Papeles del Psicólogo, 2005. Vol. 26, pp. 59-77

PSYCHOLOGICAL EVIDENCE IN COURT ON STATEMENT CREDIBILITY, PSYCHOLOGICAL INJURY AND MALINGERING: THE GLOBAL EVALUATION SYSTEM (GES)

Ramón Arce* and Francisca Fariña**

** Department of Social Psychology, University of Santiago de Compostela*

***Basic Psychology Department, University of Vigo*

This paper presents a protocol adapted to the Spanish judicial system for the assessment of malingering and deception in testimony, the Global Evaluation System (GES). The GES is the product of a combination of different strategies of assessment of the reliability (CBCA and Reality Monitoring) and validity (SRA and SVA) of statements, as well as the assessment of psychological injury and malingering. Moreover, this procedure has been validated for different cases and types of witness (children, adults and the disabled), and we review the productive and effective interview protocols for the assessment of credibility and psychological injury.

El presente trabajo presenta un procedimiento de detección de la mentira y el engaño en procesos judiciales adaptado al sistema judicial español: el Sistema de Evaluación Global. El S.E.G. resulta de un compendio de diversas estrategias de análisis de la fiabilidad (CBCA y Reality Monitoring) y validez de la declaración (SVA y SRA) a la vez que de la huella psicológica y de la simulación de la misma. Además, para poder obtener unos protocolos susceptibles del presente análisis se revisan y presentan los diversos modos productivos y efectivos en la obtención de la declaración con adultos, menores, discapacitados así como la entrevista clínico-forense para la valoración del daño psíquico y de la simulación.

The formation of legal judgements and their subsequent materialization in judicial sentences, which constitute the cornerstone of the Judicial System (Sallmann and Willis, 1984), rest on the two basic dimensions formulated by the *Information Integration Models*: reliability and validity (Ostrom, Werner & Saks, 1978). Briefly, a judgement is an assessment of the evidence in one dimension. Judgements are based on a set of beliefs about the evidence (e.g., inferences about the accused's motives, or capacities) that are relevant to the assessment dimension, so that each belief has a weight that affects assessment of the evidence for the judgement dimension. This weight is known as the scalar value of the belief. But not all beliefs contribute in equal measure to the assessment of the evidence: this contribution derives from the estimation of the reliability and validity of the belief. Reliability in the courtroom is basically defined by the credibility of the witnesses; validity is determined by the relevance of the evidence to the judgement to be made. However, credibility of witnesses contributes the highest scalar value in the verdict reached by both juries (Arce, Fariña & Real, 2000) and judges (Arce, Fariña, Novo & Seijo, 2001), and is the key to reaching conflicting verdicts (e.g., attribution of high credibility to a victim's testimony is a highly robust predictor of guilt, while lack of credibility for a victim's testimony reliably predicts a not-guilty verdict).

Estimation of the credibility of a testimony would be the appreciation of the accuracy or correctness inspired in the assessor by the witness or a part of his/her statement, which induces the assessor to believe that events occurred as stated (Mira, 1989). Two models have been formulated for the estimation of credibility: the subjective or social model and the objective or scientific model (Vrij, 2000). The social model is understood as the recourse to subjective indicators –not scientifically validated– of credibility, while the objective model rests on the assumption of empirical criteria of credibility. In everyday judicial practice, both experts, that is, judges and magistrates (Piñeiro, 2005), and lay persons in law (juries) (Arce, Fariña & Seijo, 2003) use subjective criteria for estimating the credibility of testimonies, which are ineffective for the correct classification of their accuracy. Therefore, courts find it necessary to employ qualified agents for assessing the credibility of witnesses in an empirical manner by means of productive and scientifically validated techniques. In relation to

Correspondence: Ramón Arce. *Facultad de Psicoloxía. Departamento de Psicoloxía Social, Básica e Metodoloxía. Universidade de Santiago de Compostela. 15706 Santiago de Compostela. Spain. E-mail: psarce@usc.es*

this, a review of the literature revealed that the procedures carried out by psychologists based on content analysis of statements were the most effective, classifying correctly, in simulation contexts (that is, with simulated witnesses in the laboratory), between 65% and 85% of statements, while in field studies (i.e., in real conditions) effectiveness reached 100% (Vrij, 2000). Nevertheless, these instruments present some limitations that should be addressed. First, the categories of the diverse procedures proposed for the analysis of content are not homogeneous across systems. Second, they do not usually cover all types of witness (they tend to be proposed for children). Third, these procedures are not accompanied by protocols suitable for assessing evidence in each context (minors, adults, the disabled). Fourth, these procedures do not analyze the entire judicial task (they tend to focus on the alleged victim, ignoring the alleged aggressor, who may be the true victim). Fifth, they are not oriented to the search for psychological injury resulting from the crime: post-traumatic stress disorder (PTSD). Sixth, they do not include appropriate methodologies for clinical assessment in the forensic context in which malingering, or clinical lying, has to be ruled out. And seventh, systems based on statement content analysis proposed for the study of credibility are semi-objective techniques that need adjustment for achieving greater objectivity.

On the basis of the literature on the empirical assessment of statement credibility, and with the aim of addressing the limitations referred to above, we have created (Arce & Fariña, 2002, 2005, in press) a forensic psychological protocol, the result of extensive research and professional experience, which attempts to deal comprehensively with the task, the *Global Evaluation System*. In its description we shall begin with a brief review of the protocols for obtaining a statement, which constitute essential requirements for applying the categorial systems of content analysis. We shall then present the most productive and effective systems for the analysis of credibility based on content analysis. Thirdly, we shall describe a protocol for assessment of psychological injury resulting from criminal acts and for the detection of malingering. Finally, we shall combine all of the above in a procedure adapted to the Spanish legal context for assessing the reality of memory trace and psychological injury with control of malingering: the *Global Evaluation System*.

OBTAINING THE STATEMENT

The basic tools for obtaining information from witnesses are interrogations and interviews. It is well known that the success of the interview or interrogation will depend on factors such as interviewer expertise, interviewee's degree of cooperation, time elapsed since the event, and obviously, type of interview (see Memon & Bull, 1999). Interrogations constitute the instrument par excellence for obtaining statements in police and judicial inquiries, but the testimony obtained through them is not productive for the application of systems for the analysis of credibility based on the content of the statements. Likewise, not all types of interview are valid for these purposes. Indeed, structured or semi-structured interviews may introduce misleading information in the accounts of truthful persons (e.g., Loftus, Korf & Schooler, 1988), so that the distortions would derive not from lying, but from the information introduced by the interviewer. Consequently, what is required are interviews in which the interviewer does not influence the information to be recalled by the witness. Interviews in free narrative format, such as cognitive interviews, fulfil this criterion. With regard to these, Köhnken, Milne, Memon and Bull (1999), after a review of the literature, observed that they facilitated higher rates of retrieval of information, especially correct (36%), but also incorrect information (17.5%). However, this does not mean that overall accuracy is greater in one type of interview than in another. In fact, average percentage of correct information for police interviews (interrogations) is 82%, while in the cognitive interview it is 84%. In this line, researchers in the field have proposed narrative interview formats adapted to the context and to the interviewee's capacities. We shall now briefly review each one of these formats.

COGNITIVE INTERVIEW[1]

The cognitive interview comprises four general techniques for retrieval of memories:

a) The first of these techniques consists in mentally reconstructing the physical and personal contexts existing at the time of the crime (or the event to be recalled), that is, the *reconstruction of contexts*. This involves asking the witness to situate him/herself mentally at the scene of the event, taking into account:

[1] Interviews must be recorded for subsequent content analysis

-Emotional elements (e.g., try to recall how you felt).
- Sequential elements (e.g., think what you were doing at the time).
- Perceptual characteristics (e.g., think of the scene of the crime and draw the room. What did it smell of? What did you hear?).

The reasoning behind this first technique is Tulving's principle of specific encoding, that is, the contextual information of an event is encoded together with the event and connected in an associative way (Tulving & Thompson, 1973). In turn, verbal recall of the event depends on the degree to which the contextual cues of the situation to be recalled overlap with properties previously encoded (Tulving, 1983). This first cognitive interview technique is similar to that used by judges and the police, the so-called reconstruction of the events, except that the reconstruction in the cognitive interview is carried out mentally.

b) The second technique, *free recall*, consists in asking the witness to recount everything that happened, absolutely everything, including partial information and seemingly trivial or insignificant details, since these can lead to others, associated in the memory, which are indeed relevant. This strategy is especially important when it is necessary to combine the information from different witnesses. Moreover, small details, in certain cases, can produce substantial clues.

By means of these first two techniques, retrieval of the mental image and reporting of all accessible details, an initial version of events is obtained. This statement, therefore, is of a narrative type, leaving the witness to speak without interruptions or questions. It is important to point out the need to ensure, throughout the interview, an appropriate environment for witnesses to concentrate, without noise or people to distract them, and the interviewer must obviously gain witnesses' confidence so that their testimony is as truthful and productive as possible.

c) The third technique, *change of perspective*, involves trying to encourage witnesses to put themselves in the position of the victim, or of another witness to the event –even the suspect–, and to report what they would have seen if they were in the position of that other person. This technique is based on the work of Bower (1967), who found that subjects, on imagining themselves as characters in a story, recalled more details pertaining to the perspective of the character with whom they had identified than those pertaining to other characters. In this way a second version of the interview is obtained, from a different perspective.

d) The final component is the instruction that invites recall from different starting points, *reverse-order recall*. In other words, the subject is asked to recount the event in a different order (e.g., from the end to the beginning, from the middle, backwards), with the aim of recovering small details that might have been lost in a narration of events simply in the sequence that they occurred. It is attempted through this technique to reduce the effects of previous knowledge, expectations and schemata on recall, and it may also be effective for eliciting additional details (Memon, Cronin, Eaves & Bull, 1993). In support of the use of this technique Bower and Morrow (1990) point out that we tend to recall the schema or mental model we form of an event, more than the event itself.

Application of the cognitive interview is not restricted to the reproduction of a single event, but can be extended to for recall of events that occur frequently in a similar manner (Mantwill, Köhnken & Ascherman, 1995). The cognitive interview includes, in turn, supplementary techniques such as:

a) Memory gymnastics for recall of physical appearance: Did the intruder remind of anyone you know? Was there anything unusual about his/her appearance?
b) Names: Try to remember the first letter of the name, going through the alphabet one letter at a time.
c) Objects: Describe the objects that were inside and outside the room. Did they look as if they were heavy?
d) Conversations and speech characteristics: Did you hear foreign or unusual words? Did you notice any accent? Did they stutter?
e) Car registration plates: Did the numbers or letters of the registration plate remind you of anything? Were they high or low numbers?

Apart from this standard version of the cognitive interview, Fisher & Geiselman (1992) proposed an improved version, adapted to the judicial context. Nevertheless, the effectiveness and procedure in cognitive terms are the same. The following phases summarize its general improved structure.

- Phase 1. Introductions and personalization of the interview (introductions, addressing interviewee by his/her name).
- Phase 2. Establishing communication (creating a pleasant atmosphere and a rapport by means of neutral questions).

- Phase 3. Explanation of the purpose of the interview.
- Phase 4. Reconstruction of contexts.
- Phase 5. Free recall.
- Phase 6. Preparation for the interrogation (interviewee is asked to concentrate hard, to say what comes to mind just as it comes, without *making it up*; to say, if necessary, *I don't understand, I don't know, I don't remember*; to activate and compare images).
- Phase 7. Interrogation compatible with the witness (each witness has a different memory sequence of the event, to which the interviewer must adapt).
- Phase 8. Recall from different perspectives.
- Phase 9. Reverse-order recall.
- Phase 10. Summary (made by the interviewer according to what the interviewee has reported).
- Phase 11. Close (emotional warming-down and dissipation of tensions in the interviewee).

INTERVIEWING CHILD WITNESSES

When witnesses are children, some important points should be borne in mind on using the cognitive interview procedure. First of all, one of the most consistent findings with regard to children's memory refers to performance. Briefly, in free recall tasks children retrieve significantly less information than adults: levels of detail and accuracy in recall of an event increase with age (Davies, Tarrant & Flin, 1989). Second, when the recall task is related to a meaningful and familiar context, the memory abilities displayed by children are greater than when the context is unfamiliar and without meaning for them (Bauer & Mandler, 1990). In other words, in these contexts the information they retrieve is not necessarily less productive and accurate. Given that the main objective of the cognitive interview is to increase the quantity of information recalled, this is the most appropriate procedure to employ with children (Memon & Bull, 1991). Nevertheless, it is necessary to modify the instructions so that the child understands what is being asked of him/her and to adapt the demands of the task to the child's mental capacities.

With the aim of adapting the cognitive interview for children, the British authorities (Home Office and The Department of Health, 1992) appointed Professors Diane Birch and Ray Bull to draw up a protocol for obtaining statements from child witnesses. On the basis of the draft prepared by these professors and discussions with technicians, a specific protocol was drawn up for working with children. Prior to the interview itself, it is recommended to collate information on the child's developmental state, level of language, and physical, social and sexual maturity. The protocol consists of four phases to be implemented by the interviewer in the following specific order: understanding and rapport, free recall, interrogation and conclusion. The first phase, *understanding and rapport* with the child, is of fundamental importance for obtaining the statement. If the child is not relaxed and comfortable, there will not be a good flow of communication. Moreover, it should be borne in mind that children are generally brought up not to speak to strangers. Therefore, it is recommended to begin by talking about *neutral* topics, such as hobbies, friends or school. Interviewers should also be especially careful with children who might *feel guilty* or think they have done *something wrong*. Thus, the interviewer must insist upon and make very clear the need for the interviewee to tell the *truth*. Finally, the object of this first phase is to inform the child about what is expected of him/her from that point on.

In the second phase, *free recall*, witnesses are asked to recount everything that happened (e.g., Is there anything you'd like to tell me? Do you know why we're here?). The interviewer can act as a facilitator, but must never formulate specific questions. Particular attention should be paid to the child's possible cognitions. In fact, child witnesses often think the adults already know what happened, or believe they should not give information about the events (as pointed out previously, children are generally told not to talk to strangers). Therefore, it is important to stress to them that they must tell everything. Throughout the interview, the interviewer should adopt a posture of active listening, resisting the temptation to intervene during pauses and long silences.

In the third phase, *interrogation*, the following order of priority should prevail in the formulation of questions: open questions, specific but not leading questions, closed questions and deep questions. Once the interviewer is totally satisfied that the free recall is finished, he/she may, where appropriate, pose open questions so that the child provides more information on some points that need clarification. However, it is important to bear in mind that questions of the form *why?* may cause feelings of guilt in the child, so that some caution should be exercised. Moreover, both reformulation of questions and requests for repetition of a response should be avoided, given that these can be interpreted as criticism, or that the answer was wrong,

respectively. The purpose of specific questions will be to clarify certain responses previously obtained. Even so, effective control should be exercised over inherent *suggestion* in this type of question –that is, the question should not imply its own answer. Likewise, questions with bipolar response alternatives (such as *yes* vs. *no*) should be excluded at this point of the interview. In any case, the content of the questions will be mediated by the child's level of development. Closed questions, on the other hand, will be employed if the previous types have failed to produce the desired results. Questions with only two response alternatives should be avoided as far as possible, given that children display a tendency to choose the first available option, especially if it is *yes*; therefore, if this type of question becomes unavoidable, witnesses should be provided with an escape route, such as *don't know* or *don't remember*. Finally, the interviewer may formulate deep questions, which are those whose reply is implicit in them. In contrast to the proposal of the original protocol, we feel it inadvisable to ask about the identity of the perpetrator of the crime, for two reasons. First of all, it is not the expert investigator's task to identify the criminal, but rather to reach a judgement on the reliability of the facts described. Second, it would be imprudent to give names, since our procedures for the assessment of reality are for events, so that the intentional transposition of persons cannot be detected by them.

The fourth phase, the *close of the interview*, consists in a recapitulation in which the interviewer inquires, using language adapted to the child's level of development, whether what has been reported in the interview is correct; and a rounding off, in which the aim is to bring down the levels of anxiety and tension (as at the beginning of the interview, the interviewer employs neutral questions, thanks the interviewee for his/her cooperation, and lets him/her know that he/she has been of help).

To this general protocol, we have added some complementary procedures for specific cases. First, when dealing with information on concepts that the child had not yet operativized adequately, such as the estimation of time or length, we resort to comparisons with well-established referents. Thus, if we want to know the duration of a particular event, we can compare it with school breaktime (e.g., was it shorter, longer or the same length as breaktime? If the answer is that it was shorter, it probably lasted less than half an hour). Second, we employ complementary types of language for those cases in which the child's level of linguistic development advises it; that is, there are times when the minor has the images of the facts in mind, but lacks the linguistic skills to transform them into words, or this lack of ability makes the statement very short, and thus unsuitable for our purposes. In these cases, using other communication procedures with the child is not only appropriate, but also advantageous. The communication systems we employ are drawings or acting out of the mental images in question. Obviously, these types of communication are complementary, and should not be used in isolation, since they can easily lead to misinterpretation. However, when child witnesses describe or try to describe events or actions they do not understand, and for which they lack vocabulary, they can draw them, point to them on a model or doll or on themselves, act them out, or represent them using objects. It should be borne in mind that it is inadvisable to use *anatomically correct dolls*, since these have a high probability of causing errors in the expert's interpretations (Dammeyer, 1998), and because the memory elicited through this type of figure also increases the number of incorrect responses among younger children (Goodman et al., 1997). In any case, it should be clear that the information obtained by these means can only be for clarification and complementary, and would never in itself constitute evidence for prosecution.

How can we decide on whether to use the cognitive interview or the children's protocol? The basic difference between the two procedures resides in the cognitive abilities demanded of the child. The cognitive interview requires, for example, the capacity for empathy, for change of perspective. Therefore, if this capacity is assumed to be acquired gradually from age 8 or 9 (Vrij & Winkel, 1996), it is advisable to use the protocol for children with minors under 8. In any case, children under 7 have difficulty following the techniques involved in the cognitive interview (Memon, Cronin, Eaves & Bull, 1996). And even if the cognitive interview may be effective for those over age 7, there are some risks, such as those deriving from the fact that the responses can be demand-led (Memon, Wark, Bull, & Köhnken, 1997). In cases of doubt, it is advisable to begin with the cognitive interview and, if a lack of cognitive skills is noted in the interviewee, to change to the protocol for children. We have found this system to be practicable, and that it does not lead to distortions.

INTERVIEWING DISABLED PERSONS

The scientific literature on interviewing disabled persons is extremely scarce, and there is a glaring need for more studies on the specific techniques to be applied (Bull, 1995). In fact, research has found that through the cognitive interview, 32% more correct information is retrieved, but that at the same time, there is a significant increase in the quantity of confabulation. In any case, of even more concern are the problems related to leading questions, to closed questions and to deep questions. In this line, Cahill et al. (1988, cited in Bull, 1995) have drawn up a list of aspects to be avoided by interviewers:

a) Acquiescence of witnesses with leading or suggestive questions, so that the reply is that which was asked for.
b) Undue pressure that leads witnesses to confabulate (e.g., to feel as though they were part of an event they did not in fact witness).
c) Repeatedly asking questions about a particular point, leading witnesses to make conjectures or deviate from their initial response (repeated questions lead interviewees to assume that their first answer was not correct).
d) Haste in labelling the language used by these witnesses as ambiguous or inadequate.
e) Offering descriptions to witnesses with difficulties for finding their own words (e.g., "if the jacket was not dark or light, then, would you say it was a kind of brown?").
f) Providing witnesses with closed response alternatives (e.g., "Did he have a revolver or a rifle?).
g) When the witness uses a tag such as *you know?*, the interviewer should proceed in a way that allows the extraction of information (one possibility is for the interviewer to tell the witness directly that he/she does *not* know, and that he/she wants the witness to explain).
h) Ignoring a previous fragment of information from the witness that does not fit with the interviewer's assumption of how events occurred.
i) Failure to understand everything the witness relates.
j) Failure to check, using all appropriate means, that the witness has been understood.

Given the current state of the literature, and in response to demands from the courts, Arce, Novo and Alfaro (2000) drew up some guidelines and a protocol for these cases. Prior to evaluating the testimony of a supposedly disabled person, it is necessary to determine their capacity as a witness. Under the Spanish judicial system, there is in fact a general obligation to testify: the LECrim (Art. 410) states that all persons resident in Spain, be they Spanish or foreign, are obliged to respond to the call to testify in judicial proceedings. At the same time, however, it establishes a series of exemptions from this obligation, among which are physical or moral incapacity (V. Art. 417, para. 3).

The procedure begins with the application of the corresponding Wechsler scales, generally the WAIS. These provide highly reliable information on the person's intellectual capacity, and are also a robust indicator of brain lesion. Specifically, in the WAIS, we are alerted to possibly relevant lesions by those results with a significant difference between the verbal and manipulative subscales (some authors use a criterion of >10 points, while others use >15) (see Wechsler, 1976, for a review) and, in the case in question here, the manipulative coefficient should be higher than the verbal one (possible organic lesion in the left hemisphere).

Likewise, interviews and the subject's antecedents will contribute important information for defining whether or not a lesion is a possibility, as well as data on the extent of incapacity. However, not all brain lesions incapacitate subjects for giving information, even in those cases in which they may be incapacitated as courtroom witnesses.

Thus, once a possible lesion or deficit has been detected, the experts set out to identify the areas –above all the cognitive ones– affected. The recommended tools are the Test Barcelona (Peña-Casanova, 1990) and the ERFC (Gil, 1999), for their reliability and because they combine in a single instrument the evaluation of the areas that a permit a testimony to be adapted in accordance with the witnesses' limitations.

Initially, the operative lesions are classified into aphasias, alexias and acalculias. Obviously, the lesions tend to relate to one of these. However, each grouping is subdivided into areas. After the global assessment, it is necessary to make an assessment of the adequacy of the witness's testimony (i.e., identification of accessible areas, deficient areas and forms of obtaining the testimony), and the conditions under which a reliable testimony can be obtained (e.g., a statement made to a forensic psychologist expert in this type of assessment may be valid, but not one given to judges, police or lawyers). Thus, for example, a person with anomia and orophonatory praxis may display difficulties in verbally identifying behaviours and may mix up words (e.g., bandage and

bondage), though this does not necessarily render their testimony unreliable.

It is also common to find witnesses who present disorders with clear implications for obtaining a statement in the *verbal memory of texts* task. Specifically, it may be that they are incapable of direct recall of the event, but still capable of responding to an interrogation on it, insofar as they lack episodic memory but not memory of the event if this is guided by questions. Consequently, they can reply perfectly well to an interrogation, and lack the capacity to create a coherent false testimony on lacking episodic memory (this condition has been observed in habitual drug users). In the most adverse of cases, when the deficit is severe, subjects may also provide useful information, since people with neurological damage, as long as their vision is intact, can store and recover visual information (Freed et al., 1989; Hart & O'Shanick, 1993; Winograd, Smith & Simon, 1982). In fact, all that is necessary is the reproduction of the context (bearing in mind that in many of them their processing is slow, so that extreme patience is required to avoid interrupting them in the middle of their search). Under this contingency, effective lying is impossible, and the crucial task is to find a system of communication that will be defined by the neurological analysis, the most accessible being reconstruction of the events. This procedure proved to be productive with different types of deficit, and has been validated for the judicial context in various cases.

THE CLINICAL-FORENSIC INTERVIEW

A final source of information with implications for the reliability of a testimony comes from the clinical context. The instruments normally used in clinical measurement are developed with a view to dealing with a *patient*. The study of malingering is not relevant, and therefore, neither structured/semi-structured interviews nor symptom lists and psychometric instruments are suitable for the purpose of checking the simulation of a mental disorder, since they actually provide information that facilitates such malingering.

For example, the question *Do you have headaches?* (first question from the SCL-90-R, Derogatis, 2002) makes it easy for the malingerer to produce a response consistent with the simulation. This type of question provides subjects with *a guiding path* for selection of the symptoms associated with a given mental disorder, so that all they need is the ability to discriminate between items belonging to one pathology or another.

The available data show that there are no references for the traditional clinical interview in which a diagnosis of simulation is reached (e.g., Rogers, 1997), while subjects are capable of effectively simulating an illness and discriminating it from others (e.g., Arce, Carballal, Fariña & Seijo, 2004; Arce, Fariña & Pampillón, 2002). Even though some psychometric instruments have scales for controlling the *validity* of the data registered, these are not sufficient for reliably establishing malingering, because: a) the malingering diagnosis is compatible with the formulation of alternative hypotheses (e.g., Graham, 1992; Roig-Fusté, 1993), b) not all malingerers are correctly classified (e.g., Bagby, Buis & Nicholson, 1995), and c) what are provided are not diagnoses, but diagnostic impressions.

Consequently, decisions based solely on this type of instrument leave the door open to the systematic commission of two types of error: false positives (classifying those who are actually ill as malingerers) and errors of omission (failing to detect as malingerers those subjects who are in fact simulating).

In view of these problems, and with the aim of minimizing the sources of error, a multi-method assessment strategy has been proposed (e.g., Rogers, 1997). This context makes room for an interview of a clinical nature that permits a diagnosis, and whose data can be compared with those obtained by other methods.

Thus, we have developed the so-called *Clinical-Forensic Interview*. This interview, carried out by a trained interviewer with psychopathological expertise, consists in asking subjects to list, in free narrative format, their current symptoms, behaviours and thoughts, compared with their state prior to the crime (i.e., GAF on axis V of the DSM-IV-TR). If the subjects do not respond of their own accord, they are asked by means of open questions, in accordance with axis V of the DSM-IV-TR (American Psychiatric Association, 2002), to talk about their family relationships, social relationships and workplace relationships, these being assessed on the appropriate scales. Through this procedure, subjects are given a task relating to knowledge of the symptoms they present, whilst for structured and semi-structured interviews, lists of symptoms and psychometric instruments they perform a task of symptom *recognition*. For this reason, the interview is not in the form of an interrogation, but rather *non-directive*, and oriented to the reconstruction of contexts. In other words, we use the procedure of open, free-narrative interview followed by reconstruction of

contexts. This kind of interview procedure showed itself to be reliable, valid and productive in the detection of simulation of post-traumatic stress disorder in cases of alleged sexual assault and harassment (Arce, Fariña & Freire, 2002), gender violence (Arce et al., 2004) and road traffic accident (Arce, Fariña, Carballal & Novo, 2006), in the detection of a non-imputable mental disorder (Arce, Fariña & Pampillón, 2002).

The clinical information obtained should be recorded and its content analyzed. The categories of analysis are the symptoms described in the DSM-IV-TR. Thus, we created a mutually exclusive, reliable and valid categorial system, of the type Weick (1985) refers to as methodical category systems. Once the registration sheets have been drawn up, the different symptoms detected are noted. While the vast majority of the symptoms, including the most adverse ones, can be reported directly by subjects, (Lewis & Saarni, 1993), some can only be observed. Consequently, two complementary methods are involved in the detection of categories: direct report from the subject and inferences made by the coders after analyzing the protocols. For example, deterioration of memory can be reported, or reflected, directly by the subject or inferred by the coder after the interview.

IN SEARCH OF THE TRUTH: CONTENT ANALYSIS OF THE STATEMENT

The review of the literature on content analysis of statements led us in the *Global Evaluation System* to consider the assessment of the credibility of witness statements according to two parameters: validity and reliability. Validity serves to establish the admissibility of the evidence for the content analysis, while reliability is related to the indicators of reality contained in the statement.

ANALYSIS OF THE VALIDITY OF THE STATEMENT

The study of the validity of the statement as evidence is made on the basis of the complete procedure (e.g., statements to the police or a judge, other testimonies, other evidence provided) and the recordings of the statements given to the expert investigators. There are two systems for the analysis of statement validity: Statement Reality Analysis (SRA) and Statement Validity Analysis (SVA).

The system known as SRA (Undeutsch, 1967, 1988), analyzes the validity of the testimony by means of the following categories:

a) Negative or Control criteria:
- Lack of internal consistency (contradictions).
- Lack of consistency with the laws of nature or science.
- Lack of external consistency (discrepancy with other incontrovertible facts).

b) Criteria derived from the sequences of statements:
- Lack of persistence (stability in time and contexts).
- Statement inconsistent with a previous statement.

On the other hand, SVA (e.g., Steller, 1989), employs the following assessment categories:

a) Psychological characteristics:
- Appropriateness of language and knowledge.
- Appropriateness of emotional expression.
- Susceptibility to suggestion.

b) Interview characteristics:
- Coercive, suggestive or leading questions.
- Global appropriateness of the interview.

c) Motivation:
- Reasons for making a statement.
- Context of the original statement.
- Pressure to present a false statement.

d) Investigation issues:
- Consistency with the laws of nature.
- Consistency with other statements.
- Consistency with other evidence.

As a criterion for global assessment of the statement, this is indicated by its best fit to one of the following categories: *credible*, *probably credible*, *indeterminate*, *probably incredible or incredible*.

RELIABILITY OF THE STATEMENT

The study of the reliability of statements –the search for criteria of reality in their content–, carried out from recordings of the statements given to expert investigators, constitutes the major contribution of Forensic Psychology to the assessment of evidence. Three categorial systems, based on content analysis, have been proposed and shown to be productive and effective for assessing the reliability of evidence: Reality Monitoring, SRA and CBCA.

Reality Monitoring, in its seminal proposal (Johnson & Raye, 1981), asserts that true statements contain more contextual (spatio-temporal) and sense-related (sounds, smells, etc.) attributes, while fabricated testimony includes more cognitive operations, i.e., idiosyncratic information (for example, I thought, I remember seeing, I felt nervous). Spörer (1997) extended the list of criteria to eight: clarity (as opposed to vagueness), perceptual information (sense-related information, such as sounds, tastes or visual details), spatial information (places, locations), temporal information (location of the event in time,

description of event sequences), affect (expression of emotions and feelings experienced during the event), reconstruction of the story (plausibility of reconstruction of the event based on the information given), realism (plausibility, realism and sense of the story) and cognitive operations (descriptions of inferences made by others during the event). The first seven of these are linked to truth, and the eighth to falsity, making this new categorization more effective. Validation of memory attributes is usually carried out by means of comparison between the results of the statement and the prescriptions of the model, but it can also be done through a process of reasoning that involves analysis of the qualitative characteristics of the memory trace, the characteristics of related traces and mnesic assumptions.

SRA (Undeutsch, 1967, 1988) also uses categories for assessing the credibility of the statement. These are as follows:

a) General, basic criteria:
- Spatio-temporal anchorage (fixing of the action in a space and time).
- Concreteness (clarity, vividness).
- Richness of detail (large quantity of details in the narration).
- Originality of the narrations (as opposed to stereotypes or clichés).
- Internal consistency (logical and psychological coherence).
- Mention of specific details of a particular type of sexual aggression.

b) Special manifestations of the above criteria:
- Reference to details that exceed witnesses' capacity (that go beyond their imagination or capacity for understanding).
- Reference to subjective experiences (feelings, emotions, thoughts, fears).
- Mention of unforeseen events or unexpected complications.
- Spontaneous corrections, specifications and additions during the statement.
- Statements that negatively affect their own interests.

With all these decision criteria a global assessment is made, in which the two factors *general criteria* and *special manifestations of the general criteria* are weighted positively towards truth, i.e., the presence of these criteria indicate that the statement is true, but their absence does not imply that it is false. For their part, the presence of the validity criteria *Control criteria* and *Criteria derived from the sequences of statement* are detrimental to the truth value of the statement. In any case, it should be borne in mind that each criterion has a limited weight in the determination of category (true vs. false) or the extent to which a statement represents a situation actually experienced by the witness. Moreover, it prescribes adherence to four maxims in the determination of whether the story relates a real event or not:
- Intensity of the comments in the different criteria.
- Number of details in the story that are related to a criterion (or more).
- Witnesses' capacity for giving evidence (age, intelligence, suggestibility).
- Characteristics of the narrative event (e.g., complexity, relevance).

In 1994, Steller and Köhnken proposed, on the basis of previous work, an integrated system of categories whose purpose was to assess the statements of minors who were the alleged victims of sexual abuse. This system, Criterion-Based Credibility Assessment (CBCA) consists of five main categories with 19 criteria for assessment:

a) General characteristics:
- Logical structure (coherence and internal consistency).
- Lack of structure (disorganized presentation).
- Number of details (abundance of different details or facts).

b) Specific content:
- Contextual machinery (situation of the narrative in space and time).
- Description of interactions (chain of actions between witness and other actors).
- Reproduction of conversations.
- Unexpected complications during the incident (e.g., unexpected interruption).

c) Peculiarities of content:
- Unusual details (details with low probability of occurrence).
- Superfluous details (irrelevant details that do not contribute significantly to the body of facts).
- Incomprehension of details accurately recounted (provision of details that the child does not understand, but that indeed make sense).
- Related external associations (inclusion of information external to the events in question, but related to them, e.g., in a case of sexual assault, recalling previous conversations about the subject).
- Reference to subjective mental state (references to

one's own feelings, emotions or cognitions).
- Attributions about the perpetrator's mental state (references to aggressor's mental state and attribution of motives).

d) Content related to motivation:
- Spontaneous corrections (spontaneous correction or improvement of the statement).
- Admission of lack of memory (acknowledgement of gaps in memory).
- Doubts about one's own testimony.
- Self-disapproval (critical attitude to one's own behaviour).
- Forgiveness of the perpetrator of the crime (victim's statement favours the accused, or avoidance of more accusations).

e) Specific elements of the assault:
- Characteristic details of the offence (descriptions that contradict the usual beliefs about the crime).

These criteria of content can be analyzed as present or absent, or can be scored according to the strength or degree with which they appear in the statement. In any case, the presence of these aspects will favour the interpretation that the statement is true, while from their absence it cannot be inferred that it is false. With regard to the cut-off point for discriminating between statements based on truth and the rest, Steller (1989) found true statements to contain at least 7 truth criteria.

IN SEARCH OF PSYCHOLOGICAL INJURY FROM THE CRIME: THE DETECTION OF MALINGERING IN CLINICAL ASSESSMENT

Being a victim is understood as the fact of having been the object of a crime, but it also involves the whole set of consequences of the criminal act. These can be of various types: physical, economic, social or mental. Psychological injury, like memory trace, may in an expert assessment come to constitute evidence for prosecution. However, in a context such as the one we are concerned with here, the medical-legal context, it is not sufficient to diagnose a disorder or disorders: it is also necessary to rule out malingering (American Psychiatric Association, 2002). For this dual objective –clinical diagnosis and control of malingering– ordinary clinical assessment is not effective. Indeed, traditional clinical assessment has never given information on malingering (e.g., Rogers, 1997). For the measurement of psychological injury and control of malingering (hypothesis to be checked in the measurement of psychological injury caused by a crime), Arce, Fariña and Pampillón (2002) have created and validated a protocol in accordance with the responses and strategies employed by malingerers. This is based on the operative distinction between positive criteria, which validate the protocol, and negative criteria, which invalidate or mitigate its validity, indicating simulation or malingering. Positive criteria would be those not detected in the protocols of malingerers, and these are *non-avoidance of responses* and *social desirability*. Specifically, those subjects assessed by the MMPI (Minnesota Multiphase Personality Inventory) scales and who significantly refrain from responding (? Scale) and tend to give responses of social desirability do not follow the typical strategies of the malingerer, so that this should be interpreted more as an indication of truth of the protocol than as an attempt at simulation. It should be borne in mind that lack of cooperation in the assessment (non-response) had been proposed as a reliable indicator of malingering (e.g., Rogers, 1992; Lewis & Saarni, 1993; Bagby et al., 1997), but this contingency was never observed among malingerers in a forensic assessment.

The negative criteria, that is, observed in the protocols of the malingerers, were: 1) the measurement systems of MMPI, interview or others do not detect, in valid protocols, mental illness (in other words, if the measurement instruments fail to detect any mental disorder, no such disorder can be imputed in the legal context); 2) detection of malingering by the validity control scales of the MMPI and its combinations; 3) detection of some malingering strategy in the interview; and 4) lack of inter-measure agreement. The first criterion is eliminatory: if the mental disorder is not measurable, no psychological injury can be imputed as evidence. The others, in themselves, are not determinant, so that fulfilment of at least two criteria and the study of alternative hypotheses would be necessary to draw a conclusion in relation to simulation of psychological injury.

It is for these latter indicators of non-validity that we formulated the concept of *convergent invalidity*, which requires at least two totally independent indicators of for assessing a protocol as invalid. In accordance with these criteria, we drew up the following proposal for an action protocol:

a) Use of complementary and concordant systems of measurement, which presuppose the performance of different tasks involving validity control systems. Thus, it is proposed to subject the witness to a psychometric assessment, involving a symptom-recogni-

tion task, and also to apply a knowledge task, the *Clinical-Forensic Interview*. As regards the psychometric instrument to be used, the MMPI is the instrument of reference for the assessment of psychological injury in forensic practice (Butcher & Miller, 1999), but it requires a high level of comprehension on the part of the subject. Should subjects have difficulty filling out the MMPI, the SCL-90-R checklist permits the assessor to circumvent this problem, and has measures for validity control of the protocol. Therefore, we recommend this instrument as a substitute for the MMPI or, in case of doubt, as a complement. The first measure is taken through the interview to control the effect of learning the psychometric task in the knowledge task. With regard to inter-measure agreement, it must borne in mind that this will not be total. We should take into account that even test-retest measures fall short of perfection.

b) Analysis of the internal consistency of the measures: scales for control of the psychometric instruments, and, in the interview, content analysis seeking common malingering strategies. The validity control scales of the MMPI-2 (Hathaway & McKinley, 1999) with implications for the study of malingering according to this protocol are the original validity scales (no-response scales, L, F and K), the additional indicators of protocol validity (F posterior, TRIN, VRIN), and the indices that have proved effective in the detection of disorder simulation, the *F-K index* and the *inverted-V profile* (Duckworth and Anderson, 1995). If the psychometric assessment is obtained through the SCL-90-R (Derogatis, 2002), the validity scales would be PST, PSI, GSI and PSDI. As regards the interviews, these are subjected to content analysis taking as categories the strategies followed by malingerers in the interviews: avoidance of response, strange symptoms, combination of symptoms, obvious symptoms, consistency of symptoms, improbable symptoms, indiscriminate grouping of symptoms and severity of symptoms.

c) It is advisable for two assessors to separately carry out the assessment, so that inter-assessor consistency can be measured. Briefly, this safeguard serves to control for possible biases of measurement and interpretation in the assessor.

d) Study of reliability of the assessment: internal, inter-measure, inter-context (antecedents, documentary evidence, etc.) and inter-assessor consistency (Wicker, 1975).

e) Control of false positives, that is, real disorder sufferers, through the study of the subject's antecedents and general history, of the alternative hypotheses in each non-validity indicator (see Roig Fusté, 1993, Graham, 1992), and of fulfilment of the Clinical Decision Model criteria for the establishment of malingering (Cunnien, 1997).

f) Anamnesis or study of antecedents. The aim here is to reinforce the assessment with the subject's antecedents, data from his/her social context, a study of his/her behaviour, compilation of documentary evidence, other testimonies, and so on.

g) Psychological study of psychological injury. The clinical measures provide data with respect to what is legally referred to as the *biological assessment*, but also required, according to the legal demands, is the *psychological assessment*, which clarifies the relationship between the psychological injury measured and the psychological injury expected for that case.

h) Finally, the discriminant validity can also be tested. In other words, assessors can apply a measure unrelated to the case, such as on values or personality (16-PF, SIV), with the expectation of no relationship with the objective assessment, in order to rule out an attempt by the subject at manipulation of his/her image, either positively or negatively.

The resulting impression about malingering must be fitted to one of the following categories: *probable malingerer and probable non-malingerer*. It is important to avoid attempting to establish certainty (e.g., situating the impression on a scale of several points), as this creates confusion in decision-makers (e.g., sentence of the Spanish Supreme Court, 29 October, 1981, RA 3902), and it is crucial to use probabilistic terms, since psychological assessment is subject to error.

THE GLOBAL EVALUATION SYSTEM

The *Global Evaluation System* (GES) is structured around 9 tasks which we shall briefly describe and explain below: obtaining the statement, repeating the statement, checking of the statements obtained in the course of the judicial process, content analysis of the statements, reliability analysis of the measures, measurement of clinical effects of the traumatic event, assessment of statements from persons involved, analysis of personality and capacities of those involved, and finally, implications for presentation of the report. The tasks to be performed are mediated by the case to be assessed (e.g., if in a given

case it is not possible to assess the accused, this phase is not implemented). The phases in the most complete version of the system are as follows:

a) *Obtaining the statement (memory trace).* For the forensic psychological procedure for assessment of statements and psychological injury to be productive, reliable and valid, assessors need instruments for obtaining the statement and measuring the clinical condition that permit their subsequent analysis. Therefore, the statements must be obtained, depending on whether the subjects are adults, minors or disabled persons, through the following procedures: Improved Cognitive Interview (Fisher & Geiselman, 1992), Memorandum of Good Practice (Bull, 1997), or the Forensic Interview for the Disabled (Arce, Novo & Alfaro, 2000). Psychological injury in the *knowledge task* is measured through the *Clinical-Forensic Interview* (Arce & Fariña, 2001; Arce, Fariña & Freire, 2002; Arce, Pampillón & Fariña, 2002).

b) *Repetition of obtaining the statement.* The methods proposed are based on a single measure of the memory trace. However, with a single measure, we lose the possibility of analyzing the temporal consistency of the statement (below we discuss the validity of the other statements obtained in the course of the judicial process). In other words, we leave out one of the forms of checking the validity of the information: temporal or intra-witness consistency (e.g., Wicker, 1975; Schum, 1977). Likewise, legal doctrine has defined the reliability of a testimony according to opportunity criteria (opportunity to observe, etc.), bias (control of possible interests), temporal consistency, plausibility, inter-witness consistency and credit (Schum, 1977). Also, our own jurisprudence (e.g., sentence of the Spanish Supreme Court, 29 April, 1997) establishes, when the testimony of the victim is the sole or central evidence for the prosecution, that the testimony must display the following three characteristics: absence of subjective incredibility, some peripheral corroboration of an objective nature, and persistence over time without ambiguities or contradictions. In this line, sentences have already been pronounced that annul the evidence value of content analysis (CBCA and SVA) of statements based on a single statement (e.g., AP, Pontevedra, Sección 6ª, 21 January, 2004). In sum, both scientific methodology and legal doctrine and jurisprudence demand more than one statement for the study of temporal consistency. In this regard, it has been found that the repetition of obtaining the statement need not contaminate the data from an interview not contaminated from outside (e.g., Campos & Alonso-Quecuty, 1999), as is the case for the protocols for obtaining the statement mentioned previously. Consequently, in the first measure the assessor must by no means interrogate the subject, employing solely the reconstruction of contexts, free recall, change of perspective and reverse-order recall. Interrogation, where necessary, is to be left for the second measure, so as not to contaminate the memory of events with the interrogation. From a second measure, the assessors obtain a consistency analysis, which, according to the Undeutsch hypothesis (1967, p. 125), should be understood as a function of the centrality/periphery of the contradictory material. It should be pointed out here that the contradiction is only relevant if it affects central details for the act of judgement. Inconsistency in peripheral information or the omission of certain information is only important if that information is crucial to the construction of a real event. In order to leave room for interferences (theory of the interference of forgetting), the entry of new information (constructive hypothesis of forgetting) and the forgetting curve, we estimate the time that should elapse between interview and interview at over 1 week (but not much more). We establish three axioms with respect to this. First, since the criminal act constitutes a *stressful life event*, the *obsolescence effect* will be weaker (in reference to the testimony of both plaintiff and defendant, and contiguous with the facts). Second, a *theory of rationality* on the part of the fabricator, so that the lie is planned, learned and, by extension, consistent in time; hence, the lie will not be mediated by post-event interference and information (*constructive hypothesis*). Here, it is essential to obtain the first statement in the free narrative format, without any kind of interrogation, to avoid letting in post-event information that the subject would fit into the new reconstruction. Interrogation would only take place after obtaining the second statement in free narrative. Third, the subject who is telling the truth narrates images, so that the description of the facts, though quite similar, will be constructed differently, as it does not correspond to episodic schemata. In sum, and in free recall format, the true statement will be less consistent, and although the event is the same, the narration will be significantly different, in

terms of both its retrieval and its content (omissions, elicitation of events other those under investigation but related to them, inconsistency in peripheral information, retrieval of new information of little relevance to the events). For their part, simulating subjects narrate learned stories, so that they repeat them more or less the same each time, guided by an episodic schema. It is important to bear in mind that this second statement should always be considered from the perspective that does not contribute significantly to secondary victimhood.

c) *Checking of the statements obtained in the course of the judicial process*. Similarly, an analysis is made, according to the procedure of the validity study in SRA and SVA, of the other statements made in the course of the judicial process (e.g., indictment, inquiry). However, the value of these is relative. It should be borne in mind that many of them are transcriptions of what the witness has said, so that they do not reliably reflect the testimony. Moreover, the type of interrogation may have influenced the response. In this regard, one should bear in mind, in line with SVA, the effects on statement validity of interview characteristics (types of question formulated and suitability of the interview) and motivation (motives, context and pressure). For example, in the case of interrogations of children we have found many expressions and concepts of which the child, when asked, does not know the meaning (e.g., in the minor's statement there appears the expression *semen came out*; if the child, on being asked what *semen* is, does not know, then this expression does not belong to his/her statement). In turn, statements often refer to expressions (e.g., he raped me), rather than to narrations of events, so that the reliability and validity cannot be checked. Thus, the lack of consistency of the statements given to the expert investigators and others included in the judicial proceedings has a quite relative value. Where appropriate, it should be explained that this lack of consistency is not relevant for analysis of the plausibility of the statement. Furthermore, it is important to exercise more caution than we might initially think on considering confessions by the accused, and even more so, incriminations in exchange for benefits accruing to the informer. The source of bias can be found in the interrogations. Thus, the usual techniques for obtaining a confession are based on strategies such as: threats; attribution of responsibility to external causes, such as provocation by the victim; minimization of the seriousness of the crime; or the development of a personal relationship with the suspect (i.e., the typical "good cop, bad cop" strategy, with two interviewers, one hostile and the other friendly and protective). Finally, the strategy based on the Prisoner's Dilemma for obtaining the statement may lead to either cooperation strategies or competition strategies that distort the expression of the testimony (e.g., Kelley & Stahelski, 1970). In this regard, a decision by the US Supreme Court (Miranda v. Arizona, 1966) declared this type of interrogation as coercive.

d) *Content analysis of statements referring to the events.* Content analysis of the statements addresses two dimensions: the validity and the reliability of the testimony. According to the *Global Evaluation System*, the first task of the expert assessment consists in estimating the validity of the statement, not as judicial evidence per se –which is the business of the Judicial System–, but as evidence whose reliability is to be analyzed. In this regard, there are two types of potential attack on validity. First, the statement may be of insufficient length to be subjected to a reality analysis (Raskin & Steller, 1989); and second, the statement may be considered invalid as evidence on the basis of the validity criteria of the SRA and SVA (e.g., lack of internal consistency; lack of external consistency with other robust or incontrovertible evidence, such as that obtained by experts during the course of the judicial process; statement inconsistent with a previous one; lack of persistence in statements; inconsistency with the laws of science and nature) and, in the case of minors, indicators that limit the validity (indicators of suggestibility, inappropriateness of affect, inappropriateness of language and knowledge). If the evidence is deemed invalid, it is concluded that the statements do not constitute admissible or sufficient evidence; if it is deemed valid, the reliability (consistency with criteria of reality) of the statements is analyzed. As a categorial list of reference we use the categories of the CBCA. This analysis procedure, created in principle for the testimony of minors who were victims of sexual assault, is equally effective with adults (Landry & Brigham, 1992; Zaparnuik, Yuille & Taylor, 1995; Spörer, 1997; Vrij, Edward, Roberts & Bull, 1999), in sequences of measures, and in cases other than those of sexual assault (Porter & Yuille, 1996; Spörer,

1997; Arce, Fariña & Freire, 2002). In these new contexts, obviously not all the categories are productive. Thus, Landry and Brigham (1992) restrict use to 14 categories with adults, since three are applicable only to children (incomprehension of details related to accuracy; forgiveness of the perpetrator of the crime; and details characteristic of the offence), while another two (lack of structure and related external associations) were not productive. However, we (Arce, Fariña & Freire, 2002) found that the category *forgiveness of the perpetrator of the crime* was productive, in statements by adults, both for sexual assault and threats –that is, the productivity is subject to a context effect. In sum, all the criteria should in principle be considered in the analysis, since productivity depends on type of case, peculiarities of the action under examination and interviewee's sociodemographic profile. In turn, the combination of the CBCA and RM criteria is possible and effective, as their effects can be summed (Spörer, 1997; Vrij et al., 1999). Specifically, the combination of the two assessment systems, through the addition to the CBCA of the RM criteria *perceptual information* and *cognitive operations* (Vrij, 2000), slightly improves the reliability of the system. Thus, these two new criteria can be added to those of the CBCA. This procedure can be applied in repeated measures (see the hypotheses to be tested in the section *repetition of obtaining the statement*).

e) *Reliability analysis of the measures*. The original systems of statement content analysis constitute semi-objective techniques because they only examine the reliability and validity of the instruments, and do not contain procedures for control of the specific measure, i.e., the expert assessor's measure. With a view to addressing this methodological shortcoming and approaching an objective system, we propose a method that makes it possible to verify the reliability of the measure through the analysis of inter- and intra-measure, inter-assessor and inter-context consistency (Wicker, 1975). Inter-context reliability is addressed through recourse to a trained assessor who has been effective and consistent in other, previous contexts, that is, in previous expert investigations. By using two assessors (at least one of whom has been trained and showed reliability in previous assessments) performing the tasks separately, it is possible to make an assessment of inter-assessor consistency. As a statistical tool for the analysis of inter-assessor consistency, we propose the Agreement Index [AI= Agreements/(agreements+disagreements)], which is more restrictive than the kappa values, taking as cut-off point a value of .80 (Tversky, 1977). In other words, the results are only considered reliable if two assessors, separately, agree on more than .80 of the total assessments in each analysis category. Inter- and intra-measure consistency are checked by means of: internal consistency of the measures (e.g., the validity scales of the MMPI, the statements or the study of malingering strategies in the clinical interview); consistency between different measures (e.g., agreement between MMPI and clinical interview, between statements over time); and consistency –i.e., complementariness or its lack (one may present indicators of truth and the other indicators of falsity, or none at all)– between the assessments obtained for the plaintiff and the defendant.

f) *Measurement of clinical effects of the traumatic event.* The criminal act causes a series of injuries to the victim that are basically of a physical, psychological and economic nature. The psychological damage constitutes the so-called psychological injury of the crime and, as such, can be adduced as evidence for the prosecution. In relation to the assessment of psychological injury and the subsequent judicial evidence, criminal acts (e.g., lesions, breaking and entering, abuse, sexual abuse, kidnapping) can produce a psychological response corresponding to a diagnosis of post-traumatic stress disorder (PTSD) (Blanchard & Hickling; 2004, Echeburúa & Corral, 1998; Echeburúa, Corral, Zubizarreta & Sarasúa, 1995). Therefore, the measurement of PTSD is crucial for the detection of psychological injury. Special care should be exercised with indirect measures of PTSD (e.g., hypochondria, hysteria, depression, anxiety, dysthymia, social isolation, social maladjustment), which can serve as enhancers of PTSD, but are not substitutes for it. Furthermore, it is necessary to rule out causes other than the criminal act. For example, the combination of a process of divorce or separation with abuse may make it difficult to distinguish the source of the disorder, since the two contingencies produce similar psychological injury. In any case, the expert assessor must take into account the following maxims: not all criminal acts produce PTSD in the victim; and the absence of PTSD does not imply that the assault did not take place. At the

same time, once the psychological injury has been identified, it is necessary to check whether it is real or simulated, and to this end assessors can use the protocol previously described for the measurement of psychological injury with control of malingering.

g) *Assessment of statements from persons involved.* Although in principle the techniques of content analysis of statements and assessment of psychological injury were designed for the assessment of the victim or plaintiff's testimony, the same procedure can be applied to the defendant, thus making possible a study of the two versions together. The inquisitorial justice procedure, as employed in Spain, permits this confrontation (though this would not be the case with an adversarial system). With this procedure we can obtain an estimation of convergent validity.

h) *Analysis of personality and capacities of those involved.* The study of the personality of the actors can be crucial to an explanation of the accusation itself, the assault or any mental disorder of the defendant with relevant judicial implications; in other words, where applicable, the imputability of the accused is examined [see Arce, Fariña & Pampillón (2002) for a description of how the study of imputability is carried out]. Since a clinical assessment is not sufficient in the forensic context, the clinical assessment protocol with control of malingering (Arce, Fariña & Pampillón, 2002) is followed. Actors' cognitive capacities are measured by means of the corresponding Wechsler Scale, and as a source of contrast or for samples with language difficulties, poor education or from other countries, the non-verbal intelligence test TONI-2 (Brown, Sherbenou & Johnsen, 1995) is taken. Reliability of this last-named measure is checked through correspondence of responses with the difficulty gradient of the questions and inter-measure consistency. We use the assessment of cognitive capacities to rate the capacity to testify, and, where appropriate, to indicate their effects on criminal responsibility.

i) *Implications for presentation of the report.* The system of statement credibility in 5 response categories, as proposed in SVA, does not meet the requirements of the Spanish Judicial System. The Supreme Court demands complete certainty, not merely high probability (e.g., sentence of the Spanish Supreme Court, 29 October, 1981, RA 3902). However, all measures, and particularly psychological ones, are subject to error, so that we should acknowledge this, but refraining from establishing degrees of certainty which, in accordance with the considerations of the Supreme Court, lead only to greater confusion. Thus, the most appropriate categories would be *probably true*, *probably untrue* and, where applicable, *indeterminate* (interested forensic psychologists can obtain from the authors an expert assessment format based on the GES).It should also be borne in mind that the system is more robust in the identification of truth than of lies. Likewise, it is advisable not to make a description of events based on phrases, but rather on complete actions, since the procedure validates events, and not isolated parts. Therefore, in no case is it recommended to identify the alleged perpetrator, as the procedure does not validate this point.

FINAL CONSIDERATIONS

The reliability of the entire procedure is ultimately the responsibility of the interviewer/assessor. It is for this reason that the intervention be carried out by well trained and experienced professionals with a high capacity for objectivity (Alonso-Quecuty, 1993). Thus, exhaustive training is essential. This should include: a) training in all forms of obtaining all types of information (a procedure can be seen in Fisher et al., 1987); b) training in statement analysis [a structured programme can be found in Köhnken (1999)]; c) training in the assessment of personality and psychological injury, not for clinical, but for forensic purposes (see Arce, Fariña & Freire, 2002; Arce, Fariña & Pampillón, 2002; Echeburúa, Corral & Amor, 2002; Rogers, 1997); d) training in the detection of malingering (see the steps to be followed in Arce, Fariña & Pampillón, 2002); and e) first forensic assessments to be performed in the company of an expert investigator with experience. Finally, our experience suggests that the material used in training in content analysis and clinical assessment should be real, rather than simulated, since the task executed in the two contexts is different, and the effectiveness of the procedure also (Vrij, 2000). The Forensic Psychology Unit at the University of Santiago de Compostela periodically organizes training courses in these techniques.

ACKNOWLEDGEMENTS

This research was financed by the *Secretaría Xeral de Investigación e Desenvolvemento* (General Secretariat for Research and Development) of the Galician Regional

Government, within the *Excelencia Investigadora* (Research Excellence) project, Code: PGIDIT03CS037401PR.

REFERENCES

Alonso-Quecuty, M. L. (1993). Interrogando a testigos, víctimas y sospechosos: La obtención de información exacta. In M. Diges, & M. L. Alonso-Quecuty (eds.), *Psicología forense experimental* (pp. 85-98). Valencia: Promolibro.

American Psychiatric Association, (2002). *Manual diagnóstico y estadístico de los trastornos mentales* (IV Edición-Texto Revisado). Barcelona: Masson.

Arce, R. & Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de delitos: La entrevista clínico-forense*. Unpublished manuscript, Universidad de Santiago de Compostela.

Arce, R. & Fariña, F. (2002). *El Sistema de Evaluación Global para la evaluación de la credibilidad y el engaño en el testimonio judicial*. Unpublished manuscript, Universidad de Santiago de Compostela.

Arce, R. & Fariña, F. (2005). El Sistema de Evaluación Global (SEG) de la credibilidad del testimonio: Hacia una propuesta integradora. In R. Arce, F. Fariña & M. Novo (eds.). *Psicología jurídica* (pp. 101-118). Santiago de Compostela: Consellería de Xustiza, Interior e Administración Local.

Arce, R. & Fariña, F. (in press). Evaluación de la credibilidad y de la huella psíquica en el contexto penal. In Consejo General de Poder Judicial (ed.), *Psicología del testimonio y valoración de la prueba pericial*. San Sebastián: Centro de Documentación Judicial.

Arce, R., Carballal, A., Fariña, F. & Seijo D. (2004). Can mock battered women malinger psychological evidence in a recognition task? In A. Czerederecka, T. Jaskiewic- Obydzinska, R. Roesch & J. Wójcikiewicz (eds.), *Forensic psychology and law. Facing the challenges of a changing world* (pp. 327-336) Kraków: Institute of Forensic Research Publishers.

Arce, R., Fariña, F. & Freire, M. J. (2002). Contrastando la generalización de los métodos empíricos de detección del engaño. *Psicologia: Teoria, Investigação e Prática*, 7(2), 71-86.

Arce, R., Fariña, F. & Real, S. (2000). The assessment of information integration theory and confirmatory bias hypothesis on judicial proceedings. A case of rape and murder. In A. Czerederecka, T. Jaskiewic-Obydzinska & J. Wójcikiewicz (eds.), *Forensic psychology and law* (pp. 296-303) Kraków: Institute of Forensic Research Publishers.

Arce, R., Fariña, F. & Seijo, D. (2003). Laypeople's criteria for the discrimination of reliable from non-reliable eyewitnesses. In M. Vanderhallen, G. Vervaeke, P. J. Van Koppen & J. Goethals (eds.), *Much ado about crime* (pp. 105-116). Brussels: Uitgeverij Politeia NV.

Arce, R., Fariña, F., Carballal, A. & Novo, M. (2006). Evaluación del daño moral en accidentes de tráfico: Desarrollo y validación de un protocolo para la detección de la simulación. *Psicothema*, 18.

Arce, R., Fariña, F., Novo, M. & Seijo, D. (2001). Judges' decision-making from within. In R. Roesch, R. R. Corrado & R. J. Dempster (eds.), *Psychology in the courts: international advances in knowledge* (pp. 195-206). New York: Routledge.

Arce, R., Novo, M. & Alfaro, E. (2000). La obtención de la declaración en menores y discapacitados. In A. Ovejero, M. V. Moral & P. Vivas (eds.), *Aplicaciones en psicología social* (pp. 147-151). Madrid: Biblioteca Nueva.

Arce, R., Pampillón, M. del C. & Fariña, F. (2002). Desarrollo y evaluación de un procedimiento empírico para la detección de la simulación de enajenación mental en el contexto legal. *Anuario de Psicología*, 33(3), 385-408.

Bagby, R. M., Buis, T. & Nicholson, R. A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: replication and extension. *Psychological Assessment*, 7, 84-92.

Bagby, R. M., Rogers, R. Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., Schuller, D. R. & Seeman, M. V. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, 68(3), 650-664.

Bauer, P. J. & Mandler, J. M. (1990). Remembering what happened next: Very young children's recall of event sequences. In R. Fivush, & J. A. Hudson (eds.), *Knowing and remembering in young children* (pp. 9-29). New York: Cambridge University Press.

Blanchard, E. B. & Hickling, H. J. (2004). What are the psychosocial effects of MVAs on survivors? In E. B. Blanchard, E. J. Hickling (eds.), *After the crash: Psychological assessment and treatment of survivors of motor vehicle accidents* (2ª Ed., pp. 57-97). Washington, DC: American Psychological Association.

Bower, G. (1967). A multicomponent theory of memory trace. In K. W. Spence & J. T. Spence (eds.), *The psychology of learning and motivation* (Vol. 1) (pp. 229-325). New York: Academic Press.

Bower, G. H. & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science*, 247, 44-48. Brown, L., Sherbenou, R. & Johnsen, S. (1995). *Toni-2*. Madrid: TEA Ediciones.

Bull, R. (1995). Interviewing people with communicative disabilities. In R. Bull & D. Carson (eds.), *Handbook of psychology in legal contexts* (pp. 247-260). Chichester: John Wiley and Sons.

Bull, R. (1997). Entrevistas a niños testigos. In F. Fariña & R. Arce (eds.), *Psicología e investigación judicial* (pp. 19-38). Madrid: Fundación Universidad Empresa.

Butcher J. N. & Miller, K. B. (1999). Personality assessment in personal injury litigation. In A. K. Hess & I. B. Weiner (eds.), *The handbook of forensic psychology* (2ª Ed., pp. 104-126). New York: John Wiley and Sons.

Campos, L. & Alonso-Quecuty, M. L. (1999). The cognitive interview: Much more than simply "try again". *Psychology, Crime and Law*, 5(1-2), 47-59.

Cunnien, A. J. (1997). Psychiatric and medical syndromes associated with deception. In R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 23-46). New York: Guilford Press.

Dammeyer, M. D. (1998). The assessment of child sexual abuse allegations: Using research to guide clinical decision making. *Behavioral Sciences and the Law*, 16, 21-34.

Davies, G., Tarrant, A. & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology*, 80, 415-429.

Derogatis, L. R. (2002). *SCL-90-R. Manual*. Madrid: TEA Ediciones.

Duckworth, J. & Anderson, W. P. (1995). *MMPI and MMPI-2: Interpretation manual for counsellors and clinicians.* Bristol: Accelerated Development.

Echeburúa, E. & Corral, P. (1998). *Manual de violencia familiar*. Madrid: Siglo XXI.

Echeburúa, E., Corral, P. de & Amor, P. J. (2002). Evaluación del daño psicológico en víctimas de crímenes violentos. *Psicothema*, 14, 139-146.

Echeburúa, E., Corral, P. de, Zubizarreta, I. & Sarasúa, B. (1995). *Trastorno de estrés postraumático crónico en víctimas de agresiones sexuales*. A Coruña: Fundación Paideia.

Fisher, R. P. & Geiselman, R. E. (1992). *Memory-enhancing techniques for investigative interview*. Springfield: Charles C. Thomas.

Fisher, R. P., Geiselman, R. E., Raymond, D. S., Jurkevich, L. M. & Warhaftig, M. L. (1987). Enhancing eyewitness memory: Refining the cognitive interview. *Journal of Police Science and Administration*, 15, 291-297.

Freed, D. M., Corkin, S., Growdon, J. H. & Nissen, M. J. (1989). Selective attention in Alzheimer's disease: Characterizing cognitive subgroups of patients. *Neuropsychologia*, 27, 325-339.

Gil, R. (1999). *Neuropsicología*. Barcelona: Masson.

Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M. & Kuhn, J. (1997). Children's reactions to and memory for a stressful event: Influences of age, anatomical dolls, knowledge, and parental attachment. *Applied Developmental Science*, 1, 54-75.

Graham, J. R. (1992). *Interpretation of MMPI-2. Validity and clinical scales*. Bruges, workshop.

Hart, R. P. & O'Shanick, G. J.(1993). Forgetting rates for verbal, pictorial, and figural stimuli. *Journal of Clinical and Experimental Neuropsychology*, 15, 245-265.

Hathaway, S. R. & McKinley, J. C. (1999). *MMPI-2. Inventario multifásico de personalidad de Minnesota-2. Manual*. Madrid: TEA Ediciones.

Home Office and The Department of Health (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: HMSO.

Johnson, M. K. & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67-85.

Kelley, H. H. & Stahelski, A. J. (1970). Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 16, 66-91.

Köhnken, G. (1999). *Assessing credibility*. Pre-conference of the EAPL Programme of Applied Courses, Dublin.

Köhnken,G., Milne, R., Memon, A. & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, 5, 3-27.

Landry, K. L. & Brigham, J. C. (1992). The effect of training in criteria-based content analysis on the ability to detect deception in adults. *Law and Human Behavior*, 16, 663-676.

Lewis, M. & Saarni, C. (1993). *Lying and deception in everyday life*. New York: Guilford Press.

Loftus, E. F., Korf, N. L. & Schooler, J. W. (1988). Misguided memories: Sincere distortions of reality. In J. C. Yuille (ed.), *Credibility assessment* (pp. 155-172). Dordrecht: Kluwer.

Mantwill, M., Köhnken, G. & Ascherman, E. (1995). Effects of the cognitive interview on the recall of familiar and unfamiliar events. *Journal of Applied Psychology, 80*, 68-78.

Memon, A. & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation and practical implications. *Journal of Community and Applied Social Psychology*, 1, 291-307.

Memon, A., Cronin, O., Eaves, R. & Bull, R. (1993). The cognitive interview and the child witness. In N. K. Clark & G. M. Stephenson (eds.), *Issues in criminology and legal psychology: Vol. 20. Children, evidence and procedure*. Leicester: British Psychological Society.

Memon, A. Cronin, O., Eaves, R. & Bull, R. (1996). An empirical test of mnemonic components of cognitive interview. In G. M. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (eds.), *Psychology and law: Advances in research* (pp. 135-145). Berlin: Walter de Gruyter.

Memon, A., Wark, L., Bull, R. & Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology, 88*, 179-198.

Memon. A. & Bull, R. (1999). Handbook of the psychology of interviewing. Chichester: John Wiley and Sons.

Mira, J. J. (1989). *Estudios de psicología en ambientes procesales: Un análisis de metamemoria*. Doctoral Thesis, Universidad Autónoma de Madrid.

Miranda v. Arizona, 384 U.S. 436 (1966).

Ostrom, T. M., Werner, C. & Saks, M. J. (1978). An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology*, 36, 436-450.

Peña-Casanova, J. (1990). *Programa integrado de exploración neuropsicológica. Test Barcelona*. Barcelona: Masson.

Piñeiro, A. (2005). Criterios empíricos de credibilidad y profesionales de la justicia. In R. Arce, F. Fariña & M. Novo (eds.). *Psicología jurídica* (pp. 213-221). Santiago de Compostela: Consellería de Xustiza.

Porter, S. & Yuille, J. C. (1996). The language of deceit: An investigation of the verbal clues in the interrogation context. *Law and Human Behavior*, 20(4), 443- 458.

Raskin, D. C. & Steller, M. (1989). Assessing credibility of allegations of child sexual abuse: Polygraph examinations and statement analysis. In H. Wegener, F. Lösel & J. Haisch (eds.), *Criminal behaviour and the justice system: Psychological perspectives* (pp. 290-302). Heidelberg: Springer-Verlag.

Rogers, R. (1992). *Structured interview of reported symptoms*. Odessa, Fl.: Psychological Assessment Resources.

Rogers, R. (1997). Researching dissimulation. In R. Rogers (ed.), *Clinical assessment of malingering and deception* (pp. 398-426). New York: Guilford Press.

Rogers, R. & Mitchell, C. N. (1991). *Mental health experts and the criminal courts: A handbook for layers and clinicians*. Scarborough, ON: Thomson Professional Publishing Canada.

Roig-Fusté, J. M. (1993). *MMPI y MMPI-2 en la exploración de la personalidad*. Barcelona: Agil Offset.

Sallmann, P. & Willis, J. (1984). *Criminal justice in Australia*. Melbourne: Oxford University Press.

Schum, D. A. (1977). The behavioral richness of cascaded inference models: Examples of jurisprudence. In N. J. Castellan, D. B. Pisoni & G. R. Potts (eds.), *Cognitive theory* (vol. 2). Hillsdale, N.J.: LEA.

Spörer, S. L. (1997). The less traveled road to truth: Verbal cues in deception detection in accounts of fabricated and self-experienced events. *Applied Cognitive Psychology,* 11, 373-397.

Steller, M. (1989). Recent developments in statement analysis. In J. C. Yuille (ed.), *Credibility assessment*. (pp. 135-154). Dordrecht: Kluwer Academic Publishers.

Steller, M. & Köhnken, G. (1994). Análisis de declaraciones basados en criterios. In D. C. Raskin (ed.), *Métodos psicológicos en la investigación y pruebas criminales* (pp. 217-245). Bilbao: Desclée de Brouwer (Orig. 1989).

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tulving, E. & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 353-370.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

Undeutsch, U. (1967). Beurteilung der glaubhaftigkeit von zeugenaussagen. In U. Undeutsch (ed.), *Handbuch der psychologie, Vol. II: Forensische psychologie* (pp. 26-181). Göttingen: Verlag für Psychologie.

Undeutsch, U. (1988). The development of statement reality analysis. In J. Yuille (ed.), *Credibility assessment* (pp. 101-119). Dordrecht: Kluwer.

Vrij, A. (2000). *Detecting lies and deceit*. Chichester: Wiley.

Vrij, A. & Winkel, F. W. (1996). Detection of false statements in first and third graders: The development of a nonverbal detection instrument. In G. Davies, S. Lloyd-Bostock, M. McMurray & C. Wilson (eds.), *Psychology, Law, and Criminal Justice* (pp. 221-230). Berlin: Walter de Gruyter.

Vrij, A., Edward, K., Roberts, K. P. & Bull, R. (1999, July). *Detecting deceit via criteria based content analysis, reality monitoring and analysis of non verbal behavior*. The First Joint Meeting of the American Psychology- Law Society and the European Association of Psychology and Law, Dublin.

Wechsler, D. (1976). *WAIS: Escala de inteligencia para adultos (2ª ed.).* Manual. Madrid: TEA Ediciones.

Weick, K. E. (1985). Systematic observational methods. In G. Lindzey & E. Aronson (eds.), *The handbook of social psychology* (vol. 1, pp. 567-634). Hillsdale, N.J.: LEA.

Wicker, A. W. (1975). An application of a multipletrait-multimethod logic to the reliability of observational records. *Personality and Social Psychology Bulletin*, 4, 575-579.

Winograd, E., Smith, A. D. & Simon, E. W. (1982). Aging and the picture superiority effect in recall. *Journal of Gerontology*, 37, 70-75.

Zaparniuk, J. Yuille, J. C. & Taylor, S. (1995). Assessing the credibility of true and false statements. *International Journal of Law and Psychiatry*, 18, 343-352.