A validation study of
the Novaco Anger Inventory

MATTHEW T. HUSS
Emporia State University, Emporia, Kansas

GARY K. LEAK
Creighton University, Omaha, Nebraska

and

STEPHEN F. DAVIS
Emporia State University, Emporia, Kansas

Many individuals have attempted to construct tests to measure different aspects of anger. Although the Novaco Anger Inventory (NAI) was constructed over 16 years ago, viable validity research is still lacking. The Buss–Durke Hostility Inventory, three subscales of the PRF, and the Burt Scale were administered to assess the validity of the NAI. The Balanced Inventory for Desirable Responding was also used to measure the participants' degree of socially desirable responding. With the exception of the Burt Scale, all measures were significantly correlated with the NAI. The correlation coefficients, ranging from .413 to .503, suggest a significant degree of construct validity for the NAI. Several gender effects are also reported.

Most individuals can relate to anger, which may be manifested in a variety of forms that range from the grinding of teeth and yelling to the smashing of furniture. Because it is a difficult concept to define, it has also been a difficult concept to measure. An individual may display anger only during certain times or in the presence of particular people; others may become angry when they see a particular injustice but not when they are physically confronted. In addition, anger may be an internal feeling, whereas aggression and hostility are its outward manifestations.

In view of these considerations, the construction of useful devices to measure anger has faced obstacles. Some instruments claim to measure hostility, and others profess to measure aggression, the external expression of anger. The introduction of the Buss–Durke Hostility Inventory (Buss & Durke, 1957) was one of the first attempts to construct an independent assessment tool to measure traits related to anger, hostility, or aggression. Since the introduction of this scale, other attempts have been made to measure this construct (Evans & Strangeland, 1971; Novaco, 1975; Spielberger, Jacobs, Russell, & Crane, 1983; Zelin, Adler, & Myerson, 1972).

The purpose of this study was to assess the validity of one measure, the Novaco Anger Inventory (NAI). Evidence regarding the validity of the NAI is limited and inconsistent (Biaggio, Supplee, & Curtis, 1981; Novaco, 1975; Novaco, 1977). In order to test the validity of the NAI, the

following selected instruments were used: the Buss–Durke Hostility Inventory (BDHI), the Personality Research Form-E (PRF-E), and the Burt Scale. The aggression, abasement, and defendence subscales were selected from the PRF-E; the adversarial sexual beliefs (BADV), acceptance of interpersonal violence (BVIO), and rape myth acceptance (BRMA) scales of the Burt Scale were focused on.

It is difficult to collect accurate data on a personality trait such as anger because of the desire of subjects to present themselves in a favorable manner. Hence, the Balanced Inventory of Desirable Responding (BIDR) was used to measure the socially desirable response rate (any tendency of subjects to appear more favorable than is factual).

Because the questions are intuitively derived and partially based on student interviews, the validity of the NAI is questionable. In the original study by Novaco (1975), correlations with physiological measures of anger were no higher than .35, whereas correlations on other self-report measures ranged widely from −.262 to .611. In an independent study (Biaggio et al., 1981), correlations no higher than .14 were reported. Some studies have shown that the NAI is able to distinguish between normal samples and anger-prone samples (Hamberger & Hastings, 1986; Lohr, Hamberger, & Bonge, 1988; Mizes, Morgan, & Bader, 1990; Novaco, 1977). Hence, a crude measure of validity may be attributed to these studies.

Novaco (1975) has reported internal reliability of the NAI at an acceptable level (Cronbach alpha of .96). On the other hand, a very low test–retest reliability of .17 has been reported by Biaggio et al. (1981). Clearly, future research is needed before the NAI is accepted as a measure of anger.

Address correspondence to S. Davis, Department of Psychology, Emporia State University, 1200 Commercial, Emporia, KS 66801-5087.

METHOD

Subjects

The volunteers consisted of 72 (46 women and 26 men) undergraduate psychology students (M = 19.36 years, SD = 1.19) at a private midwestern university. All the students volunteered in order to receive 1 h of extra credit toward a psychology class and gave their informed consent.

Measures

The NAI is composed of 90 descriptions of anger-provoking incidents. The participants were asked to respond on a 5-point scale (1 = *not at all*, 5 = *very much*) according to the degree of anger or provocation they would feel if placed in the situation.

The BDHI consists of 75 true-false questions with seven subscales, and is probably the most widely used and well-known measure of its kind (Buss & Durke, 1957). Sarason (1961) found significant relationships between the BDHI and two other measures of hostility. Likewise, Leibowitz (1968) reported a significant correlation (r = .36) between the total BDHI score and a role-playing aggression score. Although the BDHI may discriminate between groups, it often lacks predictive validity (Edmunds, 1976; Edmunds & Kendrick, 1980).

The PRF-E (Jackson, 1974) consists of 352 true-false items, ordered into 22, 16-item scales, and attempts to measure personality traits in areas of normal functioning. For the purpose of the present study, the abasement (PRFAB), aggression (PRFAG), and defendence (PRFDE) need scales were used. The abasement scale has been shown to correlate negatively with the defendence scale (r = −.51) and the aggression scale (r = −.50). The aggression and defendence scales correlate at a .59 level. The PRF manual also reports validity coefficients for the PRF scales that vary widely (abasement = .19-.33, aggression = .36-.73, defendence = .23-.58). On the basis of this evidence, the aggression and defendence subscales should correlate positively with the scores on the NAI, whereas the abasement subscale should be negatively correlated with the NAI scores.

The Burt Scale (Burt, 1980) has 41 items that attempt to measure respondents' attitudes about the acceptance of rape myths. It is composed of different subscales: acceptance of interpersonal violence, adversarial sexual beliefs, and rape myth acceptance. Participants respond on a 5-point scale (1 = *strongly agree*, 5 = *strongly disagree*) about the degree to which they agree or disagree with the statements. Data from Burt (1980) suggest that the measure correlates with adversarial sexual beliefs and acceptance of interpersonal violence. The rape myth scale also has been shown to correlate with aggressive tendencies, coercive sexual tendencies, and likelihood of rape, yielding a multiple R of .44 (Greendlinger & Byrne, 1987).

The BIDR (Paulhus, 1984) is a 40-item questionnaire that is scored on a 7-point scale (1 = *not true*, 7 = *very true*) and is intended to measure the social desirability of respondents. The BIDR has two subscales: self-deception and impression management. Each subscale measures slightly different aspects of socially desirable responding. Paulhus believes that the impression-management subscale is the best measure of socially desirable responding.

Procedure

Data were collected in seven separate classrooms with participants completing all questionnaires in one session. Each subject was given a packet containing all the questionnaires and a consent form. The questionnaires were arranged randomly to insure that the order would have no effect on the subjects' responses. The participants were informed that the questionnaires measured certain aspects of anger and were instructed to answer the questions as truthfully and honestly as possible.

RESULTS

Table 1 presents the means and standard deviations for each measure. It is of importance that the sample contained more women (44) than men (26). The mean age (19.37 years) is also slightly higher than that reported in the majority of associated studies conducted with college students because the study was not confined to introduc-

Table 1
Means (M) and Standard Deviations (SD) for Each Variable

Variable	M	SD
Age	19.37	1.20
GPA	3.24	.501
NAI	316.81	46.74
BDHI	113.42	10.57
BIDRSD	82.93	13.74
BIDRIM	96.83	12.98
PRFAB	23.33	2.60
PRFDE	23.29	2.84
PRFAG	24.93	3.01
FES	12.58	2.34
BADV	21.32	4.43
BVIO	18.03	2.58
BRMA	26.32	5.75

NAI = Novaco Anger Inventory; BDHI = Buss-Durke Hostility Inventory; BIDRSD = Balanced Inventory of Desirable Responding-Social Desirability; BIDRIM = Balanced Inventory of Desirable Responding-Impression Management; PRFAB = Personality Research Form-Abasement Scale; PRFDE = Personality Research Form-Defendence Scale; PRFAG = Personality Research Form-Aggression Scale; FES = Family Environment Conflict Scale; BADV = Burt Scale-Adversarial Sexual Beliefs; BVIO = Burt Scale-Interpersonal Violence; BRMA = Burt Scale-Rape Myth Acceptance.

tory students. This difference is likely due to the wider population from which the present sample was drawn.

Table 2 shows the Pearson correlations between the respective measures and the NAI. NAI scores for the full sample were significantly correlated with the BDHI (r = .462), the PRFAB (r = −.368), the PRFDE (r = .503), and the PRFAG (r = .413) subscales. NAI scores in the full sample were significantly correlated with the BIDR impression-management subscale (BIDRIM) (r = −.256). NAI scores for the full sample were not related to sex, age, grade point average, or whether or not the participants lived on campus. Finally, the NAI scores for the full sample did not significantly correlate with the social desirability subscale of the BIDR or any of the subscales of the Burt Scale (BADV, BVIO, or BRMA).

Table 2 also indicates the presence of gender effects. Although there were no significant correlations between age and grade point average, only the women showed a significant correlation between the NAI and living on campus (r = −.344). The men's NAI scores significantly correlated with the PRFAB (r = −.448), the PRFDE (r = .541), and the PRFAG (r = .541), whereas the women's scores related significantly only to the PRFAB (r = .350) and the PRFDE (r = .490). Likewise, there were differences between men and women on the relationship of the subscales of the Burt Scale and the NAI. The NAI scores of the men were not significantly correlated to the BADV, BVIO, or BRMA but there were correlations between the NAI scores of the women and the BADV (r = .309) and the BRMA (r = .364).

DISCUSSION

The significant correlation coefficients ranging from .413 to .503 appear to suggest some construct validity for the NAI. The correlation between the NAI and the PRFDE could be explained by the original construction of the NAI. Novaco (1975) felt that one of the key compo-

Table 2
Pearson Correlations Between Respective Variables and the NAI

Measure	Full Sample	Male	Female
Age	−.178	−.078	−.248
GPA	.018	−.048	−.057
NAI			
BDHI	.462†	.446†	.466†
BIDRSD	−.110	−.284	−.041
BIDRIM	−.256*	−.313	−.240
PRFAB	−.368†	−.448†	−.350†
PRFDE	.503†	.541†	.490†
PRFAG	.413†	.560†	.280
FES	.031	.051	.013
BADV	.060	−.177	.309†
BVIO	.198	.344	.103
BRMA	.141	−.013	.364†

*$p < .05$.   †$p < .01$.

nents to anger was defensiveness. Because defensiveness is essential to the definition in the PRFDE, it would be expected that this scale of the PRF-E would correlate highly. The correlation with the PRFAG was expected because this subscale measures similar traits such as aggressive, hot-tempered, easily angered, and irritable.

Because another independent measure of anger (BDHI) correlated positively with the NAI, the present data provide a measure of construct validity. This result is empirically relevant because other studies have found that the BDHI correlates less well with the NAI (Biaggio, 1980; Biaggio et al., 1981).

The negative correlation of the NAI and the abasement (PRFAB) subscale was expected because that scale also has strong negative correlations with the defendence (PRFDE) and the aggression (PRFAG) subscales (Jackson, 1974). The abasement subscale consists of trait-defining adjectives such as yielding, deferential, resigned, and subordinating, which would tend to be uncharacteristic of an individual who scores high on the NAI. An individual scoring high on the abasement scale is defined as meek, humble, and subservient—the type of individual who would be predicted to score lower on a measure of anger.

Two measures that were predicted to be not significantly correlated with the NAI were age and grade point average. As expected, the measures did not strongly correlate with the NAI.

It is interesting to note that the correlations of the NAI with the Burt Scale, though unexpected, are not surprising. It is possible that the Burt Scale measures more abstract degrees of anger or aggression; it does not measure an aspect of anger, but how this anger may be channeled. Though the subscales of the Burt Scale did not correlate with the NAI for the overall sample, there were significant positive correlations between the BADV ($r = .309$) and BRMA ($r = .364$) subscales among the women completing the NAI. These results could be explained by the fact that women may be more sensitive to the topic of rape.

Another aspect of the present data that deserves consideration is the measure of social desirability. As expected, there was a significant negative correlation between the impression-management scale (BIDRIM) of the BIDR and the NAI. It is significant that the BIDR correlation was less than that of previously reported measures of social desirability. In particular, Novaco (1975) reported a correlation of −.502 using the Marlowe–Crowne Need For Approval Scale in his development of the NAI. Though the correlation between the BIDRIM and the NAI was significant, it was also lower than might be expected when studying a trait such as anger, because individuals are less likely to be honest when evaluating themselves with regard to a negative trait. This fact potentially lends additional credibility to the present results in that they may not have been influenced by social desirability as much as in previous studies (Biaggio, 1980; Buss & Durke, 1957; Heyman, 1977; Leibowitz, 1968; Novaco, 1975; Sarason, 1961).

Though further research is still needed on both the reliability and validity of the NAI, the present study does demonstrate that the NAI signif-

icantly correlates with a variety of anger-related measures. Considering the role that anger plays in different areas of contemporary society, it seems a worthwhile endeavor to discover and utilize proper measures of this behavioral trait. Valid assessment devices are invaluable to the psychologist today, as researcher and practitioner. Effective measures of anger enable professionals to delve into the inner workings of human cognition and gain some degree of insight. This study has attempted to help identify the Novaco Anger Inventory as one of these useful measures.

REFERENCES

BIAGGIO, M. K. (1980). Assessment of anger arousal. *Journal of Personality Assessment, 44,* 289-298.

BIAGGIO, M. K., SUPPLEE, K., & CURTIS, N. (1981). Reliability and validity of four anger scales. *Journal of Personality Assessment, 45,* 639-648.

BURT, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality & Social Psychology, 38,* 217-230.

BUSS, A. H., & DURKE, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21,* 343-349.

EDMUNDS, G. (1976). The predictive validity of the Buss-Durke Inventory. *Journal of Clinical Psychology, 32,* 818-820.

EDMUNDS, G., & KENDRICK, D. C. (1980). *The measurement of human aggressiveness.* New York: Halsted.

EVANS, D. R., & STRANGELAND, M. (1971). Development of the reaction inventory to measure anger. *Psychological Reports, 29,* 412-414.

GREENDLINGER, V., & BYRNE, D. (1987). Coercive sexual fantasies of college men as predictors of self-reported likelihood to rape and overt sexual aggression. *Journal of Sex Research, 23,* 1-11.

HAMBERGER, L. K., & HASTINGS, J. E. (1986). Personality correlates of men who abuse their partners: A cross-validation study. *Journal of Family Violence, 1,* 323-341.

HEYMAN, S. R. (1977). Dogmatism, aggression, and gender roles. *Journal of Clinical Psychology, 35,* 8-12.

JACKSON, D. N. (1974). *Personality research form manual.* Goshen, NY: Research Psychologists Press.

LEIBOWITZ, G. (1968). Comparison of self-report and behavioral techniques of assessing aggression. *Journal of Consulting & Clinical Psychology, 32,* 21-25.

LOHR, J. M., HAMBERGER, L. K., & BONGE, D. (1988). The relationship of factorially validated measures of anger-proneness and irrational beliefs. *Motivation & Emotion, 12,* 171-183.

MIZES, J. S., MORGAN, G. D., & BADER, J. (1990). The relationship of cognitions, assertion, and anger arousal. *Journal of Cognitive Psychotherapy: An International Quarterly, 4,* 369-377.

NOVACO, R. W. (1975). *Anger control: The development of an experimental treatment.* Lexington, KY: Lexington.

NOVACO, R. W. (1977). Stress inoculation: A cognitive therapy for anger and its application to a case of depression. *Journal of Consulting & Clinical Psychology, 45,* 600-608.

PAULHUS, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality & Social Psychology, 46,* 598-609.

SARASON, I. G. (1961). Intercorrelations among measures of hostility. *Journal of Clinical Psychology, 17,* 192-195.

SPIELBERGER, C. D., JACOBS, G., RUSSELL, S., & CRANE, R. S. (1983). Assessment of anger: The State-Trait Anger Scale. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 2). Hillsdale, NJ: Erlbaum.

ZELIN, M. L., ADLER, G., & MYERSON, P. G. (1972). Anger self-report: An objective questionnaire for the measurement of aggression. *Journal of Consulting & Clinical Psychology, 39,* 340.