

Criteria-based Content Analysis in True and Simulated Victims with Intellectual Disability

Antonio L. Manzanero^a, M. Teresa Scott^b, Rocío Vallet^a, Javier Aróztegui^a, and Ray Bull^c

^aUniversidad Complutense de Madrid, Spain; ^bUniversidad del Desarrollo, Chile; ^cDerby University, United Kingdom

ARTICLE INFO

Article history:

Received 28 August 2018
Accepted 26 November 2018
Available online 9 January 2019

Keywords:

Credibility assessment
Intuitive judgments
Intellectual disability
CBCA Content criteria
Big data

ABSTRACT

The aims of the present study were to analyse people's natural ability to discriminate between true and false statements provided by people with intellectual disability ($IQ_{TRUE} = 62.00$, $SD = 10.07$; $IQ_{FALSE} = 58.41$, $SD = 8.42$), and the differentiating characteristics of such people's statements using criteria-based content analysis (CBCA). Thirty-three people assessed 16 true statements and 13 false statements using their normal abilities. Two other evaluators trained in CBCA evaluated the same statements. The natural evaluators differentiated between true and false statements with somewhat above-chance accuracy, even though error rate was high (38.19%). That lay participants could not effectively discriminate between false and true statements demonstrates that such assessments cannot be considered useful in a forensic context. The CBCA technique did discriminate at a better level than intuitive judgements. However, of the 19 criteria, only one significantly discriminated. More procedures specifically adapted to the abilities of people with intellectual disabilities are thus required. The presence of structured production, quantity of details, characteristics details and unexpected complications increased the probability that a statement would be considered true by non-expert evaluators. The classification made by the non-expert evaluators was independent of the participants' IQ. A big data analysis is performed in search for better classification quality.

El análisis de contenido basado en criterios en víctimas reales y simuladas con discapacidad intelectual

RESUMEN

Este trabajo tiene como objetivos examinar la capacidad natural para discriminar entre declaraciones verdaderas y falsas de personas con discapacidad intelectual ($CI_{VERDADERO} = 62.00$, $DT = 10.07$; $CI_{FALSO} = 58.41$, $DT = 8.42$) y las características diferenciales de tales declaraciones utilizando los criterios de análisis de contenido de la técnica CBCA. Treinta y tres personas valoraron 16 declaraciones verdaderas y 13 falsas utilizando su intuición. Otros dos evaluadores entrenados en CBCA valoraron las mismas declaraciones. Los evaluadores no expertos diferenciaron entre declaraciones verdaderas y falsas con una precisión por encima del azar, aunque el índice de errores fue elevado (38.19%). El hecho de que los evaluadores no entrenados no pudieran discriminar eficazmente entre declaraciones falsas y verdaderas demuestra que la intuición no puede considerarse útil en un contexto forense. La técnica CBCA discriminó mejor que los juicios intuitivos. No obstante, solo uno de los 19 criterios permitió discriminar de modo significativo, por lo que se necesitan más procedimientos adaptados específicamente a las aptitudes de los testigos con discapacidad intelectual. La presencia de producción estructurada, cantidad de detalles, características específicas de la agresión y complicaciones inesperadas incrementaba la probabilidad de que una declaración fuera considerada verdadera por los evaluadores no expertos. La clasificación realizada por los evaluadores no entrenados fue independiente del cociente intelectual de los participantes. Se lleva a cabo un análisis de macrodatos para mejorar la calidad de la clasificación.

People with intellectual disability (ID) are, for some crimes, victimised more than the general population and are involved in many police/legal proceedings as victims for such crimes (González, Cendra, & Manzanero, 2013). A large proportion of these proceedings never reach trial. Probably, one of the main reasons for these rejections stems

from the lack of adaptations of police and judicial procedures to the characteristics of these people (Bull, 2010; Milne & Bull, 2001, 1999), as well as from the actual myths that society has about the limited ability of people with ID to testify with accuracy (Henry, Ridley, Perry, & Crane, 2011; Peled, Iarocci, & Connolly, 2004; Sabsey & Doe, 1991;

Cite this article as: Manzanero, A. L., Scott, M. T., Vallet, R., Aróztegui, J., & Bull, R. (2019). Criteria-based content analysis in true and simulated victims with intellectual disability. *Anuario de Psicología Jurídica*, 29, 55-60. <https://doi.org/10.5093/apj2019a1>

Correspondence: antonio.manzanero@psi.ucm.es (A. L. Manzanero).

ISSN: 1133-0740/© 2019 Colegio Oficial de Psicólogos de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Stobbs & Kebbell, 2003; Tharinger, Horton, & Millea, 1990; Valenti-Hein & Schwartz, 1993). In many cases, the testimonies associated with people with ID have been considered less credible (Peled et al., 2004). On the other hand, one myth implies that people with ID may be more believable (Bottoms, Nysse-Carris, Harris, & Tyda, 2003).

Some studies (Manzanero, Contreras, Recio, Alemany, & Martorell, 2012) have shown that people with ID may perform approximately the same as others in forensic contexts. Moreover, their autobiographical memories may be quite stable over time, being their ability to describe an event independent of the degree of disability (Morales et al., 2017). Indeed, Henry et al. (2011) found no correlation between credibility assessment and either witness mental age or anxiety.

For eyewitnesses with ID, the key may be the lack of studies regarding differentiating characteristics of their true/false statements. With other types of population (mainly children), forensic psychology has proposed useful procedures for assessing credibility by analyzing the content of statements. One of these procedures is Statement Validity Assessment (SVA) (Köhnken, Manzanero, & Scott, 2015; Steller & Köhnken, 1989; Volbert & Steller, 2014), a technique that assesses the credibility of statements given by minors who are alleged victims of sexual abuse. SVA is a comprehensive procedure for generating and testing hypotheses about the source and validity of a given statement. It includes methods of collecting relevant data regarding such hypotheses and techniques for analyzing these data, plus guidelines for drawing conclusions regarding the hypotheses.

Criteria-based content analysis (CBCA) is a method included in SVA for distinguishing truthful from fabricated statements. It is not applicable for distinguishing statements experienced as real memories, which are actually the result of suggestive influences (Scott & Manzanero, 2015; Scott, Manzanero, Muñoz, & Köhnken, 2014), but may be applied complementarily to other procedures (Blandón-Gitlin, López, Masip, & Fenn, 2017). The use of the CBCA content criteria in the absence of a detailed analysis of the moderator variables would produce rather low percentages of discrimination between true and false statements, where around 30% of false alarms have been found (Oberlader et al., 2016). Previous research has shown that the level of accuracy in the classification of true and false statements can sometimes be low even when evaluators are specifically trained in this technique, which could indicate that CBCA has basic problems (Akehurst, Bull, Vrij, & Köhnken, 2004).

the statement, peculiarities of content, motivation-related contents, and offence-specific elements. The basic assumption of the CBCA is that statements based on memories of real events are qualitatively different from statements not based on experience (Undeutsch, 1982). According to his original proposal, each content criterion is an indicator of truth; its presence in a given statement is viewed as an indicator of the truth of that statement, but its absence does not necessarily mean the statement is false. This assumption has been shown to be incomplete, because it does not consider false memories as a source of incorrect statements, nor the effects of liars knowing about the criteria (Vrij, Akehurst, Soukara, & Bull, 2004a). However, not all the criteria are always relevant when it comes to discriminating (Bekerian & Dennett, 1992; Manzanero, 2006, 2009; Manzanero, López, & Aróztegui, 2016; Porter & Yuille, 1996; Sporer & Sharman, 2006; Vrij, 2005; Vrij, Akehurst, Soukara, & Bull, 2004b); the presence of these criteria depends on a host of moderator variables (Hauch, Blandón-Gitlin, Masip, & Sporer, 2015; Oberlader et al., 2016).

Among these variables are preparation (Manzanero & Diges, 1995), time delay (Manzanero, 2006; McDougall & Bull, 2015), the individual's age (Comblain, D'Argembeau, & Van der Linden, 2005; Roberts & Lamb, 2010), and the asking of questions and multiple retrieval (Strömwall, Bengtsson, Leander, & Granhag, 2004). Also, fantasies, lies, dreams, and post-event information do not each involve the same differentiating characteristics. Furthermore, changing a small detail, however important it may be, of a real event—such as whether the role played in the event was witness or protagonist (Manzanero, 2009)—is not the same as fabricating an entire event. Indeed, false statements rarely are entirely fabricated but originate, in part, from actual experiences that are modified to create something new. In addition, the characteristics of statements vary depending on the person's ability to generate a plausible statement. This is relevant to people with ID, it having been proposed that lying would usually be cognitively more complex than telling the truth (Vrij, Fisher, Mann, & Leal, 2006) and, therefore, would involve a greater demand for cognitive resources (Vrij & Heaven, 1999).

The aims of the present study were (i) to use CBCA in order to analyze the statements given by true and simulating witnesses with intellectual disability, (ii) people's intuitive ability to discriminate between the two types of statements, and (iii) the ability to discriminate through big data analysis.

Table 1. Content Criteria for Statement Credibility Assessment

GENERAL CHARACTERISTICS
1. Logical structure.
2. Unstructured production.
3. Quantity of details.
SPECIFIC CONTENTS
4. Contextual embedding.
5. Descriptions of interactions.
6. Reproduction of conversation.
7. Unexpected complication during the incident.
PECULIARITIES OF CONTENT
8. Unusual details.
9. Superfluous details.
10. Accurately reported details misunderstood.
11. Related external associations.
12. Accounts of subjective mental state.
13. Attribution of perpetrator's mental state.
MOTIVATION-RELATED CONTENTS
14. Spontaneous corrections.
15. Admitting lack of memory.
16. Raising doubts about one's own testimony.
17. Self-deprecation.
18. Pardoning the perpetrator.
OFFENCE-SPECIFIC ELEMENTS
19. Details characteristic of the offence.

CBCA takes into account 19 content criteria grouped into five categories (see Table 1): general characteristics, specific contents of

Method

Video recorded accounts provided by 32 people with mild to moderate, non-specific intellectual disability were used as material to be analyzed. Fifteen participants were true witnesses to a real event that took place two years ago when the bus they were travelling during a day trip caught fire. Those participants had an average IQ of 62.00 ($SD = 10.07$) and were 33.93 years old ($SD = 6.49$). Seventeen other participants who provided simulated accounts of the same event had an average IQ of 58.41 ($SD = 8.42$) and were 31.75 years old ($SD = 7.07$). No significant differences were found in IQ as a function of condition, $F(1, 30) = 1.204, p = .281, \eta^2 = .039$. The IQ scores were obtained by the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008).

All of these 32 participants provided informed consent. The statements were obtained with a procedure similar to that used in other studies (Vrij et al., 2004a, 2004b), as follows:

All the participants who did not go on the day trip knew the event beforehand, because they knew the people involved as they belong to the same care centre for people with intellectual disabilities. The event was very commented by everyone when it took place and it was even informed in the media. In any case, a verbal summary of the most important information about the day trip, such as its location, the main complication on the day trip, and the course of the day was given to all participants of either condition. To increase the ecological

validity of the study, all 32 participants were encouraged to give their testimonies as best they could. While they were not put under the stress of trying to make the interviewer believe their testimony (to prevent undue tension in the interview), we told them they would be invited to a soda if they succeeded in convincing the interviewer that they had, in fact, experienced the event (all of them actually received this invitation).

Two forensic psychologists, experts on interviewing and taking testimony, from the Unit for Victims with Intellectual Disability, interviewed each of these 32 participants individually. An audiovisual recording was made of all interviews. The same instructions were followed: "We want you to tell us, with as much detail as you can, from the beginning to the end, what happened when you went on the day trip and the bus caught fire. We want you to tell us even the things you think are not very important." Once a free-recall statement was obtained, all participants were asked the same questions: Who were you with? Where was it? Where did you go? What did you do? What happened afterwards? The forensic psychologists who conducted the interviews were blind to the groups (true vs. false experience) the participants belonged to.

Once the testimonies were obtained, the videos were evaluated using two different procedures: a) intuitive analysis carried out by people without knowledge of forensic psychology and b) technical analysis performed by forensic psychologists using CBCA criteria.

Of the 32 statements discussed above, two videos of the true condition and one of the false condition were removed from the intuitive judgments. This was due to communication problems that prevented the evaluators from understanding what the participants said in the conditions in which the intuitive evaluation was carried out.

Intuitive Credibility Assessment

There were 33 participants as evaluators (6 men and 27 women; age average 23.54, $SD = 4.04$), recruited among psychology students in Spain, who wanted to voluntarily participate in the study. They did not receive any compensation for participating, and had no specific knowledge of credibility analysis techniques and no specific understanding of intellectual disability.

The video recordings of sixteen true and thirteen false statements were shown on a large-format screen at the university. All evaluators attended the showing at the same time, but they were prevented from interacting so that they did not bias each other while making their individual assessments. The instructions were as follows: "Next, a series of videos will be shown in which people with intellectual disability are talking about an event related to a bus accident. Some of the statements were given by individuals who experienced that event; the others were given by individuals who, although they were not there, were told about the event, and they have given their statement with the intention of making us believe they were there. The task is to decide who is telling the truth and who is lying to us. As you are assessing each statement, bear in mind that the interviewees are all people with intellectual disability, so their way of telling things may be special." The twenty-nine videos were shown in random order to prevent a learning effect from impacting the ability to evaluate true and false statements. After each video was shown, the evaluators were asked to categorize the statement as true or false. In the first evaluations, it was observed that the viewing of 29 videos produced saturation and fatigue in the evaluators. To avoid this circumstance leading to random decisions, it was decided to submit to each evaluator a maximum of 15 videos, taking care that finally all the videos were evaluated. In any case, the evaluators were warned that when they felt very tired, they should warn the experimenters. A total of 197 evaluations of the true condition and 256 evaluations of the false condition were collected.

Analysis of Phenomenological Characteristics of the Statements Using CBCA Criteria

The interview video recordings were transcribed to facilitate analysis of the phenomenological characteristics of the statements. Two trained CBCA evaluators each made their own criteria assessment of each statement and then reached an interjudge agreement. To assess the CBCA criteria codings for inter-coder reliability, an agreement index was computed as follows: $AI = \text{agreements} / (\text{agreements} + \text{disagreements})$. For all the variables, this was greater than the cut-off of .80 (Tversky, 1977), except for "logical structure" and "unstructured production", where it was .67.

Each criterion was assessed in terms of its absence or presence in the statement, as was originally defined by Steller and Köhnken (1989). To measure the degree of presence of each criterion, the evaluators quantified how many times the criterion was present throughout the report. For the criteria of "quantity of details", the micropropositions that described, as objectively as possible, what happened in the actual event were used, which is a better measure than counting words because it is not influenced by the descriptive style used by participants.

Criterion 13, "attribution of perpetrator's mental state", was modified to be "attribution of other's mental state". Criterion 19, "details characteristic of the offence", was modified to be "details characteristic of the event". Criteria 17 (self-deprecation) and 18 "Pardoning the perpetrator", were not taken into consideration, because of the nature of the event.

Results

CBCA Characteristics of the Statements

An ANOVA test was conducted to assess the effects of the type of statement on the number of times each CBCA criterion was present in each report. As multiple comparisons were conducted, the significance level was adjusted with a Bonferroni adjustment to .003. Table 2 shows only "quantity of details" was significant in determining truth. The remaining 16 criteria (some of which rarely occurred) produced no significant differences.

Big Data Analysis of Characteristic Features of Statements

Big data techniques aim towards complex data exploration and analysis. High-Dimensional Visualization (HDV) graphs facilitate the visualization of complex data. This technique displays all the data at once, enabling researchers to graphically explore in search of data distribution patterns (for more information see Manzanero, Alemany, Recio, Vallet, & Aróztegui, 2015; Manzanero, El-Astal, & Aróztegui, 2009; Vallet, Manzanero, Aróztegui, & García-Zurdo, 2017). The graphs are similar to scatter plots. The different variables corresponding to a subject's responses on questionnaire items are represented as a point in a high-dimensional space (17 values or dimensions in this study). When there are more than three variables, as in this study, mathematical dimensionality reduction techniques are used to build a 3D graph (Buja et al., 2008; Cox & Cox, 2001). Each point in the hyperspace has a distance to each of the other points. Multidimensional scaling will search 3D points, preserving the distances between points as much as possible (Barton & Valdés, 2008). Sammon's error (Barton & Valdés, 2008) is used to calculate the 3D transformation error.

3D points are represented using Virtual Reality Modelling Language (VRML). VRML files allow graphical rotation and exploration to facilitate graphical data analysis. 3D graphs permit visual exploration of the data in search of its distribution patterns.

Figure 1 represents all criteria, regardless of whether their discriminating values were statistically significant. The quality of

Table 2. Means, Standard Deviations, and ANOVA Values for Each Dependent Variable

	False Statement N = 17		True Statement N = 15		F(1, 30)	P	η ²
	Mean	SD	Mean	SD			
Logical structure	5.67	2.74	6.86	2.32	1.726	.199	.054
Unstructured production	6.11	2.54	5.46	2.82	0.470	.498	.015
Quantity of details *	7.35	3.60	13.93	4.74	19.800	.000	.398
Contextual embedding	2.52	1.06	3.93	1.90	6.812	.014	.185
Interactions	1.23	1.98	1.53	2.26	0.158	.694	.005
Conversations	0.41	0.50	1.40	1.59	5.878	.022	.164
Unexpected complications	0.47	0.79	0.40	0.50	0.086	.771	.003
Unusual details	0.58	0.79	0.80	0.86	0.523	.475	.017
Superfluous details	0.00	0.00	0.40	0.82	3.984	.055	.117
Details misunderstood	0.23	0.56	0.00	0.00	2.616	.116	.080
External associations	0.05	0.24	0.13	0.35	0.496	.487	.016
Subjective mental state	0.88	0.99	1.00	1.25	0.088	.769	.003
Other's mental state	1.47	1.28	1.13	1.50	0.469	.499	.015
Corrections	0.17	0.39	0.13	0.35	0.106	.747	.004
Lack of memory	2.38	3.47	2.16	3.08	0.034	.855	.001
Doubts	0.29	0.58	0.50	0.62	0.919	.345	.030
Characteristic details	1.88	1.45	2.13	1.18	0.281	.600	.009

*p < .003 (Bonferroni adjustment for pairwise comparisons).

the dimensionality reduction through multidimensional scaling (Buja et al., 2008) was very good, with a small Sammon's error of .03. The dotted line graphically dividing true statements from false statements shows correct classification of 81.25 percent (simulated statements were classified as being true 29.42%).

A possible explanation for several of the CBCA criteria not discriminating could stem from the variability among participants. As can be seen in Figure 1, the cloud of dots that graphically represents each type of statement is very dispersed and overlapped.

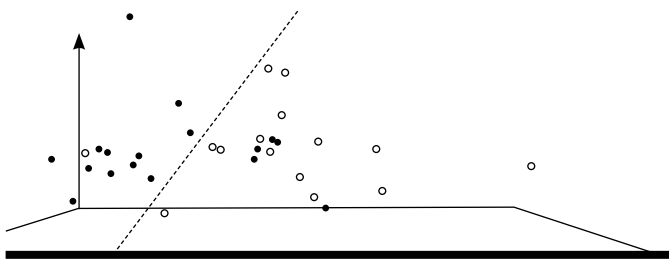


Figure 1. HDV Graph of Content Criteria in True (Light) and False (Dark) Statements, Including all CBCA Criteria.

Note. Sammon's error = .030; correct classification = 81.25%.

Intuitive Credibility Assessment

Considering the 197 evaluations of the true- and the 256 evaluations of the false videos, discriminability accuracy (hits, false alarms, omissions, and correct rejections), discriminability index (d'), and response criterion (c) as specified by Signal Detection Theory (MacMillan & Kaplan, 1985; Tanner & Swets, 1954) were measured.

Analysis of the credibility assessments based on lay participants' natural ability found above chance accuracy for the discriminability index (d') was .626 (SD = .121), Zd = 5.159, p < .05.

The response criterion (c) reached a score of .086 (SD = .061), Zc = 1.412, p = ns. The subjects had a neutral response criterion (scores equal to 0 indicate a neutral criterion, greater than 0 a conservative criterion, and less than 0 a liberal criterion). The proportion of statements correctly classified was 61.81 percent (see Table 3), with 65.48% of false statements being correctly assessed and 58.98% of the true ones.

Table 3. Intuitive Responses for Each Type of Statement

		Statement Type	
		False	True
Assessment	False	CR: 129 (65.48%)	O: 105 (41.02%)
	True	FA: 68 (34.52%)	H: 151 (58.98%)
		197 (100%)	256 (100%)

Note. CR = correct rejection; O = omission; FA = false alarm; H = hit.

Depending on the number of times a story was considered true or false by the intuitive judges, the probability of "truthfulness" was established (number of times considered true / number of evaluations made for that testimony). The average probability of truthfulness assigned to the false testimonies was 36.37 (SD = 31.64), while that assigned to the true ones was 64.00 (SD = 23.93), F(1, 28) = 6.750, p < .05, η² = .200.

The levels of disabilities of the persons with ID could be one of the indicators on which the evaluators based their intuitive assessments. However, no significant effects were found when participants' IQ was analysed based on how their statements had been classified, considering the four possible types of response (H, FA, O, and CR), F(3, 26) = 0.498, p = ns, η² = .056. As can be seen in Table 4, IQ means were similar for all groups.

Table 4. IQ Means and Standard Deviations of the Subjects according to the Type of Response Issued by the Evaluators

	Mean	SD
Hits	63.00	10.39
False alarms	59.80	6.42
Omissions	58.00	10.23
Correct rejections	58.09	9.74
Total	59.90	9.31

Relationship between CBCA Criteria and Intuitive Credibility Assessment

The Pearson correlation (bilateral) between the degree of presence of each CBCA criteria in the testimonies and the probability of truthfulness indicates that the evaluators' natural ability may have been mediated by the following criteria: "structured production",

$r(29) = .546, p < .01$; “quantity of details”, $r(29) = .618, p < .01$; “unexpected complications”, $r(29) = .526, p < .01$; and “characteristic details”, $r(29) = .437, p < .05$. No significant correlations were found for the remaining 13 criteria (see Table 5). The greater presence of these criteria would imply greater intuitive truthfulness.

Table 5. Pearson Correlations between Content Criteria and Intuitive Assessments of “True”

	Pearson Correlation ($N = 29$)	p (bilateral)
Logical structure**	.546	.002
Unstructured production	-.346	.066
Quantity of details**	.618	.000
Contextual embedding	.238	.214
Interactions	-.013	.946
Conversations	.111	.567
Unexpected complications**	.526	.003
Unusual details	.347	.065
Superfluous details	.150	.436
Details misunderstood	-.149	.441
External associations	.166	.389
Subjective mental state	.014	.944
Other's mental state	.098	.614
Corrections	-.124	.522
Lack of memory	.024	.902
Doubts	.039	.839
Characteristic details*	.437	.018

* $p < .05$ (bilateral); ** $p < .01$ (bilateral).

Discussion

In line with many other studies (not involving truth tellers/liars with ID), the lay participants could not discriminate between false and true stories at a level to be considered useful in a forensic context (Rassin, 1999), this being one of the reasons why CBCA was developed. The CBCA technique did indeed discriminate at a better level. However, of the 19 criteria, only one (“quantity of details”) was found significant. This criterion, which is present in some lies, also deemed “richness in detail”, has also been identified as potential biases which may lead to incorrect veracity judgements (Nahari & Vrij, 2015). “Quantity of details” was found in the present study to be significant for people who have ID, even though when truly narrating an event, they tend to give fewer details than the general population (Dent, 1986; Kebbell & Wagstaff, 1997; Perlman, Ericson, Esses, & Isaacs, 1994).

ID is a component of certain syndromes that have associated deficits in language development and articulation. This might explain why several of the CBCA criteria were rarely present in the current study. In Down's Syndrome, for example—the most common genetic syndrome with an ID component—language disorders are one of the effects. In spontaneous conversation, the speech of people with ID is less intelligible, and they have more difficulty with grammatical structuring (Rice, Warren, & Betz, 2005)—in fact, their problems with sentence structuring are similar to those of individuals diagnosed with language development disorder (Laws & Bishop, 2003).

Thus, if the criteria that help us to determine the true statements of people with ID indeed is “quantity of details”, what could happen if their true accounts are compared with true accounts from the general population? For those with ID who have reduced vocabulary, semantic, and autobiographical memory deficits (rendering them unable to detail the event), we could run the risk that such people will suffer an erroneous judgment of their credibility, and thus, revictimisation could result.

However, since the natural ability evaluators were capable of discriminating between true and false statements at only 12% above-chance accuracy, a procedure that achieves better accuracy is needed.

If we were to extrapolate such natural ability data to a law enforcement setting, for example, we could predict that the testimonies of people with ID would be correctly assessed in only 60 percent of cases, resulting in many true accounts not being believed. This percentage is not far from what the police (and others) usually reach when judging the statements of people with standard development (Manzanero, Quintana, & Contreras, 2015).

To analyse what the possible basis is for intuitively assessing the testimonies of people with ID—which, in turn, is going to determine the credibility assessments granted in forensic and legal settings—we correlated the probability of a “true” assessment with the IQ and the CBCA content criteria. As in the study by Henry et al. (2011), the results showed that IQ did not account for the lay evaluators' decisions. In relation to the different CBCA criteria, only four criteria appear to mediate intuitive truthfulness (structured production, quantity of details, unexpected complications, and characteristic details).

On the other hand, big data analysis reached a better classification score. It must be taken into account that, surprisingly, these results were obtained after considering all CBCA variables, not only the ones yielding significant differences, although, initially, it was expected that the variable showing significant differences should lead to a better classification in comparison with the rest. Because that was not the case, it seems that useful information is held by those other variables not showing significant differences and the big data technique is able to profit from it, providing better classification quality. This approach could maybe allow to find, in a near future, an improved way of distinguishing true and false statements.

Conflict of Interest

The authors of this article declare no conflict of interest.

References

- Akehurst, L., Bull, R., Vrij, A., & Köhnken, G. (2004). The effects of training professional groups and lay persons to use criteria-based content analysis to detect deception. *Applied Cognitive Psychology, 18*, 877-891. <https://doi.org/10.1002/acp.1057>
- Barton, A. & Valdés, J. J. (2008). Hybrid unsupervised/supervised virtual reality spaces for visualizing gastric and liver cancer databases: An evolutionary computation approach. In A. An, S. Matwin, Z. W., Raś, & D. Ślęzak (Eds.), *Foundations of intelligent systems* (pp. 256-261). Berlin, Germany: Springer.
- Bekerian, D. A., & Dennett, J. L. (1992). The truth in content analyses of a child's testimony. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law. International perspectives* (pp. 335-344). Berlin, Germany: W de Gruyter.
- Blandón-Gitlin, I., López, R., Masip, J. & Fenn, E. (2017). Cognición, emoción y mentira: implicaciones para detectar el engaño [Cognition, emotion, and lying: Implications to detect deception]. *Anuario de Psicología Jurídica, 27*, 95-106. <https://doi.org/10.1016/j.apj.2017.02.004>
- Bottoms, B. L., Nysse-Carris, K. L., Harris, T., & Tyda, K. (2003). Jurors' perceptions of adolescent sexual assault victims who have intellectual disabilities. *Law and Human Behavior, 27*, 205-227. <https://doi.org/10.1023/A:1022551314668>
- Buja A., Swayne D.F., Littman M., Dean N., Hofmann H., & Chen L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics, 17*, 444-472. <https://doi.org/10.1198/106186008X318440>
- Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology, 15*, 5-23. <https://doi.org/10.1348/014466509X440160>
- Comblain, C., D'Argembeau, A., & Van der Linden, M. (2005). Phenomenal characteristics of autobiographical memories for emotional and neutral events in older and younger adults. *Experimental Aging Research, 31*, 173-189. <https://doi.org/10.1080/03610730590915010>
- Cox, T. F., & Cox, M. A. A. (2001). *Multidimensional scaling* (2nd ed.). London, UK: Chapman and Hall.
- Dent, H. (1986). An experimental study of the effectiveness of different techniques of interviewing mentally handicapped child witnesses. *British Journal of Clinical Psychology, 25*, 13-17. <https://doi.org/10.1111/j.2044-8260.1986.tb00666.x>
- González, J. L., Cendra, J., & Manzanero, A. M. (2013). Prevalence of disabled people involved in Spanish Civil Guard's police activity. *Research in*

- Developmental Disabilities*, 34, 3781-3788. <https://doi.org/10.1016/j.ridd.2013.08.003>
- Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, 19, 307-342. <https://doi.org/10.1177/1088868314556539>
- Henry, L., Ridley, A., Perry, J., & Crane, L. (2011). Perceived credibility and eyewitness testimony of children with intellectual disabilities. *Journal of Intellectual Disability Research*, 55, 385-391. <https://doi.org/10.1111/j.1365-2788.2011.01383.x>
- Kebbell, M. R., & Wagstaff, G. F. (1997). Why do the police interview eyewitnesses? Interview objectives and the evaluation of eyewitness performance. *Journal of Psychology*, 131, 595-601. <https://doi.org/10.1080/00223989709603841>
- Köhnken, G., Manzanero, A. L., & Scott, M. T. (2015). Análisis de la validez de las declaraciones: mitos y limitaciones [Statement validity assessment: Myths and limitations]. *Anuario de Psicología Jurídica*, 25, 13-19. <https://doi.org/10.1016/j.apj.2015.01.004>
- Laws, G., & Bishop, D.V.M. (2003). A comparison of language abilities in adolescents with Down's syndrome and children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 46, 1324-1339. [https://doi.org/10.1044/1092-4388\(2003\)103](https://doi.org/10.1044/1092-4388(2003)103)
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, 98, 185-199. <https://doi.org/10.1037/0033-2909.98.1.185>
- Manzanero, A. L. (2006). Do perceptual and suggested accounts actually differ? *Psychology in Spain*, 10, 52-65.
- Manzanero, A. L. (2009). Análisis de contenido de memorias autobiográficas falsas [Criteria content analysis of false autobiographical memories]. *Anuario de Psicología Jurídica*, 19, 61-72.
- Manzanero, A. L., Alemany, A., Recio, M., Vallet, R., & Aróztegui, J. (2015). Evaluating the credibility of statements given by people with intellectual disability. *Anales de Psicología*, 31, 338-344. <https://doi.org/10.6018/analesps.31.1.166571>
- Manzanero, A. L., Contreras, M. J., Recio, M., Alemany, A., & Martorell, A. (2012). Effects of presentation format and instructions on the ability of people with intellectual disability to identify faces. *Research in Developmental Disabilities*, 33, 391-397. <https://doi.org/10.1016/j.ridd.2011.09.015>
- Manzanero, A. L., & Diges, M. (1995). Effects of preparation on internal and external memories. In G. Davies, S. M. A. Lloyd-Bostock, M. McMurrin, & C. Wilson (Eds.), *Psychology, law and criminal justice. International developments in research and practice* (pp. 56-63). Berlin, Germany: W. De Gruyter & Co.
- Manzanero, A. L., El-Astal, S., & Aróztegui, J. (2009). Implication degree and delay on recall of events: An experimental and HDV study. *The European Journal of Psychology Applied to Legal Context*, 1, 101-116.
- Manzanero, A. L., López, B., & Aróztegui, J. (2016). Underlying processes behind false perspective production. *Anales de Psicología*, 32, 256-265. <https://doi.org/10.6018/analesps.32.1.194461>
- Manzanero, A. L., Quintana, J. M., & Contreras, M. J. (2015). (The null) Importance of police experience on intuitive credibility of people with intellectual disabilities. *Research in Developmental Disabilities*, 36, 191-197. <https://doi.org/10.1016/j.ridd.2014.10.009>
- McDougall, A. J., & Bull, R. (2015). Detecting truth in suspect interviews: the effect of use of evidence (early and gradual) and time delay on criteria-based content analysis, reality monitoring and inconsistency within suspect statements. *Psychology, Crime & Law*, 21, 514-530. <https://doi.org/10.1080/1068316X.2014.994631>
- Milne, R., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*. Chichester, UK: Wiley.
- Milne, R., & Bull, R. (2001). Interviewing witnesses with learning disabilities for legal purposes: A review. *British Journal of Learning Disabilities*, 29, 93-97. <https://doi.org/10.1046/j.1468-3156.2001.00139.x>
- Morales, C., Manzanero, A. L., Wong, A., Gómez-Gutiérrez, M., Iglesias, A. M., Barón, S., & Álvarez, M. A. (2017). Stability of autobiographical memory in children with intellectual disabilities. *Anuario de Psicología Jurídica*, 27, 79-84. <https://doi.org/10.1016/j.apj.2017.05.002>
- Nahari, G., & Vrij, A. (2015). Systematic errors (biases) in applying verbal lie detection tools: richness in detail as a test case. *Crime Psychology Review*, 1, 98-107. <https://doi.org/10.1080/23744006.2016.1158509>
- 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, 440-457. <https://doi.org/10.1037/lhb0000193>
- Peled, M., Iarocci, G., & Cannolly, D. A. (2004). Eyewitness testimony and perceived credibility of youth with mild intellectual disability. *Journal of Intellectual Disability Research*, 18, 669-703. <https://doi.org/10.1111/j.1365-2788.2003.00559.x>
- Perlman, N. B., Ericson, K. I., Esses, V. M., & Isaacs, B. J. (1994). The developmentally handicapped witness: Competency as a function of question format. *Law and Human Behaviour*, 18, 171-187. <https://doi.org/10.1007/BF01499014>
- Porter, S., & Yuille, J.C. (1996). The language of deceit: an investigation of the verbal cues to deception in the interrogation context. *Law and Human Behavior*, 20, 443-458. <https://doi.org/10.1007/BF01498980>
- Rassin, E. (1999). Criteria-based content analysis: The less scientific road to truth. *Expert Evidence*, 7, 265-278. <https://doi.org/10.1023/A:1016627527082>
- Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics*, 26, 7-27. <https://doi.org/10.1017/S0142716405050034>
- Roberts, K. P., & Lamb, M. E. (2010). Reality-monitoring characteristics in confirmed and doubtful allegations of child sexual abuse. *Applied Cognitive Psychology*, 24, 1049-1079. <https://doi.org/10.1002/acp.1600>
- Sabsey, D., & Doe, T. (1991). Patterns of sexual abuse and assault. *Sexuality and Disability*, 9, 243-259. <https://doi.org/10.1007/BF01102395>
- Scott, M. T., & Manzanero, A. L. (2015). Analysis of the judicial file: Assessing the validity of testimony. *Papeles del Psicólogo*, 36, 139-144.
- Scott, M. T., Manzanero, A. L., Muñoz, J. M., & Köhnken, G. (2014). Admisibilidad en contextos forenses de indicadores clínicos para la detección del abuso sexual infantil [Admissibility of clinical indicators for the detection of child sexual abuse in forensic contexts]. *Anuario de Psicología Jurídica*, 24, 57-63. <https://doi.org/10.1016/j.apj.2014.08.001>
- Sporer, S. L., & Sharman, S. J. (2006). Should I believe this? Reality monitoring of accounts of self-experienced and invented recent and distant autobiographical events. *Applied Cognitive Psychology*, 20, 837-854. <https://doi.org/10.1002/acp.1234>
- Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY: Springer.
- Stobbs, G., & Kebbell, M. (2003). Juror's perception of witnesses with intellectual disabilities and influence of expert evidence. *Journal of Applied Research in Intellectual Disabilities*, 16, 107-114. <https://doi.org/10.1046/j.1468-3148.2003.00151.x>
- Strömwall, L. A., Bengtsson, L., Leander, L., & Granhag, P. A. (2004). Assessing children's statements: The impact of a repeated experience on CBCA and RM ratings. *Applied Cognitive Psychology*, 18, 653-668. <https://doi.org/10.1002/acp.1021>
- Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-409. <https://doi.org/10.1037/h0058700>
- Tharinger, D., Horton, C., & Millea, S. (1990). Sexual abuse and exploitation of children and adults with mental retardation and other handicaps. *Child Abuse and Neglect*, 14, 301-312. [https://doi.org/10.1016/0145-2134\(90\)90002-B](https://doi.org/10.1016/0145-2134(90)90002-B)
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352. <https://doi.org/10.1037/0033-295X.84.4.327>
- Undeutsch, U. (1982). Statement reality analysis. In A. Trankell (Ed.), *Reconstructing the past* (pp. 27-56). Stockholm, Sweden: Norstedt and Soners.
- Valenti-Hein, D.C., & Schwartz, L.D. (1993). Witness competency in people with mental retardation: implications for prosecution of sexual abuse. *Sexuality and Disability*, 11, 287-294. <https://doi.org/10.1007/BF01102173>
- Vallet, R., Manzanero, A. L., Aróztegui, J., & García-Zurdo, R. (2017). Age-related differences in phenomenal characteristics of long-term memories for the attack of March 11, 2004. *Anuario de Psicología Jurídica*, 27, 85-93. <https://doi.org/10.1016/j.apj.2017.03.002>
- 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. <https://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. <https://doi.org/10.1037/1076-8971.11.1.3>
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004a). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science*, 36, 113-126. <https://doi.org/10.1037/h0087222>
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004b). Detecting deceit via analysis of verbal and nonverbal behaviour in children and adults. *Human Communication Research*, 30, 8-41. <https://doi.org/10.1111/j.1468-2958.2004.tb00723.x>
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences*, 10, 141-142. <https://doi.org/10.1016/j.tics.2006.02.003>
- Vrij, A., & Heaven, S. (1999). Vocal and verbal indicators of deception as a function of lie complexity. *Psychology, Crime, & Law*, 5, 203-215. <https://doi.org/10.1080/10683169908401767>
- Wechsler, D. A. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Psychological Corporation.