Incarcerated Violent Offenders' Ability to Avoid Revealing Their Potential for Violence on the Rorschach and the MMPI–2

Peder Chr. Bryhn Nørbech, Lars Fodstad, Irene Kuisma, Ketil Berge Lunde & Ellen Hartmann

To cite this article: Peder Chr. Bryhn Nørbech, Lars Fodstad, Irene Kuisma, Ketil Berge Lunde & Ellen Hartmann (2016) Incarcerated Violent Offenders' Ability to Avoid Revealing Their Potential for Violence on the Rorschach and the MMPI–2, Journal of Personality Assessment, 98:4, 419-429, DOI: 10.1080/00223891.2015.1129613

To link to this article: http://dx.doi.org/10.1080/00223891.2015.1129613

Published online: 28 Jan 2016.
Incarcerated Violent Offenders’ Ability to Avoid Revealing Their Potential for Violence on the Rorschach and the MMPI–2

Peder Chr. Brynh Nørbech, Lars Fodstad, Irene Kuisma, Ketil Berge Lunde, and Ellen Hartmann

ABSTRACT
Recently, Hartmann and Hartmann (2014) found that psychiatric outpatients, both with and without access to Internet-based information about the Rorschach Inkblot Method (RIM; Weiner, 2003) and the MMPI–2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), were unable to imitate healthy test performance on these tests. We replicated the study by administering the RIM and the MMPI–2 to 63 incarcerated violent offenders using similar testing conditions. As in the previous study, comparisons were made not only among the 3 subgroups of incarcerated offenders, but also between these offender groups and the group of nonpatients examined in the previous study. On the RIM, Internet-coached and uncoached “faking good” offenders produced records with significantly higher F% and X-% and significantly lower M, m, SumC, X+i%, P, AG, and COP than nonoffenders under standard instructions (effect sizes between d = 0.24 and d = 2.39). For AgC, AgPot, AgPast, and TCI% there were no significant differences between the faking good groups and the nonoffenders under standard instructions. On the MMPI–2 clinical scales, there were no significant differences between the faking good groups and the nonoffenders under standard instructions, except on Hs, Pd, and Sc. Both faking groups were identifiable by their high L scale scores. Although both faking groups managed to avoid giving responses with aggressive and generally psychopathological content on the RIM, they were unable to produce test profiles demonstrating healthy test performance on any of the tests; nevertheless, Internet-based test information might weaken test validity.

Forensic, clinical, and child custody settings represent assessment conditions where the simulation of healthy adjustment is most likely to occur (e.g., Baer & Sekirnjak, 1997; Ganellen, 2008; Hartmann & Hartmann, 2014; Schultz & Brabender, 2013). For felons convicted of violent crimes seeking a leave from prison, or under evaluation for probation, the ability to avoid responses indicative of hostility and aggression is likely to be crucial to gaining a positive evaluation. Under such circumstances, offenders might be inclined to present themselves as more mentally stable than they are, and thus be tempted to prepare for testing by seeking out information about the particular assessment. Relatively little is known about the effects of such preparation and practice.

Tutorial material about how to respond in line with particular psychological standards on the Rorschach Inkblot Method (RIM; Weiner, 2003) and the Minnesota Multiphasic Personality Inventory–2 (MMPI–2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) is available on Internet Web sites. Instances of attorneys coaching their clients about validity scales and appropriate test response patterns are not uncommon (Bury & Bagby, 2002). Forensic examiners should, therefore, be cognizant of the possibility that the examinee might have prepared for testing to shift his or her test responses in a more well-adjusted and non-aggressive direction.

Ruiz, Drake, Glass, Marcotte, and van Gorp (2002) and Schultz and Loving (2012) found that the personality test tutorial information on most Internet Web sites was of minimal help. Even so, some sites present detailed test material that could still undermine the validity of psychological assessment. Wikipedia (2004), for instance, has released all 10 Rorschach cards online as well as the guiding principles for recognizing adequate and poor responses (e.g., that using movement and color is “safe,” whereas including aggressive and other provocative content is risky).

Randall (2010) showed that test takers who completed the RIM twice—first without knowledge of the Wikipedia Internet information and later after having been exposed to such material—altered their manner of responding minimally from the first to the second assessment. Similarly, Schultz and Brabender (2013) also reported few differences between the results of individuals who had or had not read the Wikipedia Rorschach article, beyond a significantly lower number of responses and higher number of Populars among those who had read the article. Participants were recruited from a parent–teacher association, from online parenting message boards, and from graduate programs in social work, education, nursing, and psychology at a small university.

Recently, Hartmann and Hartmann (2014; identified as the index study hereafter) found that psychiatric outpatients with
access to Internet-based test information related to the RIM and the MMPI-2 were unable to utilize this information when asked to present themselves as mentally healthy on these tests. The researchers used a between-groups design with four groups: one experimental group of psychiatric outpatients who had been exposed to Internet advice on faking healthy responses to these tests, and who had been instructed to try to do this; two control groups of outpatients, one under uncoached “faking healthy” instructions and one under standard conditions; and a comparison group of students under standard conditions. The faking participants were asked to complete the tests as if they were job applicants trying to present themselves as well-adjusted individuals with few psychological symptoms and problems. On the RIM, both the Internet-coached and uncoached patients under faking healthy instructions, similarly to the patients under standard instructions, revealed significantly more perceptual and reality-testing disturbances (X+% and X−%) than the nonpatients under standard instructions did. The Internet-exposed faking patients produced, in addition, more constricted protocols with a significantly higher Pure Form percentage (F%) and fewer aggressive and other provocative contents than the other groups. Thus, on the RIM both patient groups under faking good instructions showed better ability to control and censor what they saw, as opposed to how they perceived it. The patients under standard instructions produced psychopathological scores on most of the RIM variables. On the MMPI-2 clinical scales, the faking patients, both with and without Internet coaching, managed to appear healthy, but had higher L scale scores. The patients under standard instructions revealed substantial psychological problems and symptoms on the clinical scales. It can be said, therefore, that providing Internet coaching did not improve the test takers’ ability to imitate a mentally healthy manner of test response, but resulted in constricted, yet perceptually and cognitively disturbed, RIM records and invalid pseudo-normal MMPI-2 profiles.

Research has regularly shown that people can consciously manipulate their scores on personality instruments (e.g., Baer & Miller, 2002; Ganellen, 2008). It is also well documented that it is easier to simulate wellness on transparent tests, like self-report tools, than on less transparent tests such as the RIM, from how they actually feel, think, and behave. In this way, the validity of self-report tools can be seen to be influenced by the degree to which the respondents are inclined to portray themselves positively versus negatively and by their level of self-knowledge and capacity to define themselves and their problems accurately (Meyer, 1997; Westen & Shedler, 2007). The MMPI-2 includes many validity scales designed to identify underreporting of psychological problems and symptoms. Baer and Miller (2002) conducted a meta-analytic review reporting on positive simulation studies in which the test takers were randomly assigned to two groups, with one group asked to respond as if they were healthy individuals and one tested under standard instructions. In five of these studies, respondents instructed to fake healthy were coached briefly about the validity scales designed to identify test takers trying to present an unrealistically positive picture of themselves. The efficacy of the various MMPI-2 validity scales to identify underreporting response style bias was impressive for uncoached participants and moderate for coached participants. The review showed that the Lie and the K-scales were the most effective for
detecting faking good responding. Regrettably, only three of the reviewed studies compared clinical samples instructed to appear mentally well-adjusted with nonclinical samples under standard instructions. All of the other studies compared healthy students under both the faking good and standard conditions. However, it cannot be assumed that healthy students trying to simulate being exceptionally healthy will underreport symptoms in the same manner as violent offenders or psychiatric patients (Baer & Sekirnjak, 1997). To this point Bagby et al. (1997) found that both schizophrenic patients and healthy students, who initially completed the MMPI–2 under standard instructions, were able to underreport problems on the clinical scales when instructed to conceal their problems during a second test. Even so, both patients and students were identified by the validity scales. The L scale was found to be a superior parameter for distinguishing between faking good and honest respondents. We have not been able to locate more recent studies than these comparing coached and uncoached participants in positive simulation settings.

To summarize, existing research shows that test takers are able to suppress RIM responses with aggressive and provocative content when motivated to do so. Nonetheless, relative to self-report tests, the Rorschach has proven to be less vulnerable to positive impression management or false self-presentation.

**Purpose of the study**

The index study included only psychiatric outpatients. As a result, generalizability to other patient groups, prisoners, and different evaluation settings might be limited. For example, for the outpatients in the index study, reading the Wikipedia information about the Rorschach resulted in more deviant scores on the perceptual codes, whereas in the study carried out by Schultz and Brabender (2013) the nonpatients who had read the Rorschach article scored higher on measures linked to perceptual accuracy. For this reason, we wanted to replicate the design and personality assessment part of the index study using three corresponding groups of incarcerated violent offenders. In terms of these groups of test takers, we were especially interested in understanding whether the groups faking their responses were able to avoid including potentially aggressive and hostile content when responding to the tests. As in the previous study, comparisons were made not only among the three subgroups of incarcerated offenders, but also between each of these offender groups and the group of nonpatients (and non-offenders) examined in the previous study.

**Hypotheses**

1. In accordance with research indicating that content variables are easier to manipulate than other RIM variables (Meyer et al., 2011; Mihura, 2012; TaVoularis Benjestorf et al., 2013; Viglione et al., 2012), we expected that both Internet-coached and uncoached offenders tested under instructions to fake healthy would manage to avoid aggressive and other provoking contents on the RIM, and thus score low on the measures of aggressive content and on the TC1%. However in line with most forensic RIM studies (e.g., Gacono & Meloy, 1994; Hartmann et al., 2006; Nørbech et al., 2013; Nørbech et al., 2015), we expected that violent offenders under standard instructions would score high on these measures.
2. Based on the results of Franks et al. (2009) and the index study, we suspected that although Internet-coached and uncoached offenders would manage to underreport violent and provocative responses, they would not be able to use available Internet-based information to imitate healthy test responding on the RIM successfully. Consequently, they would score high on F%, within the psychopathological range on the perceptual and cognitive measures, and low on color and movement.
3. Offenders tested under standard instructions would score high on F% and within the psychopathological range on all the RIM outcome measures.
4. On the MMPI–2, we expected that both Internet-coached and uncoached offenders tested under instructions to fake healthy would manage to underreport psychological symptoms and problems and score within the normal range on most of the clinical scales, but would be detected by the L scale.
5. Offenders tested under standard instructions would score within the pathological range on the clinical scales and within the normal range on the validity scales.

Because we used the same comparison group of students under standard instructions as in the index study, we knew that this group scored within the range of the Comprehensive System’s (CS; Exner, 2003) international reference data (Meyer, Erdberg, & Shaffer, 2007) on the RIM and all the mean T scores on the MMPI–2 were within 1 SD of T = 50, indicating that this group is suitable as a comparison group for the analyses of the RIM and the MMPI–2 data.

**Method**

**Participants**

A total of 85 individuals (74 men and 11 women, age 18–52, \( M = 29.48, SD = 8.91 \); 83 of the participants were ethnic Norwegian) took part in the study. We used a between-groups design with four conditions. Sixty-three men serving a prison sentence for violence were recruited from two Norwegian prisons and randomly allocated to three groups: the Internet-faking offenders group (\( n = 21 \); age 22–42, \( M = 27.19, SD = 5.15 \)) tested under Internet-coached instructions to fake healthy responses; the faking offenders group (\( n = 21 \); age 19–52, \( M = 34.19, SD = 10.20 \)) tested under uncoached instructions to fake healthy responses; and the standard offenders group (\( n = 21 \); age 19–52, \( M = 33.81, SD = 9.63 \)) tested under standard instructions. Recruitment was achieved with the help of correctional officers and the psychiatric teams in the two prisons, who informed inmates about the study and asked whether they were willing to participate. Participation was strictly voluntary with no incentives. Any inmates with psychotic symptoms, with mental retardation, taking psychoactive medications, or unable to speak Norwegian were excluded from the study.

We used the standard nonpatients group supplemented with one subject (\( n = 22 \), 11 men and 11 women, age 18–34, \( M = 23.05, SD = 3.99 \)) from the index study as a comparison group.
of nonoffenders under standard instructions. This group was nonrandomly recruited at the University of Oslo. Students in various bachelor degree programs (no psychology programs) were asked to take part in a normative study on personality assessment methods. All potential participants were screened using the Inventory of Interpersonal Problems–Circumplex (Alden, Wiggins, & Pincus, 1990), which provides a brief test of interpersonal problem severity and was used to exclude individuals scoring within the clinical range. No incentives were offered for participation and all students were tested under standard instructions.

The mean age was significantly lower for the standard nonoffenders ($M = 23.05, SD = 3.99$) compared to the Internet-faking offenders ($M = 30.19, SD = 6.15$), the faking offenders ($M = 34.19, SD = 10.20$), and the standard offenders ($M = 33.81, SD = 9.63$). All comparison t-tests between the mean age of the standard nonoffenders and each of the offender subgroups were significant ($p < .01$), and the effect sizes were all large (varying between $d = -1.38$ and $d = -1.44$). There were no significant differences in mean age among the three offender subgroups and all effect sizes were small, varying between $d = -0.04$ and $d = -0.47$.

**Information packages about the Rorschach Inkblot Method and the MMPI–2**

We used the same Internet-based information packages about the RIM and the MMPI–2 as used in the index study, which were retrieved from the following Web sites: http://en.wikipedia.org/wiki/Rorschach_test (August 23, 2010), http://en.wikipedia.org/wiki/Minnesota_Multiphasic_Personality_Inventory (August 23, 2010), www.democraticunderground.com/discuss/duboard.php?az=view_all&address=04×2748618 (August 23, 2010), and http://deltabravo.net/custody/rorschach.php (June 15, 2010). We offered the following package of RIM information to each Internet-faking offender:

Presentation of each of the cards accompanied by the 13 Populums from the CS, some sexual responses, and the following written advice: “Stick to ‘seeing’ healthy, friendly images, and avoid dark and violent answers. Butterflies, people holding hands, leaves, and mountains are ‘safe’ responses. Responses based on movement and color may be ‘good’ answers, whereas sexual and provocative answers should be avoided. If you see something that you think really resembles something, answer saying that. Do not use either too short or too long a time. Try to give an answer to all the cards, preferably two or three to each card.”

The Internet-based MMPI–2 information was presented in a very general way. We offered the following package of MMPI–2 information to each Internet-faking offender:

Presentation of the traditional clinical and validity scales and the fact that the validity scales may detect dishonest responding, plus the following written advice: “Extreme words like ‘never’ and ‘always’ may be statements from the so called Lie-scale. Be honest, but try to respond in a moderate manner and avoid extreme statements.”

**Realistic scenario**

The faking participants were given a description of a realistic scenario in which incarcerated offenders might try to fake healthy. We asked these test takers to complete the RIM and the MMPI–2 as if it were part of an evaluation where they might receive more attractive prison conditions and earlier release from prison if the test results indicated that they were harmless individuals with few emotional and interpersonal difficulties, and no problems with anger and aggression.

**Personality instruments**

The RIM (Meyer & Viglione, 2008; Rorschach, 1921/1942; Weiner, 2003) is one of the most commonly used psychological assessment methods (Camera, Nathan, & Puente, 2000; Hogan, 2005). It is a performance-based personality assessment method that requires participants to use available resources when responding to an unfamiliar and somewhat stressful situation (Mihura, Meyer, Dumitrascu, & Bombel, 2013). We used the same 17 RIM variables as in the index study (for rationale for the selection of the measures, see the index study):

1. Number of Responses (R; related to productivity, constriction, and defensiveness).
2. Six determinant variables: Pure Form% (F%; related to constriction and defensiveness), Human Movement (M; capacity to relate to and understand others, using ideational activity channeled into purposeful human activity), Inanimate movement (m; related to stress-related anxiety), all Color responses (SumC; related to emotional expressiveness), all Achromatic responses (SumC%; related to painful internalized affect), and all Shading responses (SumShd; related to anxiety, stress, hopelessness, and self-critical attitudes).
3. Four perceptual and cognitive variables: Conventional Perception (X+%; related to conventional perception and social adjustment), Distorted Form (X−%; related to erroneous judgments that lead to inadequate adjustment), Popular responses (P; related to conventional perception), and the Weighted Sum of six Cognitive Special Scores, level 2 (WSum6-Lv2; related to illogical and incoherent thinking).
4. Four aggressive variables: Aggressive Movement (AG; related to assertive and competitive behavior and verbal and physical aggression and hostility), Aggressive Content (AgC; related to concern with aggressive imagery; e.g., objects that are dangerous or harmful), Aggressive Potential (AgPot; related to tension around aggression, fear of impending aggression, or sadistic aggression), Aggressive Past (AgPast; related to seeing oneself as damaged, hurt, or victimized or having masochistic features).
5. Cooperative Movement (COP; related to positive and caring interpersonal interaction).
6. The Traumatic Content Index (TCI%, Armstrong & Lowenstein, 1990; consisting of the CS variables Sex [Sx], Blood [Bl], Anatomy [An], Morbid [MOR], and Aggressive Movement [AG] and related to history of trauma). Armstrong and Lowenstein developed the TCI% as an index of traumatic history. However Meyer et al. (2011) commented that this index cannot be reliably differentiated from the Critical Content Index in the new Rorschach Performance Assessment System (R–PAS; Meyer et al., 2011). Thus, in accordance with the R–PAS
interpretation, we view the TCI% as a marker of general psychopathology.

The MMPI–2 (Butcher et al., 1989) has been comprehensively studied with regard to the impact of underreporting (e.g., Baer & Miller, 2002). We used the same scales as were used in the index study: The nine clinical scales (Hypochondriasis [Hs], Depression [D], Hypochondriasis [Hs], Psychopathic Deviate [Pd], Paranoia [Pa], Psychasthenia [Pt], Schizophrenia [Sc], Hypomania [Ma], and Social Introversion [Si]) and five validity indexes: the Lie (L), Correction (K), and the Infrequently (F) scales; the Variable Response Inconsistency (VRIN); and the True Response Inconsistency (TRIN) scales. High scores on the L scale designate denial of minor personal shortcomings, unso-

phisticated attempts to describe oneself positively, or both. High K scores are associated with the tendency to deny psychological problems. The F scale is used to indicate a rare type of response in the form of admitting to having extreme and unlikely symptoms. Thus, low F might indicate underresponding. The VRIN and TRIN scales were designed to reveal inconsistent responding.

**Procedure**

The collection of demographic and diagnostic data in the index study was cut to a minimum as the recruitment of participants took a long time, due to the fact that many potential partici-

pantS who had originally been receptive to the study did not want to take part when they realized that they would have to attempt to simulate being well-adjusted. Therefore, we decided to condense the collection of background data. As a result, diagnosis and educational and socioeconomic status were not well documented.

Four trained assistants (one woman and three men) admin-

istered the tests. We administered the RIM individually, as out-

lined in CS, differing only in that we used the Norwegian instructions for Rorschach administration (Hartmann, 2001). This instruction differs from the CS mainly in that it is longer and asks the respondent to say what the inkblots look like, remind you of, or make you think of. It also contains information about the number of cards, the fact that the inkblots do not resemble anything in particular, how the inkblots are cre-

ated, that there are no right or wrong responses, and that differ-

ent people see different things. The two groups of faking offenders were asked to complete the tests in a manner that might assure prison officers and members of the psychiatric team that they were nonviolent individuals and would not reof-

fend. In addition, the Internet-faking offenders examined the packages of the RIM and the MMPI–2 Internet-based information for 15 minutes prior to the administration of the tests. We administered the RIM as outlined in the index study. The partici-

pants completed the MMPI–2 some days after the RIM and according to standard instructions.

At the termination of the testing, participants who had been under instructions to fake their responses were questioned about their comprehension of and compliance with the faking instructions. The Internet-coached participants were asked about how they had used the packages of information to try to fake being nonaggressive and mentally well-adjusted, whereas the uncoached group was asked about how they had attempted to modify their test results. The local medical research ethics committee approved the study. The participants were informed of the procedure of the study prior to their informed consent and were treated in accordance with the Ethical Principles of Psychologists and Code of Conduct (American Psychological Association, 1992).

**Data analyses**

We analyzed the data in the same manner as in the index study. A one-way between-groups analysis of variance (ANOVA) was used to evaluate whether the means of R differed among the four groups and to explore the effects of the instruction on sim-

ulating healthy responding in the faking groups. We calculated effect sizes as eta-squared ($\eta^2$), and evaluated their levels according to Cohen’s (1988) guidelines as small (.01), moderate (.06), and large (.14). For post-hoc comparisons, we used Fisher’s least significant difference (LSD) test and calculated effect sizes as Cohen’s $d$ (1988), evaluating their levels according to his guidelines with effects around $d = 0.20$ being classified as small, effects around $d = 0.50$ being medium; and effects around $d = 0.80$ being large. An alpha level of .05 was used for all statistical tests.

**Control analyses**

Table 1 presents statistics for the RIM findings for the total sample. Descriptive analyses showed that various RIM variables did not meet the assumption of normal distribution. Meyer, Viglione, and Exner (2001) argued that distributions having a skew < 2.0 and kurtosis < 7.0 could be viewed as moderately nonnormal distributions. All our RIM scores fell within these criteria and thus were fit for parametric tests.

The range of R varied from 12 to 44 ($M = 22.46, SD = 7.20$). Two protocols in the Internet-faking group were brief ($R < 14$). As in the index study, we did not exclude brief protocols. ANOVAs (see Table 2) showed that there was no significant

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics for the Rorschach Inkblot Method outcome measures and protocol-level interrater reliability (ICC) on summary scores.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>F%</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>SumC</td>
</tr>
<tr>
<td>SumC*</td>
</tr>
<tr>
<td>SumShd</td>
</tr>
<tr>
<td>X+i</td>
</tr>
<tr>
<td>X-%</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>WSum6-Lv2</td>
</tr>
<tr>
<td>AG</td>
</tr>
<tr>
<td>AgC</td>
</tr>
<tr>
<td>AgPot</td>
</tr>
<tr>
<td>AgPast</td>
</tr>
<tr>
<td>COP</td>
</tr>
<tr>
<td>TCI%</td>
</tr>
</tbody>
</table>

*Note. N = 85. ICC = intraclass correlation coefficient. ICC(2,1) is only calculated for the outcome measures used in this study.*
main effect for R between the four groups. The effect size was moderate ($\eta^2 = .08$). Thus, we did not control for R.

Scoring and interrater reliability of the RIM variables

The first and the last authors, who did not participate in the data collection, scored the Rorschach protocols. As in the index study, scoring followed the guidelines presented in the Revised Norwegian Rorschach scoring system (Hartmann, Grønnerød, Krog, Vanem, & Høstmark Nielsen, 2003), which also includes Gacoco and Meloy’s (1994) aggression scores. The other variables selected for this study have identical definitions to the corresponding CS variables and are therefore interchangeable. The scorers were kept blind to the group assignment.

We calculated two-way intraclass correlation (ICC, 2,1) for single judge reliability on protocol level scoring (see Table 1). The ICC values ranged from .72 (X−%) to .96 (P). Meyer et al. (2002) interpreted ICC values between .60 and .74 as good reliability, and values between .75 and 1.00 as excellent reliability. According to these criteria, the values varied from good to excellent. As most values were excellent, we kept the original scoring for all protocols.

Results

The RIM

Table 2 presents statistics for the RIM measures in the four subgroups (Internet-faking offenders, faking offenders, standard offenders, and standard nonoffenders) separately, together with the results of the ANOVAs, and the means and standard deviations for the composite international normative CS reference data (Meyer et al., 2007).

As can be seen from Table 2, the ANOVA revealed a statistically significant main effect between the groups on 13 out of 17 RIM measures: F%, M, m, SumC, X+, X−, X%, P, AG, AgC, AgPot, AgPast, COP, and TCI%. The p values ranged from <.001 to .05, and the effect sizes ($\eta^2$) varied between .13 (P) and .39 (F%). For R, SumC%, SumShd, and WSum6-Lv2 there were no significant main effects, and the effect ranged between .06 (WSum6-Lv2) and .08 (R). Thus, we omitted these measures in the post-hoc comparison analyses. We obtained the same pattern of results when we ran the ANOVA excluding the two brief protocols.

LSD post-hoc subgroup comparisons are presented in Table 3. In Table 3, positive mean values indicate that the means of the first group listed were larger than the means of the second group listed. Negative values indicate the opposite pattern. As shown in Table 3, there were no significant differences between the means of the Internet-faking offenders and the faking offenders (column 2) on any of the RIM findings, and all the effect sizes ($d$) ranged between <.001 (M) and 0.52 (X−%). Furthermore, the two faking offender groups’ scores were significantly different from the standard nonoffenders (columns 4 and 6) on F%, M, m, SumC, X+, X−, X%, P, AG, and COP. The p values were between $p < .05$ and $p < .001$, and the effect sizes ranged between 0.24 (P) and 1.77 (X−%), with the faking offenders showing more deviant scores than standard nonoffenders. However, for the aggression measures—AgC, AgPot, and AgPast and the trauma measure TCI%—there were no significant differences between the means of the two faking groups and the means of the standard nonoffenders. All the means were within the normal range (columns 4 and 6). The effect sizes varied in absolute value from <0.01 (AgC; FO vs. SNO) and 0.63 (AgC; IFO vs. SNO).
Table 3. Effect sizes (Cohen's d) and significance levels for post-hoc group comparisons (least significant difference) of the Rorschach Inkblot Method outcome measures across the Internet-faking offenders (IFO), the faking offenders (FO), the standard offenders (SO), and the standard nonoffenders (SNO) subgroups.

<table>
<thead>
<tr>
<th>Determinants codes</th>
<th>F%</th>
<th>M</th>
<th>m</th>
<th>SumC</th>
</tr>
</thead>
<tbody>
<tr>
<td>F%</td>
<td>0.01</td>
<td>0.70</td>
<td>0.126</td>
<td>0.71</td>
</tr>
<tr>
<td>M</td>
<td>0.10</td>
<td>0.67</td>
<td>0.083</td>
<td>0.15</td>
</tr>
<tr>
<td>m</td>
<td>0.09</td>
<td>0.83</td>
<td>0.102</td>
<td>0.86</td>
</tr>
</tbody>
</table>

Cognitive and perceptual codes

| X+-% | 0.52 | 0.37 | 1.366 | 0.52 |
| X-%  | 0.48 | 0.15 | 1.777 | 0.44 |
| P    | 0.41 | 0.52 | 0.71  | 0.06 |

Content codes

| AG   | 0.19 | 0.66 | 0.777 | 0.88 |
| AgC  | 0.46 | 0.63 | 0.81  | 0.01 |
| AgPot| 0.01 | 0.31 | 0.114 | 0.09 |
| AgPast| 0.02 | 0.32 | 0.114 | 0.34 |
| COP  | 0.29 | 0.31 | 0.58  | <0.01 |
| TCI% | 0.06 | 0.15 | 0.106 | 0.20 |

Note. Effect sizes have a positive sign when the mean for the first group listed is larger than the mean for the second group listed; a negative sign indicates the mean for the second group is larger than the first.

*p < .05. **p < .01. ***p < .001.

Compared to the standard offenders, there were also no significant differences between the means of the Internet-faking offenders (column 3) and the faking offenders (column 5) on F%, X+-%, X-%, P, and COP. The effect sizes ranged in absolute value between < 0.01 (COP: IFO vs. SO) and 0.62 (F%; FO vs. SO). On all the other measures (M, m, SumC, AG, AgC, AgPot, AgPast, and TCI%), the means of both faking groups were significantly lower than the means of the standard offenders (column 3 and 5). The effect sizes ranged in absolute values between 0.67 (m; FO vs. SO) and 1.40 (AgC; IFO vs. SO).

Finally, the standard offenders had significantly different scores from the standard nonoffenders (column 7) and in the psychopathological direction on F%, M, X+-%, X-%, P, AgC, AgPot, AgPast, COP, and TCI%. The effect sizes ranged in absolute value between 0.58 (M) and 1.65 (X-%).

The MMPI–2

Table 4 presents the statistics for the MMPI–2 scales in the four subgroups separately, together with the results of the ANOVAs. All the scales met the assumption of normal distribution. None of the MMPI–2 protocols had a Cannot Say score ≥10.

As in the index study, the TRIN scores were elevated in all our subgroups, with means varying between 59.35 and 62.33. High raw scores indicate a "yay-saying" bias; low raw scores indicate a "nay-saying" bias. As shown in Table 4, both the Internet-faking offenders and the faking offenders had low average scores on the TRIN raw score scale, indicating a nay-saying bias, whereas the standard offenders had a high average raw score, indicating a yay-saying bias. The standard nonoffenders had an average raw score in the middle of the range, indicating no bias.

ANOVA results for the MMPI–2 scores (Table 4) demonstrated a significant main effect between the subgroups on all validity and clinical scales, except for K, VRIN, TRIN, and Ma. The p levels ranged from < .001 to .05. The effect sizes (η²) ranged between .09 (Si) and .37 (Pd) for the significant differences. Thus, K, VRIN, TRIN, and Ma were excluded from the subgroup comparisons.

As can be seen from Table 5, the subgroup comparisons indicated that on all the clinical scales but Hs and Pd the means of the Internet-faking offenders and the faking offenders (column 2) did not differ significantly from one another. The effect sizes (d) varied in absolute value between 0.18 (Pa) and 0.51 (D) for the nonsignificant differences and between 0.63 (Pd) and 0.77 (Hs) for the significant differences. Furthermore, there were no significant differences between the means of the Internet-faking and the faking offenders compared to the standard nonoffenders (columns 4 and 6) except on Hs, Pd, and Sc. The effect sizes ranged in absolute value between 0.05 (D; IFO vs. SNO) and 0.67 (Pa; IFO vs. SNO) for the nonsignificant differences and between 0.71 (Sc; FO vs. SNO) and 1.63 (Pd; FO vs. SNO) for the significant differences.

Compared to the standard offenders responding under standard instructions, the means of the Internet-faking offenders were significantly different on all the clinical scales (column 3). The effect sizes ranged in absolute value between 0.73 (D) and 1.10 (Pd). For the faking offenders compared to the standard offenders, the differences in means were significant only on Pa, Pt, and Sc (column 5). The effect sizes ranged in absolute value between 0.55 (Pt) and 1.46 (Pa). Finally, the standard offenders produced significantly different scores compared to the standard nonoffenders, and in the psychopathological direction on all the clinical scales (column 7). The effect sizes ranged in absolute value between 0.60 (Si) and 2.18 (Pd).

For the validity scales, the picture was different. There were no significant differences in means between the two faking groups on L, F, and TRIN raw score (column 2). However, the...
three groups of offenders scored significantly higher than the standard nonoffenders on L (columns 4, 6, and 7). The effect sizes (d) varied in absolute value between 0.99 (IFO vs. SNO) and 1.74 (FO vs. SNO). On F, both faking groups scored significantly higher than the standard nonoffenders (columns 4 and 6) and significantly lower than the standard offenders (columns 3 and 5). The effect sizes ranged in absolute value between 0.73 (FO vs. SO) and 2.25 (SO vs. SNO). On the TRIN raw score scale, the Internet-faking offenders and the faking offenders scored significantly lower than the standard offenders (columns 3 and 5). The effect sizes were −1.39 and −1.65, respectively. Finally, the standard offenders, compared to the standard nonoffenders (column 7), scored significantly higher on L (d = 1.01), F (d = 2.25), and TRIN raw score scale (d = 0.67).

The interviews

During the interviews at the end of the assessments, both faking groups gave very limited reports and seemed somewhat reluctant to say anything about their strategies. A few said that they had held back aggressive and sexual images. On the RIM, their main strategy for simulating healthy adjustment seemed to be to give few responses. Some participants said that they had forgotten to fake good. About half of the Internet-faking offenders stated that they had tried to remember the Populars, but most of them found it too difficult. On the MMPI–2, most of the faking offenders stated that it was difficult to fake good as the test lasted for such a long time.

Discussion

Our study reexamined a topic of substantial currency, evincing several interesting findings. Most important, and in accordance with our hypotheses and the findings of the index study, neither the Internet-faking offenders nor the faking offenders managed to present as mentally well-adjusted on the RIM and the MMPI–2. However, like the faking patients in the index study, both groups of faking offenders were able to avoid giving responses with aggressive and psychopathological contents on the RIM. Their RIM profiles were constricted, defensive, and thematically impoverished but still revealed serious perceptual and reality-testing deficiencies and their MMPI–2 profiles were invalid. On the RIM, both faking offender groups produced fewer responses with aggressive and other psychopathological content than the offenders and nonoffenders under standard conditions. This result is consistent with previous research showing that content responses are relatively easy to modify (Meyer et al., 2011; Mihura, 2012; Viglione et al., 2012). It is also in line with the TaVoularis Benjestorf et al. (2013) study, which showed that violent offenders were able to suppress responses with aggressive content when asked to present as nonviolent individuals. Thus, forensic RIM examiners should be cognizant of the possibility that low scores on the aggression variables and on the TCI% could be markers of simulation and suppression.

The standard offenders produced, as expected, RIM records with significantly higher levels of Agc, AgPot, AgPast, and TCI% than both faking groups and the standard nonoffenders. Although contrary to the results from the TaVoularis Benjestorf et al. (2013) study, our findings are consistent with most forensic RIM research, thus providing further support for the utility of the aggression variables in discriminating between offenders and nonoffenders (e.g., Gacono & Meloy, 1994; Hartmann et al., 2006; Nørbech et al., 2015). Notwithstanding this promising trend, we concur with TaVoularis Benjestorf et al.’s appeal for further research on the association between RIM aggression and subsequent violent acts.

As predicted, all three offender groups scored significantly higher on F% than the nonoffenders, indicating constriction and defensiveness. This is in line with Franks et al. (2009), who found that a very high F% score was the most characteristic finding in the RIM records of imprisoned psychopathic violent offenders. As in the index study, both faking groups, just like the standard offenders, showed more perceptual and reality testing disturbances (low X +% and high X −%) than the nonoffenders, indicating perceptual, reality testing, emotional, and relational deficits and inability to modify their perception of the inkblots. These results diverge from the findings of Schultz and Brabender (2013), which showed that test takers who were acquainted with the Wikipedia information about the RIM produced records with significantly more markers of accurate perception (P, X +%, XA%, and WDA%) than test takers who had not read this information. Yet, such preparation did not result in superior perceptual functioning by the Internet-coached participants in our study, nor in the index study. However, the high scores on X +%, XA%, and WDA% in the Schultz and Brabender study were due to an above-average number of P and disappeared when P was controlled for. Thus, the apparent differences between the findings of the U.S. study and the Norwegian studies might be misleading. Furthermore, Schultz and Brabender’s participants, in contrast to our participants and those of the index study, were solely nonprisoners and nonpatients. In accordance with Hartmann and Hartmann (2014), we think that Internet-based RIM information could be useful for individuals with little or no perceptual problems, but not so for test takers with disturbed perceptual and reality-testing
functioning. In addition, both faking groups scored low on movement and color. Despite having been advised to use movement and color, the Internet-faking offenders, like the Internet-faking patients, failed to do so.

Whereas the findings from this study and the index study provide support for Ganellen’s (1994) and Weiner’s (2003) assertion that a guarded approach to the RIM results in high F% scores, we found no evidence supportive of their contention that this would also result in elevated P and low R scores.

On the MMPI–2, the Internet-faking offenders and the faking offenders succeeded in underreporting symptoms and problems on most of the MMPI–2 clinical scales. This was expected. Thus, independent of receiving the Internet-based information, both groups of violent offenders, like the faking outpatients in the index study, were able to respond within the normal range on most of the clinical scales.

On the validity scales, only L successfully identified both faking groups. This finding accords with our prediction and is consistent with the results from the index study. More surprising was the elevated L among the standard offenders. Whether this was a result of intentionally trying to present themselves positively remains unclear; in any case, they did not succeed in this. Within both faking groups, and in accordance with the index study, F scores were significantly higher than the standard nonoffenders and significantly lower than the standard offenders, suggesting that the faking groups, and especially the standard offenders, were admitting rather than underreporting psychopathology. Although the Internet-faking offenders were coached on the existence and aim of the validity scales, their response bias was detected by the L scale, although not by K. The standard offenders, when compared to the standard nonoffenders, scored significantly higher on all the scales. Thus, although scoring high on L, the results for F and the clinical scales indicated little or no tendency to minimize personal problems and deficiencies. Our findings concur with Baer and Miller’s (2002) conclusion that L is a solid detector of a faking good bias, but does not support their statement that K is a strong validity scale.

All our subgroups also produced high TRIN T scores. Examination of the raw scores on TRIN revealed the same pattern of answering as in the index study. Both faking groups had a mean score below the average of 9, suggesting that they might be inconsistently disposed to answer falsely; the standard offenders had a mean score above average, indicating that they were inconsistently apt to answer honestly; and the standard nonoffenders had an average mean score of 9.09, indicating no bias. Our results provide further support for Hartmann and Hartmann’s (2014) hypothesis that a low raw TRIN score might be a marker of attempting to simulate a more positive presentation on the MMPI–2.

Similar to the index study, high F% scores and L scale scores among the Internet-faking participants provided further empirical support for Ganellen’s (1996/2012) assertion that elevated scores on F% are conceptually comparable to high scores on the Lie scale. Although F% and L are not correlated (Meyer, 1997), both might be indicators of a defensive response bias that could have a lowering effect on test markers of psychopathology.

We concur with the recommendations of the RR-PAS (Meyer et al., 2011) that test administrators should regularly ask test takers whether they have prepared for the assessment by searching for test information on Wikipedia and other Internet sites. This will, of course, never eliminate the problems associated with the accessibility of information about psychological tests on the Internet, as individuals motivated to look up such information on Web sites would likely not admit to it.

**Limitations**

Finally, the limitations of this study should be considered. First, our instructions to fake being mentally healthy might have shaped a response set that varies from what might occur in settings where test takers are genuinely motivated to seek out information about personality tests on the Internet to appear well-adjusted on these tests. Positive simulation might only be identified in the context of a clear inducement to respond in a manner that appears well adjusted. The MMPI–2 validity scales might work more effectively in more realistic settings. Future studies could examine whether presenting a reward to the best simulant would result in more successful simulation.

Second, the experimental and control groups included solely incarcerated violent offenders and the participants in the comparison condition were exclusively students whose mean age was significantly lower than that of the other groups. Given the heterogeneity of the offender population, generalization from our sample to other kinds of violent offenders and to new evaluation settings might be problematic. We cannot rule out the probability that the results would have been different in samples of individuals with different demographics to ours. The findings of Schultz and Brabender (2013) suggest that well-educated nonoffender and nonpatient parents and graduate students handled the Wikipedia information differently than our offenders, as well as the psychiatric patients in the index study.

Third, the sizes of our subgroups were quite small. Future research should try to replicate our positive simulation design using test information from Internet sites with larger and more diverse samples, in both clinical and forensic settings, as well as settings where educational and socioeconomic status are well documented and individuals might have real motivation to intentionally adjust their test responses in a well-adjusted and nonthreatening manner.

Fourth, we used the Norwegian instructions for Rorschach administration (Hartmann, 2001), which is longer than the standardized CS (Exner, 2003) instructions and invites the test taker to say what the inkblots look like, remind you of, or make you think of. These instructions represent a broader invitation than CS’s “What might this be?” and, therefore, might yield different results than the CS. However, Hartmann (2001) and Hartmann and Vanem (2003), comparing the Norwegian and the CS instructions in a nonpatient and an inpatient psychiatric sample, respectively, did not find any reliable differences on any of the traditional Rorschach summary scores. The new R-PAS R-Optimized Administration procedure (Meyer et al., 2011), which asks the respondent to “give two, or maybe three responses to each card” and stresses prompts and pulls to obtain a sufficient number of responses and limit the range of responses, has resulted in a higher mean number of responses (M = 24.2) with a lower standard deviation (SD = 4.7) than the International CS reference data (M = 22.3, SD = 7.9).
this reason, we think our results would have been very similar had we used the CS instructions. How our results would have turned out using the R-PAS R-Optimized Administration procedure is less clear, and a case for future studies.

Notably, the design of the study was cross-sectional, which does not provide direct information about how each individual would respond to a different condition. A cross-sectional study is an indirect measure of longitudinal changes.

Despite these limitations, our study also has its strengths. We reexamined the impact of Internet-based information about the RIM and the MMPI–2 on test takers’ ability to simulate being mentally healthy in a sample of incarcerated violent offenders, and our study produced very similar results to Hartmann and Hartmann (2014). The between-groups design, with four groups used in both studies, provided opportunities to compare acquaintance versus no acquaintance with Internet-based test information in simulation groups, as well as target subject and nontarget groups, under standard instructional conditions. We believe that this design helps to overcome criticisms of the generalizability of former simulation studies carried out with less complex designs. Another methodological asset was the fact that the comparison group of nonpatients and nonoffenders, although all students, scored within the range of the international RIM norms and within normal ranges on the MMPI–2.

Finally, coached and uncoached test takers with perceptual and cognitive deficiencies were not able to alter or inhibit response markers of inaccurate perception and incoherent cognition, although they managed to downplay aggressive and provocative contents. These findings provide further support to the notion that although response content is easy to modify, the structure of the Rorschach response (e.g., the structural variables) appear fairly robust against positive simulation. The finding that the MMPI–2 Lie scale is an effective indicator of effects to simulate more positive responses than are genuine also seems to be an important finding. Further research is needed to decide whether low raw scores on TRIN is an indicator of trying to fake good on the MMPI–2.

References


Norcross, & S. Hill (Eds.), Psychologists

Norcross, & S. Hill (Eds.), Psychologists

Norcross, & S. Hill (Eds.), Psychologists


