Assessment
Volume XX Number X
Month XXXX xx-xx

10.1177/1073191109338815
http://asmnt.sagepub.com
hosted at
http://online.sagepub.com

Reliability and Validity of the Dutch Version of the Behavioural Status Index

A Nurse-Rated Forensic Assessment Tool

Farid Chakhssi
Forensic Psychiatric Centre "de Rooyse Wissel"
Corine de Ruiter
Maastricht University
David Bernstein
Maastricht University
Forensic Psychiatric Centre "de Rooyse Wissel"

The Behavioural Status Index (BEST-Index) has been introduced into Dutch forensic psychiatry to measure change in risk level of future violence. The BEST-Index is a structured observational measure that assesses aggressive behavior, degree of insight, social skills, self-care, and work and leisure skills during inpatient treatment. Thus far, limited information regarding the psychometric properties of the Dutch version of the BEST-Index is available. The present study examines the reliability and validity of the Dutch BEST-Index in a sample of 291 mentally disordered offenders admitted to a forensic psychiatric hospital. Interrater reliability was investigated in a sample of 182 raters. Findings show that the Dutch BEST-Index can be used reliably and is significantly associated with risk of future violence and institutional aggression. Furthermore, this study revealed a different and clearer factor structure compared with the original one. Further research is needed to examine how these derived factors predict future recidivism.

Keywords: *Behavioural Status Index; forensic assessment; risk assessment; observational measure; staff rating; violence; factor analysis; interrater reliability*

The primary focus of treatment of offender patients in forensic psychiatric settings is the reduction of the risk of future (violent) offending. The ability to measure change in risk level of future violence is a prerequisite to adequately manage risk in mentally disordered offenders (Douglas & Kropp, 2002). An individual's violence risk level may be seen as changing over time, depending on context and in response to interventions (Dvoskin & Heilbrun, 2001; Heilbrun, 1997). Self-report is often used as a method of assessment in treatment outcome research. However, in forensic settings, self-report methods need to be used with prudence. Forensic patients are prone to use deception and impression management (Cima, 2003; Rogers, 1997) to influence decision making regarding release. To address these challenges with self-report, several observational measures have been developed, for example, for measuring aggression (Hornsveld, Nijman, Hollin, & Kraaimaat, 2006; Nijman, Evers, Merckelbach, & Palmstierna, 2002) and for measuring psychiatric condition (Timmerman, Vastenburg, & Emmelkamp, 2001).

A particularly promising development in the forensic field is the design of instruments to assess change in recidivism risk based on objective risk factors. For example, the Historical Clinical Risk-Management–20 (HCR-20; Webster, Douglas, Eaves, & Hart, 1997), the most widely used and extensively validated of these measures, assesses both static (i.e., historical) and dynamic (i.e., changeable) risk factors for future violence (e.g., de Vogel & de Ruiter, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004;

Authors' Note: This research was conducted at Forensic Psychiatric Centre "de Rooyse Wissel," The Netherlands. Correspondence should be addressed to Farid Chakhssi, MSc, P.O. Box 433, 5800 AK Venray, The Netherlands; e-mail: f.chakhssi@dji.minjus.nl.

Strand & Belfrage, 2001). The HCR-20 has shown good ability to predict moderate- to long-term recidivism (Douglas, Guy, & Weir, 2006). However, the HCR-20 was not originally intended to measure change in recidivism risk during forensic treatment (Douglas & Skeem, 2005). Only 10 of the HCR-20's items measure dynamic risk factors, which may limit its suitability for measuring change in risk levels. For this reason, instruments are needed that focus primarily on assessing change in dynamic risk factors during forensic treatment.

Reed, Woods, and Robinson (2000) adapted the Behavioural Status Index (BEST-Index), a broad-spectrum structured observational measure applied by nurses, for use in forensic psychiatric samples (Ross et al., 2008; Woods, 2000). The BEST-Index was first developed as an instrument for assessing change during general psychiatric treatment and during patients' transition from a psychiatric hospital to community-based care (Mahgoub, 1988). For use in forensic settings, a risk scale was added to the BEST-Index, and the original scales were modified for forensic use (Robinson, Reed, & Lange, 1996). Recently, two subscales on self-care and work skills were added (Reed et al., 2000). BEST-Index assessment is now specifically aimed at measuring five behavioral domains, including aggressive behavior, degree of insight, social skills, self-care, and work and leisure skills.

Thus far, few studies have investigated the psychometric properties of the BEST-Index. Research on the factor structure of the BEST-Index in a sample of mentally disordered offenders (N = 503) yielded an ambiguous factor solution (Woods, Reed, & Collins, 2005). In this study, the authors investigated the factor structure of the first three subscales: Risk, Insight, and Social Skills, which resulted in an overfactored model consisting of 11 factors, where one factor had no salient loadings and one factor consisted of only one item. Other attempts to investigate the factor structure were based on the same sample and more explorative in nature. For example, factor analyses were limited to a single subscale (Woods, Reed, & Collins, 2001a; Woods, Reed, & Collins, 2001b; Woods, Reed, & Collins, 2001c) or to two subscales (Woods, Reed, & Collins, 2003a; Woods, Reed, & Collins, 2003b; Woods, Reed, & Collins, 2004). Although these studies resulted in several factors per subscale or per two subscales, the identified structures did not emerge in a recent factor analytic study on the same sample containing all three subscales (Woods et al., 2005).

Internal consistency, test–retest stability, and interrater reliability on the item level of the BEST-Index subscales Risk, Insight, and Communication have been reported by Woods, Reed, and Robinson (1999). Interrater reliabilities were weak to good, test–retest stabilities were good, and, as a measure of internal consistency, item-to-subscale correlations were weak to good. With regard to the convergent validity of the BEST-Index, statistically significant correlations were found between BEST-Index items and HCR-20 items (ranging from .41 to .44, Spearman ρ correlations) in a forensic sample (Ross et al., 2008).

The possibility of using the BEST-Index as an inpatient observational measure of future violence risk served as an impetus to the introduction of the instrument in several Dutch forensic psychiatric hospitals. The instrument was translated into Dutch (van Erven, 1999), and nursing staff were trained in the use of the assessment tool. The present study examines the reliability and validity of the Dutch version of the BEST-Index in a sample of forensic psychiatric inpatients. First, interrater reliability of the BEST-Index total score and subscales was studied. Second, the internal reliability was examined by means of internal consistency and item homogeneity parameters. Third, the goodness of fit of the three original subscales of the BEST-Index was examined by means of confirmatory factor analysis. Because of inadequate fit (described later), the underlying latent constructs were investigated by means of principal components analysis (PCA). Finally, convergent and predictive validity was examined by measuring the association between the BEST-Index and the HCR-20 and institutional aggressive behavior.

Method

Setting

This study was conducted at Forensic Psychiatric Centre "de Rooyse Wissel" (dRW), a Dutch maximum security hospital for the treatment of mentally disordered offenders who were hospitalized under the Dutch judicial measure of *TerBeschikkingStelling* (TBS order). A TBS order can be imposed on offenders who have committed serious offences, carrying a punishment of at least 4 years imprisonment, and who suffer from a mental disorder according to American Psychological Association's *Diagnostic Manual and Statistical Manual of Mental Disorders*, text revision (*DSM-IV-TR*) criteria (American Psychiatric

Table 1
Sample Characteristics (N = 291)

Age (years)	38.2 (*SD* = 9.5)
Main index offense	
Homicide offense (including attempt)	39.4%
Sexual offense	23.3%
Violent theft, robbery, and assault	23.0%
Property offense	1.0%
Arson	9.1%
Criminal history	
Prior convictions	80.1%
Age at first conviction	22.8 years (*SD* = 8.2)
Prior TBS	8.2%
Mental health history	
Prior contact with mental health services (n = 287)	72.5%
Age at first mental health contact (n = 205)	17.6 years (*SD* = 9.6)

Note: TBS = *TerBeschikkingStelling*.

Association, 2000). The TBS order is prolonged as long as the court deems the patient a danger to society. The hospital has 229 residential treatment beds for male offenders, divided over three geographical locations.

Sample

The study sample consists of 291 male mentally disordered offenders admitted to dRW. These patients where admitted under the TBS order during the period between March 1, 2000, and December 1, 2007. *DSM-IV-TR* Axis I and Axis II diagnoses were extracted from the patient files. These are based on clinical diagnoses made by psychiatrists, supported by findings from psychological assessments performed by clinical psychologists.

The characteristics of the study sample are presented in Table 1. Mean age was 38.2 years (*SD* = 9.5 years). In the sample, 39.5% of the patients had committed (attempted) homicide; 23.3% had committed sexual offenses; 23% had committed violent theft, robbery, or assault; 9.1% had committed arson; and 1% had committed property offenses. Axis I and/or Axis II diagnoses were available for 287 patients. Classifications for four patients were deferred, awaiting additional (neuro-)psychological assessments. Patients were classified according to *DSM-IV-TR* Axis I disorder categories, if they met the diagnostic criteria for at least one disorder belonging to that category. Two thirds of the patients (65.9%) met the diagnostic criteria for at least one substance-related disorder. Seventy-one patients (24.7%) met criteria for schizophrenia or other psychotic disorders. Fifty-seven patients (19.6%) met criteria for a paraphilic disorder, 17 patients (5.9%) for mood disorders, 19 patients (6.6%) for anxiety disorders, 23 patients (8.0%) for impulse control disorders, 21 patients (7.3%) for pervasive developmental disorders, 25 patients (8.7%) for attention deficit and disruptive behavior disorders, and 14 patients (4.9%) for other *DSM-IV-TR* Axis I disorder categories. In terms of Axis II personality disorders (PD), the most prevalent PD (56.1%) was a mixed PD or PD not otherwise specified (NOS). The prevalence of having any PD (including PD NOS) in this sample was 81.1% (236 patients). Subsequent to mixed PD and PD NOS, the most prevalent PD was antisocial PD (18.5%). The other PDs were much less prevalent: three with paranoid PD (1%), two schizoid PD (0.7%), one schizotypical PD (0.3%), 16 borderline PD (5.6%), four histrionic PD (1.4%), 25 narcissistic PD (8.6%), two avoidant PD (0.7%), and four obsessive-compulsive PD (1.4%). No patient met the criteria for dependent PD.

Measures

BEST-Index. The BEST-Index (Reed et al., 2000) contains 70 items divided a priori among three subscales: Risk, Insight, and Communication and Social Skills. Each item can be scored on a 5-point scale ranging from 1 (*worst case*) to 5 (*optimal case*). The Risk scale contains 20 items related to dangerous behaviors, such as violence to others and general disruptive and antisocial behaviors. The Insight scale consists of 20 items measuring the level of insight into the nature of the patient's problems, antecedent events leading to their current situation, and attribution of responsibility. The Communication and Social Skills subscale contains 30 items on adaptive social behavior, social skills, and interpersonal relationships. The BEST-Index is scored over the previous 6-month period, and the time needed to complete an assessment varies between 1 and 2 hours. An example of an item, item description, and the scoring categories are given in Table 2.

Test–retest reliabilities of the BEST-Index scales and items were satisfactory in an earlier study in a U.K. forensic psychiatric sample (N = 100; Woods et al., 1999). Reliabilities (Spearman ρ correlations) over a 2-week interval ranged from .84 to .89 for the subscales and from .77 to .95 for the items. Item–to-subscale correlations (Spearman ρ) varied between .14 and .81. Interrater reliabilities (N = 37; Spearman

Table 2
Illustration of a BEST-Index Item and Scoring Categories

Insight Subscale Item 2: Description of Tension				
2. Description of Tension				
2.1 Unable clearly to describe such feelings	2.2 *With support and encouragement* occasionally describes such feelings reasonably clearly	2.3 Occasionally *spontaneously* describes such feelings reasonably clearly	2.4 Usually *spontaneously* describes such feelings reasonably clearly	2.5 Always *spontaneously and clearly* describes such feelings and their meanings

Note: BEST-Index = Behavioural Status Index. *Definition for Description of Tension:* In conversation, the patient is able to describe and illustrate his/her experience while undergoing attacks of increased tension and nervousness.

ρ correlations) for the BEST-Index items varied from .03 to .77, with a mean of .43 for all items. In addition, items from the BEST-Index subscales Risk and Insight have been shown to be significantly associated (ranging from .41 to .44, $p \leq .01$; Spearman ρ correlations) with the HCR-20 Clinical items Lack of insight, Negative attitudes, Active symptoms of major mental illness, Impulsivity, and Unresponsive to treatment as well as with the HCR-20 Risk-management item Lack of personal support (Ross et al., 2008). The BEST-Index was translated into Dutch (van Erven, 1999), in collaboration with the original authors, independent from the authors of the present study.

HCR-20. The HCR-20 (Webster et al., 1997) is specifically designed for assessing the risk of future violence among persons with mental disorders. It consists of 20 items, each rated from 0 to 2 (0 = *does not apply*; 1 = *applies somewhat*; 2 = *definitely applies*), divided among three scales: Historical (H), Clinical (C), and Risk Management (R). The H scale focuses on previous violent and antisocial behavior, the C scale on clinical aspects related to violence risk, and the R scale on future situational factors. In contrast to the 10 items of the H scale, the 5 items of the C scale and the 5 items of the R scale are amenable to change over time. The HCR-20 is intended to structure professional judgment about violence risk, and raters are asked to make a final risk judgment of low, moderate, or high (HCR-20 manual; Webster et al., 1997). Research studies within diverse populations and in several countries have shown that the HCR-20 scores and final risk judgments can be derived reliably and are related to long-term violent recidivism (see Douglas et al., 2006).

START Outcome scale. The START Outcome scale is a modified version of the Overt Aggression Scale (OAS; Yudofsky, Silver, Jackson, Endicott, & Williams, 1986) for use in forensic clinical practice (Nicholls, Brink, Desmarais, Webster, & Martin, 2006). The modified version consists of 11 categories: (1) verbal aggression, (2) physical aggression against objects, (3) physical aggression against self, (4) physical aggression against other people, (5) self-neglect, (6) substance abuse, (7) victimized by others, (8) sexual aggression, (9) unauthorized leave, (10) suicidal behavior, and (11) stalking. Each category of adverse outcome is rated according to its severity on a 4-point scale from *least severe* (1) to *most severe* (4). Prior research shows that the START Outcome scale can be reliably coded from patient files (intraclass correlation coefficient [ICC] = .70; Nicholls et al., 2006).

Raters and Training

BEST-Index raters were 182 psychiatric nurses (43.9% male) with an average of 3.6 years of experience (SD = 2.5) at the dRW and a mean age of 35.8 year (SD = 8.8). All raters attended a half-day workshop, given by the first author and/or a coteacher (researcher or a masters-level research assistant). During the workshops, nurses were provided with the BEST-Index manual and informed about the nature and purpose of the BEST-Index. Second, the nurses were informed about the procedure for observational assessments of behavior. In addition, the nurses were presented with case vignettes and asked to rate several items of the BEST-Index. Finally, the scores were reviewed and discussed between the nurses and the trainer(s).

Procedure

Approval for the study was obtained from the hospital's executive board and the institutional research board.

BEST-Index assessments. Preferably, two psychiatric nurses assessed every patient using the BEST-Index 6 months after admission. Assessments were performed by at least the primary nurse on the ward where the patient resided and one other nurse. On average, there were 4.7 days (SD = 8.0) between the two independent assessments. The scores were recorded in an electronic database with a Web-based interface that had the same layout as the paper version of the BEST-Index. Scores from this database could be imported in the data analysis software.

HCR-20 assessments. Ratings were performed by mental health professionals, who also coordinated the treatment of these patients. All the raters were trained and experienced in using the HCR-20. HCR-20 assessments are performed regularly when the TBS order of the patient is evaluated (mostly every year, sometimes every 2 years). We conducted convergent validity analyses on patients who had at least one HCR-20 assessment (n = 224). For reliability analysis, 48 cases were coded twice (by two trained masters-level psychology research assistants), and the ICC for the HCR-20 total score was satisfactory at .76. For the individual HCR-20 scales, the ICCs were .72 for the H scale, .75 for the C scale, and .66 for the R scale (single measure ICCs, absolute agreement type).

START Outcome scale assessments. Official registration of institutional misconduct was used to detect aggressive behaviors during treatment. Possible antecedents, motive, and a detailed description of the observed misconduct are registered by psychiatric nurses at each occurrence. To classify this misconduct in terms of type and severity, all misconducts during a patient's stay were recoded by three raters (a masters-level student and two masters-level research assistants) with the English version of the START Outcome scale (Nicholls et al., 2006). To ensure a prospective design, all misconducts that occurred prior to the first assessment of the BEST-Index were eliminated from the predictive analyses. Although coded retrospectively, because of the detailed official registration, this procedure is not likely to underestimate the actual prevalence of inpatient misconduct. More than 300 incidents were coded by two raters to permit reliability analysis. In this study, the ICC for two raters was satisfactory at .83 (mean ICC; n = 369).

Statistical Analysis

Interrater reliability. For estimating the interrater reliability, the two-way random effects model, with measures of absolute agreement, of the ICC was used (Shrout & Fleiss, 1979). Nurses in this study are a random sample from all possible raters, and patients are also a random factor. In this study, 262 patients were rated twice because of rater attrition for several reasons (e.g., absence of the second rater due to illness or leave, the patient was transferred to another ward while a second rating was pending). Interrater reliability for the BEST-Index total and subscale scores as well as for the individual BEST-Index items was examined.

Internal consistency and item homogeneity. Internal consistency was examined using Cronbach's alpha for the BEST-Index total score as well as for the subscales. Mean interitem correlations were used as a measure of item homogeneity. Mean scores between raters, or when unavailable the scores of the single raters, were used for these and subsequent analyses.

Confirmatory factor analysis. The goodness-of-fit of the original three-factor model of the BEST-Index (Woods, 2000) was tested using confirmatory factor analysis. The model was fit using AMOS 7.0, employing the maximum likelihood procedure (Arbuckle, 2006). Each item was assumed to load only on its own subscale. Given the expected correlations between the subscales (Woods, 2000), the three subscales were allowed to correlate. Model fit was evaluated using the chi-square test, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker Lewis index (TLI; also known as the nonnormed fit index; Byrne, 2001). A CFI and TLI value above .90 and an RMSEA value below .05 are indicative for an adequate fit (Hu & Bentler, 1999).

Principal components analysis with varimax rotation. Because of an inadequate fit (described later), we examined the latent constructs underlying the BEST-Index using exploratory PCA. The extracted communalities, which measure the percentage of variance in the variables explained by the extracted factors, were used to select variables to include in the analysis. Items with low communalities, meaning that the components explain little variance for these variables, are indicative of items that are unrelated to the domains of interest and should be removed from

further analyses (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Items with communalities below .20 were removed from further analyses. Finally, varimax rotation was used to test the relative suitability of the resulting factor structure.

Convergent and predictive validity. Pearson product–moment correlations were used to measure the association between the BEST-Index scores and the level of violence risk (HCR-20). Convergent validity would be supported when the BEST-Index is stronger and statistically significantly correlated with the dynamic risk factors of the HCR-20 (i.e., the C and R scales) compared with the static risk factors of the H scale. Furthermore, predictive validity of the BEST-Index was examined by computing Pearson correlation coefficients between the BEST-Index scales and institutional adverse outcomes (START Outcome scale). To test the BEST-Index relation to violence, only the categories related to aggression to others will be used in this study. For the aggressive behavior, total score on START Outcome categories verbal aggression, physical aggression against objects, and physical aggression toward other people was computed by summing the incidences for each category during a patient's stay (mean length of stay in hospital = 3.2 years; *SD* = 2.2 years). Analyses are performed using the Statistical Package for the Social Sciences, Version 13.0 (SPSS Inc., 2005).

Results

Confirmatory Factor Analysis

The goodness-of-fit indices for the original BEST-Index three-factor model did not meet the required cutoff values. The chi-square test yields a statistic of 7050.81 (df = 2342; N = 279), which has a corresponding p value of .000. This p value means that the null hypothesis of a good fit should be rejected. Additionally, the RMSEA is .09, the CFI is .65, and the TLI is .64, offering further evidence that the original model does not fit the present data well. In addition, inspection of the parameter estimates revealed that Risk subscale item 1 (critical ratio = 1.68; p = .093) and Item 16 (critical ratio = 1.59; p = .111) did not significantly predict the Risk subscale, suggesting that removing these items could improve model fit. Also, the error terms covariances between items revealed that seven pairs of items had high modification indices (>80). These items pairs were items 13-14 and 2-4 from the Risk subscale and items 2-3, 13-14, 19-20, 19-23, and 20-23 from the Insight subscale. The high modification indices suggest that the model fit will improve if the model is respecified by allowing these error terms to covary. A stepwise approach by first removing the Risk subscale Items 1 and 16 from the model and subsequently allowing the items with the highest modification indices to covary resulted in a significantly improved model, that is, change in $\chi^2(df = 142; N = 279) = 1094.5$, $p < .001$. However, the model still did not meet the cutoff values indicating a good model fit, $\chi^2(df = 2{,}200; N = 279) = 5956.3$, p = .000, RMSEA = .08, CFI = .72, and TLI = .71. Thus, an exploratory PCA was deemed necessary to reveal the instrument's factor structure.

Principal Components Analysis

Initial PCA with varimax rotation yielded 14 components with eigenvalues greater than 1, accounting for 70.0% of the total variance. In addition, the scree plot shows a steep drop over the first four components, followed by a leveling off for the remaining components. The first four components accounted for 50.5% of the total variance. The first component accounted for 31.7% of the total variance, whereas the other three components accounted for 9.5%, 5.2%, and 4.2%, respectively. These findings suggest that the four components offer the optimal solution. The extracted communalities of the items for the four components demonstrated low communalities (<.20) for seven items from the Risk subscale. These items and their communalities were as follows: Risk Item 1 (*Family support*, .04), Item 6 (*Serious self-harm*, .08), Item 7 (*Superficial self-harm*, .10), Item 15 (*Inappropriate sexual behaviors*, .11), Item 16 (*Sadomasochistic behaviors*, .08), Item 17 (*Macho gear and adornment*, .05), and Item 19 (*Substance abuse*, .06). These items were subsequently removed. PCA with varimax rotation on the remaining 63 items now yielded 11 components with eigenvalues greater than 1, accounting for 70.4% of the total variance. Again, the scree plot shows a steep drop over the first four components, followed by a leveling off for the remaining components. Now, the first four components accounted for 55.5% of the total variance. The first component accounted for 35.1% of the total variance, whereas the other three components accounted for 10.2%, 5.7%, and 4.5%, respectively. The pattern matrix with loadings after rotation, with the items

sorted by the size of their loadings, is displayed in Table 3. Factor 1 comprises 23 items and seems to address *Social skills*. It consists primarily of items from the original Social Skills subscale. The highest loadings were found for Social Skills Item 28 (*Ease of communication*, .76), Social Skills Item 23 (*Assertiveness*, .74), and Social Skills Item 19 (*Expressing opinions*, .73). The weakest loading was found for Social Skills Item 4 (*Body posture*, .49). Factor 2 consists of 21 items and seems to address *Insight*, as demonstrated by items primarily from the original Insight subscale. The highest loadings were found for Insight Item 16 (*Self-appraisal*, .76), Insight Item 2 (*Description of tension*, .76), and Insight Item 17 (*Prioritization of problems*, .75). The weakest loading was found for Insight Item 3 (*Tension reducing strategies*, .50). Factor 3 consists of 14 items, and the highest loadings were found for Social Skills Item 13 (*Turn-taking*, .72), Risk Item 9 (*Verbal aggression following trigger event*, .71) and Social Skills Item 14 (*Listening skills*, .67). The weakest loadings were found for Social Skills Item 14 (*Imitative disruption*, .46). These items are related to aggressive and dominant behavior in interpersonal communication and, hence, Factor 3 was named *Interpersonal hostility*. Finally, Factor 4 consists of seven items related to *Physical violence*. The highest loadings were found for Risk Item 2 (*Serious violence to others without apparent trigger event*, .79), Risk Item 4 (*Minor violence to others without apparent trigger event*, .78), and Risk Item 3 (*Serious violence to others following trigger event*, .72). The weakest loadings were found for Risk Item 10 (*Attacks on objects without apparent trigger event*, .53).

The mean scores on the factors for the sample are displayed in Table 4. The mean scores were 85.34 (*SD* = 16.39) for the Social Skills factor, 56.42 (*SD* = 15.63) for the Insight factor, 47.16 (*SD* = 7.59) for the Interpersonal Hostility factor, and 33.31 (*SD* = 2.91) for the Physical Violence factor. Internal consistency[1] for the factors is also displayed in Table 4, and ranged from α = .74 for the Physical Violence factor to α = .96 for the Social Skills factor. Finally, homogeneity of the items ranged from .38 for the Interpersonal Hostility factor to .51 for the Insight factor.

Interrater Reliability

The single-measure ICC based on absolute agreement for the BEST-Index total score containing 63 items was .71. For the BEST-Index factors extracted in the present study (Table 3), the coefficients were .71 (Social Skills factor), .66 (Insight factor), .69 (Interpersonal Hostility factor), and .68 (Physical Violence factor). The coefficients for the individual items varied from .31 to .63 (median = .50) for the Social Skills factor, from .35 to .65 (median = .47) for the Insight factor, from .35 to .73 (median = .44) for the Interpersonal Hostility factor, and from .12 to .66 (median = .47) for the Physical Violence factor. The lowest single-measure coefficient was obtained for Item 2 of the Risk subscale (*Serious violence to others without apparent trigger event*, .12). Item 9 of the Risk subscale (*Verbal aggression following trigger event*, .73) obtained the highest coefficient.[2]

Convergent and Predictive Validity

In this sample, the mean HCR-20 score was 25.1 (*SD* = 6.19). For the HCR-20 scales, the mean score for the H scale was 13.19 (*SD* = 3.39), for the C scale 4.78 (*SD* = 2.19), and 7.17 (*SD* = 2.59) for the R scale. Most patients (*n* = 146; 65.1%) received a final risk judgment of high risk, whereas 46 (20.5%) patients were judged to pose a moderate risk and 32 (14.3%) patients a low risk. Regarding the convergent validity of the BEST-Index, the correlations of the original BEST-Index and the revised BEST-Index (containing 63 items) with the HCR-20 are displayed in Table 5. From Table 5, it can be seen that for the original BEST-Index, out of 20 possible correlations, 17 were significant. Significant correlations ranged from .15 (for the BEST-Index Risk subscale and HCR-20 final risk judgment) to .45 (for original BEST-Index total score with HCR-20 C scale). For the revised BEST-Index, out of 25 possible correlations, 21 were significant. Significant correlations ranged from .15 (for the revised BEST-Index Physical Violence factor and HCR-20 R scale) to .45 (for revised BEST-Index total score with HCR-20 C scale). Overall, the differences in the correlation matrix between the original and the revised BEST-Index original were small, with the exception of the higher correlations between the revised BEST-Index Interpersonal Hostility and Physical Violence factors with the Historical Scale of the HCR-20.

During their stay, most of the patients (*n* = 200; 68.7%) displayed acts of verbal aggression, 100 patients (34.4%) displayed acts of physical aggression against objects, and 134 patients (46.0%) displayed physical aggression toward others. The Pearson correlations between the revised BEST-Index scores, the HCR-20 scores, and the START Outcome scales are presented in Table 6.[3] The highest correlation coefficient between the revised BEST-Index scores and the START Outcome scales is found for the association

Table 3
Item Loadings of BEST-Index Four-Factor Solution With Varimax Rotation

BEST-Index		Factor			
Items	Description	1	2	3	4
SOCIAL28	Ease of communication	**.76**	.34	.15	
SOCIAL23	Assertiveness	**.74**	.16	–.26	
SOCIAL19	Expressing opinions	**.73**	.27	–.18	
SOCIAL20	Disagreement	**.72**	.15	–.30	
SOCIAL22	Making requests	**.71**	.24		
SOCIAL12	Fluency	**.71**	.35		
SOCIAL16	Conversational topics	**.70**	.39	.32	
SOCIAL10	Conversational initiative	**.69**	.40		
SOCIAL11	Amount of speech	**.68**	.40	.36	
SOCIAL29	Sociability and support	**.67**	.41	.11	
SOCIAL1	Facial expression	**.66**	.28	.20	
SOCIAL7	Tone of voice	**.66**	.29		
SOCIAL25	Social activities	**.65**	.35		.11
SOCIAL5	Expressive gestures	**.64**	.29	.12	
SOCIAL2	Eye contact	**.62**	.30	.27	
SOCIAL9	Verbal delivery	**.58**	.16	.31	
SOCIAL15	Response to questions	**.58**	.35	.41	
SOCIAL3	Orientation to others	**.57**	.27	.28	
SOCIAL24	Self-presentation	**.56**	.16	.14	
SOCIAL26	Emotional control	**.53**	.19	.47	.20
SOCIAL27	Relationship with others	**.53**	.28	.42	
SOCIAL8	Voice modulation	**.51**	.14	.35	
SOCIAL4	Body posture	**.49**	.17	.38	
INSIGH16	Self-appraisal	.19	**.76**	.29	
INSIGH2	Description of tension	.24	**.76**		
INSIGH17	Prioritization of problems	.17	**.75**	.31	
INSIGH6	Tension-producing events	.31	**.74**		
INSIGH1	Awareness of tension	.19	**.73**		
INSIGH5	Tension-producing thoughts	.28	**.73**		.13
INSIGH12	Events producing insecurity	.30	**.72**		
INSIGH4	Recognition of negative feelings	.25	**.70**	–.22	
INSIGH14	Antecedent events leading to treatment	.15	**.69**	.23	
INSIGH15	Ascription of responsibility	.13	**.68**	.32	
INSIGH18	Goal planning	.25	**.68**	.37	
INSIGH20	Expectations	.25	**.67**	.40	
INSIGH8	Identifying relaxing thoughts	.34	**.64**	.15	.14
INSIGH13	Events producing security	.44	**.62**	.12	
INSIGH7	Personal strategy for reducing tension	.32	**.61**	.13	.16
INSIGH11	Attributes liked in others	.47	**.56**		
INSIGH9	Identifying relaxing activities	.39	**.56**		
SOCIAL18	Frankness	.34	**.55**	.19	
INSIGH10	Attributes disliked in others	.44	**.52**	–.17	.13
INSIGH19	Compliance with therapy	.28	**.51**	.42	.14

(continued)

Table 3 (continued)

BEST-Index		Factor			
Items	Description	1	2	3	4
INSIGH3	Tension-reducing strategies	.34	**.50**		.12
SOCIAL13	Turn-taking			**.72**	
RISK9	Verbal aggression following trigger event	–.13		**.71**	.43
SOCIAL14	Listening skills	.40	.14	**.67**	
SOCIAL17	Egocentric conversation	.24	.14	**.67**	
SOCIAL21	Arguments	.20	.12	**.62**	.33
RISK13	Disruptive episodes	–.19		**.58**	.17
SOCIAL6	Social distance	.31	.11	**.57**	.11
SOCIAL30	Deferring to others	.17	.30	**.55**	
RISK8	Verbal aggression without apparent trigger event			**.54**	.52
RISK18	Obsessive compulsive behaviors			**.49**	.48
RISK20	Psychiatric disturbance	.29		**.47**	.12
RISK14	Imitative disruption			**.46**	.20
RISK2	Serious violence to others without apparent trigger event		.11		**.79**
RISK4	Minor violence to others without apparent trigger event				**.78**
RISK3	Serious violence to others following trigger event			.11	**.72**
RISK11	Attacks on objects following trigger event	.10		.26	**.63**
RISK5	Minor violence to others following trigger event			.19	**.60**
RISK12	Breaches of security			.42	**.56**
RISK10	Attacks on objects without apparent trigger event	.17		.16	**.53**
Proportion of variance explained per factor (%)		35.1	10.2	5.7	4.5

Note: BEST-Index = Behavioural Status Index. Factor loadings <.10 are suppressed. Salient loadings are in boldface.

Table 4
Labels, Mean Scores, Cronbach's Alpha, Item Homogeneity, and Interrater Reliability for the BEST-Index Four-Factor Solution in a Sample of Forensic Psychiatric Patients

	Total Sample (N = 291)				
Scale	i	α	Mean (SD)	Item Homogeneity	ICC[a]
Revised BEST total[b]	63	.97	222.69 (34.74)	.31	.71
Social skills	23	.96	85.34 (16.39)	.49	.71
Insight	21	.95	56.42 (15.63)	.51	.66
Interpersonal hostility	12	.86	47.16 (7.59)	.38	.69
Physical violence	7	.74	33.31 (2.91)	.40	.68

Note: BEST-Index = Behavioural Status Index; i = number of items; ICC = single measure intraclass correlation coefficient, absolute agreement.
a. N = 262.
b. The BEST-total is the sum score of 63 items; 7 items were omitted because of low communalities.

Table 5
Pearson Correlation Coefficients Between the Four Factors of the BEST-Index and the HCR-20 Scores (n = 224)

	HCR-20				
Scale	H Scale	C Scale	R Scale	Total Score[a]	Risk Judgment
Revised BEST total[b,c]	.01	.45**	.42**	.33**	.34**
Social skills	.05	.31**	.30**	.21**	.27**
Insight	.09	.43**	.44**	.29**	.35**
Interpersonal hostility	.20**	.41**	.29**	.38**	.23**
Physical violence	.21**	.29**	.15*	.28**	.13
Original BEST total	.00	.45**	.42**	.34**	.34**
Risk subscale	.28**	.40**	.23**	.39**	.15*
Insight subscale	.09	.43**	.44**	.29**	.35**
Communication and social skills subscale	.01	.34**	.32**	.25**	.29*

Note: BEST-Index = Behavioural Status Index; HCR-20 = Historical Clinical Risk-Management–20.
a. The sum of all 20 items of the HCR-20. Risk judgment is the final judgment of future violence risk.
b. BEST-Index scores have been reversed to correspond with the direction of the HCR-20 scores; higher scores on BEST-Index scales mean worse functioning on that scale.
c. The Revised BEST-total is the sum score of 63 items; 7 items were omitted because of low communalities.
*p < .05, two-tailed. **p < .01, two-tailed.

between the Interpersonal Hostility factor and START verbal aggression (r = .40; p < .01). The lowest, but still significant correlation coefficient is found for the relationship between physical aggression toward others and the Social Skills factor (r = .13, p < .05). For the relationship between the HCR-20 and the START outcome scales the highest correlation coefficient is found for the H scale and START verbal aggression (r = .30; p < .01). The lowest significant correlation is found for the association between the C scale and START physical aggression toward others (r = .19; p < .01).

Discussion

The present study examined the reliability and factor structure of the Dutch version of the BEST-Index in a sample of 291 forensic psychiatric inpatients. Confirmatory factor analysis demonstrated that the original, a priori three-factor structure could not be replicated in our sample. In addition, two items had no significant contributions to the BEST-Index Risk subscale, and several item pairs of both the Risk and the Insight subscales had strong residual correlations. Correcting for these shortcomings resulted in a significantly improved model fit, but not to an acceptable overall model fit for the data in this sample. These findings are in line with the previous factor analytic study of the BEST-Index (Woods et al., 2005), which also did not support the original a priori factor structure.

The subsequent PCA resulted in a four-factor solution, which primarily indicated that (1) the Risk and Insight subscales could be retained; (2) a number of items from the original Risk and Social Skills subscales relating to interpersonal dominance and aggressive behavior formed a new, third factor; and (3) the remaining items of the Risk subscale referring to physical violence also formed another, fourth factor. Seven items were omitted from the BEST-Index because they were unrelated to any of the factors. Our findings are in contrast with earlier factor analytic work on the BEST-Index. Woods et al. (2005) identified 11 factors by selecting the factors solely on the basis of Kaiser's eigenvalue greater than 1 criterion. Although this procedure is appealing in its simplicity, exclusive reliance on this approach can lead to solutions with too many factors (overfactoring), which have little theoretical and/or practical value. In determining the optimal number of factors, several statistical (e.g., scree test) and theoretical procedures (e.g., conceptual clarity of the solution) should be employed (Fabrigar et al., 1999). Thus, the present study extended on the previous factor analytic studies by taking these considerations into account. Earlier factor analytic studies on the separate subscales (Woods et al., 2001a; Woods et al., 2001b; Woods et al.,

Table 6
Pearson Correlation Coefficients Between the Four Factors of the Revised BEST-Index and Challenging and Aggressive Behaviors as Measured With the START Outcome Scale (n = 291)

	START Outcome Scale[a]		
Scale	Verbal Aggression	Physical Aggression Against Objects	Physical Aggression Toward Others
Revised BEST total[b,c]	.20**	.15**	.20**
Social skills	.11	.06	.13*
Insight	.11	.08	.10
Interpersonal hostility	.41**	.27**	.34**
Physical violence	.20**	.25**	.23**
HCR-20 total score[d]	.32**	.21**	.26**
H scale	.29**	.20**	.27**
C scale	.27**	.18**	.21**
R scale	.20**	.12	.12

Note: BEST-Index = Behavioural Status Index; HCR-20 = Historical Clinical Risk-Management–20.
a. Verbal Aggression was displayed by 200 patients (68.7%), Physical Aggression Against Objects by 100 patients (34.4%), and Physical Aggression by 134 patients (46.0%).
b. BEST-Index scores have been reversed; higher scores on BEST-Index scales correspond with worse functioning on that scale.
c. The revised BEST-total is the sum score of 63 items; 7 items were omitted because of low communalities.
d. n = 214.
*p < .05, two-tailed. **p < .01, two-tailed.

2001c) and on two subscales combined (Woods et al., 2003a; Woods et al., 2003b; Woods et al., 2004) yielded different results in each instance, leading to a lack of conceptual clarity surrounding the factor structure of the BEST-Index. This is unfortunate, because the BEST-Index deserves a more systematic approach toward instrument development. Our findings obtained by this more systematic approach suggest that a four-factor model is a better solution than the original three-factor model.

The interrater reliabilities for the nurse raters were good for the four factors of the BEST-Index. These interrater reliabilities are consistent with those found in the only other interrater reliability study of the BEST-Index in a sample of 37 raters (Woods et al., 1999). Note that in the current study we had a sample of 182 raters and used the stringent single-measure ICC as the parameter for interrater reliability (Shrout & Fleiss, 1979). Internal consistency of the Dutch version of the revised BEST-Index and the derived factors was excellent. Item homogeneity was acceptable for the revised BEST-Index total score and the derived factors. Similar values for internal consistency have previously been reported (Woods et al., 1999).

Good convergent validity of the Dutch version of the BEST-Index with the HCR-20 was found. As expected, the scores on the revised BEST-Index total score and the derived factors were more strongly associated with the changeable (i.e., dynamic) violence risk factors of the HCR-20, compared with the historical (i.e., static) factors of the HCR-20. Thus, it seems that the BEST-Index is especially valuable as a measure of dynamic risk factors. Values for convergent validity of the BEST-Index were in the same range as previously reported by Ross et al. (2008). Furthermore, moderate to large correlation coefficients were found between the revised BEST-Index factors with diverse aggressive behaviors during institutional treatment, supporting its predictive validity. Moreover, in this study, the predictive results for the revised BEST-Index were comparable to, if not higher than, the predictive results for the HCR-20. Especially, the derived BEST-Index factor Interpersonal Hostility shows stronger associations with the diverse inpatient aggressive behaviors than the HCR-20. Note that the assessments with the BEST-Index and the HCR-20 were independent of each other and both preceded the measurement of the START Outcome variables.

The findings of the current study should be considered with several limitations in mind. The generalizability of the findings is limited to inpatient male forensic psychiatric patients. The study should be replicated in different samples (e.g., female forensic psychiatric patients) to assess the robustness of our findings. In addition, the sample size (N = 291) was

somewhat small for the CFA and PCA analyses in relation to the number of BEST-Index items (N = 70). However, recommendations concerning the number of subjects required per item of the instrument vary widely within the literature: from 4 subjects per variable (Gorsuch, 1983) to 10 subjects per variable (Nunnally, 1967). MacCallum, Widaman, Zhang, and Hong (1999) demonstrated that the recommended sample size is influenced by the extent to which factors are overdetermined and the level of communalities of the measured variables. In our sample, the average communality was .55, and 7 to 23 variables represented each factor. According to MacCallum et al., these are moderate to good conditions for a factor analysis, and one can obtain good recovery of factors with samples of at least 200. However, further studies should be performed using the BEST-Index with larger samples to test whether the current factor structure can be replicated.

Despite these limitations, the findings of the present study suggest that the Dutch version of the BEST-Index can be used reliably in Dutch forensic psychiatric settings, with four factorial domains that have apparent theoretical and clinical relevance, as demonstrated by their relationship with violence risk level and institutional aggressive behavior. Modifying the Dutch version of the BEST-Index was not central to this study, but our findings indicate that future research should further examine the suitability of a number of the BEST-Index items. This study also shows some of the strengths of the BEST-Index as a clinical instrument for the assessment of potentially risk-relevant behaviors among forensic psychiatric patients. An instrument that (1) has a broad content coverage relevant to the treatment of forensic psychiatric patients, (2) is related to future violence risk as measured with the HCR-20 and to institutional aggressive behavior, (3) can be reliably rated by paraprofessionals who have received only half a day of training and in a relatively brief amount of time (1-2 hours) clearly has significant advantages in terms of cost-effectiveness and ease of implementation. More important, this study also shows the possibility of informing the structured assessment of future violence risk with measures such as the HCR-20, with standardized ratings from nursing staff on changes in risk-relevant behaviors during forensic treatment. Finally, future research is needed that examines the predictive validity of the BEST-Index for future (violent) offending in the community as well as studies that assess the sensitivity of the BEST-Index for detecting change in future violence risk.

Acknowledgments

The authors are grateful to Dr. Tonia Nicholls and her colleagues for giving us permission to use the START Outcome scale. We would like to thank Franca Tonnaer, Eveline van Vliet, and Andrea Braga Ferreira Soares for their assistance in this study and to Erik Schouten for his statistical advice.

Notes

1. In comparison, the internal consistencies for the scales of the BEST-Index original version were α = .96 for the total scale, α = .79 for the Risk subscale, α =.95 for the Insight subscale, and α =.96 for the Communication and Social Skills subscale. The item homogeneity of the original version was .26 for the total scale and .18 for the Risk subscale, .51 for the Insight subscale, and .42 for the Communication and Social Skills subscale.

2. In comparison, the interrater reliability for the scales of the BEST-Index original version were .72 for the total scale, .71 for the Risk subscale, .66 for the Insight subscale, and .67 for the Communication and Social skills subscale.

3. To determine whether the results of the predictive analyses involving START Outcome data would differ, if we used a more appropriate predictive model given the nature of the distribution, we also analyzed these data using a negative binomial regression model. The resulting coefficients for both the revised BEST-Index and the HCR-20 showed the same magnitude of the reported effects as the Pearson correlation analyses. The results of the negative binomial regression analyses are available from the authors on request.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Rev. 4th ed.). Washington, DC: Author.

Arbuckle, J. L. (2006). AMOS (Version 7.0) [Computer software]. Chicago: SPSS, Inc.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications and programming*. London: Lawrence Erlbaum.

Cima, M. (2003). *Faking good, bad, and ugly: Malingering in forensic psychiatric inpatients*. Unpublished doctoral dissertation, Maastricht University, The Netherlands.

Douglas, K. S., Guy, L. S., & Weir, J. (2006). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Retrieved May 26, 2008, from http://kdouglas.wordpress.com/hcr-20/

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, *29*, 617-658.

Douglas, K. S., & Skeem, J. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, & Law*, *11*, 347-383.

Dvoskin, J., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law*, *29*, 6-10.

Van Erven, A. (1999). *Behavioural Status Index, Dutch version (BSI-D). A broad risk assessment for mental health workers*. Eindhoven, The Netherlands: GGzE.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Heilbrun, K. (1997). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior*, *21*, 347-359.

Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2006). 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.

Hu, L. T., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99.

Mahgoub, N. A. (1988). *"Bridging" therapy in hospital and community-based psychiatric care: A comparative study*. Unpublished doctoral dissertation, Sheffield City Library, Sheffield, UK.

Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., & Martin, M. L. (2006). The Short-Term Assessment of Risk and Treatability (START): A prospective validation study in a forensic psychiatric sample. *Assessment*, *13*, 313-327.

Nijman, H. L. I., Evers, C., Merckelbach, H. L. G. J., & Palmstierna, T. (2002). Assessing aggression severity with the revised Staff Observation Aggression Scale (SOAS-R). *Journal of Nervous and Mental Disease*, *190*, 198-200.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Reed, V., Woods, P., & Robinson, D. (2000). *Behavioural Status Index (BEST-Index): A "life skills" assessment for selecting and monitoring therapy in mental health care*. Sheffield, UK: Psychometric Press.

Robinson, D., Reed, V., & Lange, A. (1996). Developing risk assessment scales in forensic psychiatric care. *Psychiatric Care*, *3*, 146-152.

Rogers, R. (Ed.). (1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Ross, T., Woods, P., Reed, V., Sookoo, S., Dean, A., Kettles, A. M., et al. (2008). Selecting and monitoring living skills in forensic mental health care: Cross-border validation of the BEST-Index. *International Journal of Mental Health*, *36*, 3-16.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

SPSS Inc. (2005). SPSS 13.0 for Windows. [Computer software]. Chicago: SPSS, Inc.

Strand, S., & Belfrage, H. (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime and Law*, *7*, 71-79.

Timmerman, I. G., Vastenburg, N. C., & Emmelkamp, P. M. G. (2001). The Forensic Inpatient Observation Scale (FIOS): Development, reliability and validity. *Criminal Behavior and Mental Health*, *11*, 144-162.

de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime and Law*, *12*, 321-336.

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, *3*, 149-165.

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20. Assessing risk for violence, Version 2*. Burnaby, British Columbia, Canada: Simon Fraser University, Mental Health, Law and Policy Institute.

Woods, P. (2000). *The Behavioural Status Index: Descriptive studies within a forensic context*. Unpublished doctoral dissertation, Anglia Polytechnic University, Chlemsford, UK.

Woods, P., Reed, V., & Collins, M. (2001a). Measuring insight in a high security forensic setting using the Behavioural Status Index. *British Journal of Forensic Practice*, *3*, 3-12.

Woods, P., Reed, V., & Collins, M. (2001b). Measuring risk in a high security forensic setting using the Behavioural Status Index. *International Journal of Psychiatric Nursing Research*, *7*, 793-805.

Woods, P., Reed, V., & Collins, M. (2001c). Measuring communication and social skills in a high security forensic setting using the Behavioural Status Index. *International Journal of Psychiatric Nursing Research*, *7*, 761-777.

Woods, P., Reed, V., & Collins, M. (2003a). Relationships between risk and insight in a high security forensic setting. *Journal of Psychiatric and Mental Health Nursing*, *10*, 510-517.

Woods, P., Reed, V., & Collins, M. (2003b). Exploring core relationships between insight, and communication and social skills in mentally disordered offenders. *Journal of Psychiatric and Mental Health Nursing*, *10*, 518-525.

Woods, P., Reed, V., & Collins, M. (2004). Relationships between risk and communication and social skills in a high security forensic setting. *Issues in Mental Health Nursing*, *8*, 769-782.

Woods, P., Reed, V., & Collins, M. (2005). The Behavioural Status Index: Testing a social risk assessment model in a high security forensic setting. *Journal of Forensic Nursing*, *1*, 9-19.

Woods, P., Reed, V., & Robinson, D. (1999). The Behavioural Status Index: Therapeutic assessment of risk, insight, communication and social skills. *Journal of Psychiatric and Mental Health Nursing*, *6*, 79-90.

Yudofsky, S. C., Silver, J. M., Jackson, W., Endicott, J., & Williams, D. (1986). The Overt Aggression Scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry*, *143*, 35-39.