VALIDATION OF THE DUTCH ADAPTATION OF THE
BUSS-DURKEE HOSTILITY INVENTORY

ALFRED LANGE, BAHMAN DEHGHANI and EDWIN DE BEURS

Department of Clinical Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

(Received 2 February 1994)

Summary—The validity of the Dutch adaptation of the Buss-Durkee Hostility Inventory was investigated. On the basis of the ratings of the staff of a residential treatment center for adolescents with conduct disorders, two contrast groups (high versus low aggression) were formed. In all, 67 male adolescents participated in the study. The subscales of the inventory clearly discriminated between the two groups. A discriminant function analysis revealed that the inventory could correctly classify the subjects into aggressive and non-aggressive with 96% accuracy. The pattern of correlations between the inventory and an anger scale and a social desirability scale provide further support for the construct validity of the instrument.

INTRODUCTION

The Buss–Durkee Hostility Inventory (Buss & Durkee, 1957) is a widely used questionnaire on aggression (Biaggio, 1980). The original questionnaire consists of 66 items with false-true answers, and contains seven scales: Assault, Indirect Aggression, Irritability, Negativism, Resentment, Suspicion and Verbal Aggression. Despite the use of the BDHI in numerous investigations, the scales of the original BDHI lack factorial validity. Ramanaiah, Conn and Schill (1987) cite numerous studies in which factor analyses of the inventory consistently yielded two major factors: Overt Hostility and Covert Hostility.

Buss and Perry (1992) recently developed a revised BDHI, with four scales: Verbal Aggression, Physical Aggression, Anger and Hostility. In contrast to the original BDHI, the scales of the new questionnaire were established on the basis of factor analysis. The new questionnaire is shorter (29 items), and instead of false-true categories, the respondent chooses out of five categories.

Following the results of a pilot study by Lischewsky and De Ruiter (1989), Lange, Pahlich, Sarucco, Smits, Dehghani and Hanewald (1995) found two factors underlying the Dutch version of the inventory (BDHI-D). These were similar to the factors identified in the above mentioned studies: Overt Aggression (16 items measuring the tendency to express verbal or physical aggression) and Covert Aggression (19 items tapping the emotional and cognitive components: hostility, irritability, suspicion and anger), with reliability coefficients of respectively $\alpha = 0.77$ and $\alpha = 0.78$. The reliability of the total questionnaire was $\alpha = 0.80$. These factors were used to construct Overt and Covert hostility scales.

Lange et al. (1995) also found good convergent and divergent validity for both scales. Their findings show that anxiety, agoraphobia, depression, somatization, insufficiency, neuroticism, and bitterness are associated with Covert Aggression but not with Overt Aggression.

The aim of the present study is to investigate the concurrent validity of the BDHI-D, by comparing the responses to the questionnaire with the behavioural ratings of the degree of aggressiveness of the subjects. In addition, the convergent and divergent validity will be investigated by correlating the BDHI-D with other questionnaires.

METHOD

Subjects

Seventy-one male juvenile delinquents, who were undergoing a residential rehabilitatory program, voluntarily participated in the study. The data of four subjects had to be omitted from
the analyses because they did not answer all the questions. Most of the 67 subjects had finished middle school \((n = 57)\). They varied from 13 to 19 years of age, with a mean of 16 years. The subjects knew that their responses on the BDHI-D and the other inventories would not be disclosed to the staff and would remain strictly anonymous.

**Self-report measures**

The subjects completed the following questionnaires:

(a) The BDHI-D (Lange et al., 1995), as described above.

(b) The trait anger scale of the Dutch adaptation of the Spielberger State-Trait Anger Scale (van der Ploeg, Defares and Spielberger, 1982). The questionnaire consists of two scales measuring trait anger and state anger. Each scale has ten items with four response categories, varying from 'almost never' to 'almost always'. Trait anger is viewed as 'relative stable individual differences in the tendency to experience anger as a state'. The reliability of the questionnaire is good and its validity is satisfactory (Visser, Van Liet-Mulder, Evers and Ter Laak, 1982).

(c) The Dutch adaptation of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960; Hermans, 1967). A short version of this questionnaire was used, which consists of 14 items with three response categories. This questionnaire has shown to be adequately reliable and valid (Visser et al., 1982).

Marlowe and Crowne (1961) have shown that individuals with a high need for social desirability are relatively more dependent on cultural sanctions than others. Individuals with a high need for social approval use the positive evaluations of others to defensively protect enhanced self-images. As a result of avoiding negative evaluations of others, these persons have problems in recognizing and expressing their own hostility (Crowne & Marlowe, 1964). Consistent with this theory, several studies report negative correlations between social desirability and aggressiveness scales (Biaggio, 1980; Conn & Crowne, 1964; Heyman, 1977; Liebowitz, 1968; Novaco, 1975; Selby, 1984). On the basis of these findings, Selby (1984) concluded that social desirability might be useful in establishing the construct validity of aggressiveness scales.

The social desirability scale contained some items which could be viewed as hostility items. Since this might have an inflating effect on the correlation between social desirability and the BDHI-D scales, these items were deleted in the analysis.

**Behavioural ratings of the degree of aggressiveness of the subjects**

Each subject was separately rated on aggression by two raters (one of the treating staff and one of the group leaders) on a five-point scale (from 'not aggressive at all' to 'very aggressive'). Most of the subjects had spent more than three months in the rehabilitation center. The raters were instructed to establish their ratings on the basis of their own observations of the subjects. In order to create a common frame of reference for all the raters, they were provided with a list of behaviours of the subjects (as directed at group leaders, staff, or other residents) which were indicative of aggressiveness:

- being quick to take offense
- being rude
- taking a threatening position
- being vicious
- being revengeful
- using abusive language
- being hot-tempered
- throwing things
- shouting
- beating someone.

The interrater reliability between the staff’s ratings and the mentors’ ratings was sufficiently high (Spearman correlation coefficient \(r = 0.72; \ n = 66\)).
Table I. Means, standard deviations and the p-values of the contrast groups’ scores on covert aggression, overt aggression, and total aggression

<table>
<thead>
<tr>
<th>Scale</th>
<th>Aggressive (n = 16)</th>
<th>Non-aggressive (n = 33)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>t</td>
<td>P*</td>
</tr>
<tr>
<td>Covert</td>
<td>10.63</td>
<td>3.76</td>
<td>8.00</td>
<td>3.12</td>
<td>2.58</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Overt</td>
<td>12.50</td>
<td>2.19</td>
<td>8.67</td>
<td>3.34</td>
<td>4.80</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>23.13</td>
<td>5.07</td>
<td>16.67</td>
<td>5.44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One-tailed tests.

Forming the two contrast groups

To form two contrast groups, the ratings were dichotomized. The ‘non-aggressive’ group (n = 33) consisted of the subjects who had received the ratings 1 and 2 (‘not aggressive at all’ or ‘not aggressive’) by both raters. The ‘aggressive’ group (n = 16) consisted of the subjects who had received the ratings 4 or 5 (‘aggressive’ or ‘very aggressive’) by both raters. Those subjects who had received score 3 by one or both raters (11 subjects) were excluded from the two contrast groups. Those subjects who were judged to be ‘non-aggressive’ by one rater but were considered to be ‘aggressive’ by the other rater (7 subjects), were also excluded from the contrast groups.

The mean age of both the aggressive group and the non-aggressive group was 16 years, with a standard deviation of respectively 1.44 and 1.40.

RESULTS

Comparing the groups on the BDHI-D scores

As shown in Table 1, the aggressive group scored significantly higher than the non-aggressive group on the total BDHI-D and on both scales (Overt aggression and Covert aggression).

Discriminant analysis

In order to determine the degree to which the BDHI-D predicts whether the subjects belong to the aggressive or non-aggressive group, a canonical discriminant function analysis (Norusis, 1985) was carried out. For each item of the questionnaire, a coefficient function was calculated. All 35

Classification Table

<table>
<thead>
<tr>
<th>Groups</th>
<th>Cases</th>
<th>Predicted group membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive</td>
<td>16</td>
<td>16 (100%) 0 (0%)</td>
</tr>
<tr>
<td>Not aggressive</td>
<td>33</td>
<td>2 (6%) 31 (94%)</td>
</tr>
</tbody>
</table>

Percentage of the cases correctly identified: 96%

Fig. 1. The results of discriminant analysis. A = aggressive; N = not aggressive; * = centroids of the groups.
items passed the default minimum tolerance level of 0.001. The horizontal axis of Fig. 1 shows the discriminant values which were computed for all subjects (sum of the product of the standardized coefficient functions and the subject's score on the item). The vertical axis shows the frequency of the subjects with a particular discriminant value. Only two subjects were not correctly identified. In other words, the complete BDHI-D could correctly classify the subjects with 96% accuracy.

The application of the discriminant analysis to the Covert Aggression items showed that this scale with 84% accuracy predicted to which group the subjects belonged. The application of the procedure to the Overt Aggression items yielded a correct prediction of 94%.

Entering individual items into analysis provides information about the discriminative power of every item and leads to a maximum precision in classifying the subjects. However, with the present sample size, entering individual items into the discriminant function analysis might lead to an over-fitted solution (Stevens, 1986, pp. 235–236). Moreover, clinicians will not base their estimations on the individual items but on the scale scores. Therefore, it seems suitable to also investigate the discriminative power of the BDHI-D while entering the scale scores into the analysis. As expected, the accuracy of the prediction was lower than in the item-based discriminant analysis: the Overt Aggression scale and the Covert Aggression scale could classify the subjects with respectively 76% and 71% accuracy.

**Correlation between BDHI-D and the ratings**

As a second measure of the concurrent validity, the correlation between the mean of the two ratings and the scores on the BDHI-D was calculated. In this analysis, the data of the subjects who were excluded from the contrast groups were also used. The correlations were expressed in ‘r’ coefficients, being an appropriate measure of the association between variables when one is measured on an ordinal scale and the other on an interval scale (Norusis, 1983). In this calculation, the mean of the behavioural ratings by the two judges was considered as the independent variable while the score on the BDHI-D was viewed to be the dependent variable. All correlations were high. The correlation of Covert Aggression, Overt Aggression and Total Aggression with the behavioural ratings were respectively: \( r = 0.58 \); \( r = 0.61 \), and \( r = 0.70 \), for 66 subjects (the ratings of one subject were missing).

**Association of the BDHI-D with Trait Anger and Social Desirability**

In determining the correlation of the BDHI-D with trait anger and social desirability, the data concerning all initial subjects were used \( (n = 67) \). As expected, positive significant correlations were found between the scores on the Trait Anger scale with Covert Aggression \( (r = 0.57; P < 0.001) \), Overt Aggression \( (r = 0.56; P < 0.001) \), and Total BDHI-D \( (r = 0.67; P < 0.001) \). Also as expected, there were negative significant correlations between the scores on the Social Desirability scale and Overt Aggression \( (r = -0.42; P < 0.001) \), and the Total BDHI-D \( (r = -0.41; P < 0.001) \). Covert Aggression was not significantly related to Social Desirability \( (r = -0.27) \).

**DISCUSSION**

The results of the t-test of the differences between the contrast groups (Table 1) and the results of the discriminant analysis (Fig. 1) strongly support the construct validity of the BDHI-D scales. The discriminant analysis demonstrates the high sensitivity of the Overt Aggression scale. If the individual items are used, it correctly classifies 94% of the subjects. If the sum-score is used, 76% of the subjects will be correctly classified. The Covert Aggression is less sensitive in this population (84% correct classification on the basis of the items, 71% when the sum-scores are used). This provides clinicians, in a comparable setting, with an instrument by which they may identify aggressive adolescents with a reasonable degree of accuracy right from the beginning of the (residential) treatment, by using coefficient functions of the items and the value of responses to the items of the questionnaire. However, the questionnaire might be less valuable if subjects have reason to believe that their answers will have a negative effect on the way they are treated. In our study, the subjects knew that their responses would remain anonymous. Generally, this will not be the case in institutions. To obtain valid responses, staff should make it clear to the respondents
of the BDHI-D that their responses to the questionnaire will not do them any harm, and that honest responses are in their own interest.

The correlations between the behavioural ratings and the scales demonstrate the same pattern as found by the discriminant analysis. Covert Aggression has a lower correlation with the behavioural ratings than Overt Aggression. This might be expected because it is more difficult for external observers to rate Covert Aggression than Overt Aggression. Moreover, the instruction for the raters comprised more behaviours indicating overt aggression. To be sure, these validity coefficients could have been higher, had the interrater reliability (r = 0.72) been higher.

High correlations between the Trait Anger scale of the Dutch adaptation of the Spielberger State-Trait Anger Scale (Van der Ploeg et al., 1982) and the BDHI-D scales support the convergent validity of the latter and indicate that there is a strong association between trait anger and the factors measured by the BDHI-D. This emphasizes the role of anger as a kind of psychological bridge between the cognitive component (hostility) and the motor component (verbal or physical aggression), as demonstrated before by other researchers (Buss & Perry, 1992).

The negative correlation between the Overt Aggression scale and the Dutch adaptation of the Marlowe–Crowne Social Desirability Scale (Hermans, 1967) is consistent with the studies that report a negative association between “aggression” and “social desirability” constructs. The correlation between this scale and Covert Aggression is not significant but shows the expected negative direction.

To conclude, the results of the present study strongly support the construct validity of the BDHI-D scales. The next step consists of establishing norms for men and women separately, for age groups and for various groups at risk, such as delinquents and psychiatric patients. We will report on this in the near future.

Acknowledgement—We are indebted to Gerrit Hanewald for his help with the statistical analysis.

REFERENCES