

Reliability and Factor Structure of the Dutch Language Version of Hare's Psychopathy Checklist-Revised

Martin Hildebrand, Corine de Ruiter, Vivienne de Vogel, and Pascal van der Wolf

The present study examined the interrater and internal reliability and factor structure of the Dutch language version of the Psychopathy Checklist-Revised (PCL-R; Hare, 1991) in a sample of 107 patients admitted to a forensic psychiatric hospital. In addition, we evaluated the potential role of two different information sources, real-life interview versus videotaped interview, when scoring the PCL-R. Interrater reliabilities of individual items and the PCL-R total score were good to excellent. Good agreement on the categorical diagnosis of psychopathy was also obtained (weighted Cohen's $k = .63$ for simultaneous comparison of three raters). The internal consistency of the PCL-R was found to be high, as indicated by a Cronbach's alpha of 0.87, with an alpha of 0.83 for both Factor 1 and Factor 2. Comparisons between real-life and videotaped interview demonstrated that the information source did not influence the raters' coding. Confirmatory factor analysis (CFA) indicated that the two-factor structure obtained by Hare (1991) in the standardization samples did not fit the current data well. CFA also failed to confirm the three-factor model identified by Cooke and Michie (2001). Exploratory principal components analysis using oblique rotation extracted two main factors, which accounted for 44% of the variance. It is concluded that the Dutch language version of the PCL-R can be reliably rated by trained professionals, its factor structure resembles the traditional two-factor model to some extent, and future research should include larger samples of different populations such as prisoners and general psychiatric patients.

Psychopathy is a controversial psychiatric concept (Gunn & Robertson, 1976; Lewis, 1974). Definitions of the term psychopathy have historically been both diverse and difficult to operationalize (e.g., Craft, 1965; Dolan & Coid, 1993; Hare, 1970; McCord & McCord, 1964), and research on psychopathy has been characterized by the absence of a clear and generally agreed upon conceptualization of the disorder (O'Kane, Fawcett, & Blackburn, 1996). Despite the fact that the concept of psychopathy has been obscured by a multitude of definitions, the clinical description of psychopathy provided by Cleckley (1941/1976) has received widespread acceptance among contemporary researchers and clinicians. Cleckley provided the first systematic clinical account of psychopathy, defining

psychopathy as a constellation of 16 personality traits, reflecting both the affective and interpersonal characteristics that have traditionally been considered central to psychopathy, including egocentricity, failure to form close emotional bonds, callousness, and lack of guilt.

Much of the recent interest in the construct of psychopathy is attributable to the development of the Hare Psychopathy Checklist-Revised (PCL-R; Hare, 1991) and its ability to predict future (violent) criminal behavior (e.g., Hart, 1998; Hemphill, Hare, & Wong, 1998; Salekin, Rogers, & Sewell, 1996). Hare (1980), by adapting components of Cleckley's conceptualization of psychopathy, and adding items related to antisocial behavior, developed and validated the Psychopathy Checklist (PCL), followed

Martin Hildebrand, Vivienne de Vogel, and Pascal van der Wolf, Department of Research, Dr. Henri van der Hoeven Kliniek, Utrecht, the Netherlands; Corine de Ruiter, Department of Psychology, University of Amsterdam, and Netherlands Institute of Mental Health and Addiction (Trimbos-institute), Utrecht, the Netherlands.

We are grateful to Ellen van den Broek, Henriette van der Maeden, Cécile Vandeputte-van de Vijver, Anke Weenink, Pascal Wolters and Stefan Zwartjes who, in addition to the authors, participated as interviewers/raters. We also appreciate the input of Gerard H. Maassen, Stephen D. Hart, and an anonymous reviewer, who commented on previous versions of this article. The study presented here is part of a doctoral dissertation project examining the role of the PCL-R in predicting treatment response in Dutch forensic psychiatric patients conducted by Martin Hildebrand under the supervision of Professor Corine de Ruiter.

Correspondence concerning this article should be addressed to Martin Hildebrand, Department of Research, Dr. Henri van der Hoeven Kliniek, P.O. Box 174, 3500 AD Utrecht, the Netherlands (E-mail: mhildebrand@hoevenstichting.nl).

by a revised version, the PCL-R (Hare, 1991). The PCL-R is a 20-item clinical construct scale completed on the basis of a semi-structured interview and file information. Items are scored on a three point scale (0 = *item does not apply*, 1 = *item applies to a certain extent*, 2 = *item definitely applies*). The total score can range from 0 to 40, reflecting the degree to which an individual resembles the prototypical psychopath. In the PCL-R manual, Hare (1991) suggested a cutoff score of 30 or more to assign a clinical diagnosis of psychopathy.

The Hare PCL-R was initially developed and validated with data from North American samples of male adult offenders and forensic psychiatric patients. A growing body of research has demonstrated the reliability and validity of the PCL-R for prison and forensic psychiatric samples in other countries (e.g., Grann, Långström, Tengström, & Stålenheim, 1998; Moltó, Poy, & Torrubia, 2000; Tengström, Grann, Långström, & Kullgren, 2000). In addition, reliability and validity studies with adolescent offenders (Brandt, Kennedy, Patrick, & Curtin, 1997; Forth & Burke, 1998; Forth & Mailloux, 2000), substance abusers (e.g., Alterman, Cacciolo, & Rutherford, 1993; Rutherford, Cacciola, Alterman, & McKay, 1996), female offenders (Salekin, Rogers, & Sewell, 1997), and even male and female non-criminals (Forth, Brown, Hart & Hare, 1996) have been conducted.

At least initially, factor analytic studies of the PCL (Haapasalo & Pulkkinen, 1992; Harpur, Hakstian, & Hare, 1988; Templeman & Wong, 1984) and the PCL-R (e.g., Cooke, 1995; Hare, 1991; Hare et al., 1990; Hobson & Shine, 1998; Windle & Dumenci, 1999) showed that PCL(-R) psychopathy is a higher order clinical construct composed of two distinct and moderately correlated factors. PCL-R Factor 1 consisted of a cluster of eight items reflecting the affective and interpersonal features (core personality traits) of psychopathy, and has been labeled "Selfish, callous and remorseless use of others" (Hare, 1991; Hare et al., 1990). Factor 2 consisted of nine items reflecting the social deviance features of psychopathy and has been labeled "Chronically unstable and antisocial lifestyle." The remaining three items of the PCL-R (promiscuous sexual behavior, many short-term marital relationships and criminal versatility) did not load on either factor (Hare, 1991; Hare et al., 1990). Difficulty

replicating the initial two-factor solution, however, has occurred in a number of selected samples, including female offenders (Salekin et al., 1997), African-American offenders (Kosson, Smith, & Newman, 1990), a mixed sample of community- and prison-based methadone patients (Darke, Kay, Finley-Jones, & Hall, 1998), and a large sample of substance-dependent male and female patients (McDermott et al., 2000). Furthermore, various researchers have proposed alternative conceptualizations of dimensions that might underlie the PCL and PCL-R (Darke et al., 1998; Haapasalo & Pulkkinen, 1992; Raine, 1985). Recently, Cooke and Michie (2001) reexamined the adequacy of the two-factor model of psychopathy. They arrived at the conclusion that "although the two-factor model has served as a useful heuristic device to guide research on psychopathy, it does not provide an adequate structural model for psychopathy" (p. 173). Using confirmatory factor analysis, Cooke and Michie (2001) identified a three-factor hierarchical model, based on 13 of the 20 PCL-R items, in which a coherent superordinate factor (i.e., psychopathy), was underpinned by an interpersonal (*Deceitful interpersonal style*), affective (*Deficient affective experience*), and behavioral (*Impulsive and irresponsible behavioral style*) factor. Contrary to the two-factor model, the three-factor model places little emphasis on criminality.

THE PRESENT STUDY

Implementation of the PCL-R for clinical use in any new (cultural) context should be accompanied by a thorough evaluation of the psychometric status of the instrument in that particular context. According to the cross-cultural literature, this should include examining of structural invariance (via confirmatory factor analysis) and metric invariance (via item response theory or differential item functioning; e.g., van de Vijver & Leung, 1997; Cooke, Kosson, & Michie, 2001). Different offender populations require separate norms and validity data, because findings from one sample may not be applicable to another. In the case of adoption of an instrument (i.e., the PCL-R) from one cultural group to another, one is interested in adopting the nomological network of the underlying construct tapped by the instrument,

as established in the original culture. Relevant to the issue of cross-cultural adaptation, this nomological network can be understood as based on two sets of relationship: (1) the internal network of empirical relationships within the factor structure of the assessment instrument, and (2) the relationship of the instrument to external correlates associated with the construct of interest (Ben-Porath, 1990). To establish the cross-cultural validity of the PCL-R, both internal and external sets of relationships must be demonstrated to be invariant across cultural (or ethnic) groups. Approaches based on classical test theory have limited value in this regard (e.g., van de Vijver & Leung, 1997). Cross validation of the factor structure across groups using confirmatory factor analytic techniques and item response theory (IRT) modeling (Cooke & Michie, 1997; Cooke, Michie, Hart, & Hare, 1999), for example, have become widely used in psychological assessment research as a way to study the underlying structure of the data.

The generalizability of the PCL-R's dimensional structure has not yet been established for forensic psychiatric patients in the Netherlands. The present study was designed to examine the reliability and factor structure of the Dutch language version of the PCL-R in a sample of forensic psychiatric patients involuntarily admitted to the Dr. Henri van der Hoeven Kliniek, a forensic psychiatric hospital in the Netherlands. First, we examine the interrater reliability of PCL-R item and total scores as well as the internal reliability (item homogeneity and internal consistency) of the PCL-R. In addition, we examine the role of two different information sources in rating the PCL-R. Hare (personal communication, October 1997) suggested that different information sources (i.e., real-life interview versus videotaped interview) might play a role in the scoring of the PCL-R. Face-to-face interactions in a real-life interview might result in different general impressions, and thus different scores, than impressions based on videotaped interviews. Particularly scores on PCL-R items pertaining to essentially "soft" or "impressionistic" data (Cooke, 1995, p. 111), such as Item 1 (*Glibness/superficial charm*), Item 5 (*Conning/manipulative*) or Item 7 (*Shallow affect*) may be more susceptible to a possible source effect, because they require a considerable degree of subjective judgment. Finally, we examine the factor structure of the Dutch language version of the PCL-R.

METHOD

Setting

The study was conducted at the Dr. Henri van der Hoeven Kliniek, a 110-bed forensic psychiatric hospital for the residential treatment of mentally disordered offenders who are sentenced by criminal court to involuntary commitment because of (severely) diminished responsibility for the offense(s) they committed. In terms of legal status, most patients admitted to our hospital are sentenced by criminal court to a '*maatregel van terbeschikkingstelling*' (TBS-order), a judicial measure which can be translated as 'disposal to be treated on behalf of the state'. The purpose of the TBS-order is to protect society from unacceptably high risks of recidivism, directly through involuntary admission to a forensic psychiatric hospital, and indirectly through the treatment provided there. Every one or two years the court re-evaluates the patient in order to determine whether the risk of (violent) recidivism is still too high and treatment needs to be continued. Theoretically, treatment under the TBS-order is of indefinite duration if the offender continues to pose a risk to society. (For discussions of TBS, see van Marle, 2002; de Ruiter & Hildebrand, in press). The average duration of residential treatment at the Dr. Henri van der Hoeven Kliniek is approximately four years.

Participants

In total, the study sample consisted of a mixed sample of 107 patients (98 men, 9 women) with DSM Axis I and/or Axis II disorders which can be considered representative for forensic psychiatric patients in the Netherlands. A subset of 60 patients (51 men, 9 women) took part in the study of the interrater reliability of the PCL-R. Participants in the study of the psychometric properties of the Dutch PCL-R (internal consistency, factor structure) were 98 male patients. Earlier findings (e.g., Strachan, Williamson, & Hare, 1990, cited in Hare, 1991; also Silverthorn & Frick, 1997) suggest that there may be sex differences in the behavioral manifestations of PCL-R psychopathy. To exclude possible confounding influences, we therefore chose to

examine internal consistency and factor structure in male patients only. Table 1 presents demographic and diagnostic characteristics of the total sample.

Mean age upon admission to the hospital was 31.5 years ($SD = 8.0$; range 19-50) for the total sample (men: $M = 31.3$, $SD = 7.9$; women: $M = 33.6$, $SD = 9.3$). In terms of ethnic origin, 77% of the sample was White, 14% was Afro-Caribbean, 6% was Mediterranean, and 4% other (e.g., Indonesian, Korean). Eighty-five patients (79%) were single. In terms of offenses, 52% of the sample was convicted for (attempted) murder/homicide, 23% for sexual offences (e.g., rape, pedosexual offenses), 8% for robbery with violence, 7% for arson, and the rest for (aggravated) assault or extortion.

Based on all available data, consensus DSM-IV Axis I disorders were established for all patients by MH and CdR, in cooperation with a senior-diagnostician and a senior psychotherapist of the hospital staff. Forty-nine patients (46%) met criteria for at least one substance-related disorder (e.g., alcohol, cannabis, polysubstance abuse/dependence); 14 patients (13%) met criteria for schizophrenia or other psychotic disorders and 7 for mood disorders. In addition, 18 patients (17%) received a diagnosis of paraphilia.

The Dutch language version (Van den Brink & de Jong, 1992; De Jong, Derks, van Oel, & Rinne, 1997) of the Structured Interview for Personality Disorders (SIDP-R; Pfohl, Blum, Zimmerman, & Stangl, 1989; SIDP-IV; Pfohl, Blum, & Zimmerman, 1997) was used for the assessment of DSM-III-R/DSM-IV personality disorders (PDs). Our initial use of DSM-III-R diagnoses was dictated by the duration of the data collection, which started before the SIDP-IV interview became available. The prevalence of PDs was substantial in this sample. Nineteen patients met criteria for paranoid PD, 8 for schizoid PD or schizotypal PD, 47 for antisocial PD, 29 for borderline PD, 7 for histrionic PD, 26 for narcissistic PD, 14 for avoidant PD, 6 for dependent PD, 11 for obsessive-compulsive PD and 5 for passive-aggressive PD. Co-morbidity on Axis II was more common than a single diagnosis. Of the 87 patients given a PD diagnosis, 56 (64%) received multiple diagnoses. The mean number of PDs per patient for patients with at least one diagnosis was 2.1.

Assessment

Since January 1996, newly admitted patients have been assessed upon admission (T0; baseline assessment) with a standardized psychological assessment battery. In order to provide information on treatment progress, all patients in our hospital are re-tested 18-24 months after admission (T1; follow-up 1), and again 42 months after admission (T2; follow-up 2). In November 1997, we also implemented PCL-R psychopathy assessment within our hospital.

PCL-R ratings of all patients in the study were based on the Dutch translation of the semi-structured interview schedule designed by Hare (1991) and a review of all the collateral information arriving with each patient upon admission to the hospital. The Hare PCL-R interview is a comprehensive interview concerning school adjustment, work history, future goals, finances, family background, sexual and intimate relationships, child and adolescent antisocial behavior and adult delinquency. For all patients, extensive collateral information was available, consisting of earlier psychiatric and psychological assessments for the court (at least one psychiatric and psychological evaluation per patient), police reports of past and current offense(s), prior commitments to treatment facilities, and information on family background. The authorized Dutch translation of the Hare PCL-R manual and the scoring sheet were used (Vertommen, Verheul, de Ruiter, & Hildebrand, 2002).

Raters and Training

In total, a pool of 10 raters (three men, seven women) was used to administer the PCL-R interviews. Four raters (the authors) took part in the interrater reliability study. Seven raters were M.A. (clinical) psychologists; two raters had a degree in mental health science (M.Sc.) of whom one (MH) also had a degree in law, and one rater is a Ph.D. clinical and forensic psychologist (CdR). All raters were familiar with DSM-IV Axis I/Axis II disorders (American Psychiatric Association, 1994) and had experience in psychological assessment and/or

Table 1
Sample characteristics (N = 107)

		<i>N</i>	%
Sex	Male	98	91.6
	Female	9	8.4
Age (years)	18-30	54	50.5
	31-40	36	33.6
	41-50	17	15.9
Ethnic origin	White	82	76.6
	Afro-Caribbean	15	14.0
	Mediterranean	6	5.6
	Other	4	3.8
Index offence	Murder/homicide	56	52.3
	Sexual offence	25	23.4
	Robbery	8	7.5
	Arson	7	6.5
	Other	11	10.3
Axis I diagnosis	Any Axis I disorder	91	85.0
	Any substance abuse/dependence	49	45.8
	Psychotic disorder	14	13.1
	Mood disorder	7	6.5
	Paraphilia	18	16.8
	Pathological gambling	11	10.3
Axis II diagnosis	Paranoid	19	18.4
	Schizoid	8	7.8
	Schizotypal	8	7.8
	Antisocial	47	45.6
	Borderline	29	28.2
	Histrionic	7	6.8
	Narcissistic	26	25.2
	Avoidant	14	13.6
	Dependent	6	5.8
	Obsessive-compulsive	11	10.7
	Passive-aggressive	5	4.9
	Any Axis II disorder	87	84.5

Note. DSM-III-R/DSM-IV Axis II diagnoses of 103 patients.

treatment of (forensic) psychiatric patients. All raters were trained extensively in administration and scoring of the PCL-R, either by Drs. Robert D. Hare and David Cooke in a three-day PCL-R workshop held at the Dr. Henri van der Hoeven Kliniek in October 1997; by Drs. Robert D. Hare and Stephen D. Hart in a three-day PCL-R workshop (Nijmegen, April 2000), or by Dr. Stephen D. Hart in a PCL-R workshop in Amsterdam (February 2001). In addition, the four raters who took part in the interrater reliability study (see Procedure) reviewed videotaped interviews of four patients and discussed PCL-R scores in detail to improve and sharpen the coding of the PCL-R criteria, prior to the reliability study.

Procedure

As a general rule, PCL-R interviews are videotaped in our hospital. However, patients have to give their written informed consent before videotaping the interview. Thus, patients were selected on their willingness to give informed consent and to cooperate with the interview process. If a patient refused to give informed consent for videotaping the interview, one rater conducted the interview while a second rater was present as an observer. After independent review of all available information (interview and file information), each rater scored the PCL-R and a meeting was planned to obtain a final (consensus) rating for that patient. This procedure, which is recommended by Hare (1991, 1998), was chosen to maximize scoring accuracy. The PCL-R consensus scores are used in all subsequent data-analyses. If a patient refused videotaping the interview and also the presence of a second observer during the interview, PCL-R scores were based on the judgment of a single interviewer ($n = 7$). In all other cases, PCL-R consensus scores were based on independent PCL-R ratings of at least two independent raters.

To determine the interrater reliability of the PCL-R, videotaped interviews of 60 patients were rated independently by the interviewer and by two raters who watched the videotape of the interview (video versus vis-à-vis interview). In this way we were able to study the potential influence of information source on the PCL-R score (source effect). We do not know of any prior research that investigated the role of

interview source in PCL-R ratings. Therefore, we could not draw on previous findings to set an expected effect size. However, because PCL-R ratings are only partially based on interview data, we expected the size of the source effect to be small. For analysis of variance with equal N s in three samples (i.e., videorater 1 versus videorater 2 versus interviewer) a small effect size, in terms of f , is “generally found in the .00 - .40 range” (Cohen, 1988, p. 284). The power value of the F test for the main effect of source, given a significance level α of .05, effect size $f = .20$, sample size $n = 60$ and degrees of freedom for the numerator of the F ratio = 2, is given as .67. This means that the *a priori* probability of rejecting the null hypothesis (i.e., the hypothesis that there is a source effect) is .67. For $\alpha = .05$ and $f = .40$, the power is .99.

The PCL-R interviews in the interrater reliability study were conducted by three of the four raters. PCL-R interviews with female patients ($n = 9$) were equally distributed among the interviewers. Each interviewer conducted 20 interviews and also rated 15 videotaped interviews from each of the other two interviewers. The fourth rater viewed 30 videotaped interviews conducted by the interviewers, 10 of each interviewer. All raters had access to the same collateral information. Occasionally, raters had been in contact with the patient previously for other psychodiagnostic activities (especially for patients interviewed at T1). However, none of the raters had been in contact previously with a patient for psychotherapy. The mean duration of the 60 PCL-R interviews in the reliability study was 165 minutes ($SD = 47$), varying from 50 to 296 minutes.

RESULTS

Descriptive Statistics

The mean total consensus score of the PCL-R for the 98 male patients was 21.25 ($SD = 8.41$), with a range from 3 to 38, a median score of 21.10 and a mode of 17.0 (see also Table 2). A t test revealed no significant difference between the total PCL-R scores of White ($n = 75$) versus non-White ($n = 23$) participants, $t(96) = .21$, n.s. The kurtosis of the PCL-R score was $-.77$ ($SE = .49$). PCL-R scores were normally distributed, Kolmogorov-Smirnov $Z = .60$,

Table 2

Descriptive statistics, internal consistency (Cronbach's α), item homogeneity (mean inter-item correlation), and interrater reliability (intraclass correlation coefficient; ICC) of PCL-R total and factor scores

	Mean	SD	Internal consistency	Item homogeneity	Interrater reliability
PCL-R total score	21.25	8.41	.87	.26	.88
PCL-R Factor 1 score	9.42	3.73	.83	—	.76
PCL-R Factor 2 score	9.14	5.00	.83	—	.83

Note. PCL-R = Psychopathy Checklist-Revised. PCL-R scores are adjusted sums of 98 male patients. For ICCs, $N = 60$. Single rater ICCs were calculated using a two-way random effects model. — = not calculated.

$p = .86$. The mean Factor 1 score was 9.42 ($SD = 3.73$) and the mean Factor 2 score was 9.14 ($SD = 5.0$). For female patients ($n = 9$), the mean total score was 12.24 ($SD = 6.66$; range from 2 to 22), the mean Factor 1 score was 5.22 ($SD = 2.99$), and the mean Factor 2 score was 5.19 ($SD = 3.15$).

When a cut-off point of 30 was used, 20 of the male patients (20%) were classified as 'psychopaths.' When a lower threshold of 26 was applied, which is often used in European research (Grann et al., 1998; Rasmussen, Storsæter, Levander, 1999), 32 patients (33%) were classified as psychopathic.

It can be seen from Table 3 that individual PCL-R item means ranged between 0.36 for Item 17 (*Many short-term marital relationships*) to 1.55 for Item 16 (*Failure to accept responsibility for own actions*). Table 3 further indicates that a relatively high proportion of missing values was found for Items 17 (*Many short-term marital relationships*; 11%) and 19 (*Revocation of conditional release*; 28%).

Reliability Analysis

Interrater reliability. Interrater reliability for PCL-R total and factor scores, as well as for individual PCL-R items was estimated by means of the intraclass correlation coefficient (ICC; Shrout & Fleiss, 1979; McGraw & Wong, 1996). This coefficient expresses the reliability of a rating by one rater generalized to the population of raters, from which the sample of raters was taken. In other words, ICCs estimate the equivalence of repeated measurements made on the same subjects. The following

categories are often used for evaluating the observed reliability (Fleiss, 1986): $R \geq 0.75 =$ excellent; $0.40 \leq R < 0.75 =$ fair to good; $R < 0.40 =$ poor. A two-way random effects model type was used for computing the ICC. With the use of a two-way ANOVA, it is possible to measure how much of the total variance in the observed scores is a result of between-subject variation, between-rater variation and uncontrollable (random) variation. The random effects model is appropriate when the raters involved in the study are a random sample of a population of possible raters who will later use the instrument under evaluation. Because raters were not crossed with patients in one 60 X 3 design, it is best to describe the interrater reliability study as consisting of one 30 X 3 design and three 10 X 3 designs. Given the unequal n s, weighted average single measure ICC correlations were calculated for individual PCL-R items, as well as for PCL-R Factor 1, Factor 2, and total scores.

The single measure ICC for the PCL-R total score was .88; for Factor 1, it was .76, and for Factor 2, it was .83. At the level of individual PCL-R items, in general, ICCs were good to excellent ($Mdn = .67$, range .46 to .80). Table 3 details interrater reliabilities for all PCL-R items. The highest single measure ICCs were obtained for Items 3 (*Need for stimulation/proneness to boredom*; .80), and 11 (*Promiscuous sexual behavior*; .80). Four items, including three loading on Factor 1, had reliabilities less than .60. Interrater reliabilities of PCL-R Factor 1 items tend to be slightly lower than reliabilities of Factor 2 items (Mdn ICC for Factor 1 = .63, for Factor 2 = .67).

Table 3

Frequency of item scores, descriptive statistics, corrected item-total correlations, and interrater reliability (intraclass correlation coefficient; ICC) of individual PCL-R items

Item description	Value label (<i>n</i>)				Mean	SD	Item- total <i>r</i>	ICC
	0	1	2	Omitted				
1. Glibness/superficial charm	58	24	16	0	0.57	0.76	.49	.46
2. Grandiose sense of self worth	30	38	30	0	1.00	0.79	.38	.51
3. Needs stimulation/prone to boredom	27	39	32	0	1.05	0.78	.60	.80
4. Pathological lying	29	44	25	0	0.96	0.74	.53	.65
5. Conning/manipulative	20	35	43	0	1.23	0.77	.51	.66
6. Lack of remorse or guilt	10	35	53	0	1.44	0.67	.63	.69
7. Shallow affect	13	49	36	0	1.23	0.67	.26	.60
8. Callous/lack of empathy	13	49	36	0	1.23	0.67	.45	.52
9. Parasitic lifestyle	39	32	24	3	0.84	0.80	.64	.68
10. Poor behavioral controls	22	32	44	0	1.22	0.79	.47	.65
11. Promiscuous sexual behavior	32	16	48	2	1.17	0.90	.14	.80
12. Early behavior problems	63	17	15	3	0.49	0.76	.33	.79
13. Lack of realistic, long term goals	37	24	37	0	1.00	0.87	.59	.52
14. Impulsivity	22	37	39	0	1.17	0.77	.57	.67
15. Irresponsibility	16	32	50	0	1.35	0.75	.56	.63
16. Failure to accept responsibility	10	24	64	0	1.55	0.68	.51	.67
17. Many short-term marital relations	66	11	10	11	0.36	0.68	.40	.79
18. Juvenile delinquency	47	16	31	4	0.83	0.90	.36	.78
19. Revocation of conditional release	17	9	45	27	1.39	0.85	.42	.77
20. Criminal versatility	26	30	42	0	1.16	0.82	.60	.76

Note. PCL-R = Psychopathy Checklist-Revised. Item frequencies, means and standard deviations, and item-total correlations are based on (adjusted sum) PCL-R consensus ratings from 98 male patients. For interrater reliability analyses, $N = 60$. However, due to omitted items by at least one of the raters, ICCs for item 4 ($n = 59$), item 9 ($n = 59$), item 12 ($n = 57$), item 17 ($n = 54$), item 18 ($n = 57$), and item 19 ($n = 38$) are based on smaller sample sizes. Single rater ICCs were calculated using a two-way random effects model. All ICCs were significantly greater than 0 ($p < 0.05$).

Agreement on PCL-R categorical diagnoses was assessed using generalized Cohen's kappa (k ; Cohen, 1980). This statistic indicates the agreement between raters corrected for agreement by chance, and is considered the standard index of diagnostic agreement for categorical data (Shrout, Spitzer, & Fleiss, 1987). The same diagnostic categories that are used for evaluating the ICCs are used for evaluating the kappa. Comparison of PCL-R categorical diagnoses among the three raters showed

good agreement, weighted average Cohens's $k = .63$, for the presence versus absence of PCL-R psychopathy (adjusted sum total score ≥ 30). We also examined categorical agreement between raters using a lower diagnostic cut-off score of 26. This revealed an even better agreement on categorical diagnosis between the three raters, Cohens's $k = .72$. Furthermore, in 54 of the 60 cases (90%) the raters agreed on the presence or absence of psychopathy (PCL-R ≥ 30); 11 patients (18%) received a diagnosis

of psychopathy from at least one of the three raters, six patients (10%) received a psychopathy diagnosis from at least two raters, while all three raters gave five patients (8%) a psychopathy diagnosis. In addition, in 47 of the 60 cases (78%), PCL-R scores between the three raters did not differ more than five points. In three cases, however, rather extreme differences (≥ 10 points) between raters were found.

Interviewer ratings versus ratings conducted by video-observers. A source (interview versus video) by rater (Raters 1 through 4) ANOVA was performed on all individual PCL-R items and the PCL-R total score. Neither the main effects for source and rater, nor the Source X Rater interaction, were significant for any of the PCL-R items or the PCL-R total score. This indicates that the information source (i.e., real-life interview versus a videotaped interview) did not affect the scoring of the rater to a significant degree.

Internal consistency. The internal consistency of the PCL-R was examined in the sample of 98 male patients using Cronbach's coefficient alpha, and was found to be high, $\alpha = .87$, with an alpha coefficient of .83 obtained for Factors 1 and 2. These figures are comparable to those obtained with the standardization sample (Hare, 1991). To provide a more fine-grained analysis of internal consistency, we also examined corrected item-total correlations for each PCL-R item in this sample (see Table 3). With the exception of Items 7 (*Shallow affect*), 11 (*Promiscuous sexual behavior*), and 12 (*Early behavior problems*), all items had corrected item-to-total correlations $\geq .35$, indicating that they all contribute significantly to the PCL-R total score; 10 of the 20 items had item-total correlations of .50 or higher. The highest correlations were obtained for Items 9 (*Parasitic lifestyle*; $r = .64$) and 6 (*Lack of remorse or guilt*; $r = .63$). The mean inter-item correlation¹ was .26, which is above the suggested cut-off of .20 for a scale to be considered homogeneous (Green, Lissitz, & Mulaik, 1977).

¹ Because computation of the inter-item correlation requires a score for each item, the value 1 was assigned to omitted items. Following Hare (1991), we carried out different analyses to determine the most appropriate method for assigning a value to a missing item (e.g., scoring a missing value using the mean score for that item obtained in the sample, scoring a missing value as 1). Because it made little difference which method was used, we decided to use the simplest method (assigning the value 1).

Factor Analysis

To determine whether the oblique two-factor structure of the PCL-R (Hare, 1991; Harpur et al., 1989) could be replicated in the current sample of 98 male forensic psychiatric patients, a confirmatory factor analysis (CFA) was performed. The question in CFA is whether the correlations among variables are consistent with a hypothesized factor structure. Compared to exploratory FA, CFA offers more definitive empirical evidence of the underlying factor structure of a scale (Floyd & Widaman, 1995). Model feasibility was assessed using LISREL 8 (Jöreskog & Sörbom, 1993), with maximum likelihood estimation. A simple factor structure was modeled with eight variables loading on Factor 1 and nine variables loading on Factor 2 as in the original solution (Hare, 1991; Harpur et al., 1989). Because each measure of fit has limitations and no agreed methods for absolutely determining goodness of fit exist (Kline, 1998; Tabachnick & Fidell, 2001) the quality of fit was estimated using multiple indices.

Findings indicate that the two-factor model proposed by Hare (1991) did not fit our data well, $\chi^2(118, N = 98) = 561.8, p < .001$, root mean square error of approximation (RMSEA) = .16 (90% Ci = .14 - .17), standardized root mean square residual (SRMR) = .13, non-normed fit index (NNFI) = .50, comparative fit index (CFI) = .57. The estimates for those parameters between the items and the factors they were supposed to load on, ranged from 0.55 to 0.87, while the estimate for the covariance relationship between the two factors was 0.50. Post hoc model modifications were performed in an attempt to develop a better fitting model. On the basis of the Lagrange multiplier test, a path predicting Item 14 (*Impulsivity*) from Factor 1 (i.e., allowing Item 14 to load on Factor 2 as well as on Factor 1) was added. A chi-square difference test indicated that the model was significantly improved by addition of this path, change in $\chi^2(1, N = 98) = 20.9, p < .001$. The modified model also indicated a bad fit, $\chi^2(117, N = 98) = 540.9, p < .001$, RMSEA = .15 (90% CI = .14 - .17), SRMR = .12, NNFI = .52, CFI = .59. Examination of the modification indices suggested that allowing several errors of measurement to correlate would substantially improve the fit of the modified model. To this end, six errors of measure-

ment were allowed to correlate.² Although the fit of this revised model was significantly improved, change in χ^2 ($6, N = 98$) = 102.7, $p < .001$, the model still did not fit the data, χ^2 ($111, N = 98$) = 438.2, $p < .001$, RMSEA = .13 (90% CI = .11 - .15), SRMR = .11, NNFI = 61, CFI = .68.

As the two-factor model did not appear to fit our data, the 13-item model developed by Cooke and Michie (2001) was used in a CFA. A factor structure was modeled with four items loading on the Arrogant and deceitful interpersonal style factor, four items on the factor Deficient affective experience, and five items on the Impulsive and irresponsible behavioral style factor. Examination of the fit statistics revealed that the three-factor model provided a poor fit to the data, χ^2 ($62, N = 98$) = 433, $p < .001$, RMSEA = .21 (90% CI = .19 - .23), SRMR = .13, NNFI = .53, CFI = .62. Adding additional paths did not significantly improve the model. Also, allowing several errors of measurement to correlate did not result in a good fit of the model to the data.

Given the poor fit of our data to both the two-factor solution of Hare (1991) and the three-factor model developed by Cooke and Michie (2001) using CFA, we decided to investigate the factor structure of the Dutch language version of the PCL-R in the present sample by means of an exploratory principal components analysis (PCA). This analyses would also allow us to compare findings with other European research that examined the PCL-R factor structure with PCA (e.g., Hobson & Shine, 1998; Moltó et al., 2000). Analyses were conducted using SPSS 10.0 for Windows. The analysis revealed six components (factors) with eigenvalues greater than one, accounting for 68% of the total variance. The first two components accounted for 44% of the total variance. Visual inspection of the scree plot (Cattell, 1966) revealed a noticeable eigenvalue drop and leveling off after the second component. The unrotated first PC accounted for 30% of the variance in the PCL-R. The unrotated second PC accounted for 14% of the variance. The remaining factors accounted for 7%, 6%, 6%, and 5% of the variance,

respectively, suggesting that a two-factor solution is sufficient for the data of this sample.

We assessed the relative suitability of a two-factor solution using PC extraction and oblimin rotation to account for the fact that the factors produced by the PC analysis were correlated. The results obtained showed a correlation of .25 between the two factors. Table 4 presents the two-factor solution for the present study. Variables are ordered and grouped by size of loadings to facilitate interpretation. Factor 1 consisted of nine items with loadings of .4 or above. The highest loadings were found for Items 14 (*Impulsivity*; .84), 20 (*Criminal versatility*; .78), 3 (*Need for stimulation/proneness to boredom*; .72), 10 (*Poor behavioral controls*; .68), and 15 (*Irresponsibility*; .68). Factor 2 also consisted of nine items with loadings of .4 or above. The highest loadings were found for Items 7 (*Shallow affect*; .71), 8 (*Callous/lack of empathy*; .70), 2 (*Grandiose sense of self worth*, .69), 1 (*Glibness/superficial charm*; .69), and 5 (*Conning/manipulative*; .66). The Items 11 (*Promiscuous sexual behavior*) and 17 (*Many short-term marital relationships*) did not load on either factor.

DISCUSSION

The present findings provide initial evidence for the interrater reliability of the Dutch version of the PCL-R. In a sample of 60 forensic psychiatric patients, interrater reliabilities of the individual PCL-R items were demonstrated to be good to excellent. The single measure ICC for the PCL-R total score was .88; for Factor 1, it was .76, and for Factor 2, it was .83. The high levels of reliability found in the present study are consistent with those documented by other researchers. Indeed, cross-cultural research (e.g., Cooke, 1996, 1998; Moltó et al., 2000) supports the reliability of the Hare PCL-R. The reliabilities of the PCL-R items reported in many studies are generally high, although reliability coefficients vary per item, in part depending on the ease with which particular PCL-R items can be rated (Cooke, 1998). Hare et al. (1990) assessed interrater reliability using either a joint interview approach or the second rater observed a videotape of the interview in a large sample of prisoners and forensic psychiatric patients. For subsamples of subjects, ICC coefficients for the

² The errors of measurement which were allowed to correlate were chosen on the basis of which items shared content beyond that related to the underlying construct, were as follows: Items 1-2, 2-15, 6-16, 10-14, and 12-18.

Table 4

PCL-R two factor oblique solution (Delta = 0; pattern matrix) for a Dutch sample of 98 male forensic psychiatric patients

Item number and description		PCL-R	
		Factor 1	Factor 2
14. Impulsivity	(2)	.84	-.09
20. Criminal versatility	—	.78	.02
3. Need for stimulation/proneness to boredom	(2)	.72	.09
10. Poor behavioral controls	(2)	.68	-.04
15. Irresponsibility	(2)	.68	.09
12. Early behavior problems	(2)	.63	-.17
18. Juvenile delinquency	(2)	.58	-.06
9. Parasitic lifestyle	(2)	.54	.37
19. Revocation of conditional release	(2)	.52	-.08
7. Shallow affect	(1)	-.25	.71
8. Callous/lack of empathy	(1)	-.02	.70
2. Grandiose sense of self worth	(1)	-.09	.69
1. Glibness/superficial charm	(1)	.01	.69
5. Conning/manipulative	(1)	.10	.66
6. Lack of remorse or guilt	(1)	.31	.58
4. Pathological lying	(1)	.21	.58
16. Failure to accept responsibility	(1)	.24	.52
13. Lack of realistic, long-term goals	(2)	.38	.46
17. Many short-term marital relationships	—	.35	.17
11. Promiscuous sexual behavior	—	-.06	.32
Initial eigenvalues		6.0	2.7
Percentage of variance		30.1	13.6

Note. PCL-R = Psychopathy Checklist-Revised. This solution was obtained by specifying a two-factor solution. Decimal points are omitted and loadings > 0.4 are in bold; factor denotation from Hare's factor solution as described in the PCL-R manual are in parentheses. A '—' indicates that the item does not load on either factor.

PCL-R total score ranged from .78 to .94 ($M = .86$) for a single rating. In addition, Moltó et al. (2000), using a joint interview approach to assess interrater reliability of the PCL-R in 49 adult male Spanish prisoners, reported ICC coefficients ranging from .87 to .96 for a single rating using a one-way random effects model. Note that we had the disposal of *three* independent PCL-R ratings per patient, which is an extremely thorough examination of interrater reliability, far more thorough than, for example, the often practiced joint-interview approach (Zimmerman,

1994). The categorical diagnosis of PCL-R psychopathy was also reliable (weighted average $k = .63$, for simultaneous comparison of three raters). There were considerable differences between the raters with regard to clinical experience, training and background, but this did not result in significant differences in diagnostic reliability, demonstrated by the fact that there was no rater effect.

The Dutch language PCL-R has excellent internal consistency (Cronbach's $\alpha = .88$ for the PCL-R total score). This figure is comparable to those

obtained in the standardization samples (Hare, 1991). Also, we found adequate item-total correlations. Except for Items 6 (*Lack of remorse or guilt*), 11 (*Promiscuous sexual behavior*), and 12 (*Early behavioral problems*), the corrected item-total correlations were $\geq .35$, indicating they all contributed significantly to the PCL-R total score. Both internal consistency and item-total correlations were quite similar to the findings reported by other scholars, for instance, Hare et al. (1990; Cronbach's $\alpha = .88$; mean inter-item correlation = .27) and Moltó et al. (2000; Cronbach's $\alpha = .85$; mean inter-item correlation = .22).

With regard to a possible source effect (interview versus video) in rating the PCL-R, it was hypothesized that (particularly) PCL-R Factor 1 items would be susceptible to such an effect. However, we did not find a significant effect of information source on ratings of individual PCL-R items or the total score. We believe that the absence of a difference between interviewer and videotape ratings is most likely due to the fact that the information required to score the PCL-R depends as much on the extensive collateral information available as on the information provided by the interview. It may be that impressions based on real-life interviews do in fact differ from those based on videotaped interviews, because the latter do not include face-to-face interaction. These impressions may influence judgments, particularly on the "soft" Factor 1 items of the PCL-R (Cooke, 1995, p. 111). Thus, variability between raters in their impressions is a possible source of difference, and may negatively affect interrater reliability. Indeed, ICCs of PCL-R Factor 1 items were the lowest of the individual PCL-R items, especially in comparison to several behavioral items (e.g., early behavior problem, juvenile delinquency, criminal versatility). However, many PCL-R items are largely scored on the basis of collateral information (psychiatric or psychological evaluations, and police files), which might 'overrule' the impressions gathered from the interview (either interview or video). This might explain the large overall degree of convergence between video and interviewer ratings, especially because the available file information was quite extensive in this study. Our findings corroborate those of Grann et al. (1998), who demonstrated good reliability between independent clinical PCL-R ratings (based on interview and file information) and

retrospective file-only ratings, in part because the files were of good quality.

The finding that the PCL-R can be scored reliably from a videotaped interview (in combination with collateral information) has some practical implications. If, due to limited resources in a particular (forensic) setting, it is not possible to get two independent PCL-R ratings, a trained rater from a different facility could score the PCL-R on the basis of a videotaped interview, in order to provide the necessary second rating. Furthermore, videotaped interviews could be used in court cases, for example when a second opinion is asked for.

Confirmatory factor analysis showed that both Hare's two-factor model and the three-factor model identified by Cooke and Michie (2001) did not fit our data. Although this study was not intended as a model modification study, we conducted supplementary analyses in an attempt to develop a better fitting model. However, allowing items to load on more than one factor or allowing a number of errors of measurement to correlate did not significantly improve the fit of the models tested. It should be noted that it is difficult to obtain acceptable confirmatory solutions in cases of violation of assumptions, such as small sample size and noninterval scaling of items (Floyd & Widaman, 1995; Hu & Bentler, 1998). Our sample was relatively small, although there is no generally accepted guideline for sample size in case of CFA (Floyd & Widaman, 1995; see also below). Ideally, CFA is performed on interval or quasi-interval scales, for instance 5- or 7-point Likert scales; the PCL-R's 3-point scale does not meet this ideal.

The subsequent exploratory factor analysis yielded an oblique two-factor structure, accounting for 44% of the variance. Our first factor appeared to be similar to Hare's Factor 2. However, some notable differences were found. First, Item 20 (*Criminal versatility*) loaded high on our first factor but did not load on Hare's Factor 2. Second, Item 13 (*Lack of realistic, long-term plans*) did not load on the antisocial lifestyle factor in the present study. Our second factor included the eight items of Hare's Factor 1 plus Item 13 (*Lack of realistic long-term goals*). The Items 11 (*Promiscuous sexual behavior*) and 17 (*Many short-term marital relationships*) did not to load on either factor, as was the case in Hare's two-factor model.

Even though the use of confirmatory factor analytic techniques did not result in a replication of Hare's original two-factor model, our exploratory factor analysis pointed—at least to a certain extent—to a resemblance between our two-factor solution and Hare's. Several European scholars, employing exclusively exploratory factor analytic techniques, have claimed to find support for Hare's two-factor solution in their data, although some notable differences with the Hare solution exist. Hobson and Shine (1998), for example, examined the PCL-R factor structure in a sample of 104 inmates admitted to Grendon therapeutic prison. Similar to the present findings, and in contrast with Hare's two-factor solution, they found that Item 20 (*Criminal versatility*) loaded high on the antisocial lifestyle factor. Another difference was that Item 11 (*Promiscuous sexual behavior*) loaded on Factor 1 (*Selfish, callous and remorseless use of others*), while this item is not included in the Hare model. Likewise, Moltó et al. (2000), in a sample of 117 Spanish male prison inmates, found that Item 11 loaded significantly on (Hare's) Factor 1. The picture that emerges from these three European studies is that each study on its own did not provide unequivocal support for Hare's two-factor model. Pooling of these and other European data, to increase sample size, would seem worthwhile, to further examine such fundamental questions as: (1) Which model (the two-factor or the three-factor model) fits the European PCL-R data better, and (2) What is the role of cultural differences (cf. Cooke, 1996, 1998; Cooke & Michie, 2001) in the expression of psychopathic traits?

With regard to ethnic group differences in PCL-R scores, the present study found no difference between White and other subjects. Kosson et al. (1990), using a sample of Black and White inmates, found some race-related differences in PCL-R scores. As Brandt et al. (1997) noted correctly, these differences were quite selective and have not been observed in other studies. Also, Cooke et al. (2001), using IRT methods to analyze PCL-R ratings of Caucasian and African-American adult male offenders, found no evidence of racial differences.

A few methodological limitations of the present study should be mentioned. First, interrater reliability data were available for 60 patients, which may be a relatively small sample size. A larger sample might have increased variability and therefore reliability

estimates. At the very least, a larger sample would have resulted in more stable estimates. Second, one may argue that the high levels of interrater reliability found are due to a possible training effect caused by the PCL-R consensus meetings held after the three raters had independently scored the PCL-R for a particular patient. If this were the case, one would expect less divergence (higher levels of reliability) in the second half than in the first half of the series of 60 cases. However, single measure ICC values for the adjusted sum PCL-R total score were .89 for the first 20 of the 60 PCL-R ratings and .90 for the last 20 of the series. A Z test indicated that there was no significant difference between these correlations. Thus, there is no indication for a training effect.

A further limitation includes the small sample size for the study of the factor structure of the PCL-R in relation to the number of variables. However, guidelines for sample size have always been varying, the general rule of thumb being "the more subjects, the better" (Floyd & Widaman, 1995, p. 289). Streiner (1994), for example, recommends adequate solutions would be obtained with five subjects per variable as long as there were about 100 subjects in the sample. Tabachnick and Fidell (2001) suggest using at least 300 cases for factor analysis, which can be problematic in practice. In our hospital, for example, it would take about 10 to 12 years to get 300 newly admitted patients.

To summarize, it can be concluded that the Dutch language version of the PCL-R is a reliable instrument for use with forensic psychiatric patients and that the PCL-R can be applied in the forensic psychiatric population in the Netherlands. The current study did not confirm Hare's two-factor structure nor the three-factor model identified by Cooke and Michie (2001). Exploratory principal components analysis using oblique rotation, however, identified two main factors which were, to a certain extent, comparable to those obtained by Hare (1991) in the standardization samples and in other Western European samples. As the use of the PCL-R is likely to increase in Dutch forensic psychiatric hospitals in the future, there is a need for normative Dutch data and validity research. More research with different, larger samples (e.g., female forensic psychiatric patients, prisoners) is needed to further support the scientific status of the instrument for the Dutch forensic field.

REFERENCES

- Alterman, A.I., Cacciola, J.S., & Rutherford, M.J. (1993). Reliability of the Revised Psychopathy Checklist in substance abuse patients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *5*, 442-448.
- American Psychiatric Association (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.
- American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.
- Ben-Porath, Y.S. (1990). Cross-cultural assessment of personality: The case for replicatory factor analysis. In J.N. Butcher & C.D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 8, pp. 27-48). Hillsdale, NJ: Erlbaum.
- Brandt, J.R., Kennedy, W.A., Patrick, C.J., & Curtin, J.J. (1997). Assessment of psychopathy in a population of incarcerated adolescent offenders. *Psychological Assessment*, *9*, 429-435.
- Brink, van den W., & Jong, de C.A.J. (1992). *De Nederlandse versie van het Structured Interview for DSM-III-R Personality Disorders* [The Dutch version of the Structured interview for DSM-III Personality Disorders]. Eindhoven, the Netherlands: Instituut voor Verslavingszorg Brabant
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.
- Cleckley, H. (1941). *The mask of sanity*. St. Louis, MO: Mosby.
- Cleckley, H. (1976). *The mask of sanity* (5th ed.). St. Louis, MO: Mosby.
- Cohen, J. (1980). A coefficient of agreement for nominal scales. *Education and Psychological Measurement*, *1*, 37-46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cooke, D.J. (1995). Psychopathic disturbance in the Scottish prison population: The cross-cultural generalizability of the Hare Psychopathy Checklist. *Psychology, Crime, and Law*, *2*, 101-108.
- Cooke, D.J. (1996). Psychopathic personality in different cultures: What do we know? What do we need to find out? *Journal of Personality Disorders*, *10*, 23-40.
- Cooke, D.J. (1998). Psychopathy across cultures. In D.J. Cooke, A.E. Forth, & R.D. Hare (Eds.), *Psychopathy: Theory, research, and implications for society* (pp. 13-45). Dordrecht, the Netherlands: Kluwer.
- Cooke, D.J., & Michie, C. (1997). An item response theory evaluation of Hare's Psychopathy Checklist. *Psychological Assessment*, *9*, 2-13.
- Cooke, D.J., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, *13*, 171-188.
- Cooke, D.J., Kosson, D., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item and test generalizability of the Psychopathy Checklist-Revised (PCL-R) in Caucasian and African-American participants. *Psychological Assessment*, *13*, 531-542.
- Cooke, D.J., & Michie, C., Hart, S.D., & Hare, R.D. (1999). Evaluating the Screening Version of the Hare Psychopathy Checklist—Revised: An item response theory analysis. *Psychological Assessment*, *11*, 3-13.
- Craft, M.J. (1965). *Ten studies into psychopathic personality*. Bristol: John Wright.
- Darke, S., Kaye, S., Finlay-Jones, & Hall, W. (1998). Factor structure of psychopathy among methadone maintenance patients. *Journal of Personality Disorders*, *12*, 162-171.
- Dolan, B., & Coid, J. (1993). *Psychopathic and antisocial personality disorders: Treatment and research issues*. London: Gaskell.
- Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Floyd, F.J., & Widaman, K.F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286-299.
- Forth, A.E., Brown, S.L., Hart, S.D., & Hare, R.D. (1996). The assessment of psychopathy in male and female non-criminals: Reliability and validity. *Personality and Individual Differences*, *20*, 531-543.
- Forth, A.E., & Burke, E. (1998). Psychopathy in adolescence: Assessment, violence, and developmental precursors. In D.J. Cooke, A.E. Forth, & R.D. Hare (Eds.), *Psychopathy: Theory, research, and implications for society* (pp. 205-229). Dordrecht, the Netherlands: Kluwer.
- Forth, A.E., & Mailloux, D.L. (2000). Psychopathy in youth: What do we know? In C.B. Gacono (Ed.), *The clinical and forensic assessment of psychopathy: A practitioner's guide* (pp. 25-54). Mahwah, NJ: Erlbaum.
- Grann, M., Långström, N., Tengström, A., & Stålenheim, E.G. (1998). Reliability of file-based retrospective ratings of psychopathy with the PCL-R. *Journal of Personality Assessment*, *70*, 416-426.
- Green, S.B., Lissitz, R.W., & Mulaik, S.A. (1977). Limitations of coefficient alpha as an index of unidimensionality. *Educational and Psychological Measurement*, *37*, 827-838.
- Haapasalo, J., & Pulkkinen, L. (1992). The Psychopathy Checklist and non-violent offender groups. *Criminal Behaviour and Mental Health*, *2*, 315-328.
- Hare, R.D. (1970). *Psychopathy: Theory and research*. New York: Wiley.
- Hare, R.D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, *1*, 111-119.
- Hare, R.D. (1991). *The Hare Psychopathy Checklist-Revised*. Toronto, Ontario: Multi-Health Systems.
- Hare, R.D. (1996). Psychopathy: A clinical construct whose time has come. *Criminal Justice and Behavior*, *23*, 25-54.
- Hare, R.D. (1998). The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology*, *3*, 99-119.
- Hare, R.D., Harpur, T.J., Hakstian, A.R., Forth, A.E., Hart, S.D., & Newman, J.P. (1990). The Revised Psychopathy Checklist: Descriptive statistics, reliability, and factor structure. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 338-341.

- Harpur, T.J., Hakstian, R., & Hare, R.D. (1988). Factor structure of the Psychopathy Checklist. *Journal of Consulting and Clinical Psychology, 56*, 741-747.
- Harpur, T.J., Hare, R.D., & Hakstian, R. (1989). Two-factor conceptualization of psychopathy: Construct validity and assessment implications. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 6-17.
- Hart, S.D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121-137.
- Hart, S.D., & Hare, R.D. (1989). Discriminant validity of the Psychopathy Checklist in a forensic psychiatric population. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*, 211-218.
- Hart, S.D., & Hare, R.D. (1997). Psychopathy: Assessment and association with criminal conduct. In D.M. Stoff, J. Breiling, & J. Maser (Eds.), *Handbook of antisocial behavior* (pp. 22-35). New York: Wiley.
- Hemphill, J.F., Hare, R.D., & Wong, S. (1998). Psychopathy and recidivism: A review. *Legal and Criminological Psychology, 3*, 139-170.
- Hobson, J., & Shine, J. (1998). Measurement of psychopathy in a UK prison population referred for long-term psychotherapy. *British Journal of Criminology, 38*, 504-515.
- Jong, de C.A.J., Derks, F.C.H., Oel, van C.J., & Rinne, Th. (1997). *Gestructureerd Interview voor de DSM-IV Persoonlijkheidsstoornissen* [Structured Interview for DSM-IV Personality Disorders]. Sint Oedenrode, the Netherlands: Stichting Verslavingszorg Oost Brabant.
- Jöreskog, K.G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software.
- Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.
- Klinterberg, B. af, Humble, K., & Schalling, D. (1992). Personality and psychopathy of males with a history of early criminal behaviour. *European Journal of Personality, 6*, 245-266.
- Kosson, D.S., Smith, S.S., & Newman, J.P. (1990). Evaluating the construct validity of psychopathy in Black and White male inmates: Three preliminary studies. *Journal of Abnormal Psychology, 3*, 250-259.
- Marle, van H.J.C. (2002). The Dutch Entrustment Act (TBS): Its principles and innovations. *International Journal of Forensic Mental Health, 1*, 83-92.
- McCord, W., & McCord, J. (1964). *The psychopath: An essay on the criminal mind*. Princeton, NJ: Van Nostrand.
- McDermott, P.A., Alterman, A.I., Cacciola, J.S., Rutherford, M.J., Newman, J.P., & Mulholland, J.P. (2000). Generality of Psychopathy Checklist-Revised factors over prisoners and substance-dependent patients. *Journal of Consulting and Clinical Psychology, 68*, 181-186.
- McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.
- Moltó, J., Poy, R., & Torrubia, R. (2000). Standardization of the Hare Psychopathy Checklist-Revised in a Spanish prison sample. *Journal of Personality Disorders, 14*, 84-96.
- O'Kane, A., Fawcett, D., & Blackburn, R. (1996). Psychopathy and moral reasoning: A comparison of two classifications. *Personality and Individual Differences, 20*, 505-514.
- Pfohl, B., Blum, N., & Zimmerman, M. (1995). *Structured Interview for DSM-IV Personality (DSM-IV)*. Iowa City: Department of Psychiatry, University of Iowa.
- Pfohl, B., Blum, N., Zimmerman, M., & Stangl, D. (1989). *Structured Interview for DSM-III-R Personality (SIDP-R)*. Iowa City: Department of Psychiatry, University of Iowa.
- Raine, A. (1985). A psychometric assessment of Hare's checklist for psychopathy on an English prison population. *British Journal of Clinical Psychology, 24*, 247-258.
- Rasmussen K., Storsæter O., & Levander S. (1999). Personality disorders, psychopathy, and crime in a Norwegian prison population. *International Journal of Law and Psychiatry, 22*, 91-97.
- Ruiter, de C., & Hildebrand, M. (in press). The dual nature of forensic psychiatric practice: Risk assessment and management under the Dutch TBS-order. In P.J. van Koppen & S.D. Penrod (Eds.), *Adversarial versus inquisitorial justice: Psychological perspectives on criminal justice systems*. New York: Kluwer/Plenum.
- Rutherford, M.J., Cacciola, J.S., Alterman, A.I., & McKay, J.R. (1996). Reliability and validity of the Revised Psychopathy Checklist in women methadone patients. *Assessment, 3*, 43-54.
- Salekin, R.T., Rogers, R., & Sewell, K.W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist-Revised: Predictive validity of dangerousness. *Clinical Psychology: Science and Practice, 3*, 203-215.
- Salekin, R.T., Rogers, R., & Sewell, K.W. (1997). Construct validity of psychopathy in a female offender sample: A multitrait-multimethod evaluation. *Journal of Abnormal Psychology, 106*, 576-585.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
- Shrout, P.E., Spitzer, R.L., & Fleiss, J.L. (1987). Quantifications in psychiatric diagnosis revisited. *Archives of General Psychiatry, 44*, 172-177.
- Silverthorn, P., & Frick, P.J. (1997). Developmental pathways to antisocial behavior: The delayed-onset pathways in girls. *Development and Psychopathology, 11*, 101-126.
- Streiner, D.L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry, 39*, 135-140.
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.
- Templeman, R., & Wong, S. (1984). Determining the factor structure of the Psychopathy Checklist: A converging approach. *Multivariate Experimental Clinical Research, 10*, 157-166.
- Tengström, A., Grann, M., Långström, N., & Kullgren, G. (2000). Psychopathy (PCL-R) as a predictor of violent recidivism among criminal offenders with schizophrenia. *Law and Human Behavior, 24*, 45-58.

- Vertommen, H., Verheul, R. Ruiter de, C., & Hildebrand, M. (2002). *Handleiding bij de herziene versie van Hare's Psychopathie Checklist*. [Manual of the Dutch version of the revised version of Hare's Psychopathy Checklist]. Lisse: Swets Test Publishers.
- Vijver van de, F.J.R., & Leung, K. (1997). *Methods and data-analysis for cross-cultural research*. Thousands Oaks, CA: Sage.
- Windle, M., & Dumenci, L. (1999). The factorial structure and construct validity of the Psychopathy Checklist-Revised (PCL-R) among alcoholic inpatients. *Structural Equation Modeling, 6*, 372-393.
- Zimmerman, M. (1994). Diagnosing personality disorders. A review of issues and research methods. *Archives of General Psychiatry, 51*, 225-245.