .06$, Vrij, Kneller, & Mann, 2000).

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

and fabricated events for which the Hypothesis was not formulated, and research findings are inconclusive as to the validity of the Hypothesis with lie coached subjects. Second, though the results validated CBCA as a categorical system based on the Undeutsch Hypothesis, neither were all of the criteria validated, nor were they generalizable, and some even contradicted the Hypothesis. Thus, these criteria can be used neither in all types of contexts, nor indiscriminately. Both versions of the CBCA (all criteria or 14 criteria) were exactly the same ($\delta = 0.36$) in discriminating between memories of self-experienced and fabricated events. Though the results open the door to the inclusion of new reality criteria, additional criteria have been proposed that fail to fulfil the Undeutsch Hypothesis (significant negative effect sizes i.e., not reality criteria), so they cannot be included in the CBCA. Third, in field studies the discriminating power of reality criteria was significantly higher in sexual offences and IPV cases (FDA: to bring the results in sexual offences and IPV cases down to a trivial effect, it would be necessary 62 and 69 studies with null effect for all criteria and significant criteria, respectively. It is unlikely to occur) in comparison to other types of contexts (FDA: to reduce the efficacy of the reality criteria to discriminate between real and fabricated memories in any context of field studies to a trivial effect it would be necessary 35 studies with null effect. It is unlikely to happen). Succinctly, the areas of both populations do not overlap in 54% ($U1 = 0.54$), that is, they were totally independent, thus the efficacy of the reality criteria in discriminating between memories of self-experienced and fabricated events in sexual and IPV cases was total in 54% of the evaluations of credibility. Moreover, 75% of statements of self-experienced events contained more reality criteria than fabricated events (probability of superiority, $PS = 0.75$), the probability of false positives was 28% (BESD). These results were highly robust i.e., not only establishing a positive and significant relation between reality criteria and true statements, but were also generalizable to all types of sexual offences and IPV cases, and were homogeneous (i.e., subject to little variability since the correlation between the effect sizes was .72).

As for the implications for forensic practice, the results of the present meta-analysis reveal that the reality criteria were statistically effective for discriminating between memories of self-experienced and fabricated events, but this does not imply they are directly generalizable to forensic practice. Even under the best discriminating conditions i.e., field studies in sexual and IPV cases, the probability of false positives may reach .22, whilst this probability must be zero in forensic settings (Arce, Fariña, & Fraga, 2000). In general, only significant reality criteria i.e., scientifically attested evidence, were admissible for forensic practice (see note of Table 3), since the results were generalizable, whereas for all criteria they were not. However, as the credibility interval lower limit was 0.05, the practical utility of these categories was almost negligible ($PS = .51$), that is, in only 51% of true statements there were more reality criteria than in false statements, and under what specific conditions this contingency occurred remains unknown. However, the credibility interval lower limit of the reality criteria applied to cases of sexual offences and IPV, which were also generalizable both in terms of all the criteria and the significant criteria, was larger, $PS = .73$ and .75 (Hedges and Olkin's $\delta = 0.59$ and 0.65, test value = .51), for all the reality criteria and the significant criteria, respectively. However, these conclusions are not directly applicable to forensic practice as the decision criteria which in the forensic context must the 'strict decision criterion' in which a type II error (classify a false statement as true) is not admissible i.e., must be equal to zero. Regarding the strict decision criterion, Arce et al. (2010) found up to 13 CBCA reality criteria in fabricated statements of IPV cases, which means that at least 14 reality criteria would have to be detected in a statement to conclude that the testimony was true, with a correct classification of true positives (true statements classified as such) of 36%. Succinctly, the CBCA reality criteria were a poor tool for assigning the credibility of IPV victim testimony. Thus, to enhance efficacy, CBCA reality criteria must be complemented with additional criteria. In this line, Arce and Fariña (2009), Vilariño (2010) and Vilariño et al. (2011) combined CBCA and SRA criteria, memory attributes, and additional reality criteria specific to IPV cases derived from real statements (judicial judgements as ground truth), to create and validate a categorical system specific for IPV cases, including sexual offences, with a strict decision criterion to reduce the rate of false negatives to 2%. In any way, only results with a strict decision criterion can be translated into forensic practice.

In terms of future research, the results of the present meta-analysis underscored the need for further studies with experimental designs assessing the efficacy of reality criteria in discriminating between memories of self-experienced events and video witnessed non-self-experienced events; between self-experienced witnessed events vs. fabricated events; between memories of participants coached to lie and honest; and research driven to find new reality categories (bottom-up), mainly for a specific context i.e., crime victimization.

This meta-analysis is subject to the following limitations. First, previous publications have biased the results in that the non-significant results or predictably inefficacious categories were eliminated (favouring the validation of the Undeutsch Hypothesis). Second, the feigning methodology (experimental studies) had no proven external validity (Sarwar, Allwood, & Innes-Ker, 2014), but only 'face validity' (Konecni & Ebbesen, 1992). Third, for some experimental literature, statements are insufficient material for reality content analysis (Köhnken, 2004), which favours the rejection of the Undeutsch Hypothesis. Fourth, there was no control on the effects of the interviewer on the contents of the statement, or on the reliability of the interviews, which were often carried out by poorly trained interviewers. Fifth, few studies comply with SVA standards that are a requirement for applying CBCA. Sixth, the results of some meta-analysis may be subject to a degree of variability, given that $N_s < 400$, did not guarantee stability in sample estimates (Hunter & Schmidt, 2015). Seventh, primary studies did not estimate the reliability of the codings, thus results' reliability is uncertainty.

Funding

This research has been sponsored by a grant of the Spanish Ministry of Economy and Competitiveness (PSI2014-53085-R).

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

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*, 3–12. http://dx.doi.org/10.1016/j.ejpal.2014.11.002

Anson, D. A., Golding, S. L., & Gully, K. J. (1993). Child sexual abuse allegations: Reliability of Criteria-Based Content Analysis. *Law and Human Behavior*, *17*, 331–341. http://dx.doi.org/10.1007/BF01044512

Arce, R., & Fariña, F. (2005). Peritación psicológica de la credibilidad del testimonio: la huella psíquica y la simulación: El Sistema de Evaluación Global (SEG). *Papeles del Psicólogo*, *26*, 59–77.

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). Madrid: Biblioteca Nueva.

Arce, R., & Fariña, F. (2012). Psicología social aplicada al ámbito jurídico. In A. V. Arias, J. F. Morales, E. Nouvilas, & J. L. Martínez (Eds.), *Psicología social aplicada* (pp. 157–182). Madrid: Panamericana.

Arce, R., Fariña, F., & Fraga, A. (2000). Género y formación de juicios en un caso de violación [Gender and juror judgment making in a case of rape]. *Psicothema*, *12*, 623–628.

*Arce, R., Fariña, F., & Vilariño, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género. *Intervención Psicosocial*, *19*, 109–119. http://dx.doi.org/10.5093/in2010v19n2a2

Berliner, L., & Conte, J. R. (1993). Sexual abuse evaluation: Conceptual and empirical obstacles. *Child Abuse and Neglect*, *17*, 111–125. http://dx.doi.org/10.1016/0145-2134(93)90012-T

Botella, J., & Gambara, H. (2006). Doing and reporting a meta-analysis. *International Journal of Clinical & Health Psychology*, *6*, 425–440.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: LEA.

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

Fariña, F., Arce, R., & Novo, M. (2002). Heurístico de anclaje en las decisiones judiciales [Anchorage in judicial decision making]. *Psicothema*, *14*, 39–46.

Fariña, F., Arce, R., & Real, S. (1994). Ruedas de identificación: De la simulación y la realidad. *Psicothema*, *6*, 395–402.

*Gödert, H. W., Gamer, M., Rill, H. G., & Vossel, G. (2005). Statement Validity Assessment: Inter-rater reliability of Criteria-Based Content Analysis in the mock-crime paradigm. *Legal and Criminological Psychology*, *10*, 225–245. http://dx.doi.org/10.1348/135532505X52680

Hans, V. P., & Vidmar, N. (1986). *Judging the jury*. New York: Plenum Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Höfer, E., Köhnken, G., Hanewinkel, R., & Bruhn, C. (1993). *Diagnostik und attribution von glaubwürdigkeit*. Unpublished final report. University of Kiel, Germany.

Horowitz, S. W., Lamb, M. E., Esplin, P. W., Boychuk, T. D., Krispin, O., & Reiter-Lavery, L. (1997). Reliability of criteria-based content analysis of child witness statements. *Legal and Criminological Psychology*, *2*, 11–21. http://dx.doi.org/10.1111/j. 2044-8333.1997.tb00329.x

Hunter, J. E., & Schmidt, F. L. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

*Juárez, J.R., Mateu, A., & Sala, E. (2007). *Criterios de evaluación de la credibilidad en las denuncias de violencia de género*. Retrieved from http://justicia.gencat.cat/web/.content/documents/arxius/sc-3-143-07-cas.pdf.

Köhnken, G. (1996). Social psychology and the law. In G. R. Semin, & K. Fiedler (Eds.), *Applied social psychology* (pp. 257–282). Thousand Oaks, CA: Sage.

Köhnken, G. (2004). Statement Validity Analysis and the detection of the truth. In P. A. Granhag, & L. A. Strömwall (Eds.), *The detection of deception in forensic contexts* (pp. 41–63). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511490071.003

Konecni, V. J., & Ebbesen, E. B. (1992). Methodological issues on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 413–423). Berlin, Germany: Walter de Gruyter.

*Lee, Z., Klaver, J. R., & Hart, S. D. (2008). Psychopathy and verbal indicators of deception in offenders. *Psychology, Crime & Law*, *14*, 73–84. http://dx.doi.org/10.1080/10683160701423738

McNatt, D. B. (2000). Ancient Pygmalion joins contemporary management: A meta-analysis of the result. *Journal of Applied Psychology*, *85*, 314–322. http://dx.doi.org/10.1037/0021-9010.85.2.314

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.

Raskin, D.C., Esplin, F.W., & Horowitz, S. (1991). *Investigative interviews and assessment of children in sexual abuse cases*. Unpublished manuscript, Department of Psychology, University of Utah, Utah.

Sarwar, F., Allwood, C. M., & Innes-Ker, A. (2014). Effects of different types of forensic information on eyewitness' memory and confidence accuracy. *European Journal of Psychology Applied to Legal Context*, *6*, 17–27. http://dx.doi.org/10.5093/ejpalc2014a3

Steller, M., & Böhm, C. (2006). Cincuenta años de jurisprudencia del Tribunal Federal Supremo alemán sobre la psicología del testimonio. Balance y perspectiva. In T. Fabian, C. Böhm, & J. Romero (Eds.), *Nuevos caminos y conceptos en la psicología jurídica* (pp. 53–67). Münster, Germany: 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). New York: Springer-Verlag.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In I. Olkin, J. G. Ghurye, W. Hoeffding, W. G. Madoo, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 448–485). Stanford, CA: Stanford University Press.

Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von aussagen. In U. Undeutsch (Ed.), *Handbuch der psychologie, Vol. 11: Forensische psychologie* (pp. 26–181). Göttingen, Germany: Hogrefe.

*Vilariño, M. (2010). *¿Es posible discriminar declaraciones reales de imaginadas y huella psíquica real de simulada en casos de violencia de género*? (Doctoral thesis, Universidad de Santiago de Compostela, Spain). Retrieved from http://hdl.handle.net/10347/2831.

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.

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*, 207–220. http://dx.doi.org/10.1027/1016-9040/a000200

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, *11*, 3–41. http://dx.doi.org/10.1037/1076-8971.11.1.3

* Indicates the primary studies included in the meta-analysis.

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). Chichester, England: John Wiley and Sons.

Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching and social skills on CBCA scores. *Law and Human Behavior*, *26*, 261–283. http://dx.doi.org/10.1023/A:1015313120905

*Vrij, A., Kneller, W., & Mann, S. (2000). The effect of informing liars about Criteria-Based Content Analysis on their ability to deceive CBCA-raters. *Legal and Criminological Psychology*, *5*, 57–70. http://dx.doi.org/10.1348/135532500167976

Further reading

*Akehurst, L., Easton, S., Fullar, E., Drane, G., Kuzmin, K., & Litchfield, S. (2015). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psichology*, http://dx.doi.org/10.1111/lcrp.12079. Advance online publication

*Beaulieu-Prévost, D.;1; (2001). *Analyse de validité de la déclaration (SVA), mensonge et faux souvenirs: Validité et efficacité chez les adultes*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. MQ60609).

*Bensi, L., Gambetti, E., Nori, R., & Giusberti, F. (2009). Discerning truth from deception: The sincere witness profile. *European Journal of Psychology Applied to Legal Context*, *1*, 101–121.

*Biland, C., Py, J., & Rimboud, S. (1999). Evaluer la sincérité d'un témoin grâce à trois techniques d'analyse, verbales et non verbale. *European Review of Applied Psychology*, *49*, 115–122.

*Blandón-Gitlin, I., Pezdek, K., Lindsay, D. S., & Hagen, L. (2009). Criteria-Based Content Analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, *23*, 901–917. http://dx.doi.org/10.1002/acp.1504

*Bogaard, G., Meijer, E. H., & Vrij, A. (2013). Using an example statement increases information but does not increase accuracy of CBCA, RM, and SCAN. *Journal of Investigative Psychology and Offender Profiling*, *11*, 151–163. http://dx.doi.org/10.1002/jip.1409

*Caso, L., Vrij, A., Mann, S., & de Leo, G. (2006). Deceptive responses: The impact of verbal and non-verbal countermeasures. *Legal and Criminological Psychology*, *11*, 99–111. http://dx.doi.org/10.1348/135532505X49936

*Critchlow, N. (2011). *Applying Criteria Based Content Analysis to assessing the veracity of rape statements* (Unpublished doctoral dissertation). Manchester Metropolitan University, Manchester, UK.

*Critchlow, N. (2011). [A field validation of CBCA when assessing authentic police rape statements: evidence for discriminant validity to prescribe veracity to adult narrative]. Unpublished raw data.

*Dana-Kirby, L. (1997). *Discerning truth from deception: Is Criteria-Based Content Analysis effective with adult statements?* (Unpublished doctoral thesis). University of Oregon, Oregon.

*Evans, J., Michael, S. W., Meissner, C. A., & Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a psychologically based credibility assessment tool. *Journal of Applied Research in Memory and Cognition*, *2*, 33–41. http://dx.doi.org/10.1016/j.jarmac.2013.02.002

*Godoy, V., & Higueras, L. (2008). El análisis de contenido basado en criterios (CBCA) y la entrevista cognitiva aplicados a la credibilidad del testimonio en adultos. In F. Rodríguez, C., Bringas, F. Fariña, R. Arce, & A. Bernardo (Eds.), *Psicología Jurídica: Entorno judicial y delincuencia* (pp. 117-125). Retrieved from http://gip.uniovi.es/T5EJD.pdf.

*Honts, C.R., & Devitt, M.K. (1993). *Credibility Assessment of Verbatim Statements (CAVS)*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a271575.pdf.

*Johnston, S., Candelier, A., Powers-Green, D., & Rahmani, S. (2014). Attributes of truthful versus deceitful statements in the evaluation of accused child molesters. *Sage Open*, *4*, 1–10. http://dx.doi.org/10.1177/2158244014548849

*Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, *80*, 671–684. http://dx.doi.org/10.1037/0021-9010.80.6.671

*Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, *20*, 129–146. http://dx.doi.org/10.1111/lcrp.12017

*Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, *37*, 1371–1382. http://dx.doi.org/10.1016/j.paid.2004.01.007

*Porter, S., & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues to deception in the interrogation context. *Law and Human Behavior*, *20*, 443–458. http://dx.doi.org/10.1007/BF01498980

*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. http://dx.doi.org/10.1023/A:1022344128649

*Rassin, E., & van-der-Sleen, J. (2005). Characteristics of true versus false allegations of sexual offences. *Psychological Reports*, *97*, 589–598. http://dx.doi.org/10.2466/pr0.97.2.589-598

*Schelleman-Offermans, K., & Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *Journal of Investigative Psychology and Offender Profiling*, *7*, 247–260. http://dx.doi.org/10.1002/jip.121

*Sporer, S. L. (1997). The less travelled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology*, *11*, 373–397. http://dx.doi.org/10.1002/(SICI)1099-0720(199710)11:5<373::AID-ACP461>3.0.CO;2-0

*Ternes, M. (2009). *Verbal credibility assessment of incarcerated violent offenders' memory reports*. University of British Columbia: Vancouver. Unpublished doctoral thesis.

*Vrij, A., Akehurst, L., Soukara, R., & Bull, R. (2004). Detecting deceit via analyses of verbal and nonverbal behavior in children and adults. *Human Communication Research*, *30*, 8–41. http://dx.doi.org/10.1111/j. 1468-2958.2004.tb00723.x

*Vrij, A., Evans, H., Akehurst, L., & Mann, S. (2004). Rapid judgements in assessing verbal and nonverbal cues: Their potential for deception researchers and lie detection. *Applied Cognitive Psychology*, *18*, 283–296. http://dx.doi.org/10.1002/acp.964

*Vrij, A., Edward, K., & Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *British Journal of Psychology*, *92*, 373–389. http://dx.doi.org/10.1348/000712601162248

*Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, *24*, 239–263. http://dx.doi.org/10.1023/A:1006610329284

*Vrij, A., Mann, S., & Edward, K. (2000). I think it was a green scarf but I am not sure. Raising doubts about one's own testimony during lying and truth telling. In A. Czerederecka, T. Jaskiewicz-Obydzinska, & J. Wójcikiewicz (Eds.), *Forensic psychology and law. Traditional questions and new ideas* (pp. 205–207). Krakow: Institute of forensic research, Poland.

*Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime and Law*, *5*, 203–215. http://dx.doi.org/10.1080/10683169908401767

Document downloaded from http://www.elsevier.es, day 01/07/2016. This copy is for personal use. Any transmission of this document by any media or format is strictly prohibited.

*Vrij, A., & Mann, S. (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime and Law*, *12*, 337–349. http://dx.doi.org/10.1080/10683160500129007

*Vrij, A., Mann, S., Kristen, S., & Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law and Human Behavior*, *31*, 449–518. http://dx.doi.org/10.1007/s10979-006-9066-4

*Willén, R. M., & Strömwall, L. A. (2011). Offender's uncoerced false confessions: A new application of statement analysis? *Legal and Criminological Psychology*, *17*, 346–359. http://dx.doi.org/10.1111/j.2044-8333.2011.02018.x

*Wojciechowski, B. W. (2014). Content analysis algorithms: An innovative and accurate approach to statement veracity assessment. *European Poligraph*, *8*, 119–128. http://dx.doi.org/10.2478/ep-2014-0010