Psychological Assessment
2011, Vol. 23, No. 4, 937–944

1040-3590/11/$12.00 DOI: 10.1037/a0024018

The Novaco Anger Scale–Provocation Inventory (1994 Version) in Dutch Forensic Psychiatric Patients

Ruud H. J. Hornsveld
Forensic Psychiatric Centre Kijvelanden and Erasmus Medical Centre

Peter Muris
Erasmus University Rotterdam

Floris W. Kraaimaat
Radboud University Nijmegen

We examined the psychometric properties of the Novaco Anger Scale–Provocation Inventory (NAS–PI, 1994 version) in Dutch violent forensic psychiatric patients and secondary vocational students. A confirmatory factor analysis of the subscale structure of the NAS was carried out, reliability was investigated, and relations were calculated between NAS–PI scores and other measures of personality traits and problem behaviors. The 3-subscale structure of the original NAS could not be confirmed. However, the internal consistency of the NAS and the PI was excellent, and the test–retest reliability of the NAS was good. The validity of the NAS and the PI was supported by a meaningful pattern of correlations with alternative measures of anger and personality traits. Forensic psychiatric outpatients displayed higher NAS scores than secondary vocational students, but inpatients scored even lower than this nonclinical control group. Our preliminary conclusion is that the NAS–PI is a valuable instrument for the assessment of anger in Dutch violent forensic psychiatric patients.

Keywords: Novaco Anger Scale–Provocation Inventory (1994 version), psychometric qualities, forensic psychiatric patients

In the Netherlands, forensic psychiatric patients are persons for whom the court has established a connection between a psychiatric disorder on the one hand and their felony on the other hand. Rulings are based on the evaluations of a psychiatrist or psychologist. Without treatment, recidivism is deemed probable. One of the main targets in treatment programs for this population should be the enhancement of the patient's skills to effectively manage feelings of anger, which will eventually result in a reduction of aggressive behavior (e.g., Hornsveld, Nijman, & Kraaimaat, 2008; Novaco, Ramm, & Black, 2001). Reliable and valid questionnaires for assessment of anger are urgently needed for the evaluation of such treatment programs. An instrument that might be suitable for this purpose is the original version of Novaco Anger Scale–Provocation Inventory (Novaco, 1994). However, until now the psychometric properties of the original NAS–PI have not been investigated in Dutch populations. Such an investigation seems necessary because the psychometric evaluation of other questionnaires used to assess related constructs such as aggression has yielded unsatisfactory findings in Dutch forensic psychiatric inpatients (Hornsveld, Muris, Kraaimaat, & Meesters, 2009).

The NAS–PI was initially developed in conjunction with the MacArthur Foundation Network on Mental Health and Law and is based on the theoretical model that anger is an emotional state, entailing heightened physiological arousal and cognitions of antagonism that can be predictive of violence (Novaco, 1994). The original NAS–PI (1994 version) is composed of two parts and contains 48 items that represent anger disposition on Cognitive, Arousal, and Behavioral subscales, which sum to a NAS total score. The PI is composed of 25 items that focus on situations that lead to anger. For an initial investigation of the psychometrics of the NAS–PI, Novaco (1994) administered the questionnaire to 142 patients in three California state hospitals, of whom 126 completed the form on a second occasion, some 2 weeks later. The internal consistency of the NAS and the PI turned out to be excellent with alphas of .95 and .95, respectively, whereas the test–retest reliability was also good with correlations of .84 and .86, respectively. Furthermore, the validity of the NAS could be demonstrated by means of correlations of .82 with the Buss–Durkee Hostility Inventory (BDHI; Buss & Durkee, 1957) and .84 with the trait part of the Spielberger State–Trait Anger Scale (STAS; Spielberger, 1980). Prospective analyses were then conducted with another sample of 158 patients (including 68 from the initial phase) that demonstrated good predictive validity for the NAS–PI with the Spielberger State Anger subscale and staff ratings of anger as the criterion variables.

Since Novaco's (1994) investigation of the original version of the NAS–PI, the psychometric properties of successive versions (an intermediate 1998 version and the formally published 2003

This article was published Online First June 13, 2011.

Ruud H. J. Hornsveld, Forensic Psychiatric Centre Kijvelanden, Poortugaal, the Netherlands, and Department of Psychiatry, Erasmus Medical Centre, Rotterdam, the Netherlands; Peter Muris, Department of Clinical Psychology, Erasmus University Rotterdam, Rotterdam, the Netherlands; Floris W. Kraaimaat, Radboud University Nijmegen, Nijmegen, the Netherlands.

We thank Ray W. Novaco for his valuable comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Ruud H. J. Hornsveld, Lange Dreef 52, La Rijswijk 2285, the Netherlands. E-mail: r.hornsveld@tiscali.nl

version) have been examined in various populations. An example of a study on the reliability and validity of the NAS–PI (1994 version) is the one by Mills, Kroner, and Forth (1998) in a group of 102 violent offenders (age 18–55 years, $M = 28$ years, $SD = 7.9$) and 102 nonviolent offenders (age 19–69 years, $M = 33$ years, $SD = 10.5$). The internal consistency of the NAS and the PI turned out to be .95 and .96, respectively, in the nonviolent group, and .94 and .95, respectively, in the violent group. Test–retest reliability (4-week interval), which was only examined in the nonviolent group, was .89 for the NAS and .85 for the PI. Validity could be demonstrated in the violent group, in which the NAS and the PI correlated .79 and .68, respectively, with the Aggression Questionnaire total score (Buss & Perry, 1992). Further, NAS scores correlated .46 with clinical ratings of anger, which were derived from interviews and file information. Surprisingly, the violent offenders had significantly lower scores on the NAS–PI than did the nonviolent offenders.

The reliability and validity of the original NAS-PI were also studied by Jones, Thomas-Peter, and Trout (1999). These researchers provided no data on the internal consistency of the NAS–PI for a group of 58 men (mean age = 32.47 years, $SD = 8.32$), who were referred to a clinical facility for anger management training; however, they reported Cronbach's alphas of .92 for both the NAS and the PI in a group of 79 male and 133 female employees of National Health Services (mean age = 36.57 years, $SD = 9.93$). Furthermore, the clinical group scored significantly higher on the NAS–PI than the nonclinical group.

A Swedish translation of the 1998 version of the NAS-PI was evaluated by Lindqvist, Dåderman, and Hellström (2005) in a group of 95 male violent prisoners (some of them with diagnosed personality disorders), ranging in age between 18 and 67 years. In this version of the NAS, four Cognitive subscale items concerned with "attentional focus" were replaced by four new "justification" items and a 12-item Anger Regulation scale was added. The questionnaire appeared to have good internal consistency, with subscale alphas ranging between .78 and .91. Convergent validity was excellent as demonstrated by high correlations between NAS total scores and scores on Swedish versions of the Aggression Questionnaire (AQ; Buss & Perry, 1992; $r = .86$) and the Trait Anger scale of the State–Trait Anger Expression-2 (STAXI-2; Spielberger, 1999; $r = .79$).

Using the 1994 version of the NAS-PI, Baker, Van Hasselt, and Sellers (2008) examined a large group of incarcerated offenders (638 men and 349 women), who were detained in Florida jails for domestic violence or drug-related offenses. Internal consistency coefficients for the NAS and the PI were, respectively, .93 and .92 for men and .89 and .87 for women. Convergent validity was also satisfactory, as shown by correlations of .69 between both the NAS and the PI and the STAS Trait Anger and substantial positive correlations with various BDHI (Buss & Durkee, 1957) subscales (e.g., .58 between the NAS and BDHI Irritability).

In a study on anger and aggression in male offenders with developmental abilities, Novaco and Taylor (2004) used a modified version of the NAS–PI. For the NAS, the modification consisted of the rewording or simplification of 17 items, and the previously described replacement of four Cognitive subscale items. For the PI, 17 of the 25 items were modified to improve accessibility and the relevance for patients living in a highly structured forensic environment. The internal consistency appeared to be .92 for the modified versions of both the NAS and the PI. Test–retest correlations were found to be .52 and .57 for the NAS and PI, respectively. The total score on the modified NAS was significantly related to the scores on alternative self-report questionnaires for measuring anger and to records of assaultive behavior in the hospital.

Research on the factor structure of the NAS–PI (1994 version) was carried out by Jones, Thomas-Peter, and Gangstad (2003) in a group of 354 outpatients referred for an anger management training (mean age: 34.75 years, $SD = 10.41$) and a group of 212 health care employees (mean age: 36.57 years, $SD = 9.93$). Exploratory factor analysis of the data from the combined clinical and nonclinical group ($N = 566$) yielded three factors (Retaliatory Hostility, Vigilant Arousal, and Indirect Aggression) that were not consistent with the domains originally defined by Novaco (1994). The authors suggested that this result was probably due to the 3-point scale, which "allows for little response variation and may therefore increase measurement error" (p. 436).

Another exploratory factor analysis of the 1994 version of the NAS was published in the manual for the final version of the NAS–PI (Novaco, 2003). This analysis was performed in a group of 1,101 civil commitment inpatients with severe mental disorders, who participated in the MacArthur Violence Risk Project (Monahan et al., 2001). Results did not yield a factor structure that was in accordance with the original subscales of the instrument. A similar result was obtained in the 2003 NAS standardization sample of 1,546 nonclinical, multiethnic individuals (between the ages of 9 and 84 years) in four geographic regions of the United States (Novaco, 2003) and in a study of Lindqvist, Dåderman, and Hellström (2003), who investigated the factor structure of the NAS–PI (1998 version) in a group of 100 undergraduate male students (mean age: 33.2 years, $SD = 12.5$). However, it should be noted that the sample size in the latter study was too small to draw definitive conclusions about the factor structure of the NAS–PI.

To summarize, studies that have examined the psychometric qualities of the NAS–PI in violent offenders or people with anger management problems generally have provided support for its internal consistency, test–retest reliability, and concurrent validity. In nonoffender populations, no convincing support for the three-factor structure of the NAS emerged. Until now, the factor structure of the NAS, representing the three subscales (Cognitive, Arousal, and Behavioral), has not been analyzed in offender populations. In addition, more empirical data are required about the usefulness of the original version of the NAS–PI (1994 version) to discriminate between aggressive and nonaggressive populations.

We conducted the present study to evaluate the psychometric qualities of the 1994 version of the NAS–PI in violent forensic inpatients and outpatients and in secondary vocational students. A confirmatory factor analytic approach was employed to test Novaco's original three-subscale model of the NAS. In addition, we investigated associations between the NAS–PI and scores on other scales assessing anger or related constructs (i.e., hostility and aggression) and basic personality traits (i.e., neuroticism and agreeableness). The former trait was expected to correlate positively and the latter negatively with NAS–PI scores. We further examined the discriminant validity of the NAS–PI by relating scores on this scale to the Psychopathy Checklist–Revised (PCL–R; Hare, 1991) in the patient samples. As several researchers have observed a reversed relationship between anger and

callous or unemotional traits (e.g., Frick, Cornell, Barry, Bodin, & Dane, 2003), we also investigated the link between NAS–PI scores and the Interpersonal facet scores as measured by the PCL–R.

Method

Participants

A clinical sample of 142 male inpatients was recruited from six forensic psychiatric institutions situated throughout the Netherlands. Patients were detained under hospital order for a serious violent offense (e.g., murder, manslaughter, aggravated assault, or forcible rape), which is punishable with a minimum of 4 years of imprisonment. Their average age was 33.16 years (SD = 7.65; range = 21–56 years), and 28.87% belonged to an ethnic minority. The primary diagnosis was an antisocial personality disorder on Axis II or a (chronic) psychotic disorder on Axis I in combination with an antisocial personality disorder on Axis II (*Diagnostic and Statistical Manual of Mental Disorders* [4th ed.; *DSM–IV*]; American Psychiatric Association, 1994). The chronic psychiatric condition of the psychotic patients had been stabilized to the extent that their antisocial personality disorder was most prominent.

The 194 male 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 22.79 years (SD = 8.93; range = 16–56 years), and 51.54% of them belonged to an ethnic minority. The outpatients had a conduct disorder as primary diagnosis on Axis I or, if they were 18 years or older, a main diagnosis of antisocial personality disorder on Axis II (*DSM–IV*; American Psychiatric Association, 1994).

The 320 students (160 men and 160 women) followed secondary vocational education in Rotterdam, the Netherlands. The mean age was 17.35 years (SD = 1.08; range = 16–21 years) for the male students and 18.36 years (SD = 1.78; range 16–27 years) for the female students. Among the male students, 50.63% were nonethnic and among the female students 47.50%. In the Netherlands, after students complete elementary school, they receive secondary professional education consisting of instruction at the lowest level for trades such as carpenter, housepainter, electrician, administrative assistant, or hairdresser.

Ethnic patients and students were born in the Netherlands and had parents who were both of original Dutch descent. Nonethnic participants were also born in the Netherlands but had at least one parent who had immigrated from former Dutch colonies, such as Surinam or the Netherlands Antilles, or from countries around the Mediterranean Sea such as Turkey, Morocco, or the Cape Verde Islands.

Measures

The Novaco Anger Scale–Provocation Inventory (NAS-PI; Novaco, 1994). As described in the introduction, this NAS–PI consists of two parts. The NAS contains 48 items that intend to represent three interrelated factors: Cognitive (e.g., "I feel like I am getting a raw deal out of life"), Arousal (e.g., "I feel agitated and unable to relax"), and Behavioral (e.g., "When someone yells at me, I yell back at them"). Items are rated on a 3-point scale (1 = *never true*, 2 = *sometimes true*, 3 = *always true*). The PI consists of 25 items that refer to anger-eliciting situations, rated on a 4-point scale (1 = *not at all angry*, 2 = *a little angry*, 3 = *fairly angry*, 4 = *very angry*). The NAS–PI (1994 version) was translated from English to Dutch and backward.[1]

The Psychopathy Checklist–Revised (PCL-R; Hare, 1991). The PCL–R was employed for measuring psychopathy. The checklist consists of 20 items, which are rated on a 3-point scale: 0 = *does not apply*, 1 = *applies to some extent*, and 2 = *applies*. Vertommen, Verheul, De Ruiter, and Hildebrand (2002) found support for the reliability of the Dutch version of the PCL–R in a group of 1,192 inmates. Cronbach's alpha was .87, and the average interitem correlation was .25. Tentative evidence for the 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 Minnesota Multiphasic Personality Inventory (MMPI–2, Dutch version; Sloore, Derksen, Hellenbosch, & De Mey, 1993). In the present study, we used the total score as well as the four-factor structure as it has been proposed by Hare (2003), which implies the following facets: Interpersonal (e.g., grandiose self-worth), Affective (e.g., callousness and lack of empathy), Lifestyle (e.g., impulsivity), and Antisocial (e.g., juvenile delinquency). In the present study, this measure is used to show the discriminant validity of the NAS–PI.

The NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992; Dutch Version: Hoekstra, Ormel, & De Fruyt, 1996). NEO–FFI 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 5-point Likert scale, which ranges from *entirely disagree* to *entirely agree*. In a Dutch sample of 356 nonclinical adults, Cronbach's alphas were between .69 and .82 for various subscales. In a subgroup of 135 adults, test–retest reliability after 6 months appeared to be .82, .87, .81, .75, and .80, respectively (Hoekstra et al., 1996).

The Spielberger's (1980) State–Trait Anger Scale (STAS; Van der Ploeg, Defares, & Spielberger, 1982). This Trait Anger subscale was used as a concurrent measure of the general disposition to anger. Participants rate each item (e.g., "I am quick tempered") to indicate how they generally feel using a 4-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 et al. (1982) found that internal consistency of the Trait Anger subscale was .78, and in a subgroup of 70 students, a test–retest reliability of .78 was documented. The validity of the Trait Anger subscale was also proven to be satisfactory (Van der Ploeg et al., 1982).

The Rosenzweig's (1978) Picture-Frustration Study–Adapted Version (PFS-AV; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007). We employed an adapted version of the PFS–AV for measuring hostility. Participants are asked to write down their reactions to 12 cartoon-like pictures and then are

[1] Novaco reviewed the backward translation and indicated that it was generally good for the NAS and very good for the PI. The translation procedure yielded sometimes minor differences in a few NAS items. Four items were somewhat problematic because the key elements appeared to be interpreted in a different way (Items 11, 25, 26, and 31), due to the fact that certain English expressions could not be translated exactly into the Dutch language.

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 in the blank text box that enters their mind. Answers are scored by an experienced and independent research assistant (psychologist) on a 7-point scale, which ranges from 1 = *not at all hostile* to 7 = *extremely hostile*. In a sample of 231 Dutch violent forensic psychiatric patients, the internal consistency, test–retest reliability, and interrater reliability of the PFS-AV appeared to be moderate to good ($\alpha = .76$, $r = .67$, and $r = .77$, successively). Furthermore, evidence was found for the validity of the test as scores correlated with indexes of agreeableness and aggressive behavior (Hornsveld et al., 2007).

The Aggression Questionnaire–Short Form (AQ–SF; Bryant & Smith, 2001; Dutch version: Hornsveld, Muris, Kraaimaat, & Meesters, 2009). This questionnaire is a shortened version of the Aggression Questionnaire of Buss and Perry (1992) with 12 items that can be allocated to four subscales, i.e., 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 5-point scale, which ranges from 1 = *entirely disagree* to 5 = *entirely agree*. In a sample of Dutch forensic psychiatric patients (men) and a sample of secondary vocational students (women and men), Hornsveld et al. (2009) found that the four-factor structure of the AQ–SF produced an acceptable fit. In a group of 208 violent forensic psychiatric outpatients, internal consistency for the AQ–SF total score and for the subscales turned out to be .88, .65, .74, .61, and .74, respectively. The test–retest reliability of the AQ–SF total score was modest but significant ($r = .38$) in a subsample of 90 outpatients, but the test–retest reliability of the AQ–SF subscale Physical Aggression was poor and not significant ($r = .13$). The validity of the AQ–SF could be demonstrated by meaningful correlations with alternative measures of aggression and personality (Hornsveld et al., 2009). In the present study, we only employed the scores of the total AQ–SF and the Anger subscale.

Procedure

The data from the in- and outpatients were obtained during an evaluation trial of the aggression control therapy (Hornsveld, Nijman, & Kraaimaat, 2008). Questionnaires were administered individually to the patients prior to the group therapy. One of the indication criteria for participation was sufficient command of the Dutch language through the spoken and written word. Questionnaires were completed under supervision of an experienced research assistant. When all items of the questionnaires were completed, patients received a fee of 5 euros in return for their participation. A subsample of 90 forensic psychiatric outpatients completed the NAS–PI during the intake interview, about 4 weeks before the start of the therapy, thus providing a test–retest in this subsample. During the period between intake and start of therapy, outpatients were supervised by a probation agency and received no specific treatment or training. The PI was only administered in subsamples of 97 inpatients and 80 outpatients.

PCL–R scores were assessed by certified clinical psychologists. For the outpatients, PCL–R scores were based on the structured intake interview and information from the extensive psychiatric and psychological evaluation on the basis of which the judge had decided on obligatory treatment. For the inpatients, PCL–R scores were assessed on the basis of a file study. Such files comprised detailed information about life history, committed offenses, and elaborated reports from psychiatrists or psychologists. These reports were often drawn up in a special forensic assessment center, in which the offender had to stay for observation by order of the court. Inpatients had significantly higher PCL–R total scores than did outpatients, $F(2, 333) = 6.50$, $p = .002$, while controlling for age. They had significantly higher scores on the PCL–R subscales Lifestyle, $F(2, 333) = 5.68$, $p = .004$, and Antisocial, $F(2, 333) = 36.88$, $p = .001$.

The secondary vocational students completed the set of questionnaires in their classrooms at school. Most students did not need more than 1 hr to do so. Completion was supervised by the first author and a research assistant. After the experimenters had checked for missing scores, the students received a fee of 10 euros in return for their participation.

Statistical Analysis

We performed a confirmative factor analysis on the NAS data using AMOS 16. The criteria of Schermelleh-Engel, Moosbrugger, and Müller (2003) were employed to determine the goodness of fit of the hypothesized three-factor model. These authors considered the following results as indications for a good fit: chi square/*df* $\leq$ 2, goodness-of-fit index (GFI) $\geq .95$, comparative fit index (CFI) $\geq .97$, and root-mean-square error of approximation (RMSEA) $\leq .05$. In their opinion, a model has an acceptable fit when: chi square/*df* = 23–, GFI = .90–.95, CFI = .95–.97, and RMSEA = .05–.08.

Internal consistency coefficients, mean interitem correlations, and mean item-scale correlations for the NAS–PI were calculated with SPSS Version 18.0. For comparisons between groups, we used *t* tests, while applying a Bonferroni correction for the number of comparisons that were made. As a consequence, for the comparisons between patients and students, alpha was .05: 8 = .006, between ethnic groups (i.e., ethnic vs. nonethnic): 05: 18 = .003, and between male and female students: .05: 4 = .013. When comparing between groups with a significantly difference in average age, we used analyses of covariance for control on this demographic variable.

Results

Factor Structure of the NAS

The confirmatory factor analysis was carried out in the patient and the student samples to test the three-subscale structure of the NAS (1994 version). In the in- and outpatient samples, the goodness-of-fit indices indicated that the structure of the NAS with three subscales did not provide a good fit for the data. Because of the relatively small sample sizes, we also calculated the goodness-of-fit indices for the combined patient group: chi square/*df* = 2.38, GFI = .73, CFI = .78, and RMSEA = .06. In the combined group

of students, comparable results were found: chi square/*df* = 1.93, GFI = .76, CFI = .73, and RMSEA = .05 (see Table 1).

Reliability of the NAS–PI

Internal consistency coefficients, mean interitem correlations, and mean item-scale correlations were calculated for the NAS, its subscales, and the PI in the three separate samples (i.e., inpatients, outpatients, students). For the NAS, Cronbach's alphas varied from .92 to .95, mean interitem correlations from .18 to .27, and mean item-scale correlations from .44 to .51. Internal consistency coefficients for the subscales were between .70 and .90 across the samples. Mean interitem correlations were between .13 and .35, and mean item-scale correlations were between .30 and .56. In the sample of outpatients, the test–retest reliability (4 weeks interval) of the NAS was .80 for the total score and ranged between .71 and .79 (all *p*s $< .01$) for various subscales. The internal consistency for the PI was .90 in the subsample of inpatients and .94 in the sample of outpatients. Mean interitems correlations were, respectively, .28 and .39, and mean item-scale correlations were .50 and .61 (see Table 2).

Comparison of NAS Scores Across Groups

For each of the samples, we compared the scores on the NAS–PI of ethnic individuals with those of nonethnic individuals. In the inpatient, outpatient, and female student groups, no significant differences were found, but ethnic male students appeared to score significantly higher on the NAS than the nonethnic male students, with an effect size of .23 for the NAS and between .09 and .28 for its subscales. No significant differences in NAS scores were found when male students were compared with female students, controlled for age (see Table 3).

The mean NAS scores of the inpatients and outpatients were compared with those of the male secondary vocational students. Results showed that inpatients scored significantly lower on the total NAS than male students, with effect sizes ranging from .31 to .68. However, outpatients did display significantly higher total scores on the NAS than the male students, with effect sizes between .10 and .20. Compared with the students, inpatients scored significantly lower on the Behavior subscale of the NAS. Outpatients had significantly higher scores on most subscales of the NAS than the students (see Table 4).

Table 1
Goodness-of-Fit Indices for Novaco Anger Scale (1994 Version) With Three Subscales in Inpatients (N = 142), Outpatients (N = 194), and Students (N = 320)

Group	*N*	χ^2/df	GFI	CFI	RMSEA
Inpatients	142	1.82	.63	.62	.08
Outpatients	194	1.92	.68	.75	.07
All patients	336	2.38	.73	.78	.06
Male students	160	1.52	.70	.70	.06
Female students	160	1.71	.67	.65	.07
All students	320	1.93	.76	.73	.05

Note. GFI = goodness-of-fit index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation.

Validity of the NAS-PI

We determined the convergent validity of the NAS-PI for the inpatients, outpatients, and students by computing correlations between the NAS and PI scores and various other measures. In the inpatient sample, a negative correlation was found between the NAS total score and the Interpersonal facet of the PCL–R, and a positive correlation between the NAS total score and the Antisocial facet of the PCL–R. In the outpatient sample, the total score of the NAS correlated positively with the Lifestyle facet of the PCL–R. For the other measures, the pattern of results was rather similar for students, outpatients, and inpatients. As could be expected, the NAS total score correlated positively with NEO-FFI Neuroticism, STAS Anger, AQ–SF Aggression, and PFS–AV Hostility, and negatively with NEO-FFI Agreeableness (see Table 5).

The PI score correlated negatively with the Affective facet of the PCL-R in the sample of outpatients and negatively with NEO-FFI Agreeableness in both patient samples. Positive correlations were found in the outpatient and inpatient sample between PI scores and NEO-FFI Neuroticism, STAS Anger, PFS-AV Hostility, and AQ-SF Aggression scores (see Table 5).

Discussion

Our factor analytic study of two groups of violent forensic psychiatric patients and two groups of secondary vocational students yielded no confirmation of the original three-subscale structure of the NAS. These findings suggest that a firm empirical basis of the division of the NAS in these three subscales is currently lacking. Note that even large-scale studies in the United States provided similar results with the 1994 as well as the 2003 version of the NAS. These results seem to indicate that the NAS subscales should only be used for clinical purposes in combination with other measures and clinical impressions.

On a more positive note, the internal consistency of the NAS–PI appeared to be excellent in the forensic psychiatric patient samples as well as in the student sample, whereas the test–retest reliability also proved to be good in the outpatient sample. Further, the concurrent validity of the scale could be demonstrated by significant and positive correlations with other measures of anger, hostility, and aggression. In addition, the NAS–PI correlated as expected positively with the personality trait of neuroticism but negatively with the trait of agreeableness.

The discriminant validity of the NAS–PI was further supported by the lack of significant correlations between the PCL–R total score, which was the case in both the inpatient and outpatient samples. However, a significantly negative correlation was found between the Interpersonal facet of the PCL–R and the NAS total score in the inpatient group, which is in accordance with the previously documented reversed relationship between anger and callous or unemotional traits (e.g., Frick et al., 2003). Unexpectedly, such a negative relationship was not noticed in the outpatient sample.

Comparisons of NAS–PI scores between ethnic and nonethnic patients and between ethnic and nonethnic female students did not reveal any significant differences. However, male ethnic students scored significantly higher on the NAS than the male nonethnic students, possibly because they feel more anger or are more honest

Table 2
Internal Consistency Coefficients (Cronbach's Alphas), Mean Interitem Correlations, and Mean Item-Scale Correlations of the Novaco Anger Scale–Provocation Inventory (1994 Version) for Inpatients (N = 142), Outpatients (N = 194), and Students (N = 320), and Test–Retest Reliability for Outpatients (N = 90)

	Internal consistency			Mean interitem correlations			Mean item-scale correlations			
Measure	Inpts	Outpts	Students	Inpts	Outpts	Students	Inpts	Outpts	Students	Test–retest reliability: Outpts
NAS										
Anger	.93	.95	.92	.21	.27	.18	.44	.51	.41	.80*
Cognitive	.77	.81	.70	.17	.20	.13	.36	.41	.30	.71*
Arousal	.82	.88	.81	.22	.33	.21	.43	.54	.42	.78*
Behavior	.86	.90	.85	.28	.35	.27	.49	.56	.32	.79*
PI	.90	.94	—	.28	.39	—	.50	.61	—	—

* $p < .01$.

when responding to the items of this questionnaire. Male students were not found to score higher on the NAS than female students, a finding that is more or less in accordance with the fairly small gender differences as documented by Novaco (2003) in his large standardization sample of nonclinical individuals.

The finding that inpatients had lower scores than students is in line with other researchers' findings: Mills et al. (1998) already found that violent offenders had significantly lower scores on the NAS–PI than nonviolent offenders, and Hornsveld et al. (2009) found that although outpatients scored higher on the AQ (Buss & Perry, 1992) than students, inpatients scored lower. There are several explanations for the latter findings. First, inpatients may be more inclined than outpatients to respond in a socially desirable way to self-report questionnaires in order to avoid negative clinical

Table 3
Novaco Anger Scale–Provocation Inventory (1994 Version) Scores of Ethnic and Nonethnic Inpatients (N = 142), Outpatients (N = 194), Male Students (N = 160), and Female Students (N = 160)

	Total		Ethnic		Nonethnic		Ethnic vs. nonethnic inpatients
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	
				Inpatients			
NAS	82.64	14.01	82.91	13.79	81.98	14.69	$t(140) = 0.36$ ($p = .720$)
Cognitive	29.82	4.69	29.75	4.59	29.98	4.99	$t(140) = -0.26$ ($p = .798$)
Arousal	27.36	5.13	27.62	4.96	26.71	5.53	$t(140) = 0.96$ ($p = .336$)
Behavior	25.46	5.56	25.53	5.65	29.37	5.91	$t(140) = 0.23$ ($p = .815$)
PI	50.72	10.97	50.44	10.07	51.32	12.83	$t(95) = -0.37$ ($p = .714$)
				Outpatients			
NAS	92.28	17.94	94.47	17.04	90.23	18.60	$t(192) = 1.65$ ($p = .100$)
Cognitive	31.82	5.62	32.48	5.53	31.20	5.66	$t(192) = 1.59$ ($p = .113$)
Arousal	30.10	6.69	30.39	6.41	29.83	6.96	$t(192) = 0.59$ ($p = .559$)
Behavior	30.36	6.98	31.60	6.49	29.20	7.25	$t(192) = 2.42$ ($p = .016$)
PI	56.79	15.15	56.95	16.52	56.58	13.52	$t(78) = 0.37$ ($p = .711$)
				Male students			
NAS	89.54	14.34	92.97	15.46	86.20	12.36	$t(158) = 3.07$ ($p = .003$)*
Cognitive	31.28	4.68	31.72	5.01	30.84	4.33	$t(158) = 1.19$ ($p = .235$)
Arousal	28.89	5.37	30.44	5.74	27.33	27.33	$t(158) = 3.83$ ($p = .001$)*
Behavior	29.38	5.95	30.81	6.24	27.90	5.35	$t(158) = 3.18$ ($p = .002$*)
				Female students			
NAS	89.69	14.68	89.46	14.25	89.93	15.23	$t(158) = -0.20$ ($p = .806$)
Cognitive	31.01	4.44	30.80	4.32	31.25	4.59	$t(158) = -0.64$ ($p = .454$)
Arousal	30.28	5.44	30.24	5.37	30.33	5.55	$t(158) = -0.11$ ($p = .958$)
Behavior	28.39	6.35	28.43	6.11	28.36	6.65	$t(158) = -0.07$ ($p = .656$)

Note. Nonethnic patients and students were born in the Netherlands, but had at least one parent who had immigrated from former colonies or from countries around the Mediterranean Sea. NAS–PI = Novaco Anger Scale–Provocation Inventory (1994 version).
* $p < .003$ (two-tailed).

Table 4
Comparison of Novaco Anger Scale–Provocation Inventory (1994 Version) Scores of Inpatients (N = 142), Outpatients (N = 194), and Male Students (N = 160)

	NAS–PI scores						Comparisons			
	Inpatients		Outpatients		Male students		Inpatients vs. male students		Outpatients vs. male students	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F* statistic	Effect size (*d*)	*F* statistic	Effect size (*d*)
NAS	82.64	14.02	92.28	17.94	89.54	14.34	$F(2, 299) = 9.78$ ($p = .001$)*	.49	$F(2, 351) = 5.27$ ($p = .003$)*	.17
Cognitive	29.83	4.72	31.82	5.62	31.28	4.68	$F(2, 299) = 3.74$ ($p = .013$)	.31	$F(2, 351) = 5.26$ ($p = .003$)*	.10
Arousal	27.33	5.11	30.10	6.69	28.89	5.37	$F(2, 299) = 3.22$ ($p = .021$)	.30	$F(2, 351) = 6.91$ ($p = .001$)*	.20
Behavior	25.48	5.57	30.36	6.98	29.38	5.95	$F(2, 299) = 20.93$ ($p = .001$)*	.68	$F(2, 351) = 2.35$ ($p = .048$)	.15
PI	50.72	10.97	56.79	15.15	—	—	—	—	—	—

Note. Comparisons were controlled for age. NAS–PI = Novaco Anger Scale–Provocation Inventory (1994 version).
* $p < .006$ (one-tailed).

and legal consequences. Second, it is possible that inpatients are less able to observe their own behavior than outpatients. A third, and in our opinion, the most likely explanation is that inpatients live in a structured and controlled environment, which has an attenuating effect on anger and aggression.

A limitation of this study is that although according to Novaco (personal communication, October 3, 2009) the quality of the translation was generally good for the NAS and very good for the PI, four items of the NAS were problematic. More precisely, due to the translation process, the content of three items changed in such way that they were more clearly associated with another, nonintended domain of anger (Items 25, 26, and 31). In addition, for one other item, the essential point was missed because there were no proper Dutch equivalents to cover the English words of "rough" and "rude" (Item 11).

A second limitation is that the size of the subsamples may have precluded a proper analysis of the factor structure of the NAS. The fact that we compared the inpatient scores with those of male adolescents and not with those of male adults is a third limitation, although we statistically controlled for age.

Our preliminary conclusion is that the NAS–PI is a valuable instrument for the measurement of anger in Dutch violent forensic psychiatric patients, but that the original NAS subscales should be used in this population with the necessary prudence. The scores on the NAS–PI and its subscales can be used for the evaluation of forensic psychiatric patients on an individual level, provided that they are combined with data from other measurement instruments and with clinical impressions. It has to be to be noted that the NAS–PI was in part constructed with case formulation in mind to enable the clinician to target areas for treatment. For inpatients,

Table 5
Correlations Between the Novaco Anger Scale–Provocation Inventory (1994 Version) and Other Measures for Inpatients (N = 142), Outpatients (N = 194), and Students (N = 320)

	NAS			PI	
Measures/content of scale	Inpts ($N = 142$)	Outpts ($N = 194$)	Students ($N = 320$)	Inpts ($N = 97$)	Outpts ($N = 80$)
PCL–R					
Psychopathy	−.05	.03		−.17	−.16
Interpersonal	−.29**	−.00		−.15	−.15
Affective	−.12	−.11		−.03	−.25*
Lifestyle	−.02	.16*		−.16	−.04
Antisocial	.21*	.02		.02	−.19
NEO-FFI					
Neuroticism	.48**	.49**	.27**	.49**	.58**
Agreeableness	−.33**	−.54**	−.54**	−.27**	−.53**
STAS Trait Anger	.42**	.69**	.68**	.20*	.64**
PFS–AV Hostility	.35**	.48**	.46**	.24*	.46**
AQ–SF					
Aggression	.69**	.66**	.71**	.55**	.66**
Anger	.46**	.53**	.55**	.49**	.53**

Note. NAS–PI = Novaco Anger Scale–Provocation Inventory (1994 version); 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.
* $p < .05$. ** $p < .01$.

NAS–PI scores can only be used for comparisons with scores from other inpatients and not with scores from nonclinical populations.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental Disorders* (4th ed.). Washington, DC: Author.

Baker, M. T., Van Hasselt, V. B., & Sellars, A. H. (2008). Validation of the Novaco Anger Scale in an incarcerated offender population. *Criminal Justice and Behavior, 35,* 741–754. doi:10.1177/0093854808316275

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., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21,* 343–349. doi:10.1037/h0046900

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

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.

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology, 31,* 457–470. doi:10.1023/A:1023899703866

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised.* Toronto, ON, Canada: Multi-Health Systems.

Hare, R. D. (2003). *The Hare Psychopathy Checklist–Revised* (2nd ed.). Toronto, Canada: Multi-Health Systems.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO Persoonlijkheidsvragenlijsten* [Manual NEO Personality Questionnaires]. Lisse, the Netherlands: Swets Test Services.

Hornsveld, R. H. J., Muris, P., Kraaimaat, F. W., & Meesters, C. (2009). 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 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

Hornsveld, R. H. J., Nijman, H. L. I., & Kraaimaat, F. W. (2008). Aggression control therapy for violent forensic psychiatric patients: First results. *Psychology: Crime & Law, 14,* 1–18. doi:10.1080/10683160701340569

Jones, J., Thomas-Peter, B., & Gangstad, B. (2003). An investigation of the factor structure of the Novaco Anger Scale. *Behavioural and Cognitive Psychotherapy, 31,* 429–437. doi:10.1017/S1352465803004041

Jones, J. P., Thomas-Peter, B. A., & Trout, A. (1999). Normative data for the Novaco Anger Scale from a non-clinical sample and implications for clinical use. *British Journal of Clinical Psychology, 38,* 417–424. doi:10.1348/014466599163024

Lindqvist, J. K., Dåderman, A. M., & Hellström, A. (2003). Swedish adaptations of the Novaco Anger Scale–1998, the Provocation Inventory, and the State–Trait Anger Expression Inventory–2. *Social Behavior and Personality, 31,* 773–788. doi:10.2224/sbp.2003.31.8.773

Lindqvist, J. K., Dåderman, A. M., & Hellström, A. (2005). Internal reliability and construct validity of the Novaco Anger Scale-1998-S in a sample of violent prison inmates in Sweden. *Psychology: Crime & Law, 11,* 223–237. doi:10.1080/10683160500036863

Mills, J. F., Kroner, D. G., & Forth, A. E. (1998). Novaco Anger Scale: Reliability and validity within an adult criminal sample. *Assessment, 5,* 237–248. doi:10.1177/107319119800500304

Monahan, J., Steadman, H. J., Silver, E., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., . . . Banks, S. (2001). *Rethinking risk assessment: The MacArthur Study of Mental Disorder and Violence.* Oxford, England: Oxford University Press.

Novaco, R. W. (1994). Anger as a risk factor for violence among the mentally disordered. In J. Monahan & H. J. Steadman (Eds.), *Violence and mental disorder* (pp. 21–59). Chicago, IL: The University of Chicago Press.

Novaco, R. W. (2003). *The Novaco Anger Scale and Provocation Inventory: Manual.* Los Angeles, CA: Western Psychological Services.

Novaco, R. W., Ramm, M., & Black, L. (2001). Anger treatment with offenders. In C. R. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 281–296). Chichester, England: Wiley.

Novaco, R. W., & Taylor, J. L. (2004). Assessment and aggression in male offenders with developmental disabilities. *Psychological Assessment, 16,* 42–50. doi:10.1037/1040-3590.16.1.42

Rosenzweig, S. (1978). *Aggressive behavior and the Rosenzweig Picture-Frustration Study.* New York, NY: Praeger.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8,* 23–74.

Sloore, H., Derksen, J. J. L., Helenbosch, G., & De Mey, H. R. A. (1993). *De Minnesota Multiphasic Personality Inventory–II* [Dutch translation]. Nijmegen, the Netherlands: PEN Test.

Spielberger, C. D. (1980). *Preliminary manual for the State–Trait Anger Scale (STAS).* Tampa, FL: University of South Florida.

Spielberger, C. D. (1999). *State–Trait Anger Expression–2: Professional manual.* Lutz, FL: Psychological Assessment Resources.

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, the Netherlands: 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, the Netherlands: Swets Test.

Received June 16, 2009
Revision received April 13, 2011
Accepted April 18, 2011 ■