Are validity scales useful for detecting deliberately faked personality tests? A study in incarcerated populations

Pedro Medina\textsuperscript{a}, Ausiàs Cebolla\textsuperscript{b,c,*}, Rosa Baños\textsuperscript{c,d} and Cristina Botella\textsuperscript{b,c}

\textsuperscript{a}Tarragona Prison, Tarragona, Spain; \textsuperscript{b}Universitat Jaume I, Castelló, Spain, \textsuperscript{c}CIBEROObn Fisiopatología de la Obesidad y la Nutrición, Santiago de Compostela, Spain; \textsuperscript{d}Universitat de València, Valencia, Spain

(Received 31 October 2011; final version received 9 May 2012)

Personality self-report questionnaires are frequently used in forensic settings to detect psychopathology, to predict recidivism, and to assess adaptability to life in prison. Although most personality questionnaires include validity or control scales, even with the scales most outcomes can be easily manipulated. The aim of this study is to analyze the utility of the control scales of the \textit{Situational Personality Questionnaire} (SPQ). A sample of 200 male prisoners was randomized into two groups. Both groups completed the SPQ as a part of the mandatory psychological assessment when they entered prison, and then again eight months later. In time 2, one group received instructions to falsify the results of the questionnaire. Results indicated that the feigned induction was effective. The control scales were not able to detect feigners. Results are discussed with regard to their implications for further research into assessing fake responses in forensic settings.

\textbf{Keywords:} fake-bad; assessment; prisoners

The use of personality self-report questionnaires is very common in forensic settings. These scales are usually employed to detect the presence of psychopathology, especially personality disorders or psychopathic traits, to predict recidivism, or to predict adaptability to prison life (Listwan, Van Voorhis, & Ritchey, 2007). The findings from these personality tests are then used as forensic evidence in correctional and judicial settings to help determine the subject’s (1) competence, (2) criminal responsibility, and/or (3) level of risk (Duellman & Bowers, 2004).

Although different types of interviews are frequently used for these purposes, questionnaires are generally considered as providing ‘objective’ information that is uninfluenced by the beliefs, judgments and experiences of the interviewer. However, personality questionnaires, as with all self-reports,
are susceptible to their answers’ being manipulated in distorted, biased, or false ways. There is a variety of possible distortions, one of them being respondents’ attempts to present favorable or unfavorable pictures of themselves, yielding inaccurate and misleading personality profiles. In some cases, respondents actually believe in their positive self-reports, but in other cases, respondents consciously dissemble, especially under public conditions (Paulhus, 1984). In some situations, individuals can be motivated to distort their personality scores or to exaggerate psychopathological problems in order to obtain beneficial outcomes, and therefore the effect of this distortion on self-reports should be controlled for.

Despite efforts to design reliable and valid instruments for assessing personality traits and dispositions, questionnaires remain vulnerable to lying, faking, feigning, and malingering (Sullivan & King, 2010). There are different types of faking, and the more common distinction is between ‘fake good’ and ‘fake bad’. ‘Fake good’ has been defined as a ‘conscious effort to manipulate responses to personality items to make a positive impression’ (Zickar & Robie, 1999). This bias is expressed as intentionally looking better than one might perform, and it is probably the most extensively studied bias (Mersman & Schultz, 1998). ‘Fake bad’ is expressed as intentionally looking worse than one might perform, and it has been also studied, mainly focused on the malingering of psychopathology1 (Sullivan & King, 2010). As Meehl and Hathaway (1946) pointed out, ‘one of the most important failings of almost all structured personality tests is their susceptibility to “faking” and “lying” in one way or another’ (p. 525). However, occurrence of this bias may vary in different contexts and different populations.

In forensic settings, the question of faking is of primary concern (Pierson, Rosenfeld, Green, & Belfi, 2011), particularly where the outcome of the assessment influences the legal status of prisoners, who may exaggerate various psychological problems in an effort to receive special services (e.g. psychopharmacological agents) or who may minimize their involvement in drug or alcohol use to avoid more stringent probationary terms and requirements (Morey & Quigley, 2002). Thus, the reliability of the assessment may be severely compromised when attempts to feign go undetected (Rogers, 1997). As a result, psychologists have been involved in detecting deception during psychological evaluations in legal contexts. To control for faking, validity scales have been usually included to assess the accuracy of self-reports (Mogge, Lepage, Bell, & Ragatz, 2009; Morey, 1991; Schoenberg, Dorr, & Morgan, 2006). There is a large literature on validity scales (Rogers, Sewell, Martin, & Vitacco, 2003; Singh, Avasthi, & Grover, 2007), aimed at constructing robust control strategies for detecting feigners. However, researchers recognize that it is still relatively easy to deliberately exaggerate the results without being detected (Piedmont, McCrae, Riemann, & Angleitner, 2000; Singh et al., 2007). Many researchers have come to recognize the limitations of validity scales, and
even several test authors (for instance NEO-PI-R authors) expressly omit the usual validity scales because they believe there is scarce empirical justification for their use (Piedmont et al., 2000). Piedmont et al. (2000) concluded the questionnaires are not an infallible method, and furthermore the validity scales will not improve them. They propose to use well-validated instruments with improved quality and multiple sources of data, like external criteria, separate instruments or independent sources.

The typical design for the study of feigning has been the ‘faking paradigm’ (Piedmont et al., 2000) in which participants are explicitly asked to simulate some form of distortion (fake good or fake bad) (Mogge et al., 2009; Shores & Carstairs, 1998). The scores of these ‘fakers’ are then compared to those of control groups (CGs). These experiments have been useful for studying the effectiveness of control scales. Several studies have shown that validity scales distinguish between faking and control conditions, and it has been concluded that they are able to detect the bias in question (Baity, Siefert, Chambers, & Blais, 2007; Piedmont et al., 2000). However, most of these studies have been done in the context of personnel selection or with volunteer undergraduate students (Omar & Uribe, 2000) rather than in clinical or prison settings, and they are often criticized as generalizing the results to other contexts. Regarding students, Heinze and Vess (2005) pointed out that the incentive to fake in real-world situations (such as when one is evading criminal prosecution) is much stronger than an experimental context can ethically assess. Regarding personnel selection, several researchers have claimed that faking is used by relatively few applicants and should therefore not be an important issue in this context (Griffin, Hesketh, & Grayson, 2004; Hough, 1998). On this point, it should be emphasized that sensitivity and specificity of a test to detect distortion and faking depends on the base rate of invalid responding in the sample. Sensitivity reflects the capacity of an instrument to yield true positive results, whereas specificity reflects the capacity of an instrument to yield true negative results. Both sensitivity and specificity are determined by the established cutting score of the test, but cutting scores vary for different populations. For example, Lim and Butcher (1996) showed that a cutoff score that discriminated faking bad from honest student respondents with 100% accuracy identified fully 30% of a sample of presumably honest psychiatric patients as faking bad. These results point to the importance of base rate information in understanding the accuracy of prediction methods to detect feigning, and suggest that different cutoff scores should be used for different populations. In addition, limitations also come from the problem of false positives. For example, Lim and Butcher (1996) showed that a cutoff score that discriminated faking from honest student respondents with 100% accuracy identified fully 30% of a sample of presumably honest psychiatric patients as faking bad. These results suggest that different cutoff scores should be used for different populations and point to the importance of base
rate information in understanding the accuracy of prediction methods to detect feigning.

Another option for discerning faking is to compare self-report scores with independent assessments such as observer ratings.

Yet another strategy to detect fakers involves using multivariate techniques such as discriminant factorial analysis (Cashel, Rogers, Sewell, & Martin-Cannici, 1995; Schoenberg et al., 2006). In this line, Rogers, Harrell, and Liff (1993) have developed the Rogers Discriminant Function scale (RDF) (Rogers, Sewell, Morey, & Ustad, 1996) for the Personality Assessment Inventory (PAI, Morey, 1991). This function distinguishes between malingering and nonmalingering simulators (Rogers et al., 1996) and has demonstrated an impressive detection rate across several simulation samples (Hopwood, Morey, Rogers, & Sewell, 2007; Sullivan & King, 2010). However, the use of RDF in criminal forensic settings is being increasingly questioned (Hopwood et al., 2007). Rogers, Sewell, Cruise, Wang, and Ustad (1998) applied the RDF to a forensic sample and found that the detection accuracy was near chance levels, leading these authors to issue a caution against using the RDF with forensic populations. Kucharski, Toomey, Fila, and Duncan (2007) also found that the RDF scale and the Malingering index from PAI do not have acceptable enough sensitivity and specificity to differentiate the malingering from the nonmalingering in a sample of criminal defendants. Negative results with the RDF may occur because the base rate of pathology-free individuals may be lower in forensic populations than in standard simulation studies (Hopwood et al., 2007).

Overall, the use of control scales and other associated solutions to the problems of malingering has received significant empirical attention in the general population, but there are few studies addressing the validity of the control scales when used with incarcerated populations. As mentioned above, malingering is a very relevant topic given the characteristics of incarcerated populations. The aim of this study is to investigate the effect a malingering induction has on each of the scales of a personality questionnaire and to measure the utility of the control scales in detecting faking in an experimental induction with an incarcerated sample. For this study, the Situational Personality Questionnaire (Cuestionario de Personalidad Situacional (CPS); Fernández-Seara, Seisdedos, & Mielgo, 1998) has been used because it is a common tool used to assess personality traits in Spanish forensic settings, and it is a mandatory instrument in the assessment protocol in the prison where the present study was conducted.

Method

Participants
The respondents were 200 male prisoners from the Tarragona (Spain) prison, with a mean age of 34 (SD = 9.2). The only exclusion criterion was a
low level of reading ability; potential participants were excluded when their reading level was insufficient to understanding the questionnaire. The sample was randomized into two groups, a CG and a Feigner Group (FG), with 100 participants in each. The average age was 34.6 (SD = 9) years for CG, and 33.4 (SD = 9.4) years for FG. The average duration of incarceration (in weeks) was 24.9 (SD = 34.6) for the CG and 34.5 (SD = 40.8) for the FG. In terms of educational attainment among CG participants, 35% subjects completed primary school, 61.2% completed secondary school, and 3.8% completed university studies. Among FG participants, 41.5% subjects completed primary school, 57.1% completed secondary school, and 1.3% completed university studies. For all categories (age, length of incarceration, and education level) there were no significant differences between CG and FG.

**Instruments**

*Situational Personality Questionnaire* (CPS; Fernández-Seara et al., 1998). This questionnaire contains 233 items, each with two answer options (true/false), and takes approximately 30 minutes to complete. This instrument offers scores on 15 personality scales, three control scales and five summary scales (second order factors). The 15 main scales are: *Emotional Stability* (irritable, touchy, and overexcited versus serene, stable, and balanced), *Anxiety* (relaxed, calm, and patient versus worried, anxious, and fearful), *Self-Concept* (having low self-esteem and poor self-image versus having high self-esteem and strong self-image), *Efficacy* (socially insecure and anxious, versus socially competent and confident), *Self-Confidence* (hesitant and insecure versus trusting and confident about him/herself and his/her possibilities), *Independence* (dependent versus autonomous), *Dominance* (docile, obedient, and trying to please versus energetic, assertive, organizing, and competitive), *Cognitive Control* (external attribution and impulsive versus cautious, analytical, and calculating), *Sociability* (reserved, withdrawn, shy, and distant versus friendly, sociable, enthusiastic, expressive, and participative), *Social Adjustment* (rebellious and in conflict with the rules versus socialized, dutiful, and accepting of the rules), *Aggressiveness* (peaceful and unperturbed versus warlike and critical), *Tolerance* (unyielding, rigid, dogmatic, and ‘picky’ versus understanding, permissive, flexible, and open), *Social intelligence* (socially awkward and change-avoidant versus socially comfortable and flexible with change), *Integrity/ Honesty* (informal and undisciplined versus reliable, responsible, formal, and disciplined), and *Leadership* (uninterested in giving orders or leading others versus confident in organizing tasks or leading people).

The five summary scales are: *Adjustment* (undecided, tense, capable, critical, doubtful versus balanced, relaxed, understanding, sure), *Leadership* (tolerant and permissive versus assertive and competitive), *Independence*
(submissive, doubtful, tolerant, and conventional versus autonomous, rebellious, and self-confident), and Consensus (rule-conflictive, insecure, and impulsive versus reliable, socially skillful, and organized).

The CPS incorporates three validity scales that are used to detect purposeful distortion: The Sincerity Scale is composed of 21 items referring to behaviors that social norms advise against carrying out. A low score (lower than 5) refers to a person who desires to hide personal defects. A high score (higher than 9) refers to a person who is sincere and truthful. The Social Desirability Scale is composed of 28 items assessing the distortion that can be introduced into the responses by overestimation of oneself and one’s own behavior. A low score (lower than 24) on this factor refers to a person whose social self-conception corresponds to natural and spontaneous behavior. A high score (higher than 27) refers to a person who ruminates and worries about his/her social image. The Response Control Scale is composed of 26 items, grouped in 13 pairs with similar answer direction, and it is expected that responders answer both items similarly. The objective of this scale is to detect individuals who respond the questionnaire carelessly, without attending to the items. A score of 8 or higher means coherency in the answering. A score of 7 or lower indicates incoherency, meaning that the evaluation results should be considered with caution.

Strengths of this instrument include its broad understandability, due to its simple language, its validation using a large Spanish sample (n = 39,631), and its standing, established through previous research, as a good instrument to predict conflict-seeking prisoners (Raya, Eliseo, & Medina, 2008).

**Procedure**

All participants filled out the CPS as part of a mandatory psychological assessment when entering prison (time 1). After a period of approximately eight months (time 2), participants were asked to fill out the questionnaire again (X = 251.6 days for the CG, and X = 251.2 days for FG). There were no significant differences between groups in the time spent between time 1 and time 2 (t = .99).

At time 2, the FG group received instructions to fake the results of the questionnaire, thereby presenting a different self-image; it was not specified whether they should fake good or bad. Feigner Group participants were told that they would receive a small reward for the task (cigarettes, sweets, chocolate, etc.) if their new scores different from their initial test scores at time 1, and if the test did not detect that they were faking. At the end of the experiment, all participants received the reward independent of the results. The specific instructions were made in colloquial language, were always the same, and were meant to induce participants to fake the questionnaire in such a way that the test did not detect the deception. The CG group received the same standard test instructions that they received in time 1.
**Results**

**Differences between groups and times**

A $2 \times 2$ ANOVA (FG vs. CG × time 1 vs. time 2) analysis with Sidak’s post-hoc tests was applied. The aim of this analysis was to analyze the efficacy of the fake induction across the analysis of the differences according groups (FG vs. CG) in each of the scales, before and after the fake induction. The descriptive data are shown in Table 1.

ANOVA results are shown in Table 2. Regarding group effects and time effects, significant differences have been found for almost all the scales. That is, there were differences between both groups (FG and CG), and scores changed from Time 1 to Time 2. In order to analyze groups differences Sidak’s post hoc analysis between groups (FG vs. CG) were applied for both times (Time 1 and 2). Although groups were randomized, in time 1 there were significant differences in Dominance ($p = .03$), Independence ($p = .03$), Social Adjustment; ($p = .02$), Aggressiveness ($p < .01$), and Tolerance, ($p < .05$) scales. In time 2, after the fake induction, there were significant differences between both groups in almost all of the scales (Emotional

Table 1. Descriptives before and after the fake induction.

|       | Time 1 |       |       | Time 2 |       |       |
|-------|--------|-------|-------|--------|-------|-------|
|       | FG     | CG    |       | FG     | CG    |       |
|       | M      | SD    | M      | SD    | M      | SD    |
| E stab| 13.6   | 5.7   | 13.6   | 5.4   | 12.1   | 5.6   | 14.8   | 5.7   |
| Anx   | 14.3   | 5.3   | 15.0   | 5.2   | 12.7   | 4.9   | 13.9   | 5.6   |
| S-con | 15.9   | 4.3   | 15.1   | 4.0   | 13.1   | 5.2   | 16.3   | 4.1   |
| Effi  | 18.9   | 4.1   | 19.1   | 3.5   | 11.4   | 6.1   | 19.6   | 3.7   |
| S-conf| 16.7   | 4.7   | 16.3   | 4.2   | 11.7   | 5.3   | 17.4   | 4.2   |
| Indep | 9.6    | 3.5   | 8.6    | 3.3   | 10.5   | 3.3   | 9.1    | 3.5   |
| Dom   | 9.7    | 4.8   | 8.2    | 4.2   | 14.7   | 5.5   | 8.2    | 4.4   |
| Cog   | 16.6   | 3.9   | 17.5   | 3.2   | 9.9    | 5.4   | 18.6   | 3.4   |
| Sociab| 15.2   | 5.5   | 15.5   | 5.1   | 12.4   | 4.9   | 15.8   | 5.0   |
| S adj | 12.3   | 3.5   | 13.1   | 3.3   | 8.5    | 4.0   | 12.8   | 3.1   |
| Aggre | 9.2    | 5.8   | 6.9    | 5.4   | 12.8   | 6.1   | 7.7    | 5.7   |
| Tol   | 12.3   | 4.3   | 13.7   | 3.6   | 9.9    | 5.4   | 18.6   | 3.4   |
| S Intel| 15.4  | 2.9   | 20.1   | 3.7   | 8.7    | 4.7   | 16.1   | 3.3   |
| Int/Hon| 20.8 | 2.8   | 8.9    | 3.6   | 10.8   | 6.4   | 21.1   | 2.8   |
| Lead  | 8.9    | 3.6   | 8.2    | 3.3   | 11.3   | 4.3   | 7.9    | 3.4   |
| Sinc  | 8.6    | 4.0   | 7.9    | 3.1   | 12.2   | 3.6   | 7.82   | 3.1   |
| S Des | 20.3   | 6.1   | 21.7   | 5.0   | 11.5   | 6.9   | 22.0   | 5.5   |
| R cont| 8.6    | 1.8   | 8.9    | 1.5   | 8.1    | 2.0   | 9.2    | 1.6   |

Note: E stab, Emotional Stability; Anx; Anxiety; S-con, Self-Concept; Effi, Efficacy; S-conf, Self-Confidence; Indep, Independence; Dom, Domination; Cog, Cognitive Control; Sociab, Sociability; S adj, Social Adjustment; Aggre, Aggressiveness; Tol, Tolerance; S Intel, Social Intelligence; Int/Hon, Integrity/Honesty; Lead, Leadership; Sinc, Sincerity; S Des, Social Desirability; R cont, Response Control.
Stability, \( p < .001 \); Efficacy, \( p < .001 \); Self-Confidence, \( p < .001 \); Dominance, \( p < .01 \); Independence, \( p < .01 \); Cognitive Control, \( p = < .01 \); Sociability, \( p < .001 \); Social Adjustment, \( p < .01 \); Aggressiveness, \( p < .01 \); Tolerance, \( p < .01 \); Social Intelligence, \( p < .01 \); Integrity/honesty, \( p < .01 \); Leadership, \( p < .01 \); Sincerity, \( p < .01 \); Social Desirability, \( p < .01 \); Response Control, \( p < .01 \). There were no significant differences between groups in time 2 except in Self-Concept and Anxiety scales.

More important, results showed significant interaction effects (group \times time) for almost all the scales (see Table 2, Figure 1), except for the Anxiety, Independence, and Sincerity scales. In order to analyze these interaction effects Sidak’s post hoc analysis between time 1 and 2 were applied for both groups. For the CG group, there were no differences between time 1 and 2 for any scale. However, for the FG group, there were significant differences in all the scales (Emotional Stability, \( p < .005 \); Anxiety, \( p < .005 \); Self-Concept, \( p < .05 \); Efficacy, \( p < .001 \); Self-Confidence, \( p < .001 \); Dominance, \( p = .040 \); Independence, \( p = .040 \); Cognitive Control, \( p < .01 \); Sociability, \( p < .01 \); Social Adjustment, \( p < .01 \); Aggressiveness, \( p < .01 \); Tolerance, \( p < .01 \);
Social Intelligence, $p < .01$; Integrity/Honesty, $p < .01$; Leadership, $p < .01$; Sincerity, $p < .01$; Social Desirability, $p < .01$; Response Control, $p < .05$).

**Classification of the participants’ as ‘reliable’ respondents**

Normative data for the CPS in the Spanish sample allow for the categorizing of respondents as ‘reliable’ in function of the three control scale scores (Fernández-Seara et al., 1998). According to the cutoff scores, at time 1, the percentage of the sample that could be categorized as ‘reliable’ was between 60.5% and 97.5% (see Table 3). At time 2, after the experimental induction, the percentage varied between 60.5% and 97.5%.

![Figure 1. Efficacy of the fake induction in CPS scales.](image)

Table 3. Respondents categorized as ‘reliable’in time 1 and 2.

| Control scale            | Time 1       | Time 2       |
|--------------------------|--------------|--------------|
|                          | FG           | CG           | FG           | CG           |
| Sincerity                | 78 (96.3%)   | 68 (84%)     | 73 (90.1%)   | 69 (85.2%)   |
| Social desirability      | 79 (97.5%)   | 73 (90.1)    | 78 (96.3%)   | 63 (77.8)    |
| Response control         | 49 (60.5%)   | 65 (80.2%)   | 58 (71.6%)   | 70 (86.4%)   |
the percentages remained similar in both groups, FG and CG. In fact, the percentage of FG participants categorized as ‘reliable’ increased in the Response Control scale by more than 10%.

According to normative data (Fernández-Seara et al., 1998), average scores on control scales (see Table 1) indicated that participants were not dishonest and did not hide personal defects, as mean scores on Sincerity were higher than 5. These scores increased at time 2, where FG participants scored higher than 9, meaning that they were assumed to be highly ‘sincere’. Regarding Social Desirability, all average scores were lower than 24, indicating that participants were ‘natural and spontaneous in their social image’, and FG participants even showed better scores in Time 2. Finally, Response Control scores were higher than 8, meaning that participants’ answers were reliable and coherent.

Discussion
The present study was aimed at analyzing the efficacy of control scales at detecting faking in a commonly used personality questionnaire (CPS) administered in a Spanish prison. For this purpose, an experimental ‘fake’ induction was used, and the subsequent data were compared to a CG. In general, results from this study do not support the utility of validity scales for the CPS questionnaire.

First, results indicated that ‘fake’ induction was successful, as FG participants changed their scores in all scales after the induction. Data did not show differences between times 1 and 2 for the CG, indicating that participants’ scores remain relatively stable after eight months. This result was expected since the questionnaire measures stable dispositions and good test–retest reliability data have been reported (Fernández-Seara et al., 1998). However, the FG group did show differences for every subscale, meaning that their scores changed at time 2 when faking was requested. Furthermore, group × time interaction effects were significant for almost all scales, indicating that changes were bigger for the FG than for the CG group. Only two personality scales did not show significant interaction effects: Anxiety and Independence. Anxiety scores were lower at time 2 for all participants (although post analysis showed only significant differences for FG), and Independence scores were higher at time 2 for all participants (although again post analysis showed only significant differences for FG). It is possible that anxiety and independence have positive social values in prison contexts, making high independence and low anxiety highly valued traits among prisoners, hence, these dimensions might be more easily affected by unconscious, self-deceptive enhancement (Paulhus, 1984).

Regarding the efficacy of validity scales in detecting respondents’ attempts to manipulate the answers, the results did not support their utility. According to normative data (Fernández-Seara, et al., 1998), after fake
induction, 90.1% of FG participants were sincere, 96.3% showed a natural and spontaneous self-image and 71.6% answered the questionnaire with interest and attention, avoiding answering at random. Taking average scores into consideration, FG participants were even more ‘sincere’ and ‘natural and spontaneous in their social image’ in Time 2 than in Time 1. These data clearly point out that the control scales were not able to detect the feigned manipulation.

Several limitations of the present study should be noted. First, the current population was exclusively male. Future studies might question whether these results are applicable to females. Second, faking instructions did not direct the faking into ‘good’ or ‘bad’ orientation. Further research should be address to distinguish between fake good and fake bad inductions, in order to explore different characteristics related to the direction of the faking. Third, the CPS questionnaire does not give a psychopathological cut off scores, thus there is not data about the faking effect over the diagnostic status of the participants, further research should use another personality questionnaire which measuring this factor.

Furthermore, as previously mentioned, the CPS is the most common personality instrument used in Spanish forensic settings, but it is only available in the Spanish language, so it is not possible to compare the present results with those from other countries, and to determine the degree to which the findings of the study are generalizable to other particular measures.

Finally and surprisingly, there were significant differences between groups at time 1 in some factors. Compared to CG participants, FG participants showed higher scores on Dominance, Independence, and Aggressiveness and lower scores on Social Adjustment, and Tolerance. This was not expected since the sample was randomized and there were no differences with respect to age, length of incarceration, and educational level.

Despite these limitations, the design of this study is unique in that a fake induction was used with an incarcerated population; traditionally faking and malingering has been studied in the context of personnel selection or with volunteer undergraduate students. The present study has notable implications for future research and for the use of validity scales in forensic contexts. In these situations, individuals may be motivated to distort their personality scores or to exaggerate psychopathological problems in order to obtain beneficial outcomes.

In conclusion, results indicate that validity scales are not effective tools to detect feigning in a jailed sample, although, as previously mentioned, findings of the present study are not generalizable to other measures. It should be highlighted that psychological testing is one of several strategies used in forensic decision-making, but it is not the only source used to answer forensic question. Assessment also includes other strategies, such as observation, interviews, and the use of collateral information. Nonetheless, personality questionnaires remain very useful assessment procedures in forensic
questions, but it is important to know the relative vulnerability of these personality measures to being feigned. More research is needed to establish systems that are effective at detecting feigning. One alternative entails using a combination of relevant scales grouped in one factor, rather than using additional scales that are easy to detect and circumvent by participants (Schoenberg et al., 2006). These topics are fundamental when the assessment is done in a context associated with higher feigning prevalence.

Acknowledgements
CIBERobn is an initiative of ISCIII. This study has been supported in part by the Ministerio de Ciencia e Innovación (Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008–2011) in the project (PSI2008-04392/PSIC).

Note
1. Although both ‘malingering’ and feigning are use in this paper, there are differences between the terms. Malingering is defined as the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding military duty, avoiding work, obtaining principal compensation, evading criminal prosecution or obtaining drugs (APA, 2000). Feigning means to deliberately fabricate or to grossly exaggerate a clinical condition. Psychological tests can be used to assess whether an individual may be feigning but tests cannot establish the motivations required to categorize such deception as malingering (Rogers & Correa, 2008).

References
American Psychiatric Association (APA). (2000). Diagnostic and Statistical manual of mental disorders, text revision (4th ed.). Washington, DC: Author.
Baity, M.R., Siefert, C.J., Chambers, A., & Blais, M.A. (2007). Deceptiveness on the PAI: A study of Naïve Faking with psychiatric inpatients. Journal of Personality Assessment, 88, 16–24.
Cashel, M.I., Rogers, R., Sewell, K.S., & Martin-Cannici, C. (1995). The personality Assessment Inventory (PAI) and the detection of defensiveness. Assessment, 2, 333–342.
Duellman, R.M., & Bowers, T.G. (2004). Use of personality assessment inventory (PAI) in forensic correctional settings: Evidence for concurrent validity. International Journal of Forensic Psychology, 1, 42–57.
Fernández-Seara, J.L., Seisdedos, N., & Mielgo, M. (1998). CPS, Cuestionario de Personalidad Situacional [Situational Personality Questionnaire]. Madrid: TEA Ediciones.
Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. Personality and Individual Differences, 36, 1545–1558.
Heinze, M.C., & Vess, J. (2005). The relationship among malingering, psychopathy, and the mmpi-2 validity scales in maximum security forensic psychiatric inpatients. Journal of Forensic Psychology Practice, 5, 35–53.
Hopwood, C.J., Morey, L.C., Rogers, R., & Sewell, K. (2007). Malingering on the personality assessment inventory: identification of specific feigned disorders. Journal of Personality Assessment, 88, 43–48.
Hough, l.M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. Human Performance, 11, 209–244.

Kucharski, L.T., Toomey, J.P., Fila, K., & Duncan, S. (2007). Detection of malingering of psychiatric disorder with the personality assessment inventory: An investigation of criminal defendants. Journal of Personality Assessment, 88, 25–32.

Lim, J., & Butcher, J.N. (1996). Detection of faking bad on the MMPI-2: Differentiation among faking-bad, denial, and claiming extreme virtue. Journal of Personality Assessment, 67, 1–25.

Listwan, S.J., Van Voorhis, P., & Ritchey, P.N. (2007). Personality, criminal behavior, and risk assessment. Criminal Justice and Behavior, 34, 60–75.

Meehl, P.E., & Hathaway, S.R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. Journal of Applied Psychology, 30, 525–564.

Mersman, J.L., & Shultz, K.S. (1998). Individual differences in the ability to fake on personality measures. Personality and Individual Differences, 24(2), 217–227.

Mogge, N.L., Lepage, J.S., Bell, T., & Ragatz, L. (2009). The negative distortion scale: A new PAI validity scale. Journal of Forensic Psychiatry & Psychology, 21, 77–90.

Morey, L. (1991). Professional manual for the Personality Assessment Inventory. Odessa, FL: Psychological Assessment Resources.

Morey, L.C., & Quigley, B.D. (2002). The use of the Personality Assessment Inventory in assessing offenders. International Journal of Offender Therapy and Comparative Criminology, 46, 333–349.

Norris, M.P., & May, M.C. (1998). Screening for malingering in a correctional setting. Law and Human Behavior, 22, 315–323.

Omar, A., & Uribe, H. (2000). Tendencia al falseamiento y temor a ser descubierto [Tendency to distortion and fear of discovery]. Acta Psiquiátrica y Psicológica de América Latina, 46, 67–73.

Paulhus, D.L. (1984). Two-component models of socially desirable responding. Journal of Personality and Social Psychology, 46, 598–609.

Piedmont, R.L., McCrae, R.R., Riemann, R., & Angleitner, A. (2000). On invalidity of validity scales: Evidence from Self-reports and Observer Rating in volunteer Samples. Journal of Personality and Social Psychology, 78, 582–593.

Pierson, A.M., Rosenfeld, B., Green, D., & Belfi, B. (2011). Investigating the relationship between antisocial personality disorder and malingering. Criminal Justice and Behavior, 38, 146–156.

Raya, D., Eliseo, A., & Medina, P. (2008). Validació creuada en població penitenciaria de criteris psicomètrics i tècnics per a la prediccí de conduites adaptatives i factors de risc [Cross validation of psychometric and technical criteria to predict adaptive behaviour and risk factors in jailed population]. Centre d’estudis jurídics i formació especialitzada. Retrieved from: http://www20.gencat.cat/docs/Justicia/Documents/ARXIUS/SC-3-152-08.pdf Barcelona.

Rogers, R. (1997). Current status of clinical methods. In R. Rogers (Ed.), Clinical assessment of malingering and deception (pp. 391–410). New York: Guilford.

Rogers, R., & Correa, A.A. (2008). Determinations of Malingering: Evolution from case-based methods to detection strategies. Psychiatry, Psychology and Law, 15, 213–223.

Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. Clinical Psychology Review, 13, 255–274.

Rogers, R., Sewell, K.W., Cruise, K.R., Wang, E.W., & Ustad, K.L. (1998). The PAI and feigning: A cautionary note on its use in forensic-correctional settings. Assessment, 5, 399–405.
Rogers, R., Sewell, K.W., Martin, M.A., & Vitacco, M.J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment, 10*, 60–77.

Rogers, R., Sewell, K., Morey, L., & Ustad, K. (1996). Detection of feigned mental disorders on the Personality Assess Inventory: A discriminant analysis. *Journal of Personality Assessment, 67*, 629–640.

Schoenberg, M.R., Dorr, D., & Morgan, C.D. (2006). Development of discriminant function to detect dissimulation for the Millon clinical Multiaxial Inventory (3rd ed.). *The Journal of Forensic Psychiatry & Psychology, 17*, 405–416.

Shores, A., & Carstairs, J.R. (1998). Accuracy of the MMPI–2 computerized Minnesota Report in identifying fake-good and fake-bad response sets. *The Clinical Neuropsychologist, 12*, 101–106.

Singh, J., Avasthi, A., & Grover, S. (2007). Malingering of psychiatric disorders: A review. *German Journal of Psychiatry, 10*, 126–132.

Sullivan, K., & King, J. (2010). Detecting faked psychopathology: A comparison of two tests to detect malingered psychopathology using a simulation design. *Psychiatry research, 176*, 75–81.

Zickar, M.J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551–563.