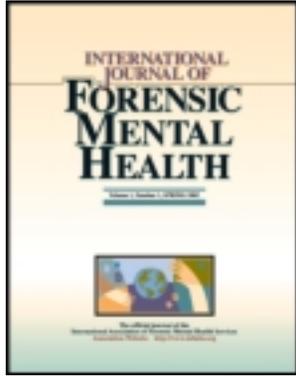


This article was downloaded by: [Ruud H.J. Hornsveld]

On: 07 November 2012, At: 00:23

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Forensic Mental Health

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ufmh20>

The Adapted Version of the Sociomoral Reflection Measure (SRM-AV) in Dutch Forensic Psychiatric Patients

Ruud H.J. Hornsveld^a, Floris W. Kraaimaat^b & Almar J. Zwets^c

^a FPC de Kijvelanden, Poortugaal, NL; Erasmus University Medical Centre, Rotterdam, The Netherlands

^b Radboud University, Nijmegen, The Netherlands

^c FPC de Kijvelanden, Poortugaal, NL; Erasmus University, Rotterdam, The Netherlands

Version of record first published: 06 Nov 2012.

To cite this article: Ruud H.J. Hornsveld, Floris W. Kraaimaat & Almar J. Zwets (2012): The Adapted Version of the Sociomoral Reflection Measure (SRM-AV) in Dutch Forensic Psychiatric Patients, *International Journal of Forensic Mental Health*, 11:3, 218-226

To link to this article: <http://dx.doi.org/10.1080/14999013.2012.723667>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The Adapted Version of the Sociomoral Reflection Measure (SRM-AV) in Dutch Forensic Psychiatric Patients

Ruud H.J. Hornsveld

FPC de Kijvelanden, Poortugaal, NL; Erasmus University Medical Centre, Rotterdam, The Netherlands

Floris W. Kraaimaat

Radboud University, Nijmegen, The Netherlands

Almar J. Zwets

FPC de Kijvelanden, Poortugaal, NL; Erasmus University, Rotterdam, The Netherlands

We developed a new instrument for measuring moral maturity in Dutch forensic psychiatric patients with the *Sociomoral Reflection Measure-Short Form* as a starting point. This instrument contains 11 propositions to which respondents indicate their own perceived level of importance and justifications of their choices. Twenty new statements were added to these 11 propositions. Justifications of the responses were scored with the aid of a seven-point scale, each point representing a (transitional) level of moral maturity. In order to have a practical instrument, we deleted 11 items with the lowest interrater reliability of the 31 items. The internal consistency, interrater reliability and test-retest reliability of the new instrument with the remaining 20 items (SRM-AV) were good. Validity was supported by significant correlations with age, psychopathy, agreeableness, and verbal aggression. Patients scored significantly lower on the SRM-AV than two non-clinical control groups. For the time being, the SRM-AV appears to measure moral maturity in forensic psychiatric patients in a reliable and valid way.

Keywords: forensic psychiatry, moral maturity, self-report measure

INTRODUCTION

In the Netherlands, forensic psychiatric centers have increasingly implemented cognitive-behavioral treatment programs, such as the Aggression Control Therapy (Hornsveld et al., 2008). This therapy is an adaptation of Goldstein, Glick, and Gibbs' Aggression Replacement Training (1998), which contains three modules: Anger management, Social skills, and Moral reasoning. The aim of the Moral reasoning module is to increase the moral maturity of the patients who follow the therapy. However, the module could not be evaluated until now because a reliable and valid instrument for measuring moral maturity in Dutch populations known for their vio-

lent behavior was lacking. The current study describes the development of such an instrument.

In the last ten years several studies have demonstrated a relationship between moral maturity and delinquency, in most cases by comparing delinquents with non-delinquents. For example, Stams and coworkers (2006), in their meta-analysis of 50 studies, found that delinquents had, on average, a lower level of moral maturity than non-delinquents. Differences were greatest in detained male adolescents with psychopathic traits. Five years later, Van Vogt and coworkers (2011) also found a significant inverse relationship between moral development and recidivism based on a meta-analysis of 19 studies. In this meta-analysis, "production" measures, which ask respondents to write down their reactions to a number of pictures or propositions, showed a much larger effect than "multiple-choice" measures, which use a Likert scale for those reactions. Effects in young and adult delinquents

Address correspondence to: Ruud H. J. Hornsveld, PhD, Lange Dreef 52, 2285 LA Rijswijk, Netherlands. E-mail: r.hornsveld@tiscali.nl

hardly differed from each other; self-report questionnaires yielded larger effects than official reports on recidivism.

Other authors have reached the same conclusions after examining the literature comparing delinquents with non-delinquents on moral maturity. For instance, Palmer (2003) concluded, on the basis of 30 years of research, that moral reasoning seems to be part of a much broader framework of social-cognitive factors formed by early experiences with parents and peers. Children from neglectful and harsh backgrounds were at a greater risk of both moral developmental delay and a hostile vision on the world, which filters future experiences. This combination should make it more likely that ambiguous social cues will be interpreted as hostile, triggering negative beliefs about social interactions, and may ultimately lead to criminal behavior.

A first instrument that was designed to determine the level of moral maturity was the *Moral Judgment Interview* (MJI; Colby & Kohlberg, 1987; Colby et al., 1987). This instrument was based on Kohlberg's three levels of moral development ("preconventional," "conventional," and "postconventional"), each with two phases, and contained a number of hypothetical dilemmas to which respondents had to react by making a choice and subsequently substantiating this choice. The justifications of these reactions were scored by the interviewers on the different levels for moral development. However, a disadvantage of the MJI is that scoring the answers is complex and time-consuming. Moreover, respondents in an interview may be receptive to unwanted suggestive influences by the interviewer (Carlo, Eisenberg, & Knight, 1992).

Gibbs, Basinger, and Fuller (1992) assumed that presenting moral dilemmas is not the only way to elicit moral judgments. On the basis of the MJI they developed the *Sociomoral Reflection Measure-Short Form* (SRM-SF), a measure that contains propositions instead of moral dilemmas (for example: "Imagine: a friend of yours is in mortal danger and you are the only one who can save him"), followed by questions about moral judgments ("How important is it for a person to save the life of a friend, without losing one's own life") and, finally, by asking about justifications for the choice that was made. For the scoring of the 11 items, Basinger, Gibbs, and Fuller (1995) only applied the first four phases of Kohlberg's theory (i.e., Unilateral authority, Pragmatic solution, Basis for interpersonal relations, and Social involvement). They investigated the psychometric characteristics of the SRM-SF in a group of 165 primary school pupils, 89 high school pupils, 72 university students, and 58 adults, and in a group of 89 delinquent boys in a medium-security prison. Five assessors found an interrater reliability between .94 and .99 and test-retest reliability between .61 and .78 with 25 respondents (three to five per subgroup). Cronbach's α was .93 ($N = 384$) for the total group of non-delinquents and .70 ($N = 89$) for the group of delinquents. Convergent validity was demonstrated by significant correlations with, for instance, the Moral Judgment Interview (MJI; Colby & Kohlberg, 1987). As expected, the delinquent boys scored significantly lower than male con-

temporaries on the SRM-SF. However, as is the case with the MJI, the main disadvantage of the SRM-SF is that the scoring of the responses is rather time-consuming. This is because Gibbs and coworkers (1992) collected 25 examples of justifications for each of the responses on the eleven propositions and identified the proper level of maturity for each example. Consequently, a researcher who wants to use the SRM-SF for the evaluation of a cognitive-behavioral module needs to compare the justifications of the participants in his study with the examples set by Gibbs and coworkers (1992).

Several authors have advocated applying "production instruments" instead of "recognition instruments" with detainees (Gavaghan, Arnold, & Gibbs, 1983; Hornsveld et al., 2007; Stams et al., 2006). When applying production instruments, respondents are asked to write down their reactions to a number of pictures or propositions, after which these reactions are scored by an independent assessor. These written reactions would offer more direct information about the behavior of respondents, because they do not have to think about what their correct score is on the Likert scale, as is the case with "recognition instruments."

For the development of a more practical instrument to assess moral maturity, we therefore decided to start with Basinger, Gibbs, and Fuller's SRM-SF, but to add 20 statements that, together with the original 11 propositions would, in our opinion, create a more complete representation of moral dilemmas that forensic psychiatric patients could encounter in daily life. Secondly, we wanted to score the justification of the reactions on the statements with a seven-point scale, each point representing a (transitional) level of moral maturity (see Appendix). In this article, we describe our analysis of the reliability and validity of the Adapted Version of the Sociomoral Reflection Measure (SRM-AV). We expected at least a sufficient internal consistency, interrater reliability and test-retest reliability for the SRM-AV. Because of the negative relationship between moral maturity and delinquency, we assumed a negative correlation of moral maturity with psychopathy, a positive correlation with agreeableness and a negative correlation with aggression in a group of forensic psychiatric patients. Moreover, we supposed a significantly lower score on the SRM-AV in forensic psychiatric patients than in non-clinical men.

The current study was approved by the institutional research review committee, the Dutch Review Committee for Patient-Linked Research in Arnhem (Netherlands), and the Scientific Research and Documentation Center of the Dutch Ministry of Safety and Justice.

STUDY 1: RELIABILITY OF THE SRM-AV

First, we examined the inter-rater reliability of the original SRM-AV with 31 items. After removal of less reliable items, we studied the internal consistency and the test-retest reliability of the definitive SRM-AV with 20 items.

METHOD

Patients

The pilot study was carried out with 80 forensic psychiatric inpatients (all males) who were convicted for a serious violent offense (e.g., murder, manslaughter, aggravated assault, or forcible rape), which is punishable in the Netherlands with a minimum of four years imprisonment. Their average age was 37.7 years ($SD = 9.7$, range = 19 to 62 years). The primary diagnosis was in 56 inpatients (70% of the sample) an Antisocial Personality Disorder on Axis II and in 24 inpatients (30%) a (chronic) psychotic disorder on Axis I combined with an Antisocial Personality Disorder on Axis II. The chronic psychiatric condition of the psychotic patients had been stabilized to the extent that their Antisocial Personality Disorder was most prominent. All patients had a command of the Dutch language in speech and in writing that was sufficient for the study.

Measure

The pilot edition of the *Adapted Version of the Sociomoral Reflection Measure (SRM-AV)* comprised the 11 items of the SRM-SF (Gibbs et al., 1992) and 20 new items. The 11 SRM-SF items were translated from English into Dutch and then back-translated. This translation yielded some minor differences, because certain English expressions could not be translated exactly in Dutch. The 20 new items were based on the literature about morality and our own clinical experience (Hornsveld et al., 2008). They referred to different a priori aspects of morality: respecting others (e.g., "Imagine that two lesbian women are kissing each other. How important is it that people do not discriminate against each other?"), helping people in need (e.g., "Imagine that a disabled person does not dare to cross the street. How important is it to help this person with crossing the street?"), addressing others with regard to their misbehavior (e.g., "Imagine that you notice a friend of yours is dealing in hard drugs. How important is it that dealing is prohibited?"), and taking responsibility by parents (e.g., "Imagine that a child is confronted every day with quarreling parents. How important is it that parents consider their children by not involving them in their quarrels?").

For the scoring of the answers with regard to the importance of a proposition, a five-point scale was applied, which ranged from 1 = very unimportant to 5 = very important. The justifications for the answers were scored on a seven-point scale (see Appendix for the seven points). The points on this scale were described in conformity with the four phases and three transitional phases used by Gibbs and coworkers (1992). However, the distinction that these authors made in the two types of moral maturity (Gibbs et al., 1992; Gibbs, 2003) was abandoned in phases 3 and 4 (see Appendix). Bizarre or irrelevant answers were scored with a "1," because they do not indicate an understanding of the (moral) content of the proposition. The answers to questions regarding the importance of a proposition were not further

examined. These questions had to stimulate respondents to think about their opinion regarding the proposition before writing their opinion down.

Procedure

As is the case with the SRM-SF, respondents were first asked with all items how important they think the proposition is that is advanced in an item, followed by the question why they consider the matter that is mentioned in the proposition to be (un)important. In order to determine the interrater reliability, two experienced research assistants (psychologists) scored the answers on the propositions and the justifications of these answers independently. In advance, these assistants were instructed about the scoring method based on several examples. One of these two assistants also scored a second measurement of the SRM-AV, four days after the first one, for the assessment of the test-retest reliability. Patients participated in the study voluntarily and received € 7 for completing the self-report measures. The SRM-AV was administered in groups of six to eight patients.

RESULTS

The interrater reliability was calculated for the total score of the 31-item SRM-AV and for each item separately in a subgroup of 25 inpatients. We used the Spearman correlation coefficient because scores were not normally distributed. The total scores yielded a significant positive correlation ($\rho = .86$, $p < .01$), and the correlations of the individual items varied from .42 ($p < .05$) to .89 ($p < .01$). The total score and the item scores of both research assistants did not differ significantly from each other.

In order to have a relatively short and applicable measure, only 20 items were selected and hence eleven items with the lowest interitem reliability were deleted. Two of these eleven items came from the original SRM-SF, namely item 6: "Let's say a friend of yours needs help and may even die, and you're the only person who can save him or her. How important is it for a person to save the life of a friend?" and item 7: "What about saving the life of anyone? How important is it for a person to save the life of a stranger?" In the total group of 80 inpatients the interrater reliability of the 20-item SRM-AV was .84 ($p < .01$) and test-retest reliability was .77 ($p < .01$). We again used Spearman's correlation coefficient for the calculation of the test-retest reliability, which could only be assessed in 71 patients, because nine patients refused to complete the SRM-AV for the second time.

STUDY 2: VALIDITY OF THE SRM-AV

In a second study, the internal consistency, factor structure and validity of the 20-item SRM-AV were examined in a larger group of forensic psychiatric inpatients, a group of

forensic psychiatric outpatients, a group of members of a soccer club, and a group of secondary vocational students.

METHOD

Participants

The second study was carried out with a combined group of 132 inpatients (including 25 patients from the first study) and 33 outpatients. For a comparison with non-clinical populations, we asked 65 members of a soccer club and 87 secondary vocational students to complete the SRM-AV. The 132 inpatients had an average age of 37.5 years ($SD = 8.8$, range = 19 to 62 years), 70.5% ($n = 93$) had a primary diagnosis of an Antisocial Personality Disorder on Axis II, and 29.5% ($n = 39$) had a (chronic) psychotic disorder on Axis I in combination with an Antisocial Personality Disorder on Axis II. The chronic psychiatric condition of the psychotic patients had been stabilized to the extent that their Antisocial Personality Disorder was most prominent. The 33 outpatients were referred to a forensic psychiatric outpatient clinic in Rotterdam, the Netherlands, as part of their sentence for violent offenses (e.g., assault, robbery with violence, or serious threats with violence). Their average age was 31.9 years ($SD = 7.9$; range: 19 to 46 years). The primary diagnosis of three outpatients (9.1%) was a Conduct Disorder on Axis I, and 30 outpatients (91.0%), who were 18 years or older, had a primary diagnosis of Antisocial Personality Disorder on Axis II.

The 65 members of a soccer club had an average age of 31.9 years ($SD = 11.0$; range: 18 to 64 years), and the 87 secondary vocational students of 19.2 years ($SD = 2.2$; range: 16 to 26 years). All participants of the study were male and had a sufficient command of the Dutch language in speech and in writing.

Measures

To measure the characteristics of forensic psychiatric patients in this study, we used a standard set of measures for several aspects of antisocial behavior. This set comprised:

The *Psychopathy Checklist-Revised* (PCL-R; Hare, 1991) was employed for measuring psychopathy. The checklist consists of 20 items, rated on a three-point scale with 0 = "does not apply," 1 = "applies to some extent," and 2 = "applies." Vertommen and colleagues (2002) found support for the reliability of the Dutch version of the PCL-R in a group of 1192 inmates. Cronbach's α was .87 and the average interitem correlation .25. Tentative evidence for the convergent validity was found in a subgroup of 98 forensic psychiatric inpatients, as there were modest, but meaningful correlations with self-report questionnaires such as the Dutch version of the Minnesota Multiphasic Personality Inventory (MMPI-2; Sloore et al., 1993). In the present study we used the total score as well as the four-factor structure proposed by Hare

and Neumann (2006), which includes the following facets: Interpersonal (e.g., "Grandiose self-worth"), Affective (e.g., "Callous and lack of empathy"), Lifestyle (e.g., "Impulsivity"), and Antisocial (e.g., "Juvenile delinquency").

The *NEO Five Factor Inventory* (NEO-FFI; Costa & McCrae, 1992; Dutch version: Hoekstra, Ormel, & De Fruyt, 1996) has 60 items and measures the Big Five personality domains of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Participants score items of the NEO-FFI on a five-point Likert scale ranged from "entirely disagree" to "entirely agree." In a Dutch sample of 356 non-clinical adults, Cronbach's alphas ranged between .69 and .82 for various subscales (Hoekstra et al., 1996). In a subgroup of 135 adults the test-retest reliability after 6 months ranged from .75 to .87 (Hoekstra et al., 1996). In the present study, we only used the scores on the Neuroticism and Agreeableness subscales, since these subscales differentiate between forensic psychiatric patients and non-clinical men (Hornsveld et al., 2008).

The Trait Anger subscale of Spielberger's (1980) *State-Trait Anger Scale* (STAS; Van der Ploeg, Defares, & Spielberger, 1982) was used to measure the general disposition to anger. Participants rate each item (e.g., "I am quick tempered") based on how they generally feel using a four-point Likert scale: 1 = "almost never," 2 = "sometimes," 3 = "often," and 4 = "almost always." In a group of 150 Dutch male university students, Van der Ploeg and colleagues (1982) found that internal consistency (coefficient alpha) of the Trait Anger scale was .78, and a test-retest reliability of .78 was documented in a subgroup of 70 students. The convergent validity of the Trait Anger scale also proved to be satisfactory (Van der Ploeg et al., 1982).

The *Adapted Version* of Rosenzweig's (1978) *Picture-Frustration Study* (PFS-AV; Hornsveld et al., 2007) was employed for measuring hostility. The test asks participants to write down their reactions to 12 cartoon-like pictures. Subjects are instructed to examine the situations as shown in the pictures (e.g., to a shopkeeper: "This is the third time that this watch has stopped") and to write the first appropriate reply that enters their mind in the blank text box. Answers are scored by an experienced research assistant (psychologist) on a seven-point scale, ranging from 1 = "not at all hostile" to 7 = "extremely hostile." In a sample of 231 Dutch violent forensic psychiatric patients, the internal consistency ($\alpha = .76$), test-retest reliability ($r = .67$), and interrater reliability ($r = .77$) of the PFS-AV were moderate to good. Furthermore, evidence was found for the convergent validity of the test, as scores correlated with indices of agreeableness and aggressive behavior (Hornsveld et al., 2007).

The *Aggression Questionnaire-Short Form* (AQ-SF; Bryant & Smith, 2001; Dutch version: Hornsveld et al., 2009) is a shortened version of Buss and Perry's (1992) Aggression Questionnaire, with 12 items that can be allocated to four subscales: Physical Aggression (e.g., "Once in a while I can't control the urge to strike another person"), Verbal

Aggression (e.g., “My friends say that I’m somewhat argumentative”), Anger (e.g., “I have trouble controlling my temper”), and Hostility (e.g., “Other people always seem to get the breaks”). Respondents score the items using a five-point scale ranging from 1 = “entirely disagree” to 5 = “entirely agree.” In a sample of Dutch forensic psychiatric patients (males) and a sample of secondary vocational students (females and males), Hornsveld and colleagues (2009) found that the four-factor structure of the AQ-SF produced an acceptable fit. In a group of 208 violent forensic psychiatric outpatients the internal consistency (alpha coefficient) for the AQ-SF total score and for the subscales was .88, .65, .74, .61, and .74, respectively. The test-retest reliability of the AQ-SF

total score in a subsample of 90 outpatients was modest but significant ($r = .38$). The convergent validity of the AQ-SF was supported by correlations with alternative measures of aggression and personality (Hornsveld et al., 2009). In the present study, we only employed the scores of the Physical Aggression and Verbal Aggression subscales.

Procedure

All patients were informed by the researchers about the purpose of the study and that participation was voluntarily. Those who participated received € 7 for their effort. The patient sample administered the SRM-AV between 2009 to

TABLE 1
Means, Standard Deviations, and Factor Loadings of the SRM-AV Items in the Patient Group (132 Inpatients and 33 Outpatients)

Item	Description	<i>M</i>	<i>SD</i>	Factor loadings			
				1	2	3	4
10	How important do you think it is that others treat you with respect?	2.61	1.09	.79	.09	-.14	.27
2	What about keeping a promise to people in general? How important is it for people to keep promises, even to someone they hardly know?	2.55	1.13	.65	.09	.31	.13
12	How important do you think it is that others do not gossip about you?	2.05	0.89	.63	.22	.18	.10
11	Imagine that two children nag another child. How important is it to see to it that these children learn to respect each other?	2.71	1.13	.60	.32	.01	.20
1	How important is it for people to keep promises to friends?	2.82	1.14	.60	.09	.30	-.01
9	How important is it for judges to impose a punishment on people who break the law?	2.62	1.24	.56	.50	.21	-.05
3	What about keeping a promise to a child? How important is it for parents to keep promises to their children?	3.24	1.15	.53	.13	.43	.25
7	How important is it for people not to steal things?	2.32	1.12	.49	.02	.48	.22
8	How important is it for people to obey the law?	2.44	1.28	.40	.29	.33	.05
16	Imagine that your friend would call his girl friend “bitch” in the presence of others. How important is it that these other people say something about this?	2.57	1.32	.12	.73	.08	.22
18	Imagine that you notice a friend of yours is dealing in hard drugs. How important is it that dealing is prohibited?	2.66	1.32	.05	.68	.22	.09
17	Imagine that two men maltreat another man while bystanders just stand and watch. How important is it that the bystanders say something about the behavior of these two men?	2.54	1.24	.34	.59	-.08	.03
19	Imagine that two lesbian women are kissing each other. How important is it that lesbian women are not discriminated?	3.00	1.22	.37	.49	.24	.36
6	How important is it for a person to stay alive, even if that person doesn’t want this?	3.05	1.30	.04	.18	.70	.06
20	Imagine that a child is confronted every day with quarreling parents. How important is it that parents consider their children by not involving them in their quarrels?	3.24	1.12	.18	.50	.62	.07
5	How important is it for children to help their parent(s)?	3.13	1.16	.35	-.10	.50	.37
4	In general, how important is it for people to be honest?	2.74	1.16	.32	.09	.46	.39
15	How important is it to help a physically disabled person when necessary?	3.37	1.20	.10	.17	.09	.83
14	Imagine that a disabled person does not dare to cross the street. How important is it to help this person to cross the street?	3.22	1.17	.08	.28	.08	.79
13	How important is it that people give their opinion in a direct way?	2.82	1.05	.26	.13	.35	.51

Note: Factor 1 = Expecting decent behavior from others, factor 2 = Addressing others with regard to their behavior, factor 3 = Exhibiting decent behavior to others, and factor 4 = Being helpful to others. Factor loadings $\geq .40$ are printed in bold.

2011, on several different occasions. In the summer of 2011 we asked 111 inpatients to participate in our study, from which 38 (34.23%) refused. The participants were significantly younger than the patients who refused, $t(109) = -2.04$, $p < .05$, but the groups did not differ in PCL-R total score, $F(2,108) = 1.02$, $p = .37$.

The members of the soccer club had been informed in advance about the study in the club magazine and folders with information were distributed during a game day. They received € 10 after completing the measures individually. The secondary vocational students completed the set of self-report measures in their classrooms at school. After a check on missing scores, students received € 10 in return for their participation.

The refund for the soccer players and the students was larger than the one for the patients, because they had to complete a few additional self-report questionnaires for another study.

RESULTS

Regarding the inpatients, the ratings of the two research assistants for determining the interrater reliability were used to calculate average SRM-AV total and item scores for the 80 participants in study 1. The scores of the first assistant yielded an internal consistency (alpha coefficient) of .91, whereas the internal consistency of the second assistant was .95. Mean scores had an internal consistency of .94, a mean interitem Spearman correlation coefficient of .44 and a mean item-total correlation coefficient of .64.

Factor Structure

In order to explore the factor structure of the SRM-AV, the data from the patient group ($n = 165$) was factor analyzed using the Principal Axis method with Varimax rotation. This

resulted in four factors that were provisionally interpreted as follows: Factor 1 = Expecting decent behavior from others, factor 2 = Addressing others with regard to their behavior, factor 3 = Exhibiting decent behavior to others, and factor 4 = Being helpful to others (Table 1). The interpretations for each factor were based on the items with the highest loadings.

Convergent Validity

The mean scores and standard deviations for various measures in the patient group are shown in Table 2.

A significant positive correlation was found between the total SRM-AV score on the one hand and age and Agreeableness (NEO-FFI) on the other (Table 2). The total SRM-AV score was significantly negatively correlated with total PCL-R score, and in particular, with the Affective, Lifestyle, and Antisocial facets, and the Verbal Aggression subscale of the AQ-SF.

Differences between Groups

We compared the patient sample with the soccer players and, in a separate analysis, with the secondary vocational students. The mean scores on the total SRM-AV and on the provisional four factors are presented in Table 3. Because of the exploratory nature of these comparisons, we did not apply a Bonferroni correction.

The patient sample had significantly lower total scores on the SRM-AV than the soccer players, $F(2,227) = 6.50$, $p = .002$, and the secondary vocational students, $F(2,249) = 16.28$, $p < .001$, controlling for age. Regarding the four provisional factors of the SRM-AV, a significant difference between patients and soccer players was found on factor 1, "Expecting decent behavior from others" [$F(2,227) = 3.87$, $p = .022$]; factor 2, "Addressing others with regard to their behavior" [$F(2,227) = 7.41$, $p = .001$]; and factor 3, "Exhibiting decent behavior to others" $F(2,227) = 4.89$, $p =$

TABLE 2
Correlations Between SRM-AV and Other Measures in the Patient Group (132 Inpatients and 33 Outpatients)

Measure	Content of the scale	<i>M (SD)</i>	SRM-AV
Age		36.04 (9.23)	.22**
PCL-R	Psychopathy	20.55 (8.41)	-.13*
	Interpersonal	3.12 (2.45)	.04
	Affective	5.70 (1.89)	-.15*
	Lifestyle	5.43 (2.97)	-.14*
	Antisocial	4.85 (2.81)	-.20**
NEO-FFI	Neuroticism	33.04 (7.84)	.01
	Agreeableness	41.38 (5.18)	.26**
STAS	Trait anger	17.62 (5.74)	-.10
PFS-AV	Hostility	23.39 (9.40)	-.13
AQ-SF	Physical aggression	8.37 (3.21)	-.02
	Verbal aggression	6.13 (2.50)	-.16*

Note: SRM-AV = Adapted Version of the Sociomoral Reflection Measure; PCL-R = Psychopathy Checklist-Revised; NEO-FFI = Five Factor Inventory; STAS = State-Trait Anger Scale; PFS-AV = Adapted Version of the Picture-Frustration Study; AQ-SF = Aggression Questionnaire-Short Form.

* $p < .05$, ** $p < .01$ (one-tailed).

TABLE 3
SRM-AV Total and Factor Scores in Patients ($N = 165$), Soccer Players ($N = 65$), and Secondary Vocational Students ($N = 87$)

Group	Total	Factor 1	Factor 2	Factor 3	Factor 4
Patients	55.74 (13.80)	23.41 (6.84)	10.77 (3.76)	12.15 (3.40)	9.41 (2.76)
Soccer players	58.22 (14.01)	24.29 (6.26)	11.71 (3.99)	12.71 (3.45)	9.51 (3.08)
Secondary vocational students	64.00 (12.04)	28.33 (5.84)	11.33 (3.44)	13.60 (3.05)	10.74 (2.64)

.008]. The secondary vocational students scored significantly higher than the patients on all four factors [factor 1: $F(2,249) = 19.49, p < .001$; factor 2: $F(2,249) = 4.32, p = .014$; factor 3: $F(2,249) = 7.46, p = .001$; and factor 4 (“Being helpful to others”): $F(2,249) = 12.25, p < .001$, controlling for age.

DISCUSSION

For the evaluation of the Moral reasoning module of the Aggression Control Therapy (Hornsveld et al., 2008), an adaptation of the Aggression Replacement Training of Goldstein, Glick, and Gibbs (1998), we developed an instrument for measuring moral maturity on the basis of the Sociomoral Reflection Measure (SRM-SF). The SRM-AV with 20 items showed a good internal consistency, interrater reliability and fair test-retest reliability. Exploratory factor analysis yielded four provisional factors: “Expecting decent behavior from others,” “Addressing others with regard to their behavior,” “Exhibiting decent behavior to others,” and “Being helpful to others.” As expected, the convergent validity of the SRM-AV was supported through significant, but relatively low positive correlations with age and Agreeableness, and negative correlations with psychopathy and Verbal Aggression in the patient sample. Patients scored significantly lower on the SRM-AV than the soccer players and the secondary vocational students. These findings support the findings of other researchers, that low moral maturity is related to delinquency.

Our study had a number of limitations. First, the patients participated in the study voluntarily and were all males. It is therefore not clear whether the result is representative for all forensic psychiatric patients. Secondly, the factor structure as found in this study has to be replicated in larger patient groups and in the general population, in part because some items did not load not exclusively on one factor. However, the four provisional factors showed some similarity with four of the five moral categories of Haidt (2007): “harm/care,” “fairness/reciprocity,” “in-group/loyalty,” and “authority/respect” (p. 999). Our clinical experience is that these categories are related to certain personality characteristics of forensic psychiatric patients. Although they are egotistic and generally not inclined to meddle with other people’s business, they can be helpful to others who are even more underprivileged than they are in situations where their authority is not at stake. A third limitation was the relatively large percentage of inpatients in the combined patient group. Hornsveld and col-

leagues (2009), and Hornsveld, Muris, and Kraaimaat (2011) found indications that the structured and controlled environment in an institution with a relatively high patient-staff ratio probably has an attenuating effect on the patients’ behavior and therefore on the scores of self-report questionnaires. In particular psychopaths seem to know how to behave and act according to their own, short-term interests (e.g., Cleckley, 1988, p. 338). Finally, certain subscales of the completed questionnaires did not correlate significantly with the SRM-AV, namely Neuroticism (NEO-FFI), Trait Anger (STAS), hostility, and Physical Aggression (AQ-SF). The low correlation with hostility is especially surprising. One would expect that people with a high level of moral maturity are in general less hostile towards their fellow men. In our opinion, the latter result needs to be explored further in a future study of the SRM-AV.

The finding that nine of the eleven items of the original SRM-SF were retained in the SRM-AV seems consistent with the statement of Gibbs and coworkers (2007), that “multimethod convergence is found for common moral values, basic moral judgment stage development, and related social perspective-taking across cultural groups” (p. 444). However, for the other eleven items of the SRM-AV we must take into consideration that responses to the statements may also be determined by the context in which the instrument will be used (Turiel, 1998). In general, the response to the content of, for example, item 19 (“How important is it that lesbian women are not discriminated?”) in the study in question was tolerant, but it is doubtful whether this also applies to responses of inhabitants of other countries. Then such an item may make a less reliable or valid contribution to the total score of the SRM-AV. Therefore, the items of the SRM-AV may be restricted to a particular time and place in our opinion.

In sum, the SRM-AV appears to be a promising instrument for measuring moral maturity in Dutch prisoners detained under hospital order. This first evaluation shows that the psychometric characteristics are good; however, more research is required to support these preliminary results, in particular on the ability of the SRM-AV to distinguish between forensic psychiatric patients and non-clinical populations.

REFERENCES

- Basinger, K. S., Gibbs, J. C. & Fuller, D. (1995). Context and the measurement of moral judgement. *International Journal of Behavioral Development, 18*, 537–556. doi: 10.1177/016502549501800309.

- Bryant, F. B., & Smith, B. D. (2001). Refining the architecture of aggression: A measurement model for the Buss-Perry Aggression Questionnaire. *Journal of Research in Personality, 35*, 138–167. doi: 10.1006/jrpe.2000.2302.
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452–459. doi: 10.1037/0022-3514.63.3.452.
- Carlo, G., Eisenberg, N., & Knight, G. P. (1992). An objective measure of adolescents' prosocial moral reasoning. *Journal of Research on Adolescence, 2*, 331–349. doi: 10.1207/s15327795jra0204_3.
- Cleckley, H. M. (1988). *The mask of sanity*. Augusta, GA: Emily S. Cleckley.
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Colby, A., Kohlberg, L., Spechier, B., Hower, A., Candee, D., Gibbs, J., & Power, C. (1987). *The measurement of moral judgment* (Vol. 2). Cambridge, UK: Cambridge University Press.
- Costa, P. T. Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and the Five Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources Inc.
- Gavaghan, M. P., Arnold, K. D., & Gibbs, J. C. (1983). Moral judgment in delinquents and nondelinquents: Recognition versus production measures. *Journal of Psychology, 114*, 267–274. doi: 10.1080/00223980.1983.9915424.
- Gibbs, J. C. (2003). *Moral development & reality: Beyond the theories of Kohlberg and Hoffman*. Thousand Oaks, CA: Sage.
- Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). *Moral maturity: Measuring the development of sociomoral reflection*. New Jersey: Erlbaum. doi: 10.1037/031668.
- Gibbs, J. C., Basinger, K. S., Grime, R. L., & Snarey, J. R. (2007). Moral judgment development across cultures: Revisiting Kohlberg's universality claims. *Developmental Review, 27*, 443–500. doi: 10.1016/j.dr.2007.04.001.
- Goldstein, A. P., Glick, B., & Gibbs, J. C. (1998). *Aggression Replacement Training: A comprehensive intervention for aggressive youth*. Champaign, IL: Research Press.
- Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316*, 998–1002. doi: 10.1126/science.1137651.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised*. Toronto, CA: Multi-Health Systems.
- Hare, R. D., & Neumann, C. S. (2006). The PCL-R assessment of psychopathy: Development, structural properties, and new directions. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 58–88). New York: Guilford.
- Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO Persoonlijkheidsvragenlijsten [Manual NEO Personality Questionnaires]*. Lisse, NL: Swets Test Services.
- Hornsveld, R. H. J., Muris, P. E. H. M., & Kraaimaat, F. W. (2011). The Novaco Anger Scale-Provocation Inventory (1994 version) in Dutch forensic psychiatric patients. *Psychological Assessment, 23*, 937–944. doi: 10.1037/a0024018.
- Hornsveld, R. H. J., Muris, P., Kraaimaat, F. W., & Meesters, C. (2009). Psychometric properties of the Aggression Questionnaire in Dutch violent forensic psychiatric patients and secondary vocational students. *Assessment, 16*, 181–192. doi: 10.1177/1073191108325894.
- Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2007). An adapted version of the Rosenzweig Picture-Frustration Study (PFS-AV) for the measurement of hostility in violent forensic psychiatric patients. *Criminal Behaviour and Mental Health, 17*, 45–56. doi: 10.1002/cbm.638.
- Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2008). Aggression Control Therapy for violent forensic psychiatric patients: Method and clinical practice. *International Journal of Offender Therapy and Comparative Criminology, 52*, 222–233. doi: 10.1177/0306624x07303876.
- Palmer, E. J. (2003). An overview of the relationship between moral reasoning and offending. *Australian Psychologist, 38*, 165–174. doi: 10.1080/00050060310001707177.
- Rosenzweig, S. (1978). *Aggressive behavior and the Rosenzweig Picture-Frustration Study*. New York: Praeger. doi: 10.1002/1097-4679(1976-10)32:4<885::AID-JCLP2270320434>3.0.CO;2-R.
- Sloore, H., Derksen, J. J. L., Helenbosch, G., & De Mey, H. R. A. (1993). *De Minnesota Multiphasic Personality Inventory-II (Dutch translation)*. Nijmegen, NL: PEN Test Publishers.
- Spielberger, C. D. (1980). *Preliminary manual for the State-Trait Anger Scale (STAS)*. Tampa, FL: University of South Florida.
- Stams, G. J., Brugman, D., Dekovic, M., Rosmalen, L. van, Laan, P. van der, & Gibbs, J. C. (2006). The moral judgment of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology, 24*, 697–713. doi: 10.1007/s10802-006-9056-5.
- Turiel, E. (1998). The development of morality. In W. Damon (Ed.), *Handbook of child psychology*, fifth edition, volume 3 (pp. 863–932). New York: Wiley. doi: 10.1002/9780470147658.chpsy0313.
- Van der Ploeg, H. M., Defares, P. B., & Spielberger, C. D. (1982). *Handleiding bij de Zelf-Analyse Vragenlijst, ZAV [Manual for the Self-Analysis Questionnaire, ZAV]*. Lisse, NL: Swets & Zeitlinger.
- Vertommen, H., Verheul, R., De Ruiter, C., & Hildebrand, M. (2002). *De herziene versie van Hare's Psychopathie Checklist (PCL-R) [The revised version of Hare's Psychopathy Checklist (PCL-R)]*. Lisse, NL: Swets Test Publishers.

APPENDIX. SCORING OF THE ADAPTED VERSION OF THE SOCIOMORAL REFLECTION MEASURE (SRM-AV)

Instructions

The responses to questions regarding the importance of a proposition do not have to be scored. These questions have to stimulate respondents to think about their opinion regarding the proposition before writing their opinion down. The justifications of the responses have to be scored on a seven-point scale according to the following seven phases.

Score	Phase
1	<p><i>Phase 1: Unilateral and physicalistic</i></p> <ul style="list-style-type: none"> • Justification because of an authority figure • Justification because of the most salient role or status of the person(s) involved • Flat assertions in absolute terms • Gross or undifferentiated labels in moral justification • Justification because of physicalistic (punitive) consequences
2	<p><i>Transitional phase 1–2</i></p>
3	<p><i>Phase 2: Exchanging and instrumental, morality through interaction with others</i></p> <ul style="list-style-type: none"> • Quid Pro Quo • Strict equality or inequality in relations • Unfettered or unconstrained freedom as concrete rights • Justification on account of one's own wishes, desires, or inclinations • Appeal to pragmatic needs • Calculation of anticipated practical benefits or liabilities
4	<p><i>Transitional phase 2–3</i></p>
5	<p><i>Phase 3: Mutual and prosocial</i></p> <ul style="list-style-type: none"> • Awareness of the psychological meaning of interpersonal relationships • Strongly empathic references to another's psychological or emotional welfare • Awareness of consequences when violating normative expectations • Awareness of prosocial intentions or features of the normal social personality • Normative prosocial prescriptions or values beyond the context of particular relationships or roles • Applying values that refer to feelings of a clean conscience or pride.
6	<p><i>Transitional phase 3–4</i></p>
7	<p><i>Phase 4: Systematic and standard</i></p> <ul style="list-style-type: none"> • Justifications where a moral value is supported as a requirement for society or one of its institutions • Appeal to basic rights or values applicable to any viable society • Acceptance of norms such as responsibility, obligation or commitment • Justifications that appeal to considerations of responsible character or integrity • Justification of normative values because alternatives are harmful for society • Rights and duties that society owes to the individual • Social standards with regard to the individual or personal conscience