



# The European Journal of Psychology Applied to Legal Context

<https://journals.comadrid.org/ejpalc>



## Verbal Baseline: Within-Subject Consistency of CBCA Scores across Different Truthful and Fabricated Accounts

Jonas Schemmel<sup>a</sup>, Benjamin G. Maier<sup>a</sup>, and Renate Volbert<sup>a, b</sup>

<sup>a</sup>Psychologische Hochschule Berlin, Germany; <sup>b</sup>Charité – Universitätsmedizin Berlin, Institut für Forensische Psychiatrie, Germany

### ARTICLE INFO

#### Article history:

Received 12 February 2019  
Accepted 5 November 2019

#### Keywords:

Statement Validity Assessment (SVA)  
Criteria-based Content Analysis (CBCA)  
Consistency  
Verbal baselining  
Baseline statements  
Within-subject measure

### ABSTRACT

Statement Validity Assessment (SVA) proposes that baseline statements on different events can serve as a within-subject measure of a witness' individual verbal capabilities when evaluating scores from Criteria-based Content Analysis (CBCA). This assumes that CBCA scores will generally be consistent across two accounts by the same witness. We present a first pilot study on this assumption. In two sessions, we asked 29 participants to produce one experience-based and one fabricated baseline account as well as one experience-based and one fabricated target account (each on different events), resulting in a total of 116 accounts. We hypothesized at least moderate correlations between target and baseline indicating a consistency across both experience-based and fabricated CBCA scores, and that fabricated CBCA scores would be more consistent because truth-telling has to consider random event characteristics, whereas lies must be constructed completely by the individual witness. Results showed that differences in correlations between experience-based CBCA scores and between fabricated CBCA scores took the predicted direction ( $c_{\text{experience-based}} = .44$  versus  $c_{\text{fabricated}} = .61$ ) but this difference was not statistically significant. As predicted, a subgroup of event-related CBCA criteria were significantly less consistent than CBCA total scores, but only in experience-based accounts. The discussion considers methodological issues regarding the usage of total CBCA scores and whether to measure consistency with correlation coefficients. It is concluded that more studies are needed with larger samples.

### La línea base verbal: la congruencia intrasujeto de las puntuaciones del CBCA en diferentes relatos veraces e inventados

### RESUMEN

El Statement Validity Assessment (SVA) propone que las declaraciones sobre diferentes eventos pueden servir como una línea base intrasujeto de la medida de las capacidades verbales individuales de un testigo al evaluar las puntuaciones del Criteria Based Content Analysis (CBCA). Esto implica que las puntuaciones del CBCA serán congruentes en dos relatos del mismo testigo. Presentamos un primer estudio piloto sobre este supuesto. Se pidió a 29 participantes en dos sesiones que elaboraran un relato verdadero (línea base) y otro inventado, además de un relato verdadero y otro inventado (cada uno en situaciones diferentes), arrojando un total de 116 relatos. Se planteó la hipótesis de una correlación al menos moderada entre la declaración fabricada y la verdadera, que indicaría una consistencia entre las puntuaciones en el CBCA de relatos inventados y experimentados y que las puntuaciones en el CBCA inventadas serían más consistentes porque la verdad incluye las características aleatorias de los hechos, mientras que las mentiras las construye totalmente el testigo. Los resultados mostraron que las diferencias en las correlaciones entre las puntuaciones en el CBCA de relatos experimentados y fabricados iban en la dirección predicha ( $c_{\text{vivido}} = .44$  frente a  $c_{\text{inventado}} = .61$ ), pero esta diferencia no fue significativa. Como se predijo, un subgrupo de criterios de CBCA relacionados con los hechos fue menos congruente que las puntuaciones totales de CBCA, pero sólo en los relatos de hechos experimentados. Se discuten las implicaciones metodológicas relacionadas con el uso de las puntuaciones totales del CBCA y si se debe medir la consistencia mediante el coeficiente de correlación. Se concluye que se necesitan otros estudios con muestras más grandes.

#### Palabras clave:

Statement Validity Assessment (SVA)  
Criteria-based Content Analysis (CBCA)  
Congruencia  
Línea base verbal  
Declaraciones de referencia  
Medidas intrasujeto

Cite this article as: Schemmel, J., Maier, B. G., & Volbert, R. (2020). Verbal baselining: Within-subject consistency of cbca scores across different truthful and fabricated accounts. *The European Journal of Psychology Applied to Legal Context*, 12, 35-42. <https://doi.org/10.5093/ejpalc2020a4>

Correspondence: [schemmel.jonas@googlemail.com](mailto:schemmel.jonas@googlemail.com) (J. Schemmel).

ISSN: 1889-1861/© 2020 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Criteria-based Content Analysis (CBCA; Steller & Köhnken, 1989; Volbert & Steller, 2014) is based on the assumption that the quality of experience-based statements is generally higher than the quality of fabricated statements (Undeutsch, 1967). The quality of a statement is indicated by the CBCA criteria compiled by Steller and Köhnken (1989) and revised by Niehaus (2008) and Volbert and Steller (2014) (see Table 1). Generally, meta-analyses have shown the assumed difference in CBCA criteria between experience-based and fabricated accounts (Amado, Arce, & Fariña, 2015; Amado, Arce, Fariña, & Vilariño, 2016; Masip, 2017; Oberlader et al., 2016). However, the number and type of content characteristics to be found in a statement are supposed to be determined not only by the veracity of the statement but also, inter alia, by the cognitive abilities and narrating habits of the statement provider (Nahari & Vrij, 2015; Schemmel & Volbert, 2016). Thus, Statement Validity Assessment (SVA) stresses that CBCA scores need to be evaluated “against the background of an individual’s . . . verbal competencies” (Steller & Köhnken, 1989, p. 218), thereby highlighting the relevance of within-subject measures when assessing individual statements (see also Köhnken, 2004). To account for these individual verbal capabilities, “verbal baselining” has been suggested as a procedure with which to assess individual verbal behavior (Greuel et al., 1998; Köhnken, 2004; Volbert, Steller, & Galow, 2010; Vrij, 2016). This study focuses on baseline statements on different events as a verbal baselining procedure within SVA, and is the first to address CBCA score consistency—that is, the correspondence of individuals’ CBCA scores across statements on different events—as its main precondition.

### Verbal Baselining in SVA: Baseline Statements as Within-Subject Measures

Elaborations on verbal baselining as a within-subject measure within SVA are rare. Offe and Offe (2000) have provided specific instructions, but only in German. They proposed asking the statement provider to produce one or more experience-based and/or false accounts about various personally relevant events. The CBCA scores of these *baseline accounts* could then be compared with the CBCA scores of the statement on the event in question (*target account*) and used as within-subject references. In short, the idea is that similarities between the target account and the experience-based baseline account and dissimilarities between the fabricated baseline and the experience-based target account will suggest the truthfulness of the target and vice versa. Other authors suggest asking only for experience-based biographical events and then checking whether the target account deviates from the baseline truthful account (Volbert & Steller, 2014). However, the underlying notion common to both approaches is to draw on at least one baseline statement to infer an estimation of an individual CBCA score (or profile) that a person is expected to achieve. Thus, the use of verbal baselining will generally require CBCA scores of individuals obtained from statements on different events to correlate, that is, to be consistent.

### CBCA Score Consistency as the Prerequisite of Verbal Baselining Approaches in SVA

Verbal baselining approaches are therefore based (implicitly) on the assumption that a content analysis of one truthful or fabricated statement can to some extent predict the content quality of another truthful or fabricated statement. Thus, drawing on key concepts from personality psychology (“past behavior predicts future behavior”), CBCA scores might, to a certain degree, be *person-specific* (Schemmel & Volbert, 2017). Person specificity, in turn, is expected to be indicated by the *consistency* of CBCA scores across a person’s

baseline and target statements.<sup>1</sup> That is, CBCA scores contained in two truthful statements or two fabricated statements produced by the same individual are assumed to be similar, even if these statements refer to different events. Moreover, because some authors suggest using both experience-based and fabricated baseline statements (Greuel et al., 1998; Offe & Offe, 2000), they are clearly assuming that both experience-based and fabricated statements have *CBCA score consistency*. To the best of our knowledge, however, no studies have been conducted on this baseline statement approach in the context of SVA and its main precondition, that is, CBCA score consistency.<sup>2</sup> In the following, we shall formulate three testable assumptions on CBCA score consistency across baseline and target accounts.

### Hypothesis 1: CBCA scores of experience-based and fabricated accounts will be moderately consistent

Both telling the truth and lying can be considered to be cognitive-verbal tasks that should thus be influenced by stable person variables. The literature discusses the influence of the following variables on the verbal content of statements: age (Vrij, 2008), a general tendency to narrate autobiographical experiences (Nahari & Vrij, 2014), personality characteristics (Schelleman-Offermans & Merckelbach, 2010), and abilities to deceive/deception strategies (Vrij, Granhag, & Mann, 2010) (for a brief overview, see Volbert & Steller, 2014).

However, producing a target account places different demands on the statement provider than producing a baseline account: following the verbal baselining procedure using baseline statements (cf. above), the statement provider is instructed directly by the interviewer to lie or tell the truth. Hence, the ground truth of the baseline is overt to both sender and recipient. Therefore, producing an experience-based baseline account under these conditions does not involve any special communication demands, and fabricating a baseline account does not include any deception demands (Köhnken, 1990; Sporer, 2016). We therefore assume that experience-based and fabricated CBCA scores will be fairly consistent overall, but only to a moderate degree (Cohen, 1988).

### Hypothesis 2: CBCA scores of experience-based accounts will be less consistent than CBCA scores of fabricated statements

As mentioned before, both experience-based and fabricated statements have been suggested as baseline measures. Therefore, the question arises whether CBCA scores of experience-based baseline and target accounts versus CBCA scores of fabricated baseline and target accounts might differ in their consistency. Indeed, an analysis of the main processes involved in producing experience-based or fabricated accounts of personal experiences reveals one obvious, but significant difference: an experience-based statement is based on specific experiences, whereas a lie is based on general cognitive schemata and knowledge about the event type in question (Köhnken, 1990, 2004; Sporer, 2016), the witness’s deception ability (Köhnken, 1990; Vrij et al., 2010), and individual deception strategies (Maier, Niehaus, Wachholz, & Volbert, 2018). Hence, CBCA scores of fabricated accounts are influenced mainly by personal variables. In contrast, CBCA scores of experience-based statements are determined more strongly by actual event characteristics: truth-tellers can reproduce conversations only if there was a conversation; they will point out unusual details only if there was something unusual to notice; and so forth. These event characteristics should not covary systematically between two experience-based events. Thus, they should increase the random share of variance across experience-based, but not across fabricated baseline and target accounts. This should lower the relative influence of personal variables on CBCA scores contained in experience-based accounts. Note that we neither assume the

influence of the witness' personality on perceiving, encoding, and reporting event-related criteria in experience-based accounts to be zero, nor do we deny that personality might, in part, evoke specific responses resulting in some of the event-related criteria (such as conversations). This is why we do not argue that event-related criteria in experience-based statements should not be consistent at all, but only consistent to a lesser degree. In summary, we assume that CBCA scores of experience-based baseline and target statements will be less consistent than those of fabricated statements.

### **Hypothesis 3: The differential consistency of CBCA scores will be determined mainly by event-related criteria**

Following the considerations above, we assume that event-related CBCA scores in experience-based baseline and target statements will show a weaker correlation compared to the correlation between the corresponding total CBCA scores contained in experience-based statements. In contrast, we consider that event-related CBCA criteria in "fabricated" baseline and target statements will mainly reflect person-related cognitive processes. Therefore, they should correlate at a level comparable to the total CBCA scores found in fabricated baseline and target statements. If, finally, a lower consistency between two experience-based CBCA scores is due at least partially to the unsystematic influence of the event itself, the difference in consistency of experience-based and false statements tested in Hypothesis 2 should be more pronounced when only *event-related criteria* are included.

Criteria are defined as event-related if they directly reflect incidents and details of the event itself. This is assumed for "contextual information", which is determined by the specific surroundings and temporal setting that a witness finds rather than creates him- or herself; "verbal and nonverbal interactions", which require the presence of another person who can be addressed and who shows a (verbal) reaction determined by him or her rather than by the witness; "consequences of the event", which are by definition determined by the event itself as they happen to the witness; "attribution of the other's mental state", which is, by definition, illustrated by the concerned other's behavior and not by the witness; "unexpected complications", which are defined as not foreseen by the witness before they happen to her or him; and, finally, "unusual or superfluous details and accurately reported details misunderstood", which represent special features of the surroundings or aspects of the event that are perceived and remembered by the witness, but not created. The other criteria, such as accounts of own mental state or criteria summarized as indications of memory-related processes (cf. Table 1), were categorized as more person-related.

The current study therefore aims to investigate the consistency of CBCA scores as a precondition of using baseline statements on different events as within-measures in SVA. Specifically, we derived three testable hypotheses on the consistency of experience-based and fabricated accounts, on the differential consistency of experience-based versus fabricated accounts, and on the differential consistency of event-related versus total CBCA scores. For a preliminary investigation, we designed a pilot study within the context of a larger study on CBCA scores. Thus, the results presented here can be viewed as indications, but they need to be replicated in studies with larger sample sizes.

## **Method**

### **Participants**

The sample consisted of 29 participants who completed 116 interviews (one of the initial 30 participants did not attend the target interview and had to be excluded). Each participant produced four statements (one experience-based baseline, one experience-based

target, one fabricated baseline, and one fabricated target statement) following the procedure depicted below. Participants were 16 females and 13 males with an average age of 35.8 years ( $SD = 12.02$ , range 18-57 years). Their educational background was diverse with an emphasis on higher education (21 participants had an upper secondary school and/or a university degree, one was still attending upper secondary school, four had completed a vocational training, three had completed middle school—*Realschule*). German was the native language of 26 subjects; the remaining 3 had lived in Germany since early childhood.

Participants were recruited via university mailing lists (4 participants), an online study participants' recruitment platform ("vp-gesucht.de"; 1 participant), and small advertisements on eBay (24 participants).<sup>3</sup> The study was announced as an experiment on the relationship between creativity and linguistic expression in describing personal events offering an opportunity to earn between 25€ and 50€. All participants received either 50€ or, if applicable, 25€ combined with a student credit for participating in all three sessions.

### **Procedure**

Each participant attended a preparatory session followed by two separate interview sessions (baseline and target). Within each interview session, participants had to report one event based on actual experience (experience-based condition) and one event based on fabrication (false condition).

In the *preparation session*, participants were given a list of nine negative, highly emotional events (i.e., being attacked by another person or an animal, being the victim of a criminal assault, having a serious accident, medical intervention, death of a loved one). They had to choose two actual experiences and two events they had not experienced. One experience-based and one fabricated event was randomly assigned to one of the two interviews, respectively. Participants were then asked to prepare an experience-based and a false account for the first interview. Also, in order to establish a high level of motivation, they were told that if their statements were good enough, they would be invited to participate in a second interview session with additional compensation.

The *baseline interview* followed within 2 weeks of the preparation session. Participants were asked to produce the experience-based as well as the fabricated account selected in the preparation session in as much detail as possible. Interviews were semi-structured with an open question to elaborate the central event of the statement followed by three specific questions on details of the report. Note that the interviewer was the second author who had already instructed the participants in the preparation session. Moreover, the interviewer explicitly asked for the experience-based or the fabricated story. Thus, it was obvious to the participants that the interviewer was aware of the ground truth of the respective accounts. This procedure corresponds to baseline interviews in the field where witnesses are also instructed to report an actual experience or to fabricate a lie. The order of truths and lies was randomized across the sample. After the interview, participants were told that they had "qualified" for a second interview session (*target interview*), and that this interview would be conducted by a psychological expert on credibility assessment. Participants were instructed to prepare the remaining experience-based and false event previously chosen in the preparation session. Moreover, they were told that the expert interviewers would not be informed about the ground truth of their statements. It was also announced that they would receive an additional compensation of 25 Euros if they were able to convince the expert interviewer that both of their statements were experience-based.

Within the following 2 weeks, the *target interview* took place in the office of psychologists specialized in credibility assessment. The interviewer—actually a research assistant—was announced as a psychological expert on credibility assessment whose task was

to determine the veracity status of each statement. Interviewers in the target interview were blind to the veracity status. The task for participants was to produce the experience-based and fabricated statements in as much detail as possible. Again, the order of truths and lies was randomized. Because the blinded interviewer could not tell them which statement to start with, the randomized order was emailed to the participants beforehand. The target interviews were again semi-structured like the baseline interview and the number of questions was restricted to three. Note that the task in the target interview corresponded with that in real evaluations in the field: the interviewee had to convince the interviewer of the truthfulness of her or his statement(s) while the interviewer aimed to find out the ground truth.

After the target interview, participants were fully debriefed. All participants received the full amount of financial compensation (total of 50€) regardless of their performance.

## Measures

All statements were audio-recorded, transcribed verbatim, and coded by the first author and a research assistant. Prior to rating the transcripts, the research assistant received a 2-week training in CBCA scoring by the first author to ensure a consistent and standardized approach when rating each criterion. Training included reading and discussing relevant literature containing detailed descriptions and examples of the relevant criteria. Both raters also rated several transcripts not contained in the current study until they had achieved a satisfactory interrater reliability (ICC > .40). They then independently coded each of the transcripts for CBCA criteria.<sup>4</sup> Both raters were blind to the veracity status of the accounts, did not know in which interview session (baseline or target) the account had been produced, and had no information on the participants or interviewer.

**CBCA compilation.** Raters used a revised version of the CBCA compilation based on Volbert and Steller (2014). It differs from the original compilation suggested by Steller and Köhnken (1989) in three main aspects: (a) logical consistency was removed because it was regarded as a prerequisite rather than a quality criterion (Greuel et al., 1998); (b) unusual and superfluous details were combined into one rating because their overlapping definitions have been discussed as impairing interrater reliability (cf. Anson, Golding, & Gully, 1993; Hauch, Sporer, Masip, & Blandon-Gitlin, 2017; Roma, Martini, Sabatello, Tatarelli, & Ferracuti, 2011); finally, (c) criteria were arranged along the assumed psychological processes underlying CBCA scores (Volbert, 2015) which only affected the order in which the criteria were rated, not their definition.

**Rating process.** Ratings followed a standardized procedure. Raters first coded the absolute frequency of each criterion before weighing each occurrence of single criteria (0 for nonexistent, 1 for weak, and 2 for pronounced). Subsequently, they rated each criterion on a 4-point scale ranging from 1 (*absent*) to 4 (*very pronounced*) with reference to the whole account. This stepwise procedure was chosen to ensure raters went through the same structured assessment process and based their ratings on specific text passages. The relevant CBCA assessment included in the further analysis was the final scale rating because this allowed a more fine-grained assessment of the respective criteria and has been recommended in a recent meta-analysis on the reliability of CBCA (Hauch et al., 2017).

Ratings were done in 15 stages.<sup>5</sup> Within the first 8 stages, raters independently coded 20 accounts per stage. After each rating stage, the two raters' CBCA codings were checked for each statement and—if necessary—debated. To enhance reliability, raters made sure their codings referred to the same text passage (an aspect often ignored in other studies; cf. Sporer, 2012) and checked whether they basically agreed on the relevance of a specific text passage although they

had just assigned it to a different criterion (Tully, 1998). Because agreement improved after the first 8 stages, raters coded 30 instead of 20 accounts within rating stages 9 to 15. Otherwise, the procedure remained the same.

**Reliability analysis.** As suggested in a recent meta-analysis (Hauch et al., 2017), reliabilities of the criteria ratings were estimated using intraclass correlation coefficients (ICC; Shrout & Fleiss, 1979) of type ICC (2,2) (two-way random, average measure) with consistency as the measure of conformity. ICC (2,2) was chosen because reliability results should be applicable to a (fictitious) population of trained raters and the ratings were to be averaged. The criteria “related external associations” and “accurately reported details misunderstood” were not coded indicating that they were not to be found in this sample. They were consequently excluded from further analyses. We calculated ICCs using the R package “irr” (Gamer, Lemon, Fellows, & Singh, 2012) implemented in the statistical software R (v.3.4.1; R Core Team, 2017) run in RStudio (v. 1.1.383; RStudio Team, 2017). For the 116 statements included in the current study, ICCs of the remaining criteria varied between .71 and 1.0 (cf. Table 1). Hence, reliability was good to excellent (Cicchetti, 1994).

**Table 1.** CBCA Criteria and Their Interrater Reliabilities as ICC (2,2) with 95% Confidence Intervals

	ICC [95% IC]
General characteristics	
Quantity of details	.86 [.71, .94]
Vividness of event	.82 [.62, .92]
Episodic memory criteria	
Contextual information (spatial, temporal, personal/situational) <sup>1</sup>	.76 [.48, .89]
Interactions <sup>1</sup>	1
Reproduction of conversation/verbal accounts <sup>1</sup>	.83 [.64, .92]
Accounts of subjective mental state	.77 [.51, .89]
Consequences <sup>1</sup>	.71 [.38, .86]
Attribution of the other's mental state <sup>1</sup>	.91 [.80, .96]
Script-deviant/script-irrelevant details	
Unexpected complications <sup>1</sup>	.95 [.89, .98]
Unusual/superfluous details <sup>1</sup>	.81 [.60, .91]
Accurately reported details misunderstood <sup>1</sup>	-
Related external associations	-
Special knowledge	.92 [.82, .96]
Indications of memory-related processes	
Unstructured production	.89 [.77, .95]
Spontaneous corrections	.85 [.69, .93]
Efforts to remember	.76 [.49, .89]
Insecurity about correctness of details	.86 [.71, .94]
Admitting lack of memory	.86 [.71, .94]
Admitting lack of knowledge	.95 [.89, .98]
Reasons for insecurities and lack of memory or knowledge	.87 [.73, .94]
Reality control	1
Criteria (allegedly) reducing credibility	
Raising doubts about one's own testimony/person	.89 [.76, .95]
Self-deprecation	1
Pardoning the perpetrator	1

Note. <sup>1</sup>Criteria directly encoding details of the event itself (event-related criteria).

## Results

Prior to the analyses, the two raters' scores were averaged for each criterion included in the analysis. Alpha was set as 5% for all statistical analyses. For each account, the CBCA criteria were summed up to a total CBCA score resulting in 29 total CBCA scores for each of the four conditions (experience-based baseline, experience-based target,

fabricated baseline, fabricated target). Table 2 presents the mean total CBCA score for each of these four conditions as well as absolute differences between them. Correlation coefficients were then compared with the R-package “cocor” (Diedenhofen & Musch, 2015).

**Table 2.** Mean CBCA Sum Scores and Standard Deviations by Veracity Status and Interview Session

		Experience-based	Fabricated	Absolute difference <sup>3</sup>
Baseline interview	$M_{total}^1 (SD)$	13.14 (6.03)	10.98 (4.14)	2.16
	$M_{event}^2 (SD)$	6.10 (3.23)	4.53 (1.79)	1.57
Target interview	$M_{total}^1 (SD)$	11.52 (4.26)	13.86 (5.59)	2.34
	$M_{event}^2 (SD)$	4.95 (2.30)	5.45 (2.84)	0.50
Absolute difference <sup>4</sup>	$D_{total}^1$	1.62	2.88	
	$D_{total}^2$	1.15	0.92	

Note. <sup>1</sup>Includes all criteria from Table 1; <sup>2</sup>includes event-related criteria marked with <sup>1</sup> in Table 1; <sup>3</sup>refers to differences between experience-based and fabricated accounts; <sup>4</sup>refers to differences between baseline and target statements.

**Hypothesis 1: CBCA Scores of Experience-based and Fabricated Accounts Will Be Moderately Consistent**

Pearson product-moment correlations were calculated between the total CBCA scores of the experience-based baseline and the experience-based target accounts ( $c_t$ ) as well as between the total CBCA scores of the fabricated baseline and the fabricated target account ( $c_f$ ). The total CBCA scores of the experience-based statements correlated at  $c_t(27) = .44, 95\% CI [.14, 1.0], p = .009$  (one-tailed); total CBCA scores of the fabricated statements correlated at  $c_f(27) = .61, 95\% CI [.37, 1.0], p < .001$  (one-tailed). Thus, in line with our assumptions,  $c_t$  was moderate and  $c_f$  was high in this sample (Cohen, 1988).

**Hypothesis 2: CBCA Scores of Experience-based Accounts Will Be less Consistent than CBCA Scores of Fabricated Accounts**

Even though  $c_f = .61$  and  $c_t = .44$  differed in the predicted direction on a descriptive level ( $c_f > c_t$ ), this difference did not attain statistical significance ( $z = -1.08, p = .14$ , one-tailed; cf. Steiger, 1980).

**Hypothesis 3: The Differential Consistency of Experience-based and Fabricated CBCA Scores Will Be Determined mainly by Event-related Criteria**

For each account, contextual information, interactions, conversations, consequences, attribution of the other’s mental state, unexpected complications, and unusual/superfluous details event-related CBCA-criteria were summed to an event-related CBCA subscore. Table 2 presents the mean event-related CBCA scores for baseline and target accounts. Pearson product-moment correlations were calculated between the event-related CBCA score of the experience-based baseline and the experience-based target accounts ( $c_{t,event}$ ) as well as between the event-related CBCA scores of the fabricated baseline and the fabricated target account ( $c_{f,event}$ ).

Strictly speaking, Hypothesis 3 consisted of three assumptions: first, for experience-based baseline and target accounts, we assumed that event-related CBCA scores would show a weaker correlation than total CBCA scores. Second, for fabricated baseline and target accounts, we expected event-related CBCA scores and total CBCA scores to correlate at a similar level. Third, we expected that the difference between correlations of event-related CBCA scores in experience-based versus fabricated accounts would be higher than the correlations of total CBCA scores in experience-based versus fabricated accounts.

In line with Hypothesis 3, event-related CBCA scores showed a weaker correlation than total CBCA scores with experience-based statements at  $c_t = .44$  and  $c_{t,event}(27) = .24, 95\% CI [-.08, 1.0], p = .10$  (one-tailed). This difference between  $c_t$  and  $c_{t,event}$  was statistically significant ( $z = 1.83, p = .03$ , one-tailed; cf. Steiger, 1980). As predicted, event-related CBCA scores were thus less consistent than total CBCA scores in experience-based accounts.

Second, in contrast to this finding, event-related CBCA scores with fabricated statements correlated at a similar level as the total CBCA scores at  $c_{f,event}(27) = .55, 95\% CI [.29, 1.0], p = .001$  (one-tailed). Nonetheless, the difference between  $c_f$  and  $c_{f,event}$  was not statistically equivalent with .01 as equivalence interval ( $p = .23$ ; cf. Counsell & Cribbie, 2015, p. 296). The assumption of a negligible difference between  $c_f$  and  $c_{f,event}$  could thus not be confirmed statistically in the current sample.

Third, on a descriptive level, the difference between  $c_{t,event} = .24$  and  $c_{f,event} = .55$  was clearly larger than the difference between  $c_t = .44$  and  $c_f = .61$ . However, like the difference between  $c_t$  and  $c_f$ , it did not attain significance ( $z = -1.6, p = .06$ , one-tailed) in this sample.

To summarize, only the weaker correlation of event-related CBCA scores across experience-based statements (when compared with total CBCA scores) was in line with our predictions.

**Unplanned Analyses: Within-subject Correlations**

In the current article, CBCA sum scores were used. In order to check for ecological validity of the sum score, we calculated mean within-subject correlations for experience-based and fabricated accounts. If their patterns correspond with the one by the sum score correlations, this speaks against an ecological fallacy when using the sum score. To calculate within-correlations, the criteria were treated as nested within each participant resulting in 4 columns (experience-based and fabricated baseline and experience-based and fabricated target statement) and 24 rows (one for each criterion) per participant. Using intra-class correlations, we then correlated the two columns with CBCA scores obtained from experience-based statements and the two columns with CBCA scores obtained from fabricated statements. These participant-wise within-subject ICCs were averaged. In the current data set, the pattern for between-correlations was paralleled descriptively by the one for within-correlations: Mean  $ICC_{within}$  for CBCA-total scores across experience-based vs. fabricated accounts were .53 vs. .59; mean  $ICC_{within}$  for event-related CBCA-scores across experience-based vs. fabricated accounts were .50 vs. .59. Nevertheless, the differences were not as pronounced as they were in the case of between-correlations. Because these calculations were merely explorative and meant for illustration, no tests for statistical significance were conducted.

**Discussion**

Within Statement Validity Assessment (SVA), the CBCA score of a target account is not considered to be decisive per se, but needs to be referred to the baseline verbal behavior of the statement provider (Köhnken, 2004). Textbooks on SVA suggest asking the statement provider to produce an experience-based and/or a fabricated baseline statement on a different personal event whose CBCA score can then be compared with the one contained in the target statement (cf. Offe & Offe, 2000). The implicit assumption in this approach is that CBCA scores are to some extent person-specific, thereby resulting in a consistency of CBCA scores across statements about different events (Schemmel & Volbert, 2017). The current study is the first to address this assumption. We hypothesized that even though both experience-based and fabricated accounts are at least moderately consistent (Hypothesis 1), fabricated CBCA scores are more consistent than experience-based CBCA scores (Hypothesis 2). We assumed this

to be due to the influence of event-related CBCA characteristics (such as “unexpected complications”) on experience-based but not on fabricated accounts (Hypothesis 3).

Before further discussing the results and implications of this study, a cautionary note is necessary. Research indicates that correlations can be inaccurate in small sample sizes and may tend to stabilize at  $N$ s much larger than 29 (for typical scenarios in psychology with regard to experience-based effects and confidence levels, they stabilize at approximately  $N = 250$ ; cf. Schönbrodt & Perugini, 2013). Also, equivalence tests as applied when testing Hypothesis 3 usually require more power through being based on a larger sample size (Counsell & Cribbie, 2015). Therefore, as already stressed above, the results presented here should be understood as merely tentative and as a first step toward a theory-driven and empirically based conceptualization of verbal baselining using baseline statements in SVA. Large-scale studies will be needed to draw conclusions relevant for practitioners in the field.

In the current pilot sample, CBCA scores of experience-based baseline and experience-based target statements showed a moderate correlation, whereas those contained in fabricated baseline and fabricated target statements showed a high correlation. If these results can be replicated in future research, the basic requirement of using verbal baselining in SVA, that is, the consistency of CBCA scores, would be met. Moreover, both coefficients differed in the predicted direction, even though this difference did not attain statistical significance. Given more robust empirical evidence, the higher consistency of fabricated statements would speak for their preferred use. However, SVA often deals with sexual violence cases where witnesses might be confused by the instruction to lie. Interviewers therefore will often have to rely on experience-based baselines. It might thus be useful if interviewers could differentiate more consistent criteria from less consistent criteria. One possible approach is the identification of event-related criteria which share less systematic personal variance and should thus not be used as baseline measures. Notably, event-related CBCA scores contained in experience-based accounts were in the current study significantly less consistent than total CBCA scores in experience-based statements. This was not, however, the case for event-related versus total CBCA scores in fabricated accounts, indicating that the difference between event-related versus total CBCA scores might be due to varying demands when telling the truth versus lying. Nonetheless, the difference between experience-based and fabricated event-related CBCA scores was not significant. Thus, the assumption that event-related criteria drive the different consistency scores of experience-based and fabricated accounts could not be definitely corroborated in this pilot sample.

### Some Methodological Considerations

The results of the current study need to be replicated through further research on CBCA score consistency. In the following, we discuss two methodological considerations that might be relevant for future research designs. They concern the use of total CBCA scores and the downsides of correlations as indexes of similarity.

To begin with, the current study based its calculations on total CBCA scores. Even though widely used as an index for overall content quality, this approach has been criticized mainly due to its lack of ecological validity (Hauch et al., 2017). Yet, the ecological validity of summing up CBCA scores depends largely on the underlying coding of the single criteria. Binary presence ratings not allowing for any weighting of criteria clearly do not match coding procedures in the field, whereas weighted measures and ratings do come close. Thus, total scores seem to be appropriate as long as they are based on the latter coding procedures. Nonetheless, total CBCA scores might increase the probability of ecological fallacies

by making unjustified conclusions on an aggregated data level. Two approaches can be taken to deal with possible ecological fallacies. First, rather than just computing a total score, subscales can be defined based on theoretical considerations (such as calculating a sum score consisting of event-based criteria). The second and more data-driven option is to calculate within-person next to between-person correlations like in the current article (cf. the unplanned analyses in the results section). Ecological fallacies are indicated by substantial differences between the mean sample within-ICC and the between-correlation. Thus, in order to ensure the ecological validity of total CBCA scores, future research should drop binary presence ratings, derive meaningful CBCA subscales, and possibly consider between- as well as within-correlations.

Second, only reporting correlations as measures of consistency might impair the ecological validity of consistency studies. Indeed, many studies on personality use Pearson product-moment correlations as an index for consistency. However, these refer to the stability of interindividual differences. Hence, strictly speaking, the current study investigated the relative consistency of CBCA scores (Fleeson & Nofhle, 2008) whereby the reference is always a sample/population of others. Correlations certainly contain valuable information especially if the interest is in consistency as a general phenomenon or rather as a possible general characteristic (of a given sample). Notably, however, practitioners in the field do not have sample scores at their disposal, but have to base their decisions on the few accounts produced by one current witness. This raises the question regarding how group results such as correlations can be used to infer the behavior of individuals in the field (Faigman, Monahan, & Slobogin, 2014; Nahari et al., 2019). Given the restricted amount of group information, practitioners might (intentionally or not) compare the CBCA score of a baseline directly with the CBCA score of a target statement. Consistency would then most likely be defined as the degree to which both absolute CBCA scores are the same. A low absolute difference between the CBCA scores of, for example, an experience-based baseline and the target statement would suggest the truthfulness of the latter one. This corresponds to the definition of absolute consistency formulated by Fleeson and Nofhle (2008), and consequently requires different statistical parameters. Absolute consistency can be indicated by small to zero absolute within-subject differences of CBCA scores contained in baseline and target statements (Schemmel & Volbert, 2017). However, this might yield results that differ from the correlation pattern of relative consistency. In the current data, for example, absolute differences of CBCA scores were larger in fabricated than in experience-based accounts (see absolute differences displayed in Table 2). Hence, contrary to the results on relative consistency, CBCA scores of experience-based accounts were more consistent than those of fabricated accounts when consistency was based on absolute differences. This was probably due to the unexpected finding that participants produced more CBCA criteria in their fabricated target than in their fabricated baseline statements while there was no such effect for experience-based accounts (cf. below). Whether these results can be replicated is an empirical question. Nonetheless, it can already be concluded that CBCA score consistency in a sample might vary depending on how similarity is defined (see also Schemmel & Volbert, 2017). Hence, research should examine the implementation of the baselining procedure in the field to shed more light on which comparison procedure is applied. Moreover, especially against the background of recent calls to consider group-to-individual inferences (Nahari et al., 2019), results must be communicated carefully, because correlations should not be confused with absolute measures—both might even yield contradictory results.

## Prospects

CBCA score consistency is only the precondition for using baseline statements as references when judging target statements in the field. The essential question remains whether baseline statements in fact improve CBCA judgments and credibility assessment. One way to investigate this could be to present baseline statements before having perceivers judge a target statement. Another promising research design would include more statements per person so that consistency scores could be calculated and, subsequently, baseline statements could be used to statistically predict the veracity of a target statement (Dahle & Wolf, 1997). Moreover, little is known about if and how baseline procedures are in fact applied by experts in the field. This is an important point, because the way individual baseline information is applied may well differ from the procedure presented here. Presumably, experts often do not explicitly ask for experience-based and false baseline statements. Instead, in the context of a biographical survey, the witness is often asked for experience-based experiences, not fabricated ones. These are then combined to form an intuitive experience-based baseline. However, should CBCA scores in experience-based statements indeed prove to be less consistent in better powered replication studies, this would call for caution in the use of experience-based baselines especially referring to event-related criteria (cf. above). In short, more insight is needed into whether and how baseline statements are used as a part of credibility assessment in the field. These insights might help to enhance the ecological validity of study designs and thus enable research to provide practitioners with guidelines on how to use baseline statements effectively in the field.

## Limitations

As already mentioned, due to the low sample size, the present results must be regarded as preliminary, and studies with more power are needed. Moreover, the design of the present pilot study is not without its limitations. First, an unexpected trend emerged in the data because in contrast to fabricated target statements, experience-based target statements did not contain more CBCA scores compared to experience-based baseline statements (cf. Table 2). Whereas this might be a result of the low statistical power, it might also have emerged due to an (un-)conscious trade-off. Participants might have been more motivated to prepare and present the fictitious rather than the experience-based event, simply because it needed preparing and the task was perceived as more difficult. Future studies should thus consider including just one target statement. Such a design would not only prevent such a trade-off but would also enhance the study's ecological validity. Moreover, in the current study, we could not control for "experience-based lying", in which participants change actual experiences only slightly by adding false details instead of fabricating whole new accounts. Future studies should control for partly truthful events with quasiexperimental designs (that assign participants to groups based on their actual experience of a specific event, e.g., Volbert & Lau, 2013) or by applying mock-crime paradigms (e.g., Gödert, Gamer, Rill, & Vossel, 2005).

## Conclusion

The current pilot study aimed to investigate the consistency of CBCA scores as a precondition of using baseline statements on different events as within-measures in SVA. Within-measures have largely been ignored in CBCA research so far. Thus, this study is meant as a first step that will hopefully encourage more powerful studies that highlight within-subject comparisons in CBCA research (Nahari et al., 2019). Such studies should address not only CBCA

score consistency but also—and maybe even more importantly—whether baseline statements actually improve the evaluation of the target statement.

## Conflict of Interest

The authors of this article declare no conflict of interest.

## Notes

<sup>1</sup>Note that this definition of consistency deviates from others in the credibility assessment literature in which consistency often refers to different statements by one or more persons on the same event—that is, within-statement or statement-evidence consistency (for an overview, see Vredeveltdt, van Koppen, & Granhag, 2014).

<sup>2</sup>Even in the general context of personality assessment, CBCA score consistency has—strictly speaking—no equivalent, because CBCA scores are other-ratings (by the expert) referring to the results (CBCA scores) of different tasks (baseline vs. target statements) performed by the same individual (witness). Therefore, CBCA score consistency as the basis for the baseline statement approach has yet to be investigated either conceptually or empirically.

<sup>3</sup>The data for the current study were collected as part of a larger study with 90 participants producing 360 statements (Maier et al., 2019). However, 60 participants received a training on CBCA criteria between the first and second interview sessions and were thus not included here. Participants also completed personality questionnaires that are not considered in the current study.

<sup>4</sup>As already stated above, the current study was part of a bigger data collection. Overall, 90 participants produced 360 statements. These were all assessed by the two raters, even though only 116 were included in the current study. The statements relevant for the current study were randomly distributed within the whole sample

<sup>5</sup>This refers to the total of 360 statements.

## References

- Amado, B. G., Arce, R., & Farina, F. (2015). Undeutsch hypothesis and Criteria Based Content Analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7, 3-12. <https://doi.org/10.1016/j.ejpal.2014.11.002>
- Amado, B. G., Arce, R., Fariña, F., & Vilariño, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16, 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- 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. <https://doi.org/10.1007/BF01044512>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cohen, J. (1988). *Statistical power analysis for the behavioral science*. Hillsdale, NJ: Erlbaum Associates.
- Counsell, A., & Cribbie, R. A. (2015). Equivalence tests for comparing correlation and regression coefficients. *British Journal of Mathematical and Statistical Psychology*, 68, 292-309. <https://doi.org/10.1111/bmsp.12045>
- Dahle, K.-P., & Wolf, P. (1997). Realkennzeichen im Einzelfall - Einzelfalleperimentelle Möglichkeiten einer ipsativen Kriteriengewinnung für die forensische Glaubhaftigkeitsdiagnostik [Reality criteria in single cases - Experimental potentials for an ipsative extraction of credibility criteria. In L. Greuel, T. Fabian, & M. Stadler (Eds.), *Psychologie der Zeugenaussage. Ergebnisse der rechtspsychologischen Forschung* (pp. 101-119). Weinheim, Germany: Psychologie Verlags Union.
- Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. *PLOS ONE*, 10(4), e0121945. <https://doi.org/10.1371/journal.pone.0121945>
- Faigman, D. L., Monahan, J., & Slobogin, C. (2014). Group to individual (G2i) inference in scientific expert testimony. *The University of Chicago Law Review*, 81(2). <https://doi.org/10.2139/ssrn.2298909>
- Fleeson, W., & Nofzle, E. E. (2008). Where does personality have its influence? A supermatrix of consistency concepts. *Journal of Personality*, 76, 1355-1386. <https://doi.org/10.1111/j.1467-6494.2008.00525.x>

- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). *irr: Various coefficients of interrater reliability and agreement. R package version 0.84*. Retrieved from <https://cran.r-project.org/package=irr>
- 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. <https://doi.org/10.1348/135532505X52680>
- Greuel, L., Offe, S., Fabian, A., Wetzels, P., Fabian, T., Offe, H., & Stadler, M. (1998). *Glaubhaftigkeit der Zeugenaussage [Credibility assessment of a witness's testimony]*. Weinheim, Germany: Beltz.
- Hauch, V., Sporer, S. L., Masip, J., & Blandon-Gitlin, I. (2017). Can credibility criteria be assessed reliably? A meta-analysis of Criteria-Based Content Analysis. *Psychological Assessment, 29*, 819-834. <https://doi.org/10.1037/pas0000426>
- Köhnken, G. (1990). *Glaubwürdigkeit [Credibility]*. Munich, Germany: Psychologie Verlags Union.
- 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, UK: Cambridge University Press.
- Maier, B. G., Niehaus, S., Wachholz, S., & Volbert, R. (2018). The strategic meaning of CBCA criteria from the perspective of deceivers. *Frontiers in Psychology, 9*, 855. <https://doi.org/10.3389/fpsyg.2018.00855>
- Maier, B.G., Schemmel, J., & Volbert, R. *When lying requires more efforts than just to fabricate: The effects of creative and strategic task demands on statements' content quality*. Manuscript submitted for publication.
- Masip, J. (2017). Deception detection : State of the art and future prospects. *Psicothema, 29*, 149-159. <https://doi.org/10.7334/psicothema2017.34>
- Nahari, G., Ashkenazi, T., Fisher, R. P., Granhag, P.-A., Hershkowitz, I., Masip, J., ... Vrij, A. (2019). "Language of lies": Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology, 24*, 1-23. <https://doi.org/10.1111/lcrp.12148>
- Nahari, G., & Vrij, A. (2014). Are you as good as me at telling a story? Individual differences in interpersonal reality monitoring. *Psychology, Crime & Law, 20*, 573-583. <https://doi.org/10.1080/1068316X.2013.793771>
- 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>
- Niehaus, S. (2008). Merkmalsorientierte Inhaltsanalyse [Criteria-based content analysis]. In R. Volbert & M. Steller (Eds.), *Handbuch der Rechtspsychologie* (pp. 311-321). Göttingen, Germany: Hogrefe.
- 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>
- Offe, H., & Offe, S. (2000). Die individuelle Vergleichsgrundlage für die Aussagebeurteilung [The individual baseline for the evaluation of testimonies]. *Praxis der Rechtspsychologie, 10*(2), 37-54.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Roma, P., Martini, P. S., Sabatello, U., Tatarelli, R., & Ferracuti, S. (2011). Validity of Criteria-based Content Analysis (CBCA) at trial in free-narrative interviews. *Child Abuse and Neglect, 35*, 613-620. <https://doi.org/10.1016/j.chiabu.2011.04.004>
- RStudio Team. (2017). *RStudio: Integrated development for R*. Boston, MA. Retrieved from [www.rstudio.com](http://www.rstudio.com)
- 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. <https://doi.org/10.1002/jip.121>
- Schemmel, J., & Volbert, R. (2016). *Different people, different stories - Differential aspects of Criteria-Based Content Analysis*. Annual Conference of the European Association of Psychology and Law (EAPL). Toulouse, France.
- Schemmel, J., & Volbert, R. (2017). Gibt es eine personenspezifische Aussagequalität? - Zur Konsistenz von wahren und erfundenen Aussagen einer Person zu verschiedenen Ereignissen [Is there a person-specific content quality? - Consistency of true and fabricated accounts by the same person]. *Praxis der Rechtspsychologie, 27*(1), 79-104.
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*, 609-612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Sporer, S. L. (2012). Making the subjective objective? Computer-assisted quantification of qualitative content cues to deception. In E. Fitzpatrick, J. Bachenko, & T. Fornaciari (Eds.), *Proceedings of the workshop on computational approaches to deception detection* (pp. 78-85). Stroudsburg, PA: The Association for Computational Linguistics.
- Sporer, S. L. (2016). Deception and cognitive load : Expanding our horizon with a working memory model. *Frontiers in Psychology, 7*, 1-12. <https://doi.org/10.3389/fpsyg.2016.00420>
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251. <https://doi.org/10.1037/0033-2909.87.2.245>
- 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.
- Tully, B. (1998). Reliability of criteria-based content analysis of child witness statements: Cohen's kappa doesn't matter. *Legal and Criminological Psychology, 3*, 183-188. <https://doi.org/10.1111/j.2044-8333.1998.tb00358.x>
- Undeutsch, U. (1967). Beurteilung der Glaubhaftigkeit von Zeugenaussagen [Credibility assessment of witness testimonies]. In U. Undeutsch (Ed.), *Handbuch der Psychologie - Forensische Psychologie* (pp. 26-181). Göttingen, Germany: Hogrefe.
- Volbert, R. (2015). *What can we learn from deception detection research for single case decisions?* Annual Conference of the European Association of Psychology and Law (EAPL) 2015. Nuremberg, Germany.
- Volbert, R., & Lau, S. (2013). Unspezifität des autobiographischen Gedächtnisses bei Depressiven - Konsequenzen für die Aussagequalität [Unspecificity of the autobiographical memory in patients with depression. Consequences for statement quality]. *Praxis der Rechtspsychologie, 23*, 54-71.
- Volbert, R., & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? *European Psychologist, 19*, 207-220. <https://doi.org/10.1027/1016-9040/a000200>
- Volbert, R., Steller, M., & Galow, A. (2010). Das Glaubhaftigkeitsgutachten [Credibility assessment in court]. In H.-L. Kröber, D. Dölling, N. Leygraf, & H. Sass (Eds.), *Handbuch der Forensischen Psychiatrie - Psychopathologische Grundlagen und Praxis der Forensischen Psychiatrie im Strafrecht* (pp. 623-689). Heidelberg, Germany: Springer.
- Vredevelde, A., van Koppen, P. J., & Granhag, P. A. (2014). The inconsistent suspect: A systematic review of different types of consistency in truth tellers and liars. In R. H. C. Bull (Ed.), *Investigative interviewing* (pp. 183-207). New York, NY: Springer New York. [https://doi.org/10.1007/978-1-4614-9642-7\\_10](https://doi.org/10.1007/978-1-4614-9642-7_10)
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). New York, NY: John Wiley.
- Vrij, A. (2016). Baseline as lie detection method. *Applied Cognitive Psychology, 30*, 1112-1119. <https://doi.org/10.1002/acp.3288>
- Vrij, A., Granhag, P. A., & Mann, S. (2010). Good liars. *Journal of Psychiatry & Law, 38*, 77-98. <https://doi.org/10.1177/009318531003800105>