

# Assessment

<http://asm.sagepub.com>

---

## **Psychometric Properties of the Aggression Questionnaire in Dutch Violent Forensic Psychiatric Patients and Secondary Vocational Students**

Ruud H. J. Hornsveld, Peter Muris, Floris W. Kraaimaat and Cor Meesters  
*Assessment* 2009; 16; 181 originally published online Oct 13, 2008;  
DOI: 10.1177/1073191108325894

The online version of this article can be found at:  
<http://asm.sagepub.com/cgi/content/abstract/16/2/181>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Assessment* can be found at:**

**Email Alerts:** <http://asm.sagepub.com/cgi/alerts>

**Subscriptions:** <http://asm.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** <http://asm.sagepub.com/cgi/content/refs/16/2/181>

# Psychometric Properties of the Aggression Questionnaire in Dutch Violent Forensic Psychiatric Patients and Secondary Vocational Students

Ruud H. J. Hornsveld

*Forensic Psychiatric Center De Kijvelanden, Poortugaal, Netherlands*

Peter Muris

*Erasmus University Rotterdam, Netherlands*

Floris W. Kraaimaat

*Radboud University Nijmegen, Netherlands*

Cor Meesters

*The University of Maastricht, Netherlands*

The psychometric properties of a Dutch version of Buss and Perry's Aggression Questionnaire (AQ) were examined in a sample of violent forensic psychiatric inpatients and outpatients and a sample of secondary vocational students. The internal consistency, interitem correlations, and item-scale correlations of the subscales Physical Aggression, Anger, and Hostility were good but turned out to be unsatisfactory for the subscale Verbal Aggression. The four-factor structure of the AQ could not be confirmed, but the four-factor structure of a 12-item version (short form) of the AQ, the AQ-SF, produced an acceptable fit. The test-retest reliability of the AQ and AQ-SF total scores was good, although the test-retest reliability of the AQ-SF subscale Physical Aggression was not satisfactory. The validity of both the AQ and AQ-SF could be demonstrated by meaningful correlations with alternative measures of aggression and personality, but inpatients were not found to display higher scores on the AQ or AQ-SF than the students.

**Keywords:** *Aggression Questionnaire; psychometric qualities; forensic psychiatric patients; secondary vocational students*

In recent years, cognitive-behavioral interventions have been increasingly developed and implemented in Dutch forensic psychiatric hospitals and institutions for youths with disruptive behavior problems. To evaluate these treatment programs, there is a need for specifically designed measurement instruments that have been validated in populations at risk for aggressive behavior. One of these instruments is the Dutch version of Buss and Perry's (1992) Aggression Questionnaire (AQ; see Meesters, Muris, Bosma, Schouten, & Beuving, 1996).

To develop the AQ, Buss and Perry used an initial pool of 52 items representing the main components of the Buss-Durkee Hostility Inventory (Buss & Durkee, 1957) and administered these to a total of 1,253 college students (641 female and 612 male students, age 18-20 years); they were divided into three separate samples. Data from the first sample ( $n = 406$ ) were subjected to an exploratory factor analysis, after which 23 items were eliminated because of low factor loadings. The remaining 29 items could be clearly related to four subscales: Physical Aggression (9 items), Verbal Aggression (5 items), Anger (7 items), and Hostility (8 items). Confirmatory factor analysis in the two other samples yielded further support for this four-factor structure, which was invariant across gender. Other

**Authors' Note:** Please address correspondence to Ruud H. J. Hornsveld, Lange Dreef 52, 2285 La Rijswijk, Netherlands; e-mail: [Ruud.Hornsveld@Kijvelanden.nl](mailto:Ruud.Hornsveld@Kijvelanden.nl).

psychometric properties were additionally examined in this first test of the AQ. The results indicated that the internal consistency of the four subscales was adequate, and test–retest correlations demonstrated sufficient stability over time. The concurrent validity was supported by meaningful correlations between the total score of the AQ and scores on self-report questionnaires for measuring traits such as impulsiveness, assertiveness, and competitiveness. Furthermore, significant correlations of AQ scores with peer nominations of various kinds of aggression were found, which produced evidence for the convergent validity of the scale. On the basis of these findings, Buss and Perry (1992) concluded that the AQ seems to be an adequate index for measuring the motor (Physical Aggression and Verbal Aggression), affective (Anger), and cognitive (Hostility) components of aggressive behavior. However, they noted that their findings “must be extrapolated to populations of people with less education and lower socioeconomic status” (p. 457) because they only studied college students.

Since Buss and Perry’s original investigation, the psychometric properties of the AQ have been investigated in a large number of studies that mostly relied on university student populations. For instance, Harris (1995) submitted the AQ to a group of 306 Canadian university students and found support for the hypothesized four-factor structure using confirmatory factor analysis after removing two items of the Hostility subscale because of low factor loadings. Harris (1997) further evaluated the psychometrics of the AQ in a group of 106 Canadian female undergraduate university students. The AQ Scales turned out to have moderate to good internal consistency and test–retest reliability over a period of 7 months. Evidence was obtained for the convergent validity of the scale by meaningful correlations with alternative measures of aggression and personality. Negative correlations were found between AQ scores and social desirability, which according to Harris may have an attenuating effect on validity estimates.

Bryant and Smith (2001) explored the factor structure of the AQ in a sample of 307 American, 200 British, and 306 Canadian undergraduates. Because the predicted four-factor model only produced a modest fit, the researchers deleted items that displayed low or multiple loadings in a principal components analysis and excluded a number of reverse-scored items. This procedure yielded a shortened 12-item version of the AQ, for which the hypothesized four-factor model produced an acceptable fit. Furthermore, additional

analyses in a sample of 341 American undergraduates yielded support for the construct and discriminant validity of the shortened version of the AQ (hereafter referred to as the AQ-SF).

The psychometric qualities of the AQ with 29 items has been examined in Dutch (Meesters et al., 1996), Japanese (Nakano, 2001), and Italian (Fossati, Maffei, Acquarini, & Di Ceglie, 2003) university and high school students. In those studies, confirmation of the four-factor structure of the AQ was sometimes found after removal of a few items with unacceptably low loadings (i.e.,  $<.03$ ). Internal consistency of the subscales turned out to be good with the exception of the Verbal Aggression subscale. Test–retest correlations were found to be similar to those originally reported by Buss and Perry (1992).

Only a limited number of studies have evaluated the psychometrics of the AQ in populations known for their aggressive behavior. One exception is an investigation by Williams, Boyd, Cascardi, and Poythress (1996), who examined the factor structure, reliability, and convergent validity of the AQ in a group of 200 aggressive and nonaggressive offenders (76 female and 124 male offenders; age 17–69 years). The AQ was found to be reliable in this sample, although the internal consistency of the Verbal Aggression subscale was below acceptable limits (i.e.,  $\alpha = .50$ ). A confirmatory factor analysis of the hypothesized four-factor model yielded a poor fit in this population, and therefore, an exploratory approach was adopted. This procedure pointed in the direction of a two-factor solution with Physical Aggression and Anger items loading on one factor and Verbal Aggression and Hostility items generally loading on the other factor. The correlation between the total score of the AQ and the Novaco Anger Scale (NAS; Novaco, 1994) was positive and significant, which supported the convergent validity of the AQ. However, offenders with aggressive crime charges did not display significantly higher AQ scores than offenders charged with nonaggressive crimes, which of course questions the validity of the scale.

The validity of the Dutch AQ in a clinical setting was further studied by Morren and Meesters (2002) in a group of 69 violent male offenders aged 12 to 18 years who participated in a residential rehabilitation program. They noted that AQ scores were substantially related to trait anger and (with the exception of the Hostility subscale) not to scores of general psychopathology. Furthermore, correlations between the AQ total score and behavioral ratings of aggression

were found to be rather modest ( $r$ s between .07 and .42). Although not all psychometric properties of the AQ were equally strong in this clinical sample, Morren and Meesters concluded that their study at least provided some evidence for the validity of the AQ in Dutch adolescents who were detained for violent offenses.

Diamond, Wang, and Buffington-Vollum (2005) analyzed the factor structure of the AQ and AQ-SF in a sample of 786 male, mentally ill prisoners (aged 19-68 years). In this sample, 72% had a primary axis I diagnosis, 56% a secondary axis I diagnosis, and 42% had an axis II personality disorder diagnosis, but no information was provided as to whether these offenders had committed a violent crime. The four factors of the 12-item AQ-SF produced the best fit, but Diamond et al. (2005) also found that the fit could be improved by substituting one Anger item for another item of the original subscale. Altogether, these researchers concluded that "the shorter, refined version actually has measurement characteristics superior to the full version." In a second study, Diamond and Magaletta (2006) assessed the construct validity of the AQ-SF in a sample of 916 male and 355 female offenders. No information was available about whether the offenders in this study also participated in the previous study of Diamond et al. (2005). In addition, data about age and committed offenses were not reported. Confirmatory factor analysis supported the four-factor structure across both genders, and the internal consistency coefficients of the subscales varied between .62 and .77. Concurrent validity was supported by significant correlations between AQ-SF subscales and relevant subscales of the Personality Assessment Inventory (Morey, 1991), a 344-item self-report instrument designed to measure a variety of mental health and personality symptoms, and syndromes. Unfortunately, the authors did not distinguish between subsamples of violent and nonviolent offenders. Therefore, the finding of Williams et al. (1996) that offenders who were convicted for violent crimes did not display significantly higher AQ scores than nonaggressive offenders was not tested in this study.

In summary, several studies conducted in various countries have supported the notion that the AQ and AQ-SF can be used as reliable and valid instruments for measuring different components of aggressive behavior in university students of both genders. Only a limited number of studies have investigated the two versions in (mentally ill) offenders. Clearly, more research is needed to examine the psychometric qualities of both the AQ and AQ-SF in persons with a

lower level of education and in clinical populations known for their aggressive and violent behavior. With this in mind, the present study investigated the psychometric properties of both AQ versions in Dutch violent forensic in- and outpatients with oppositional-defiant disorder, conduct disorder, or antisocial personality disorder and in secondary vocational students (only male students). We compared the mean scores of the forensic psychiatric outpatients and inpatients with those of the male students to determine whether the AQ and AQ-SF differentiate between these groups. Aggressive behavior as measured by the two versions of the AQ was expected to correlate in a theoretically meaningful way with other constructs. Our research on the validity of other aggression-related measures such as the NAS-Provocation Inventory (Hornsveld, Muris, & Kraaimaat, 2008a) and the adapted version of the Picture-Frustration Study (PFS-AV; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007a), suggested that a moderate correlation would be expected between the AQ and AQ-SF and the personality domains of neuroticism and agreeableness.

Relatively high correlations were expected between AQ and AQ-SF scores and other indexes of aggression, anger, and hostility. As aggressive behavior is thought to be partly associated with social anxiety and lack of social skills (Crick & Dodge, 1996; Loeber & Stouthamer-Loeber, 1998), we expected significantly positive correlations between the AQ or AQ-SF and social anxiety and negative links with social skills. Because authors such as Walters (2003) and Guy, Edens, Anthony, and Douglas (2005) have questioned the magnitude of the relation between aggressive behavior and psychopathy, we explored the associations between AQ and AQ-SF scores and psychopathy as measured by the Psychopathy Checklist-Revised (PCL-R; Hare, 1991) in the patient samples.

## Methods

### Participants

The study was carried out with a sample of 138 male *inpatients*, who were recruited from six forensic psychiatric institutions situated throughout the Netherlands. Patients were "detained under hospital order" for a serious violent offense, punishable with a minimum of 4 years (e.g., severe assault, manslaughter, or murder). Their mean age was 33.55 years

( $SD = 7.65$ ; range = 21-56 years), and 29.18% belonged to an ethnic minority. The primary diagnosis was an antisocial personality disorder on axis II, or a psychotic disorder on axis I in combination with an antisocial personality disorder on axis II (see the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. [DSM-IV]; American Psychiatric Association, 1994). The psychiatric condition of the psychotic patients had stabilized to the extent that their antisocial personality disorder became prominent, which means that they regularly exhibited aggressive behavior on the ward. Diagnoses and psychiatric condition were ascertained in multidisciplinary teams of psychologists and psychiatrists who determined whether a patient could be indicated for a cognitive behavioral treatment group therapy program (Hornsveld, Nijman, & Kraaimaat, 2008b). The teams based this indication on file data, recent psychiatric and psychological evaluations, and other clinical reports.

The 206 male outpatients (78% adolescents and 22% adults) were treated at a forensic psychiatric outpatient clinic in Rotterdam, the Netherlands, as part of their sentence for violent offenses (e.g., assault or armed robbery). In the Netherlands, the court can require offenders to undergo outpatient treatment as (a) an added conditional punishment for offenses to which imprisonment is for 3 years maximally, (b) an alternative punishment for offenses to which imprisonment is for at most 6 months, (c) part of a penal program, and (d) a supervision by a youth protection agency. The mean age of the outpatients was 22.50 years ( $SD = 8.86$ ; range = 16-56 years), and 51.78% of them belonged to an ethnic minority. The outpatients had an oppositional-defiant or conduct disorder as the primary diagnosis on axis I or, when they were 18 years or older, a main diagnosis of antisocial personality disorder on axis II (DSM-IV). Again, diagnoses were ascertained in multidisciplinary teams on the basis of file data and an extensive intake interview that was conducted by the first author.

The 160 male students followed secondary vocational education in Rotterdam, the Netherlands. In the Netherlands, secondary vocational education follows after elementary school and concerns a professional training at the lowest level for professions such as carpenter, housepainter, and electrician. The mean age of these students was 17.82 years ( $SD = 1.54$ ; range = 16-27 years). One half of the students was of original Dutch descent (i.e., 50.92%), whereas the others (i.e., 49.08%) were of non-Dutch descent.

In terms of similarities in educational backgrounds, the adolescents among the outpatients mostly followed secondary vocational education just as the students. Most of the adult participants in the in- and outpatient samples had followed such education in the past as well.

Patients and students of non-Dutch descent generally had at least one parent from Surinam, the Netherlands Antilles, Turkey, Morocco, or Cape Verde.

## Measures

The PCL-R (Hare, 1991; Dutch version: Vertommen, Verheul, De Ruiter, & Hildebrand, 2002) was used for measuring psychopathy. The checklist consists of 20 items, which have to be rated on a 3-point scale with 0 = *does not apply*, 1 = *applies to some extent*, and 2 = *applies*. Vertommen et al. (2002) found support for the reliability and validity of the Dutch version of the PCL-R, and they confirmed Hare's two-factor structure: (a) callous and remorseless use of others (e.g., lack of remorse or guilt) and (b) chronically unstable and antisocial lifestyle (e.g., poor behavioral controls).

The NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992; Dutch version: Hoekstra, Ormel, & De Fruyt, 1996) has 60 items and measures the big five personality domains of Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Participants score items of the NEO-FFI on a 5-point Likert Scale ranging from *entirely disagree* to *entirely agree*. In the present study, we were only interested in the neuroticism (e.g., I seldom feel lonely or sad) and agreeableness (e.g., Some people find me selfish and egotistic) scales, because these traits are considered as relevant in the context of aggression (Hornsveld et al., 2008b). In a Dutch sample of 135 nonclinical adults, evidence has been obtained to support the reliability and validity of the NEO-FFI (Hoekstra et al., 1996).

The trait items of Spielberger's (1980) State-Trait Anger Scale (STAS; van der Ploeg, Defares, & Spielberger, 1982) were used to measure the general disposition to anger. Participants were asked to indicate for each item (e.g., I feel irritated) how they generally felt using a 4-point Likert Scale: 1 = *entirely not*, 2 = *a bit*, 3 = *rather much*, and 4 = *very much*. Internal consistency, test-retest reliability, and validity of the Dutch version of the STAS have proven to be satisfactory (Van der Ploeg et al., 1982).

An adapted version of Rosenzweig's (1978) PFS-AV (Hornsveld et al., 2007a) was used for measuring

hostility. The test asks participants to write down their reactions to 12 cartoon-like pictures. They are instructed to examine the situations as shown in the pictures (e.g., to a shopkeeper: "This is the third time that this watch has stopped.") and write the first appropriate reply that enters their mind in the blank box. Answers are scored by an independent rater on a 7-point Likert Scale, ranging from 1 = *not at all hostile* to 7 = *extremely hostile*. In a sample of 231 Dutch violent forensic psychiatric patients, which overlapped with the current sample, the internal consistency of the PFS-AV appeared to be .76, the interrater reliability .77, and the test-retest reliability .66 (all  $ps < .01$ ). Furthermore, evidence was found for the validity of the test as scores correlated in a meaningful way with indexes of agreeableness and aggressive behavior (Hornsveld et al., 2007a).

Part A of the NAS (Novaco, 1994) was used to study the participants' self-reported responses to 48 anger-eliciting situations (e.g., When someone yells at me, I yell back at them). Items are scored on a 3-point Likert Scale: 1 = *never true*, 2 = *sometimes true*, 3 = *always true*. In a pilot-study of 90 violent forensic psychiatric outpatients, Cronbach's alpha of part A of the NAS was .95, and the test-retest reliability was .85.

The Inventory of Interpersonal Situations (IIS; Van Dam-Baggen & Kraaimaat, 1999) assesses how much anxiety people experience during social interactions (social anxiety) and how often they are able to actually perform the appropriate behavior in such situations (social skills). In the present study, two subscales of this inventory, namely Giving Criticism (e.g., Saying to a friend that he/she does something that troubles you) and Giving Compliments (e.g., Saying to somebody that he/she looks good) were used because the results of a previous study indicated that only these subscales differentiated adequately between violent and nonviolent participants. A psychometric evaluation of the IIS was conducted in a subsample of violent forensic psychiatric patients, some of whom were also investigated in the current study (Hornsveld, 2005). The internal consistency and test-retest reliability of the IIS can be qualified as good (i.e.,  $\alpha_s > .80$  and test-retest correlations  $> .70$ ). Moreover, significant correlations have been found between the IIS and measures of anxiety, which support the validity of the scale (Van Dam-Baggen & Kraaimaat, 1999).

The Observation Scale for Aggressive Behavior (OSAB; Hornsveld, Nijman, Hollin, & Kraaimaat, 2007b) measures behavior on the ward. The scale

comprises 40 items spread over the subscales Irritation/Anger (e.g., irritated), Anxiety/Gloominess (e.g., gloominess), Aggressive Behavior (e.g., threats toward staff), Social Behavior (e.g., gives his opinion adequately), Antecedent (e.g., restraining measure), and Sanction (e.g., must apologize). The staff scores the behavior of the inpatients in the preceding week on a 4-point scale with 1 = *no*, 2 = *seldom*, 3 = *occasionally*, and 4 = *frequently*. The psychometric qualities of the OSAB were studied in a sample of 220 violent forensic psychiatric inpatients who were also investigated in the current study. Results showed that the internal consistency of the subscales varied from .63 to .93, interrater reliability correlations were between .49 and .81 ( $ps < .01$ ), whereas test-retest reliability ranged between .48 and .79 ( $ps < .01$ ). Convergent validity was demonstrated through significant positive correlations with subscales of the Forensic Inpatient Observation Scale (Timmerman, Vastenburg, & Emmelkamp, 2001).

## Procedure

The data from the inpatients and outpatients were collected between 2002 and 2006 and were obtained during an evaluation trial of a cognitive-behavioral group therapy program (Hornsveld et al., 2008b). As such, we were especially interested in personality traits and problem behaviors and less in demographic data. Questionnaires were submitted individually to the patients prior to the group therapy. One of the indication criteria for participation in the group therapy was sufficient command of the Dutch language through the spoken and written word. Patients had to understand the information brochures and to make homework assignments in a portfolio. Questionnaires were completed under the supervision of an experienced research assistant. When all items of the questionnaires were completed, patients received a fee of €5 in return for their participation. In the week that inpatients completed the questionnaires, staff on the ward with experience in the use of observation scales were asked to score the items of the OSAB (Hornsveld et al., 2007b). A subsample of 90 outpatients filled out the AQ during the intake interview some 4 weeks before the start of the therapy, so that it became possible to examine the test-retest of the scale in a forensic outpatient sample. During the period between intake and start of the therapy, outpatients were supervised by a probation agency and received no specific treatment or training.

**Table 1**  
**Mean Scores (*M*) and Standard Deviations (*SDs*) for Various Measures as**  
**Obtained in Inpatients (*n* = 138), Outpatients (*n* = 206), and Students (*n* = 160)**

Measures	Content of Scale	Inpatients		Outpatients		Students	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PCL-R	Psychopathy	21.43	7.30	18.69	5.68	—	—
	Use of others	8.98	3.60	9.92	3.29	—	—
	Antisocial lifestyle	10.47	3.68	8.43	3.28	—	—
NEO-FFI	Neuroticism	33.23	7.80	32.14	8.37	30.57	7.13
	Agreeableness	40.77	4.85	39.32	5.33	39.58	4.79
STAS	Trait anger	19.47	8.53	21.31	7.63	18.07	5.47
PFS-AV	Hostility	31.97	9.79	34.73	11.34	28.91	5.88
NAS	Anger	82.72	13.33	90.31	21.80	89.12	14.22
IIS Social anxiety	Criticism	14.78	5.40	14.14	5.45	17.11	4.55
	Compliment	6.35	2.93	6.82	3.46	9.30	4.11
IIS Social skills	Criticism	22.52	5.09	21.45	5.25	19.94	4.12
	Compliment	15.46	2.88	14.63	3.43	13.79	3.46
OSAB	Aggression on the ward	15.98	5.53	—	—	—	—

Note: PCL-R = Psychopathy Checklist-Revised; NEO-FFI = Five-Factor Inventory; STAS = State-Trait Anger Scale; PFS-AV = Adapted version of the Picture-Frustration Study; NAS = Novaco Anger Scale; IIS = Inventory of Interpersonal Situations; OSAB = Observation Scale for Aggressive Behavior.

PCL-R scores were calculated by experienced and certified clinical psychologists. For the inpatients, PCL-R scores were calculated in 2005 and 2006 on the basis of file study. Files comprised detailed information about life history, committed offenses, and elaborate reports from psychiatrists and/or psychologists. These reports were often made in a special forensic assessment center in which the offender had to stay for observation by order of the court. After the assessment period, the judge determined whether the offender would be detained under hospital order. For the outpatients, PCL-R scores were based on the structured intake interview and information from an extensive psychiatric and psychological evaluation on the basis of which the judge had decided on obligatory treatment.

The data for the secondary vocational students were collected in 2004. The students completed a set of questionnaires in their classrooms at school. Most students did not need more than 1 hour to do so. Completion was supervised by the first author and a research assistant. After a check on missing scores, students received a fee of €10 in return for their participation.

The clinical part of this study was approved by the Dutch Review Committee for Patient-Linked Research in Arnhem, the Netherlands, and by the Scientific Research and Documentation Center of the Dutch

Ministry of Justice. The study dealing with the secondary vocational students was approved by the Board of the Albeda College in Rotterdam.

## Results

The mean scores and standard deviations for all measures in each of the samples are shown in Table 1.

### Factor Structure

A confirmatory factor analysis (AMOS 16.0) was carried out in both samples to test the original four-factor structure of the AQ with 29 items, the two-factor structure with 26 items as suggested by Williams et al. (1996), and the four-factor structure of Bryant and Smith (2001) with 12 items. We used the criteria of Schermelleh-Engel, Moosbrugger, and Müller (2003) to evaluate the goodness-of-fit indexes (GFI). These authors consider the following results as indications for a good fit:  $\chi^2/df \leq 2$ , GFI  $\geq .95$ , CFI (comparative fit index)  $\geq .97$ , and RMSEA (root mean square error of approximation)  $\leq .05$ . In their opinion, a model has an acceptable fit when  $\chi^2/df = 2$  to 3, GFI = .90 to .95, CFI = .95 to .97, and RMSEA = .05 to .08.

In the combined patient samples, the four-factor structure of the original AQ with 29 items yielded an

**Table 2**  
**Goodness-of-Fit Indexes (GFIs) for AQ and AQ-SF in Patients ( $n = 344$ ) and Students ( $n = 160$ )**

	Number of Items	Patients				Students			
		$\chi^2/df$	GFI	CFI	RMSEA	$\chi^2/df$	GFI	CFI	RMSEA
AQ with four factors	29	3.12	.80	.78	.08	2.21	.72	.65	.09
After removal of four items	25	3.28	.82	.82	.08	2.09	.77	.74	.08
AQ with two factors	26	3.78	.77	.75	.09	2.28	.73	.68	.09
After removal of one item	25	3.69	.78	.77	.09	2.29	.75	.68	.09
AQ with one factor	29	3.71	.74	.71	.09	2.49	.67	.56	.10
AQ-SF with four factors	12	3.40	.93	.91	.08	1.49	.93	.94	.06
AQ with two higher-order factors	29	.64	1.00	1.00	.00	17.23	.95	.92	.32
AQ-SF with two higher-order factors	12	1.12	1.00	1.00	.02	9.39	.97	.94	.03

Note: AQ = Aggression Questionnaire; AQ-SF = Aggression Questionnaire–Short Form;  $df$  = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation.

unacceptable fit for the data. When we removed four items with fairly low loadings (i.e.,  $<.30$ ), the fit improved but still remained unsatisfactory. GFIs for the two-factor structure with 26 items (Williams et al., 1996) were unacceptable, even after removing one problematic item. However, the four-factor structure of Bryant and Smith (2001) with 12 items provided an acceptable fit (Table 2).

In the sample of students, highly similar results were found. The GFIs indicated that the four-factor model of the AQ with 29 items or 25 items did not produce a good fit for the data. The two-factor structure with 26 items (Williams et al., 1996) or with 25 items also did not provide a good fit, but the four-factor structure with 12 items (Bryant & Smith, 2001) again yielded an acceptable fit.

Finally, a confirmatory factor analysis was carried out in both samples to test a model with two higher order factors for the AQ with 29 items (Buss & Perry, 1992) and the AQ-SF with 12 items (Bryant and Smith, 2001). In this model, the Physical Aggression and Verbal Aggression factors load on a higher Behavior Factor, and the Anger and Hostility factors on a higher Experience Factor. GFIs were good for both the AQ and the AQ-SF in the patient sample. In the student sample, the model produced an acceptable fit for the AQ-SF but not for the AQ (Table 2).

### Internal Consistency, Interitem Correlations, Item–Scale Correlations, and Test–Retest Reliability

Internal consistency coefficients, mean interitem correlations, and mean item–scale correlations were

calculated for the AQ, the AQ-SF, and each of the subscales in the three studied samples (i.e., inpatients, outpatients, and students). For the total score of the full-length AQ, Cronbach's alpha varied from .83 to .91, mean interitem correlations from .15 to .27, and mean item–scale correlations from .35 to .50. Internal consistency coefficients for the subscales were between .34 and .81, mean interitem correlations were between .09 and .34, and mean item–scale correlations were between .17 and .52. The test–retest reliability (4 weeks interval) of the AQ in the sample of outpatients was .72 for the total score and ranged between .54 and .76 (all  $ps < .01$ ) for various subscales (Table 3).

Internal consistency coefficients for the total score of the AQ-SF were between .72 and .88, mean interitem correlations were between .19 and .27, and mean item–scale correlations were between .35 and .50. Cronbach's alphas for the subscales varied from .38 to .74, mean interitem correlations from .18 to .49, and mean item–scale correlations from .23 to .57. Clark and Watson (1995) stated that for a broad higher-order construct a mean interitem correlation between .15 and .20 is probably desirable, but for a narrower construct, a mean correlation between .40 and .50 is required. The test–retest correlations (4 weeks interval) were significant for the AQ-SF total and subscale scores except for the Physical Aggression subscale (Table 3). Inspection revealed that it was one particular item of this subscale (i.e., I have threatened people I know) that had very low test–retest correlations, whereas those of the two other items were satisfactory.

Correlations among AQ total and subscale scores were all positive and significant and varied between .20 and .57 for the inpatients, between .54 and .82 for the



**Table 3**  
**Internal Consistency Coefficients (Cronbach's alpha), Mean Interitem Correlations, and Mean Item-Scale Correlations of the AQ and AQ-SF for Inpatients ( $n = 138$ ), Outpatients ( $n = 206$ ), and Students ( $n = 160$ ), and Test-Retest Reliability of the AQ and AQ-SF for Outpatients ( $n = 90$ )**

		Internal Consistency			Mean Interitem Correlations			Mean Item-Scale Correlations			Test-Retest Reliability
		Inpats.	Outpats.	Students	Inpats.	Outpats.	Students	Inpats.	Outpats.	Students	Outpats
AQ	Total	.83	.91	.87	.15	.27	.18	.35	.50	.40	.72**
	Physical Aggression	.72	.79	.78	.23	.30	.28	.41	.49	.46	.76**
	Verbal Aggression	.34	.67	.34	.09	.28	.09	.17	.42	.17	.58**
	Anger	.57	.73	.64	.17	.26	.21	.21	.44	.36	.65**
	Hostility	.81	.78	.73	.34	.31	.26	.52	.48	.43	.54**
AQ-SF	Total	.72	.88	.81	.19	.38	.26	.37	.58	.46	.38**
	Physical Aggression	.52	.65	.62	.25	.38	.36	.34	.46	.46	.13
	Verbal Aggression	.38	.74	.59	.18	.49	.33	.23	.57	.40	.49**
	Anger	.60	.61	.51	.34	.34	.27	.41	.43	.33	.48**
	Hostility	.69	.74	.74	.43	.48	.49	.51	.57	.57	.48**

Note: Inpats. = inpatients; Outpats. = outpatients; AQ = Aggression Questionnaire; AQ-SF = Aggression Questionnaire-Short Form.  
 $*p < .05$ .  $**p < .01$ .

**Table 4**  
**Correlations Among Subscales of the AQ and AQ-SF for Inpatients ( $n = 138$ ), Outpatients ( $n = 206$ ), and Students ( $n = 160$ )**

Questionnaire	Subscales	Inpatients				Outpatients				Students			
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
AQ	(1) Total												
	(2) Physical Aggression	.50**				.73**				.55**			
	(3) Verbal Aggression	.51**	.52**			.77**	.70**			.65**	.49**		
	(4) Anger	.57**	.47**	.47**		.82**	.74**	.70**		.70**	.59**	.52**	
	(5) Hostility	.36**	.28**	.20*	.38**	.68**	.54**	.62**	.67**	.53**	.34**	.57**	.54**
AQ-SF	(1) Total												
	(2) Physical Aggression	.32**				.66**				.44**			
	(3) Verbal Aggression	.56**	.34**			.74**	.61**			.57**	.40**		
	(4) Anger	.46**	.27**	.55**		.78**	.65**	.70**		.64**	.47**	.47**	
	(5) Hostility	.29**	.16	.30**	.24**	.62**	.45**	.58**	.61**	.48**	.22**	.46**	.50**

Note: AQ = Aggression Questionnaire; AQ-SF = Aggression Questionnaire-Short Form.  
 $*p < .05$ .  $**p < .01$ .

outpatients, and between .34 and .70 for the secondary vocational students. Between the subscales of the AQ-SF, correlations ranged from .16 (not significant) to .56 for the inpatients, from .45 to .78 for the outpatients, and from .22 to .64 for the students (Table 4).

### Comparison of AQ and AQ-SF Scores Across Groups

The mean scores of the inpatients and outpatients were compared with those of the male secondary vocational students and with each other. We corrected

for age by entering this variable as a covariate in the analyses of variance, as the students were significantly younger than the patients (Table 5). Results showed that the outpatients displayed significantly higher aggression scores than the students, and this was true for all AQ subscales. However, inpatients did not exhibit significantly higher AQ scores as compared with the male students. On the contrary, inpatients scored significantly lower on the Hostility subscale. The outpatients scored significantly higher than the students on the AQ-SF Scales, with the exception of the Verbal Aggression subscale. The inpatients scored

**Table 5**  
**Comparison of AQ and AQ-SF Scores Between Inpatients ( $n = 138$ ),**  
**Outpatients ( $n = 206$ ), and Students ( $n = 160$ ), Controlling for Age**

	Mean (Standard Deviation)			Comparisons	
	Inpatients	Outpatients	Students	Inpatients Versus Students	Outpatients Versus Students
AQ total	77.11 (15.44)	88.11 (20.26)	82.56 (16.31)	$F(2, 295) = 4.50$	$F(2, 363) = 12.24^*$
Physical Aggression	25.48 (7.09)	29.83 (7.66)	27.16 (7.10)	$F(2, 295) = 2.96$	$F(2, 363) = 9.06^*$
Verbal Aggression	14.33 (2.60)	15.34 (3.81)	14.60 (2.90)	$F(2, 295) = 0.87$	$F(2, 363) = 5.76^*$
Anger	17.58 (4.59)	20.01 (5.61)	17.93 (4.74)	$F(2, 295) = 0.65$	$F(2, 363) = 22.06^*$
Hostility	19.85 (6.65)	22.80 (6.50)	22.89 (5.84)	$F(2, 295) = 8.79^*$	$F(2, 363) = 4.96^*$
AQ-SF total	28.16 (7.47)	32.82 (10.32)	30.44 (8.41)	$F(2, 295) = 3.09$	$F(2, 363) = 13.29^*$
Physical Aggression	8.58 (3.06)	9.31 (3.34)	7.60 (3.09)	$F(2, 295) = 4.09$	$F(2, 363) = 18.48^*$
Verbal Aggression	5.96 (2.13)	7.33 (2.90)	6.96 (2.56)	$F(2, 295) = 6.78^*$	$F(2, 363) = 3.24$
Anger	6.44 (2.68)	8.01 (3.06)	7.35 (2.59)	$F(2, 295) = 5.74^*$	$F(2, 363) = 13.71^*$
Hostility	7.18 (3.01)	8.17 (3.09)	8.52 (3.05)	$F(2, 295) = 7.71^*$	$F(2, 363) = 8.79^*$

Note: AQ = Aggression Questionnaire; AQ-SF = Aggression Questionnaire–Short Form.

\* $p < .01$ .

significantly lower on the Verbal Aggression, Anger, and Hostility subscales as compared with the students, when controlling for age (Table 5).

### Validity

The validity of the AQ was determined for the inpatients, outpatients, and students by computing correlations between the AQ total scores and various measures. As can be seen in Table 6, the pattern of results was highly similar for inpatients, outpatients, and students. As predicted, AQ total scores correlated positively with NEO-FFI neuroticism, STAS, PFS-AV Hostility, and NAS Anger scores, and negatively with NEO-FFI agreeableness (Table 6). Correlations between AQ and IIS scores showed a less consistent pattern. That is, only in the samples of inpatients and students did the AQ total score correlate positively with Social Anxiety in situations pertaining to giving criticism and negatively with Social Skills in situations referring to giving a compliment. For the inpatients, but not for the outpatients, the AQ total score correlated negatively with Factor 1 (i.e., callous and remorseless use of others) of the PCL-R. The total AQ score correlated positively with the Aggressive Behavior subscale of the OSAB. Examination of the validity of the AQ-SF yielded a highly comparable pattern of results (Table 6).

### General Discussion

In a sample of violent forensic psychiatric patients as well as in a sample of secondary vocational students

(all males), the psychometric properties of the Dutch version of Buss and Perry's AQ was less satisfactory than previously observed in populations of college or university students (female and male students). For instance, the four-factor structure could not be confirmed in both samples, and the internal consistency, interitem correlations, and item–scale correlations of the Verbal Aggression subscale were found to be unsatisfactory in the subsample of inpatients and students. The two-factor structure of Williams et al. (1996) did not turn out to be a better alternative. In contrast, the psychometric properties of a 12-item version of the AQ, the AQ-SF (Bryant & Smith, 2001), were better than those of the original full-length version. The validity of the AQ and AQ-SF was sufficiently demonstrated by meaningful patterns of correlations with other measures of aggression, anger, and hostility. Yet, inpatients scored lower on the AQ and AQ-SF than the students when controlling for age differences across samples.

Our preliminary conclusion is that at the moment, empirical evidence seems to suggest that the AQ-SF has to be preferred over the AQ in populations known for violent behavior, although further investigation of the test–retest reliability of the Physical Aggression subscale has to be recommended. Our finding that the AQ-SF is superior to the original AQ is on the whole in accordance with the outcome of the studies by Diamond et al. (2005) and Diamond and Magaletta (2006). Because the AQ-SF has now been tested in American, British, Canadian, and Dutch populations, this probably means that this 12-item version of the AQ is applicable in other Western countries. However, further research is needed to verify this

**Table 6**  
**Correlations Between AQ and AQ-SF Total Scores and Scores on Other Measures for Inpatients ( $n = 138$ ), Outpatients ( $n = 206$ ), and Students ( $n = 160$ )**

Measures	Content of scale	AQ			AQ-SF		
		Inpatients	Outpatients	Students	Inpatients	Outpatients	Students
PCL-R	Psychopathy	-.09	-.07	—	-.05	.05	—
	Use of others	-.39**	.01	—	-.29**	-.01	—
	Antisocial lifestyle	.13	.13	—	.10	.11	—
NEO-FFI	Neuroticism	.48**	.44**	.20**	.46**	.45**	.30**
	Agreeableness	-.42**	-.51**	-.51**	-.33**	-.51**	-.51**
STAS	Trait anger	.45**	.68**	.62**	.43**	.63**	.59**
PFS-AV	Hostility	.30**	.47**	.44**	.22*	.42**	.41**
NAS	Anger	.74**	.60**	.76**	.66**	.55**	.71**
IIS anxiety	Criticism	.27**	.14	.18**	.25**	.17*	.26**
	Compliment	.10	.06	.10	.13	.10	.15**
IIS skills	Criticism	-.08	.13	.23**	-.06	.06	.17**
	Compliment	-.18*	-.12	-.12*	-.14	-.19*	-.18**
OSAB	Aggression on the ward	.21*	—	—	.28**	—	—

Note: AQ = Aggression Questionnaire; AQ-SF = Aggression Questionnaire–Short Form; PCL-R = Psychopathy Checklist–Revised; NEO-FFI = Five-Factor Inventory; STAS = State-Trait Anger Scale; PFS-AV = Adapted version of the Picture-Frustration Study; NAS = Novaco Anger Scale; IIS = Inventory of Interpersonal Situations, OSAB = Observation Scale for Aggressive Behavior.

\* $p < .05$ . \*\* $p < .01$ .

statement. In addition, more studies should be carried out on comparing the AQ-SF scores of aggressive (mentally ill) offenders with those of nonaggressive offenders, and of inmates with “normals.”

Two findings, namely that inpatients scored lower on the AQ and AQ-SF than secondary vocational students and that AQ and AQ-SF correlated negatively with Factor 1 of the PCL-R, need further elaboration. The lower scores of the inpatients on the AQ and AQ-SF may be explained by their restricted living environment, which gave them lesser opportunities to exhibit aggressive or violent behavior. Another explanation is that the inpatients completed the questionnaire in a more socially desirable way to avoid negative clinical and legal consequences. Such an explanation agrees with Harris' (1997) conclusion that “social desirability may be a response bias with respect to measuring self-report aggression” (p. 1052). Finally, the comparison of adult inpatients with adolescent students is questionable. We controlled for age, but the two samples still differed on a number of other variables such as socioeconomic status and a history of (unsuccessful) marriage. Although other indexes of validity supported the use of the AQ and AQ-SF in inpatient or inmate populations, it seems advisable to develop separate norms for these specific groups.

The second finding concerns the negative correlation in the inpatient sample between Factor 1 of the PCL-R and AQ or AQ-SF total scores, which is not in agreement with the studies of Walters (2003) and Guy et al. (2005). These authors found that institutional adjustment was mostly related to Factor 2 and less to Factor 1. However, their studies concerned not only male but also female participants, and nonaggressive offenders as well as aggressive offenders.

A possible solution to the aforementioned problems of self-report questionnaires in inpatient or inmate populations may be the use of production measures (e.g., PFS-AV; Hornsveld et al., 2007a). Production measures require respondents to write down their reaction to vignettes or statements, after which responses are scored by independent raters. These written responses may reflect the behaviors of the respondent more directly because they do not need to deliberate on what might be their most appropriate score on a Likert Scale. Stams et al. (2006) analyzed 50 studies evaluating the differences in moral judgment between juvenile delinquents and nondelinquent peers and came to the conclusion that larger effect sizes were found for production measures that require participants to respond to open questions than for recognition measures on which participants have to score preformulated judgments.

Recently, Buss and Warren (2000) developed a new 34-item version of the AQ. In this version, a number of items are more actively formulated and an Indirect Aggression subscale with 6 items has been added. Unfortunately, the reformulation of original items and the addition of an Indirect Aggression subscale hinder the comparison of scores on this new version with those on the original 29-item version. Although it was the intention of Buss and Warren that the meaning of the reformulated items should be unchanged, so far they have not investigated in different populations whether the “new” items yield the same results as the original items.

Further research should, in our opinion, focus on the application of the AQ-SF in countries other than those in which it has been studied until now. Regarding the AQ-SF, we advise the use of the originally formulated items, eventually in combination with the six new Indirect Aggression items of Buss and Warren (2000). Clearly, the use of self-report questionnaires in forensic psychiatric inpatient or inmate populations needs more exploration. Especially, the issue that self-report questionnaires of aggression do not seem to discriminate between these populations and control samples requires further study.

## References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- 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.
- Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21*, 343-349.
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology, 63*, 452-459.
- Buss, A. H., & Warren, W. L. (2000). *Aggression Questionnaire: Manual*. Los Angeles: Western Psychological Services.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319.
- 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.
- Crick, N. R., & Dodge, K. A. (1996). Social information processing mechanisms in reactive and proactive aggression. *Child Development, 67*, 993-1002.
- Diamond, P. M., & Magaletta, P. R. (2006). The short-form Buss-Perry Aggression Questionnaire (BPAQ-SF): A validation study with federal offenders. *Assessment, 13*, 227-240.
- Diamond, P. M., Wang, E. W., & Buffington-Vollum, J. (2005). Factor structure of the Buss-Perry Aggression Questionnaire (BPAQ) with mentally ill male prisoners. *Criminal Justice and Behavior, 32*, 546-564.
- Fossati, A., Maffei, C., Acquarini, E., & Di Ceglie, A. (2003). Multigroup confirmatory component and factor analyses of the Italian Version of the Aggression Questionnaire. *European Journal of Psychological Assessment, 19*, 54-65.
- Guy, L. S., Edens, J. F., Anthony, C., & Douglas, K. S. (2005). Does psychopathy predict institutional misconduct among adults? A meta-analytic investigation. *Journal of Consulting and Clinical Psychology, 73*, 1056-1064.
- Hare, R. D. (1991). *The Hare Psychopathy Checklist-Revised*. Toronto, Ontario, Canada: Multi-Health Systems.
- Harris, J. A. (1995). Confirmatory factor analysis of the Aggression Questionnaire. *Behaviour Research and Therapy, 33*, 991-993.
- Harris, J. A. (1997). A further evaluation of the Aggression Questionnaire: Issues of validity and reliability. *Behaviour Research and Therapy, 35*, 1047-1053.
- 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. (2005). Evaluation of aggression control therapy for violent forensic psychiatric patients. *Psychology, Crime and Law, 11*, 403-410.
- Hornsveld, R. H. J., Muris, P., & Kraaimaat, F. W. (2008a). *The Novaco Anger Scale-Provocation Inventory (1994 version) in Dutch forensic psychiatric patients and secondary vocational students*. Manuscript submitted for publication.
- Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2007a). 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.
- Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2007b). Development of the Observation Scale for Aggressive Behavior (OSAB) for Dutch forensic psychiatric inpatients with an antisocial personality disorder. *International Journal of Law and Psychiatry, 30*, 480-491.
- Hornsveld, R. H. J., Nijman, H. L. I., & Kraaimaat, F. W. (2008b). Aggression control therapy for violent forensic psychiatric patients: First results. *Psychology, Crime & Law, 14*, 1-18.
- Loeber, R., & Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence. Some common misconceptions and controversies. *American Psychologist, 53*, 242-259.
- Meesters, C., Muris, P., Bosma, H., Schouten, E., & Beuving, S. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour Research and Therapy, 34*, 839-843.
- Morey, L. C. (1991). *Personality assessment inventory*. Odessa, FL: Psychological Assessment Resources.
- Morren, M., & Meesters, M. (2002). Validation of the Dutch version of the aggression questionnaire in adolescent male offenders. *Aggressive Behavior, 28*, 87-96.
- Nakano, K. (2001). Psychometric evaluation of the Japanese adaptation of the Aggression Questionnaire. *Behaviour Research and Therapy, 39*, 853-858.
- 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: The University of Chicago Press.
- Rosenzweig, S. (1978). *Aggressive behavior and the Rosenzweig Picture-Frustration Study*. New York: 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.
- Spielberger, C. D. (1980). *Preliminary manual for the State-Trait Anger Scale (STAS)*. Tampa, FL: University of South Florida.
- Stams, G. J., Brugman, D., Dekovic, M., van Rosmalen, L., Van der Laan, P., & Gibbs, J. C. (2006). The moral judgment of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology*, 24, 697-713.
- Timmerman, I. G. H., Vastenburg, N. C., & Emmelkamp, P. M. G. (2001). The Forensic Inpatient Observation Scale (FIOS): Development and validity. *Criminal Behavior and Mental Health*, 11, 144-162.
- Van Dam-Baggen, C. M. J., & Kraaimaat, F. W. (1999). Assessing social anxiety: The Inventory of Interpersonal Situations (IIS). *European Journal of Psychological Assessment*, 15, 25-38.
- 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 Publishers.
- Walters, G. D. (2003). Predicting institutional adjustment and recidivism with the Psychopathy Checklist factor scores: A meta-analysis. *Law and Human Behavior*, 27, 541-558.
- Williams, T. Y., Boyd, J. C., Cascardi, M. A., & Poythress, N. (1996). Factor structure and convergent validity of the Aggression Questionnaire in an offender population. *Psychological Assessment*, 8, 398-403.