VALIDATING A TEST TO MEASURE THE AWARENESS AND EXPRESSION OF ANGER

by

RICHARD E. D. BRAHA

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science,

Saint Mary's University,

Halifax, Nova Scotia, Canada.

August 28, 1987

ISBN 0-315-48189-7

CAMP HILL HOSPITAL A Hospital of the Camp Hill Medical Centre

1763 Robie Street
Halifax, Nova Scotia
B3H 3G2

Tel (902) 420-2222

April 26, 1989

National Library of Canada
Ottawa, Ontario
c/o Norma Corbett
Reference Department
Library
St. Mary's University
Halifax, N.S.
B3H 3C3

Fax No: (902) 420-5561

To Whom It May Concern:

We, the undersigned, registered copyright holders for "The Awareness and Expression of Anger Indicator" and "Manual for the Use of the Awareness and Expression of Anger Indicator" hereby grant non-exclusive license to the National Library of Canada to microfilm Appendices A and E included in the Master's Thesis written by Richard E. D. Braha entitled, "Validating a Test to Measure the Awareness and Expression of Anger (1987)", which includes printed copies of the aforementioned copywritten material.

Richard F.H. Catchlove, M.B., B.S., F.F.A.R.A.C.S, M.Sc.,
Montreal, Quebec

Richard E.D. Braha, B.A., M.Sc.
Halifax, Nova Scotia.

Saint Mary's University
Halifax, Nova Scotia
Canada
B3H 3C3
Patrick Power Library

FAX COVERING LETTER

DATE: July 11, 1989

TO: Huguet Lussier-Tremblay or Mel Simoneau

COMPANY NAME: National Library - Canadian Theses Service

FAX# (819) 953-8508

CONFIDENTIAL YES ____ NO ✓

ACKNOWLEDGEMENT REQUIRED YES ____ NO ✓

NUMBER OF PAGES INCLUDING COVERING LETTER 2

If you do not receive all pages please call Paula Gosby at (902) 420-5400 ext. 2561.

****

FROM: Norma Corbett

DEPARTMENT: Reference Dept., Library

FAX # (902) 420-5561

REMARKS: Sorry for the delay in responding to your phone call regarding Richard Brawn's thesis. I am enclosing a revised letter granting non-exclusive license to the National Library, from the author and his co-author for appendices A and E. I hope this letter will be acceptable.

VALIDATING A TEST TO MEASURE THE AWARENESS AND EXPRESSION OF ANGER

BY

RICHARD E. D. BRAHA

Accepted in partial fulfillment of the requirements for
the degree of MASTER OF SCIENCE,
Saint Mary's University,
Halifax, Nova Scotia, Canada.

Approved: ______________
G.M.H. Pretty, Ph.D.
Thesis Supervisor

Approved: ______________
R.J. Konopasky, Ph.D.
Thesis Committee Member

Approved: ______________
B.D. Dufton, Ph.D.
Thesis Committee Member

Approved: ______________
C.J.A. Hayes, Ph.D.
External Examiner

Date: September 17, 1987

Acknowledgemments

I would like to express my deepest gratitude to my Thesis Supervisor, Dr. Grace Pretty, for her scholarly counsel and support through all phases of this research.

I would like to thank the members of the Thesis Advisory Committee, Drs. Robert J. Konopasky and Brian D. Dufton for their numerous improvements to the text. Dr. Dufton deserves recognition for his extensive contributions to the research design, data analyses and interpretation.

Also, I would like to thank Richard F. H. Catchlove, M.B., B.S., F.F.A.R.A.C.S., M.Sc., co-author of the AEAI. His commitment to clinical pain research resulted in the development of the AEAI.

Last, but by no means least, I would like to thank my wife, Andrea John, B.A., M.L.S., for her unequivocal support and relentless editing through several uninspiring readings of draft versions of this thesis.

Table of Contents

List of Tables and Figures

Figure **Page**

ABSTRACT

Validating a Test to Measure the Awareness and Expression of Anger

R. E. D. Braha

August 28, 1987

Experimental, physiological, and theoretical literature on anger is reviewed briefly. It is concluded that there is a lack of consensus on standard definitions of anger, or on the nature of anger. The Buss-Durkee Hostility Inventory, The Oken Scale, The Gottschalk-Gleser Content Analysis Scales, and, The Reaction Inventory are reviewed and dismissed as adequate instruments for the measurement of the awareness and expression of anger.

The Awareness and Expression of Anger Indicator (AEAI) is presented as a test which purports to measure different dimensions of anger. Existing psychometric data on the AEAI is reviewed and it is concluded that further psychometric study on the reliability and validity of the AEAI is needed.

A research design is developed which examines aspects of reliability and validity of the AEAI. The major predictions are: that there would be subscale homogeneity, internal consistency, and a latent factor structure which would confirm the three dimensions of anger measured by the test scales. The Multidimensional Anger Inventory, the Anger Inventory, the Marlowe-Crowne Scale of Social Desirability, and the Cognitive Attitudes Scale are used to evaluate the convergent and discriminant validity of the AEAI.

The results indicate that the AEAI demonstrates 1. adequate internal reliability; 2. a factor structure which supports a distinction between non-induced awareness, expression and induced awareness of anger; 3. that there is some evidence of convergence between

AEAI awareness measures and other awareness of anger measures, but no evidence of convergence between AEAI expression of anger measures and other measures of the same trait; and, 4. that AEAI scores show no positive relationship to scores on the social desirability scale but correlate positively with a measure of subjects' beliefs about the consequences of expressing anger. Discussion focuses on the effects of a weighting system on AEAI scores, the multidimensional nature of anger, and on more general issues in testing.

Many authors believe anger is a distinct fundamental emotion (Bowlby, 1980; Izard, 1977; Stearns, 1972; Wolf, 1970; Gottschalk & Gleser, 1969; Frank, 1950). Several studies have demonstrated physiological concommitant responses to anger which are discrete with respect to other emotions (e.g., fear, anxiety, sadness), and which can be reliably evaluated (Pennebaker, 1980; Stearns, 1972; Sternbach, 1966; Knapp, 1963). Moreover, there is research evidence which suggests that anger can be reliably differentiated from hostility, rage, contempt, disgust and aggression (Biaggio, 1980; Izard, 1977; Knapp, 1963). However, despite many attempts at developing standard definitions for anger, several workers have continued to use anger, aggression, hostility, and rage interchangeably. All are sometimes equated as the total construct (Gottschalk & Gleser, 1969). Some regard rage and anger as relative degrees in the expression of aggression (Bowlby, 1980). Still others include a combination of the above when referring to anger.

Because researchers continue to report results from studies which used undefined constructs or independently defined constructs, it has been difficult to reconcile seemingly contradictory research findings (Rubin, 1986). Recent reviews have criticized researchers for failing to acknowledge and/or attempt to resolve this problem (Diamond, 1982; Tavris, 1982). In the early seventies Bandura (1973) attempted to explain the empirical contradictions in the anger literature. For example, many theorists believe and have produced results that support the hypothesis that unless anger is expressed, clients may experience a wide range of disturbances (e.g., Kaufman & Feshback, 1963; Matarazzo, 1954). Other theorists believe that the expression of anger is risky and ineffective and that clients should be taught alternative behaviours and attitudes (Bandura, 1965; Berkowitz, 1970). Using a social learning perspective, Bandura (1973) argued:

> Studies in which the outcomes of reciprocal interchanges are systematically varied demonstrate that the same counterresponse, regardless of its content, can acquire either arousing or tranquilizing properties depending on its

> consequences (Hokanson, Willers, & Koropsak, 1968). Both kindly and aggressive responses to provocation produce cathartic-like physiological relief after they consistently elicited positive reactions in others; conversely, they become physiologically arousing when they consistently draw punitive responses. By reversing reinforcing outcomes, the tension relief value of the same mode of response is radically altered (Bandura, 1973, p.152).

Consequently, Bandura proposed that future empirical research on anger include the subjects' beliefs about the consequences of anger expression. Unfortunately, as reviewers have indicated (Rubin, 1986; Diamond, 1982; Tavris, 1982), only one study since 1973 has adopted this guideline (e.g., Van Egeren, Abelson, & Thorton, 1978). Further, even the most recent relevant research does not acknowledge this portion of the anger literature (e.g., Jansen & Muenz, 1984). Although Bandura can be commended for his attempts to clarify contradictory theories on anger, he too fails to provide a clear definition of anger (Rubin, 1986), and, as Averill (1979) points out, treats anger as merely a hypothetical construct "of little substantive interest in its own right". Proponents of the literature that regards anger as a pure affect (e.g., Zajonc, 1984) would argue that what Bandura has described is simply another intervening variable in the anger expression process. Some cognitivists would likely reply that a person's appraisal (i.e., cognitive pre-appraisal) of a stimulus and the environmental context in which the stimulus occurs is an essential precondition to the labelling of arousal as a feeling. Hence, they would argue that a person's belief about the consequences of expressing anger would play a role in that person's response. Clearly, the issue becomes circular. Whether affect or cognition is primary more likely reflects overlapping taxonomies than mutually exclusive or linearly occurring events. As Averill (1979; 1983) suggests, we are not researching the relevant aspects of anger, the emotion. The queries which need to be addressed are: What is anger? What are its functions (both positive and negative)? How does it function? How does it affect other aspects of behaviour? How is it expressed? and, central to this study, How can it be reliably assessed?

Etymology tells us that the word anger means a response to an offending stimulus. In most languages the term is not associated with hostility, aggression or rage (Stearns, 1972).

Anger does not appear to be a reaction to threatening stimuli as some have described (Yacorzynski, 1951; Rubin, 1986). Rather, it seems to be a response to a provoking stimulus. Fear or hostility may be reactions to threatening stimuli (Stearns, 1972)

Montagu (1955), explains that some stimuli may elicit different affective responses in different individuals; for example, fear, or anxiety may arouse varying degrees of anger in different individuals. Most agree with this tenet, however, others (Biaggio, 1980; Stearns, 1972; Izard, 1977; Rubin, 1986), disagree with Montagu that a single stimuli can arouse immediately different combinations of emotions in one individual: for example, much anger and a little fear or a little anger and a lot of fear. But, they argue, anger can subside, be repressed, or be replaced by anxiety, hostility, depression, etc. (Stearns, 1972). Moreover, fear is not considered to be a constituent part of anger. Fear can replace anger when, for example, the verbal stimulus which may have originally provoked the anger response is "rationalized" into its determinant contents (Bowlby, 1980). Understood through a cognitive or behavioral paradigm the initial anger-provoking stimuli and the unconditioned emotional response (i.e., anger) may become paired or associated with conditioned cognitive and behavioral responses such as negative thoughts and/or autonomic nervous responses. These threatening negative thoughts and behaviours may then elicit a fear response (Beck & Emery, 1985). Bowlby's (1980) rationalization may thus be understood in terms of the cognitive construct of negative or automatic thoughts.

<u>The anger process</u>

The individual's response to an affect-provoking external event consists of at least two reactions: a perceptual-cognitive and an affective reaction (Alexander, 1950). The affective reaction involves biological and psychological components (emotions and feelings) (Sifneos et al., 1977; Zajonc, 1984). The biological components of affect are assumed to be responsible for the concurrent somatic arousal (emotion) to the external affect-provoking event. The psychological components are considered to be expressive of thought and fantasy or feelings. To summarize, it is assumed that the internal reactions that occur in response to an external affect-provoking event include perceptual, biological and psychological systems.

These reactions can be conceptualized as occurring in a stepwise fashion through six stages: 1. perceiving the external event, which leads to 2. subsequent biological arousal; 3. a refinement of the arousal into a variety of different nuances which have the potential for conscious experience as feelings, for example, anger, fear, sadness, joy; 4. a linking of words which are descriptive of those feelings; 5. the production of fantasies which are expressive of those feelings, and finally: 6. the arousal of a network of memories and associations which are related to those feelings (Nemiah, 1977). It is assumed that somewhere along this hypothetical progression the phenomenon of cognitive pre-appraisal (Lazarus, 1984) occurs. In most cases, a conscious awareness of these elements occurs and these are expressed in the appropriate manner.

Anger resolution

However, there are no acceptable comprehensive theories about the process of expression that theoretically leads to the resolution of angry feelings. A necessary prelude to the construction of any psychometric instrument designed to measure process is the understanding of the very process one intends to measure. Several researchers have attempted to resolve this problem.

Anger control

Novaco (1975), developed a theoretical strategy for the control of anger. His work was concerned with individuals who displayed inappropriate outbursts of rage, and consequently his theory may be limited. Novaco's programme was designed to teach people to control and learn not to express their anger. The result is a series of cognitive-behavioral strategies that the subject uses to control his/her anger. This is presented as a way to resolve anger. Although Novaco's theory has been effective in helping his subjects control the behavioral consequences of their rage, I wonder whether or not there has been resolution of the affective component.

Assertiveness training

There are volumes of work on assertiveness, but again, these do not provide us with a conceptual framework from which one could evaluate the individual's processing of his or her angry feelings. Assertiveness theory has provided specific skills which may or may not be effective in resolving anger. Because these programmes commonly place little emphasis on the stages involved in the constructive expression of anger, they have provided little insight into process (Hoffman, Kirwin, & Rouzer, 1979).

Constructive expression of anger

Catchlove and Braha (1985) conceptualized the process involved in the resolution of anger as occurring through four phases. These phases describe a theoretical progression of events for the resolution of anger through effective and constructive expression.

The first phase is "Awareness". Subsumed under this heading are the six stages of arousal described above. Awareness of anger then, is defined as a subjective (phenomenological) experience of an emotional state or condition which occurs in response to a provoking stimulus. It contsists of feelings of anger with concomitant activation or arousal of the autonomic nervous system. Anger can vary in intensity and fluctuate over time as a function of perceived injustices, provocations, or frustration resultingthe conscious, unsolicited, personal experience or acknowledgement of a feeling.

The second phase is "Identification". Identification is the attempt at determining the causal agent for the anger-provoking situation, in order to attribute responsibility. This process may occur on a conscious (aware), or unconscious (unaware) level, or both. Completing identification, the process continues to the next phase, "Confrontation".

Confrontation is the provocative, constructive expression of the feeling with the purpose of extracting an apology and restitution from the causal agent. Effective confrontation also requires a containment , or mastery over the feeling. This phase has the concurrent role of allowing the maintenance of the relationship by obtaining the causal agent's motivation; as such, a normative judgement is inherent in this phase. For example, should an accident be the reason for the provoking event, then the nature of the confrontation would likely be changed.

Achieving an apology (e.g., "I'm sorry") and restitution (e.g., "I won't do it again"), the affect is presumably discharged and resolution (phase four) occurs. Resolution describes the release/discharge of the affect, in this case through effective and constructive expression. In some cases resolution can only be achieved by avoiding the anger-provoking situation. The latter may involve terminating the relationship (The *Manual for the use of the AEAI* describes this progression and provides examples of subject responses typical to each of the above phases: see Appendix E). There are, of course, other methods of expression which may be equally effective in resolving angry feelings, but it is assumed that these processes are less socially effective unless the goal is the destruction of the relationship.. For example, there is some evidence which suggests that verbal or physical aggression may be an effective, destructive method of expressing anger because it intensifies the interpersonal conflict between the aggressor and his or her target (Holt, 1970; Hoffman et al., 1979).

Part of the degree of effectiveness in the way an individual deals with the emotion of anger is related to a low discrepancy between the actual and attributed causal agent(s) and the actual subject of the confrontation. This theoretical qualifier may reconcile Attribution

Theory's contribution to the emotion literature (Kelley, 1973). For example, if an individual attributes responsibility for the anger-provoking event to himself rather than to the actual perpetrator of the event, then effective confrontation with the perpetrator or causal agent is not feasible. However, styles of misattribution have also been associated with attempts to maintain relationships: subjects may personally take the blame to avoid what they perceive to be a potentially threatening confrontation with the actual causal agent (Weiner, Amirkhan, Folkes, & Verette, 1987). Conversely, it has also been shown that subjects' may blame others in attempts to maintain a positive self-image (Russell & McAuley, 1986). Further, some have suggested that if an individual attributes causality to another and then fails to confront that individual (thus resolving the incident simply with herself or himself), there may remain a degree of residual anger and/or frustration (Holt, 1970). Finally, the attribution of blame to an unintentional or accidental occurrence should alter the confrontation process. Averill (1983) and others (Pepitone, 1976; Rule & Nesdale, 1976) have also discussed the relevance of "blame" towards the resolution of angry feelings. However, as they point out, most psychological theories have not accommodated this aspect of anger. Figure 1 illustrates the attribution of blame.

Identification of causal agent   Actual agent confronted

SELF -------------------effective route-------------------SELF

ineffective route

OTHER -----------------effective route-------------------OTHER

Figure 1. Attribution and confrontation patterns

Bandura's (1973) theory regarding the role of a subject's beliefs about the consequences of expressing anger is also, no doubt, reflected in the subject's choice of responses. However, there is evidence to suggest that such beliefs do not mitigate the initial subjective experience of anger and as such do not diminish the importance of the feeling but rather effect the expression of that feeling (Zajonc, 1984). Conversely, the ecological context in which the event occurs no doubt has a significant role in the emotional arousal of the subject (Averill, 1983). This is reflected in most behavioral theories of emotion. Unfortunately, the nature of this context defies inclusion in most research designs (Averill, 1983; Lazarus, 1984).

Anger test construction: critical factors

Several workers have criticized tests using role-playing situations for their lack of experiential realism (Aranson & Carlsmith, 1968; Miller, 1972; Holmes & Bennett,1974). This has been due to these tests' use of highly structured formats. Conversely, others have encouraged the development of tests which would have the subject actively participate in the procedure (Kelman, 1967; von Radd, 1979; Novaco, 1975). Self-report scales and inventories which have the subject select words or responses from predetermined lists have been dismissed as valid measures of affect (Merskey, 1978; Knapp,1963; Oken, 1960). Although projective tests generally involve active participation on the part of the subject, and seldom involve checklists, they are heavily reliant on the clinician's objectivity, expertise and interpretation. Also, they are not easily quantifiable. On the other hand, clinicians have long understood the value of projective tests to be in their ability to generate affect-laden material.

Since anger is a reactive emotion (Novaco, 1975; Stearns, 1972; Brown, 1945) the test would ideally employ a provoking stimulus. There has been extensive work on anger-provoking stimuli. Canfield (1949) found that noise could produce anger. Gates (1926) found that hunger and fatigue were provocative. Others (Cleghorn, 1957; Gibson, 1962; Küntz, 1951) have found that proprioceptive stimuli, as well as induced hypoglycemia (Gellhorn, 1953) can provoke anger. However, apart from being inaccessible, most of these methods are derived from animal analogue studies. Several others have used verbal statements (Stearns, 1972; Novaco, 1975) or the recall or associations with such comments as anger-provoking stimuli with human subjects. Finally, waiting scenarios are well accepted as constituting anger provoking stimuli (Cohn, 1953; Stearns, 1972; Novaco, 1975; Doré & Kirouac, 1985). However, most studies have used either deception (Novaco, 1975) or written descriptions of waiting situations where the subject rates only how angry he would have felt. Martin (1961) found poor correlations between these types of self-reports, nonverbal behaviour and physiological measures of anger arousal.

To summarize these findings, a test of anger should possess a degree of experiential realism without being projective; should not be too rigidly structured; should allow for active, verbal participation on the part of the subject and should not employ a self-report or paper and pencil format. Further, in order to evaluate process, the test design should incorporate items which assess the stages in the awareness and expression of anger while allowing for alternative response patterns.

Anger Tests

The Buss and Durkee Hostility Inventory (BDHI) (Buss & Durkee, 1957) has served as the basis for many subsequent hostility and anger scales. Although the items on this test purport to assess what people do when they are angry, closer examination reveals a forced choice true/false item format which seems to assess more what people do when they are feeling hostile towards others. This is reflected in the content of the test scales which assess resentment, assault, verbal insult, suspicion, negativism and irritability. There are also no provisions to evaluate process. The BDHI does not meet all of the conditions listed above and is, therefore, not considered an appropriate anger test.

The Oken Scale (Oken, 1960) although intended to measure experienced, aroused anger and suppression, was not properly validated and relies entirely on an adjective checklist format. As previously discussed, checklists have been dismissed as a valid measure of affect. These lists also are suggestive and do not evaluate process.

Gottschalk et al. (1963) developed hostility scales which served as the prototypes for the later Gottschalk-Gleser Content Analysis Scales (Gottschalk & Gleser, 1969). These instruments, although well designed and researched, require extensively trained personnel to spend laborious hours administering and rating the interviews. Hence, this scale is impractical. Also, Schofer (1980) found that scales which rely entirely on the content analysis of random verbal behaviour are contaminated by socio-economic and educational status. Poor scores on these scales go hand in hand with low education and socio-economic status. He goes on to explain that this type of scale often provides little insight into generalizable affective processes.

Evans and Strangeland (1971) developed the Reaction Inventory (RI) to measure anger. Apart from being suggestive (it presents the subject with predetermined response possibilities), subjects' responses on this self-administered inventory likely involve cognitive rather than affective processes. Also, there is very little reliability or validity data on this inventory. The RI "Degree of Anger" score has shown moderate positive correlations with the total score of the BDHI (Biaggio et al., 1981). However, because the test 1. is particularly suggestive, 2. was developed to assess the extent or range of specific stimulus situations which evoke anger reactions (Spielberger et al., 1983), 3. does not assess expression, and finally, 4. because the validity data that is available is based on the BDHI, this test was not considered an appropriate convergent measure of anger.

Several other anger tests have been developed: for example, The Anger Self-Report (Zelin, Adler, & Myerson, 1972), The State-Trait Anger Scale (Spielberger, Russell, & Crane, 1983), and the Harburg Scale (Harburg et al., 1973). However, the same limitations discussed above apply to these measures. Also, there is less information available about these tests than the tests presented above.

The Awareness and Expression of Anger Indicator (AEAI)

The AEAI allows active participation of the subject without being rigidly structured nor lacking in experiential realism. Further, the test was designed to elicit affect-laden responses, that is, behaviours. The test results are objective and amenable to quantification. The design generates a range of scores which may be used to evaluate process. Finally, the test itself is quick and easy to use (administration time is15 minutes maximum).

The AEAI presents the subject with a short hypothetical waiting situation (see Appendix A for a copy of the AEAI). The subject is read the following by a trained interviewer:

> You have a regular arrangement with a friend to pick you up every morning at 08.15 to go to work. He arrives at 08.35.

The statement is not suggestive, it does not imply what, if any affect the subject should feel, and it does not suggest to the subject that his or her ride is late. The subject is asked how she or he would feel. The trained interviewer remains neutral (e.g., interviewer does not respond to subject statements and maintains neutral facial and body expression) and records the subject's first five responses. This part of the test evaluates the subject's awareness of anger, for example; either a spontaneous report of angry feelings, no report of angry feelings or a report of other reactions. The subject is then asked what she or he would do after getting into the car. Again the interviewer records the subject's first five responses. This part of the test evaluates the process of expression. Subjects' responses are rated according to a ranking system which determines which, if any, of the theoretical phases of expression has been reported (i.e., identification, confrontation or resolution). Since all responses are recorded, any alternate process of expression that may be reported may also be assessed. The recurring scenario design allows for the hypothetical situation to transpire over five consecutive hypothetical mornings. This functions to increase the offensiveness of the provoking situation. Resolution or expiry completes the non-inducing part of the test.

The subject is then asked how she or he would feel the next time she or he saw his or her friend, and what she or he would do. These questions provide insight into any subsequent resolution or presence of residual angry feelings. Finally, the subject is asked whether sshe or he would have felt angry on each of the hypothetical mornings that had transpired. The *Manual for the use of the AEAI* (Braha & Catchlove, 1984; see Appendix E) provides a description of the test, its constructs and use.

Psychometric properties of the AEAI

In two independent studies, the AEAI has demonstrated good inter-rater reliability. Catchlove and Braha (1985) reported an overall inter-rater reliability coefficient of 0.94 for total test scores. These findings were based on a sample of thirty subjects. They report inter-rater reliability coefficients on the three subscales of the AEAI as ranging from 0.91 to 0.98. All of these findings were found to be significant at beyond the 0.001 level. They concluded that the AEAI is an objective assessment instrument and that there was either minimal subjectivity involved in scoring the AEAI, or that raters made common subjective decisions (Catchlove & Braha, 1985). In a later study, Braha & Catchlove (1986) replicated these findings.

In an effort to assess the internal reliability of the AEAI, Catchlove and Braha (1985) dichotomously ranked subjects' responses for each part of the test against the ideal score level for that particular test item. They then used a phi coefficient to assess the relationship between passes and failures on each Scale of the test with passes and failures on other Scales of the test. The results are reproduced in Table B-I (see Appendix B). The lack of any significant positive correlations between test items is explained as reflecting the heterogenous nature of the scales and the constructs of awareness and expression. Although the authors have examined the internal structure of the test with regards to the test's subscales and the inter-rater reliability of the test, they have not examined the internal consistency of individual items in each scale with the total scores for each respective scale.

The AEAI has demonstrated some ability to differentiate patient groups. Braha and Catchlove (1986) found significant differences between pain patients' mean scores and medical patients' mean scores. However, they add that a discriminant function analysis would help to determine more specifically which items on the AEAI best discriminate between normals and pathological groups, and whether the test itself may be of use in discriminating between diagnostic categories.

Catchlove and Braha (1985), also found a range of scores with the use of the AEAI. They reported four distinct response patterns. First, they found that some subjects reported no awareness of anger and exhibited no constructive expression as measured by the AEAI. Other subjects reported no awareness of anger but through their responses, exhibited constructive expression of anger. Third, some subjects were aware of anger but did not express it at all. Finally, a proportionally small group of subjects reported an awareness of anger and expressed it constructively in their responses.

The AEAI: an example of its relevance to clinical problems

> So neither ought you to attempt to cure the body without the soul, and this ... is the reason why the cure of many diseases is unknown to the physicians of Hellas, because they are ignorant of the whole which ought to be studied also, for part can never be well unless the whole is well. (Plato 4th c. B.C.).

Central mechanisms including psychological variables have long been recognized as playing a significant role in the perception of chronic pain (Bonica, 1983; Catchlove, 1983; Blumer & Heilbronn, 1982; Sternbach, 1974; Engel, 1959). Unfortunately, there is little information on the specific psychological deficits in chronic pain patients (Melzack & Wall, 1982). Several authors have suggested that pain patients have difficulties with the awareness and expression of affect (Catchlove et al., 1985; Blumer & Heilbronn, 1982; Pilowsky & Spence, 1976). Anger has been implicated as one of the emotions with which pain patients seem to have problems (Catchlove et al., 1985; Pilowsky & Spence, 1976; Parkes, 1973; Merskey & Spear, 1967). Some patients seem to have difficulty labelling certain social situations as being anger-provoking or labelling their own arousal as anger or even being aware of their own arousal. Others who are aware of their feelings do not express them, or when they do express their feelings, they do so in an ineffective manner. These observations have been based on clinical experience and on results from the use of non-specific psychological tests (Catchlove & Braha, 1985).

The question remains: Is there a relationship between difficulties with the awareness and expression of anger and the incidence of chronic pain? And if so, what is the exact nature of this relationship?

Without an adequate measure of anger, research investigating the relationship between the awareness and expression of anger and chronic pain will remain at the anecdotal level. If this relationship exists, as several authors have reported, then the specific nature of this difficulty needs to be elucidated and assessed more reliably. During the past decade there has been an increase in research studying the psychological aspects of the chronic pain experience. For this research to be useful, methods must be developed which reliably assess theoretical constructs.

Test validity: a note on evolving concepts

Test validation procedures have evolved dramatically over the past eighty years. Perhaps the earliest empirical approach to the assessment of tests was the age-differentiation criterion used by Binet and Simon (1908, cited in Anastasi, 1986). It was assumed that the cognitive skills which constituted intelligence increased with age. Items were retained if the frequency of correct responses increased with the age of subjects. Each task, or item, was then assigned to the age level where the task was passed by a specified percentage of children in that particular age level. This was the procedure employed in the construction of the Stanford-Binet and several other tests of the era. In retrospect, it is not too difficult to come up with any one of the number of criticisms that have since been levelled against those practices. Norm-related validity is no longer acceptable criteria of construct validity. We now understand intelligence as being a multifaceted construct, with many culturally-relative manifestations.

From this era evolved the halcyon days of psychological test validation, the 1950's and 1960's. The period has been described as one of "blind empiricism" (Anastasi, 1986, p.6). As Messick (1980) argues: tests were expected to demonstrate sometimes up to seventeen types of validity. More alarmingly, validity was regarded almost as a "state": it was a goal; once attained, never disputed. Reliability and validity coefficients abounded. There was an overemphasis on purely empirical validation procedures, and a deemphasis on knowledge, hypotheses, theoretical rationale, and construct formulation. This may have been a rebellion against the rampant armchair theorizing that characterized much of the literature during the early part of this century. But, as Anastasi (1986, p.6) argues: "Theory need not be subjective speculation". Increasingly, researchers have become concerned, once again, with theoretical rationale in all phases of the test development process.

By the end of the 1960's, psychometrists had come full circle. Test developers had pared down their expectations of tests. Three broad categories of validity emerged: criterion-related, content and construct validity (*Standards for educational and psychological tests*, 1974). It was felt that validity could be reduced to two questions: "(a) What can be inferred about what is being measured by the test? (b) What can be inferred about other behavior?" (*Standards*, 1974). There was some consensus that, depending on the type of inferences we wished to make from test scores, one needed to establish only the appropriate form of validity. Although no one would have disagreed that the three validity domains were interrelated, in practice, the three types of validity were treated as mutually exclusive entities (Messick, 1980).

There is now abreast a movement to redefine our concepts of validity. Prominent theorists have suggested that, for purposes of clarification and rigour, the term "construct validity" should be retained to represent many of the previously defined aspects of validity. Messick (1980) and others (Frederiksen, 1986; Guion, 1977) have suggested retaining all the old concepts of validity, but renaming them, and the procedures used in their name. This would be done in an effort to more accurately reflect their purpose. To use an example relevant to this study, Messick (1980) and Guion (1977) have both suggested that the old concept's of convergent, discriminant, and factorial validity can be unified. They suggest using these concepts to reflect tangible procedures which can provide evidence to support the theoretical notion of construct validity.

However, because of the influence of convention, and to avoid confusion, reference will be made throughout this study to convergent, discriminant and factorial validity. It is understood, nevertheless, that these terms and the procedures associated with them are being replaced by terms such as: convergent coherence, discriminant distinctiveness and factorial composition of tests. It is also understood that the latter are now generally considered as useful procedures which establish applied aspects of construct validity (Messick, 1980; Green, 1981; Anastasi, 1986).

Test validity and the AEAI

The AEAI has shown promise as an objective assessment instrument. It has demonstrated acceptable inter-rater reliability, an ability to discriminate between clinical and non-clinical groups, and it has elicited response patterns which provide support for the theoretical process of constructive resolution. However, there are other psychometric properties which need to be investigated.

Internal validity. Further studies need to be conducted on the internal reliability of the test. Specifically, AEAI scales need to be assessed for internal consistency, homogeneity and adequacy of content sampling. Although there would be an expected degree of practice effect, studies should be conducted which examine test re-test, or if available, alternate form reliability. Factor analytic studies need to be conducted to confirm that the reported correlations between test variables were due to the three anger dimensions the test purports to measure and were not due to other unidentified latent structures (Engelsmann, 1982). Internal item analytic studies need to be conducted to determine whether or not the response patterns that have been reported with the use of the AEAI were due to order effects inherent in the test design.

External validity. Finally, convergent and discriminant validity trials should be conducted to confirm that the test is in fact measuring anger and not other distinct constructs (Campbell, 1960; Costa & McCrae, 1983).

A final comment on test validity

All too often test developers fail to recognize that validity is a theoretical construct. Any use of the term "validity" is predicated by normative judgement of research evidence (*Standards*, 1985), and as such will always be subjective.

The purpose of this study

It was my intent in proposing and executing this thesis to address the need for an empirically testable, valid instrument which would measure the awareness and expression of anger. I wished to study whether the AEAI could fulfil this need. This test purported to measure specific and clearly defined processes. It operationalized the construct of anger in a quantifiable way. The discriminant ability and high reliability coefficients of the AEAI warranted further validation study.

The aim of this project, therefore, was to assess the validity of the AEAI as a measure of the awareness and expression of anger.

HYPOTHESES

Internal validity of the AEAI

1. *Scale homogeneity and consistency.* It was hypothesized that there would be intra-subscale homogeneity for each of the three subscales of the AEAI (i.e., Scales A, E, and I). This would be indicated by positive relationships between item scores, within each of the three AEAI scales.

2. *Mood induction.* The repeated scenario of the AEAI was presumed to induce increased levels of anger. Theoretically, with each day that transpires, subjects should report higher intensity ratings in response to Scale A questions. As well as exhibiting intra-scale homogeneity, it was therefore hypothesized that there would be a within-subjects effect for each successive item in Scale A.

3. *Confirmatory factor analysis.* It was hypothesized that the AEAI would assess three main dimensions of anger: non-induced awareness of anger, expression of anger, and, induced awareness of anger. Further, most of the test variance accounted for by the fifteen AEAI test variables would load on these latent dimensions.

4. It was hypothesized that factor analysis of Scale A would reveal that the common variance between the variables could be accounted for by a general latent factor.

5. It was further hypothesized that factor analysis of Scale E variables would elicit three latent factors which would reflect the processes of identification, confrontation and resolution.

6. It was hypothesized that factor analysis of Scale I would reveal that the common variance between the variables could be accounted for by a general latent factor.

7. Finally, it was hypothesized that factor analysis of Section I items (Scales A and E combined) would reveal two general latent factors: a general awareness factor and a general expression factor.

8. *Order effects.* It was predicted that if suggestive test items preceeded

non-suggestive test items; scores for remaining test items would be higher than when suggestive test items succeeded non-suggestive test items (i.e., there would be an order effect).

9. Moreover, it was predicted that there would be no order effect for AEAI Scales A or E.

External validity of the AEAI

10. *Convergent validity of the AEAI.* It was predicted that there would be moderate-order positive relationships between the AEAI and the two convergent tests of anger, the Anger Inventory (Novaco, 1975) and the Multidimensional Anger Inventory (Siegel, 1983).

11. *Convergent validity of AEAI scales.* a) Because of the nature of the AEAI test items, it was hypothesized that there would be a high-order positive relationship between subjects' AEAI Scale I scores, and their scores on the Anger Inventory and those subscales of the Multidimensional Anger Inventory which assess the awareness of anger. However, b) because of the non-inducing nature of AEAI Scale A and E items, it was predicted that there would be low-order positive relationships between AEAI Scale A and E scores and the other convergent tests of anger.

12. *Discriminant validity of the AEAI.* It was predicted that there would be zero or low-order correlations between AEAI total and scale scores and Marlowe-Crowne Social Desirability Scale scores.

Although it can not be assumed that any psychological research or instrument is free from the effects of response sets, certain designs would seem to be more vulnerable. Because of the partially didactic, or interactional nature of the AEAI, it seemed appropriate to determine the influence of this response set on subjects' scores. Many factors affect responses: subject perceptions of what the experimenter may expect, desires to protect one's own image, and attempts to please or frustrate the experimenter all come into play (Anastasi, 1982). However, several authors have described the tendency to choose what subjects' believe to be socially desirable responses as being the most prevalent response set (Edwards, 1957; Crowne & Marlowe, 1964).

13. Finally, it was predicted that there would be zero or low-order correlations between AEAI total and scale scores and subjects' beliefs about the consequences of expressing their anger (i.e., CAS scores).

The addition of this independent measure is intended to be consistent with Bandura's (1973) guidelines discussed earlier. The author concurs with Bandura's criticisms of anger tests. Nevertheless, because the belief factor was not incorporated in any of the anger tests used in this study, it is included here as a discriminant consruct.

METHOD

SUBJECTS

Three subject groups were used in this study. All three groups were comprised of students from Saint Mary's University in Halifax, Canada. Subject selection criteria for all groups were that subjects' mother tongue was English, that subjects were literate and between 17 and 60 years of age.

The three experimental groups were selected to assess two basic aspects of the research mandate. First, all groups were used to assess the internal validity of the AEAI. Data from Group 1 was used to determine how well the AEAI demonstrates external validity. Also, Group 1 data was collected to develop the multitrait-multimethod matrix which assessed the convergent and discriminant validity potential of the AEAI.

Group 1 was comprised of 128 subjects. The 43 males (33.6 percent) and 85 females (66.4 percent) in this group were used to test hypotheses 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, and 13. The mean age for subjects in Group1 was 21.73 years (*S.D.*=4.76, Mode=18 years, range=17-44 years.). There was no significant difference in age between males and females in Group 1 (t=0.65, 126*df*, p >.5). Subjects in Group1 had an average of 14.33 years of schooling (*S.D.*=1.35, Mode=13, range=12-18 years). There was also no significant difference in years of schooling between males and females for Group 1 (t=0.25, 125*df*, p >.8).

Group 2 was comprised of 148 subjects made up from the Group 1 total plus a further twenty subjects who were recruited exclusively for participation in this group. Data from Group 2 was used to evaluate hypotheses 3, 4, 5, 6, and 7.

Group 3 was comprised of forty subjects who were used to test hypotheses 8 and 9. Five conditions with eight subjects per cell made up the total. Although eight subjects per cell were used, five subjects per cell provided adequate power for this factorial design (Phi squared effect, $p < .05$). There were 18 males (45 percent) and 22 females (55 percent) in this group. The mean age of subjects in Group 3 was 19.83 years ($S.D.=2.48$ years, Mode=19 years, range=18-31 years). There was no significant difference in age between males and females in Group 3 ($t=0.75$, 38df, $p > .4$). Subjects in Group 3 had an average of 13.45 years of schooling ($S.D.=0.78$, Mode=13 years, range=13-16 years). There was no significant difference in years of schooling for males and females in Group 3 ($t=0.77$, 38df, $p > .4$). Also, there was no significant difference between the proportion of males to females in any of the five cells ($X^2=1.75$, 4df, $p > .4$).

MEASURES

Order modified AEAI

Five alternate forms of the AEAI (Forms 2 to 6) were developed which presented the three AEAI scale items in reversed order to comply with the counterbalanced design described earlier. Appendix C presents the five order-modified forms of the AEAI.

Convergent psychological tests

The Multidimensional Anger Inventory (Siegel, 1983) and the Anger Inventory (Novaco, 1975) were used as convergent psychological tests.

The Multidimensional Anger Inventory (MAI) is a 38 item self-report inventory designed to measure the frequency, duration and magnitude of anger. It also assesses the range of situations to which an individual responds with anger, the individual's mode of anger expression and the extent of hostility in the individual's outlook (Siegel, 1983).

A review of the literature (to July, 1987) revealed that there were no independent psychometric studies on the MAI or independent studies which employed the MAI. Nevertheless, the instrument was selected for use for several reasons. First, the test developer provided reliability and validity data which were not generally available for most other measures of anger. Second, there were data available on the origin and rationale of each item and scale of the test. Third, there were psychometric data available for each item of the test. Fourth, the author provided detailed results of factor analytic studies on the instrument. Fifth, the test included scales which purported to measure the same dimensions of anger that the AEAI purported to measure. Finally, the MAI demonstrated adequate test-retest reliability ($r = 0.75$) and high internal consistency (alpha =.84 based on a college student sample N =198) (Siegel, 1986).

The Anger Inventory (NAI) (Novaco, 1975), assesses anger- eliciting situations. It is a 90 item inventory of potentially anger- arousing situations. The subject is asked to rate the degree to which the incident described would anger them (Novaco, 1975). The items on this test were used to partially validate AEAI measures of both induced and non-induced awareness of anger (Scales A and I respectively). Novaco (1975) reports internal consistency coefficients of between .94 and .96 for the NAI. The NAI has also discriminated more clinically angry groups from less clinically angry groups (Deffenbacher, Demm, et al., 1986; Novaco, 1976). Finally, the NAI is in widespread use and is considered to be an effective measure of a subject's awareness and ability to admit feelings of anger (Biaggio, 1980).

A short version of the Anger Inventory was administered in this study. Novaco (1975) and others (Biaggio, 1980 &1981) have reported that the 45 item short version has good reliability and correlation with the longer version. The Minitab (Ryan, Joiner, & Ryan, 1976) computer generated random numbers programme was used to select 45 items between 1 and 90 which comprised the short version of the Anger Inventory.

Discriminant psychological tests

The Marlowe-Crowne Social Desirability Scale (M-C) (Crowne & Marlowe, 1960), assesses both an individual's need for attention and sympathy and his or her tendency to choose socially desirable responses rather than personally reflective responses. Individual scores on this scale have been described as enabling the examiner to evaluate the degree to which a socially desirable response set affects test scores (Anastasi, 1982) and as such individual scores were used as a test of of the AEAI's capacity to discriminate between these constructs.

Subjects also received a short questionnaire which assessed their beliefs about the consequences of expressing anger (Cognitive Attitudes Scale-CAS). The data from this test were included as an independent variable in a multitrait-multimethod matrix to determine whether this factor accounts for any portion of the variance in AEAI scores. There is no reliability or validity data available on this measure.

Finally, all subjects received a short demographic data form which requested information on the subject's sex, age, level of education, and programme of study.

Appendixes C-1 to C-10 (Appendix C) present copies of all measures used in this study.

PROCEDURES

Subject recruitment

Subjects were recruited using three methods. The majority of subjects were recruited from introductory level psychology classes. Individual course instructor's consent was obtained prior to initial contact with students solicited from these classes. Following a brief description of the nature and intent of the study, students were asked to sign their name and telephone number on a sign-up sheet (refer to Appendix D for copy of statement). Students who volunteered were subsequently telephoned and given an appointment time. Prior to testing, informed consent was obtained from the students in accordance with Saint Mary's

University Department of Psychology Ethics Committee guidelines. Students who were enrolled in either Introduction to Psychology or in the Psychology of Learning course received credit points for participation in this study. Other subjects were recruited through sign-up sheets which were posted on the Department of Psychology subject recruitment bulletin board. Finally, a small number of subjects were recruited from the ubiquitous group of students which graces the psychology department corridors.

Test administration

Following informed consent, subjects in Group 1 completed the demographic data questionnaire, the AEAI, and the validation test battery. All tests were administered according to standard procedures which were included in the respective test manuals. Convergent and discriminant tests were administered in a random order to eliminate any systematic order effects inherent in the battery. Each of the four test instruments was assigned a code between 1 and 4. The Minitab (Ryan, Joiner, & Ryan, 1976) random numbers programme was used to generate four digit numbers between 1 and 4. Individual test packets were then collated according to this random sequence representing order of administration. To ensure anonymity of subjects, only their appointment times and dates were recorded.

Following informed consent, subjects in Group 3 also received the demographic data questionnaire which was followed by the respective order modified version of the AEAI. Subjects were randomly assigned to one of the five cells, each of which comprised an experimental subject group.

For identification purposes, random numbers were assigned to each completed test prior to coding and scoring.

Scoring and coding the test battery

All tests were scored according to the method described in the respective test manual. A blind rater scored all test protocols: raters were unaware of subjects' identities or group membership.

The AEAI. Each recorded subject response is ranked from A to E according to the theoretical response progression presented earlier. Each rating is then assigned a point level corresponding to the AEAI scoring guide (Braha & Catchlove, 1984 see Appendix E for a copy of the *Manual for the use of the AEAI*). Should the last subject response occur prior to the fifth hypothetical morning, the scoring system weights the responses giving more points to response progressions closer to the idealized response progression (Catchlove & Braha, 1985), and fewer points to delayed or theoretically inappropriate responses. Scores are derived for non-induced awareness (Scale A), expression (Scale E) and induced awareness of anger (Scale I).

For the purposes of factor analysis, AEAI protocols were coded according to a nine point scale, from 0 to 8, which conformed with the nine possible ratings for each test item.

DATA ANALYSIS

Primary data investigation conformed with two broad categories of statistical analysis: correlational procedures and analysis of variance. Each category of analysis was used to test the veracity of the research hypotheses.

Correlational procedures

Correlational procedures encompass a vast array of related statistical techniques. The primary concern in correlational research is the description and measurement of the relationship between variables (Ferguson, 1981). Bivariate correlation was used to test hypotheses 1, 10, 11, 12, and 13. These analyses included Pearson Product-Moment correlations, item-mean biserial correlations, item analytic techniques as well as other related statistics. Internal factor analytic techniques are used to examine the internal structure of a set of variables (Thorndike, 1978). Hence, this class of correlational procedures was used to test the factorial validity of the AEAI (hypotheses 3, 4, 5, 6, and 7). Factor analysis was the primary tool of this analysis.

Analysis of variance

Simple analyses of the significance of the difference between means (t-tests and analyses of variance) were used to assess hypotheses 2, 8, and 9.

Computing systems

Analyses 1, 2, 3, and 5 were performed on a Digital Equipment Corporation Vax 11780 computer utilizing SPSSx (1983) statistical algorithms and software. Analysis 4 was performed on an Apple MacIntosh Plus utilizing StatView 512+ (1986) statistical software.

Analysis 1

Analysis 1 was performed on the results obtained from Group 1 to test hypothesis 1. Coefficient alpha was calculated to determine the extent of intra-scale homogeneity for each scale of the AEAI as well as overall test consistency.

Coefficient alpha is an index of inter-item homogeneity. It can be said to assess the extent to which test items measure the same thing (Lemke & Wiersma, 1978; Cronbach, 1951). Coefficient alpha can be used as an estimate of the proportion of the item variance due to common factors among the items, and, as such, is the statistic of choice for determining homogeneity and structure of test scales (Cronbach, 1951; Anastasi, 1982).

Analysis 2

Analysis 2 was performed on the results obtalned from Group 1 to test hypothesis 2. A single classification repeated measures analysis of variance was performed on Scale A of the AEAI. Each item of the scale, representing different days, was used as the independent variable. Subject scores on each day was the dependent variable for these analyses.

Analysis 3

Analysis 3 was performed on the results obtained from Group 2 to test hypotheses 3, 4, 5, 6, and 7.

Although principal components factor analysis (PCA), principal axis analysis, or, principal factor analysis, are often reported in this type of study, several characteristics of these procedures contraindicated their use. First, and perhaps foremost, PCA and related forms of analyses are based on the assumption that both subjects and variables represent populations (Daultrey, 1976; Thorndike, 1978; Cattell, 1978). Second, the statistical extraction algorithms of these procedures generally produce psychologically meaningless factors (Cattell, 1978; Jöreskog & Sörbom, 1979). Because these procedures are correlation oriented rather than variance oriented, factors will be extracted which may be comprised of variables with little common variance (Harman, 1967). Finally, these procedures always require the same number of factors as variables to account for all of the variance in the variable set. There is little utility in a ten factor solution of a ten variable set. Further, the practice of retaining only those factors which account for more than one variable equivalent of variance (i.e., one eigenvalue) may result in the loss of potentially significant factors and variables (Cattell, 1966). PCA extractions with the Kaiser criteria (formerly described) is rarely able to reproduce the original correlation matrix. This results in high residual correlations. The equation for PCA extraction, and its use in the development of a reduced rank model, includes a uniqueness term that is comprised of non-common variance. This uniqueness term is thought to be composed in part of specific variable variance, and in part of error variance. In order to extract factors composed purely of common variance, elements other than unities need to be used in the diagonals of the variable correlation matrix.

Several alternative methods for dealing with these limitations have been developed (see Harman, 1967; Gorsuch, 1974 for reviews). Unless otherwise indicated, maximum-likelihood estimates for communalities (Jöreskog & Sörbom, 1979) were used in all factor analytic procedures reported in this study. Maximum-likelihood estimates of the communalities (diagonals), produce statistical or sample estimates of the actual population factor loadings (which presumably are composed of only common variance) (Kaiser & Caffrey, 1965; Thorndike, 1978). Because there is an assumption that the subject pool is a sample (Thorndike, 1978), the PCA assumption of minimal error variance in the data is not accepted.

Varimax rotations were used where neccessary to assist in the interpretation of factors and to remain consistent with the theoretical test postulates. Unlike other rotations, the varimax procedure attempts to spread the common variance across selected factors while retaining orthogonality of the factor axes (Harman, 1967).

A maximum-likelihood factor analysis was performed on the correlation matrix for all 15 AEAI items. Varimax rotation of the three extracted factors was initiated to clarify interpretation.

Typically, an arbitrary factor loading cut-off point is selected to signify significance (Thorndike, 1978). This procedure works moderately well with extraction methods which base the analyses on an unreduced correlation matrix. However, the maximum-likelihood method is based on a reduced matrix with elements other than unities in the diagonal of the correlation matrix. An arbitrary cut-off may enhance the likelihood of significant loadings because there is a higher ratio of variance accounted for by the factor to proportion of variable variance included in the correlation matrix (Cattell, 1981). Nevertheless, the use of arbitrary cut-offs is the convention. A more conservative method of selecting cut-off criteria is to take a proportion of the average communality (variance accounted for) as indicative of significance. Thorndike (1978) suggests the use of the square root of one quarter of the average communality as a more cautious criteria. Both the conventional method and the more cautious method were employed in the analysis of this data set.

When the planned solution accounted for only a small percentage of the total variance for many items, further extractions were conducted to attain more representative factors. Also as planned, separate analyses were conducted for each scale of the AEAI, and for Section I items. Once again, when the planned factor solutions did not account for large proportions of many of the individual item variances, further extractions were conducted to attain more representative factors. In this way, meaningful solutions could be developed which would account for more item variance, while maintaining the assumptions of this analytic method.

Further issues in factor analysis

The number of factors problem. Thorndike (1978) and others (Jöreskog & Sörbom, 1979; Gorsuch, 1974) suggest that there are two types of criteria for determining the number of factors to select for a maximum-likelihood solution: criteria based on logical, theoretical considerations, and those based on statistical theory. They also discuss two schools of thought on the problem.

One school, predominated by British factor theorists, recommends selecting as parsimonious a solution as possible. Conversely, the other school, predominated by American factor theorists, has argued that "the number of real common factors operating in any set of variables is indefinitely large and can be 'infinite' (Cattell, 1958, p.801)." Most factor analysts probably stand somewhere in the middle, and adhere to a principle perhaps best described by Thorndike. Thorndike (1978, p.273), paraphrases a quotation from Tom Lehrer: "What it boils down to, . . . is that factor analysis is like a sewer; what you get out of it depends on what you put into it." He goes on to say that, "The factors that emerge from any factor analyses are a function of the variables selected for study. A particular selection will precondition certain factors to appear in the analyses and prohibit others from appearing (Thorndike, 1978, p.273)."

Finally, and perhaps the most relevant point with respect to this study, is taken again from Thorndike (1978, p.282):

> There is an increasing amount of evidence that statistical significance is not a sufficient condition on which to base a decision about importance (e.g., Hays, 1963). To be worthwhile, a factor should be statistically significant *and meaningful* [italics in original text]. This implies that the judgement must be based on both logical and statistical criteria.

This also implies that statistical significance alone does not warrant retaining a particular solution: but rather, that it suggests that a factor may be of value (Comrey, 1978).

Therefore, results were examined for data and factors that not only met the minimum standards of statistical adequacy, but that also confirmed the three dimensions of anger proposed by the AEAI.

The adequacy of factor solutions. The single most important criteria in determining the adequacy of one factorial solution over another is the magnitude of the differences between the original and the reproduced correlations (called the residuals) (Thorndike, 1978). Cattell (1978 & 1958) has said that this is the only justifiable standard. Of course, a problem arises when attempting to assess the adequacy of one solution over another of different rank. It is almost always the case that the residuals will be smaller as the number of factors increases (Gorsuch, 1974; Thorndike, 1978). Harman (1967) and others (Thorndike, 1978) suggest that the addition of another factor to an already adequate solution should result in noticeable decreases in the size of the residuals, and allow for clearer interpretation of the factors.

Analysis 4

Analysis 4 was performed on the data obtained from Group 3 to test hypotheses 8 and 9. A counterbalanced design (Keppel, 1982) was used to examine order effects manifest in AEAI Scale A, E and I scores. Figure II illustrates the five order of presentation combinations which were used. Three single classification analyses of variance were conducted to determine whether there were any significant differences between scale scores due to order of presentation of scale items. Mean scores and standard deviations for each item of each scale were examined for the five order-modified forms of the AEAI. A random sample of AEAI data from Group 1 was compared to these data to determine whether there were any significant deviations between the distribution of scores. In the event of significant differences, a random sample of data from Group 1 AEAI scores would have been included in a secondary analysis to compare the original AEAI scales to scores on Forms 2 to 6.

AIE
IEA
IAE
EAI
EIA

Figure II. Order of presentation of AEAI scales.

Analysis 5

Analysis 5 was used to determine the convergent and discriminant validity of the AEAI and test hypotheses 10, 11a, 11b, 12, and 13. Data from Group 1 were entered into a multitrait-multimethod matrix.

Pearson Product-Moment correlation coefficients were calculated between all continuous variables measured in this study. A coefficient matrix was derived which presented the relationship between inter-test items, inter-test subscales, item to test scale subtotal scores, item to total test scores, and all the latter to scales derived from extracted factors.

For the purposes of inclusion into the multitrait-multimethod matrix, MAI scales were collapsed into two broad scales: Awareness and Expression. All scales which were collapsed into the same new scale intercorrelated significantly. The following MAI subscales were collapsed into a MAI Awareness scale: Frequency, Duration, Magnitude, Range of anger-eliciting situations, and Modes of Expression subscale, anger-in. MAI Modes of Expression scales Anger-out, and Anger-discuss were collapsed into a MAI Expression scale.

MAI factor derived scales were used in another multitrait-multimethod matrix. Close examination of the individual item or variable loadings presented by Siegel (1983) revealed that MAI factors 1, 2, and, 3 (labelled General anger, Range of anger-eliciting situations, and Hostile outlook, respectively) had a fair degree of common variance. The items included in these three factor scales could be described as being concerned with assessing subjects' awareness of anger. Therefore, these three original MAI factors (factors 1, 2, &3) were collapsed into a new MAI factor derived awareness scale for inclusion into the second multitrait-multimethod matrix.

Similarly, factors 4 and 5 (labelled Anger-in/brood, and Anger-out/brood respectively) (Siegel, 1983) could be described as examining style of expressing anger. In fact, the latter were derived solely from original MAI Modes of Expression items. Hence, these two original MAI factors (factors 1 & 2) were collapsed into a new MAI factor derived expression scale also for inclusion into the second multitrait-multimethod matrix.

Multitrait-multimethod matrices were examined for evidence of the convergent and discriminant validity of the AEAI.

In order to assess the convergent validity of the AEAI, all correlations were isolated between traits measured by the AEAI and congruent traits measured by the convergent tests.

Evidence for the discriminant validity of AEAI scales was derived by two means. First, correlation coefficients were extracted between traits measured by the AEAI and the non-congruent traits measured by the discriminant tests. Second, and less crucial with respect to discriminant validity, coefficients were derived between traits measured by the AEAI and non-congruent traits measured by the convergent tests (Kavanagh, MacKinney, & Wolins, 1971; Humphreys, 1960; Zurawski & Smith, 1987).

RESULTS

The AEAI: unweighted responses

AEAI Scale A and E data presented below reflect a combination of actual subject responses and the effects of the AEAI score weighting system.

Table F-1 (see Appendix F) presents the frequency of unweighted subject responses for each hypothetical morning of the test. Over 75 percent of subjects continued beyond the third hypothetical morning of the test. Thus, items 1 to 6 are barely affected by the weighting system. However, items 7 and 9 from Scale A, and items 8 and 10 from Scale E were increasingly affected by the weighting system due to decreasing frequency of actual subject responses for these items (see Table F-1).

Analysis 1

Internal reliability of the AEAI. Results from analysis 1 appear to support hypothesis 1. Table 2 presents the results of an analysis of internal consistency and homogeneity of the AEAI and its three scales based on a sample of 128 university students (Group 1). The values presented in Table 2 suggest high intra-subscale homogeneity for Scales A and E and moderate intra-scale homogeneity for Scale I.

Scale A and Scale E demonstrated good internal reliability with unbiased alpha coefficients of .89 and .78 and standardized item alpha coefficients of reliability of .88 and .79 respectively.

Scale I had an overall unbiased alpha coefficient of reliability of .43 and a standardized item alpha coefficient of .43.

The global unbiased coefficient of internal reliability for the AEAI was .80 (see Table 2).

Item-item, item-scale, item-section, and item-total correlations. There were generally high positive correlations between scale items, item to scale scores, item to section scores, and, item to total scores. Close examination of the correlation matrices (i.e., Tables 3, 4, and 5) revealed interesting patterns of correlations. The first few items in each scale did not tend to correlate as highly with the last items in the scales as did the middle and last items of each scale (e.g., item 1 & items 7 & 9 in Table 3, versus items 5, 7, & 9). Results from the analysis of each scale are presented below.

Table 3 presents Pearson Product-Moment inter-item correlation coefficients for AEAI Scale A. Correlations ranged from .28 between item 1 and item 9, to .91 between item 7 and item 9. All inter-item correlations for Scale A were significant ($p \leq .001$).

Table 4 presents Pearson Product-Moment inter-item correlation coefficients for AEAI Scale E. Correlations ranged from .16 between items 1 and 10, to .87 between items 8 and 10. Correlations between items 2 and 8, and 2 and 10 were significant at beyond the 95 percent level ($p \leq .05$). All other inter-item correlations for Scale E were significant at beyond the 99.9 percent level ($p \leq .001$).

Table 2

Internal consistency coefficients for the AEAI and its scales (N=128)

Scale	Number of Items	Alpha	S.I.A.[a]	U.E.R.[b]
Awareness (A)	5	.8913	.8838	.8930
Expression (E)	5	.7794	.7928	.7829
Non-Induced Awareness (I)	5	.4254	.4254	.4344
Total	15	.7992	.7804	.8024

[a] Standardized item reliability. [b] Unbiased estimate of reliability.

Table 3

Inter-item correlations for Scale A

	Item 1	Item 3	Item 5	Item 7	Item 9
Item 1	---				
Item 3	.51*	---			
Item 5	.38*	.69*	---		
Item 7	.36*	.64*	.83*	---	
Item 9	.28*	.62*	.80*	.91*	---

* $p \leq .001$.

Table 4

Inter-item correlations for Scale E

	Item 2	Item 4	Item 6	Item 8	Item 10
Item 2	---				
Item 4	.34*	---			
Item 6	.22*	.36*	---		
Item 8	.17**	.44*	.72*	---	
Item 10	.16**	.43*	.63*	.87*	---

* $p \leq .001$. ** $p \leq .05$.

Table 5 presents Pearson Product-Moment inter-item correlation coefficients for AEAI Scale I. Coefficients ranged from .63 between items 11 and 13, to .90 between items 11 and 12. All inter-item coefficients for Scale I were significant ($p \leq .001$).

Table 5

Inter-item correlations for Scale I

	Item 11	Item 12	Item 13	Item 14	Item 15
Item 11	---				
Item 12	.90*	---			
Item 13	.63*	.64*	---		
Item 14	.69*	.65*	.75*	---	
Item 15	.79*	.76*	.73*	.88*	---

* $p \leq .001$.

Table 6 presents Pearson Product-Moment item to scale, item to section, and, item to total correlation coefficients for the items in AEAI Scale A. Item to scale coefficients ranged from .56 for item1, to .93 for item 7. All item to scale correlations for Scale A were significant ($p \leq .001$). Item to section coefficients for Scale A items ranged from .55 for item 1, to .69 for item 3 and were all significant ($p \leq .001$). Item to total-score coefficients ranged from .55 for item 1, to .82 for item 9. All of these correlations were also significant ($p \leq .001$).

Table 6

Item to scale, item to section, and item to total correlations for AEAI Scale A items[a]

	Scale A	Section I	Total
Item 1	.56*	.55*	.55*
Item 3	.83*	.69*	.73*
Item 5	.91*	.64*	.76*
Item 7	.93*	.66*	.81*
Item 9	.90*	.64*	.82*

[a] Pearson Product-Moment Correlation Coefficients.
* $p \leq .001$.

Table 7 presents Pearson Product-Moment item to scale, item to section, and, item to total correlation coefficients for the items in AEAI Scale E. Item to scale coefficients ranged from .54 for item 2, to .84 for item 8. All item to scale coefficients for Scale E were significant ($p \leq .001$). Item to section correlations ranged from .06 for item 6, to .31 for item 1. Correlations between items 4 and 6 and Section I subtotal scores were not significant ($p \leq .05$). All other item to section correlations for Scale E were significant ($p \leq .001$). Item to total correlation coefficients for Scale E ranged from .39 for item 2, to .50 for item 10. These were all also significant ($p \leq .001$).

Table 7

Item to scale, item to section, and item to total correlations for AEAI Scale E items[a]

	Scale E	Section I	Total
Item 2	.54*	.31*	.39*
Item 4	.70*	.08	.34*
Item 6	.79*	.06	.40*
Item 8	.84*	.19**	.47*
Item 10	.81*	.27*	.50*

[a] Pearson Product Moment Correlation Coefficients.
* $p \leq .001$. ** $p \leq .01$.

Table 8 presents Pearson Product-Moment item to scale, item to section, and, item to total correlation coefficients for the items in AEAI Scale I. Item to scale coefficients ranged from .84 for item 13, to .94 for item 15. All item to scale correlations for Scale I were significant ($p \leq .001$). Item to section correlation coefficients for Scale I items ranged fror .37 for item 15, to .58 for item 13 and were all significant ($p \leq .001$). Item to total-score coefficients ranged from .00 for item 15 to .60 for item 12. Only the correlations between Scale I items 1, 2 and 3 and total scores were significant ($p \leq .001$).

Table 8

Item to scale, item to section, and item to total correlations for AEAI Scale I items[a]

	Scale I	Section II	Total
Item 11	.89*	.53*	.54*
Item 12	.87*	.56*	.60*
Item 13	.84*	.58*	.29*
Item 14	.90*	.47*	.01
Item 15	.94*	.37*	.00

[a] Pearson Product-Moment Correlation Coefficients.
* $p \leq .001$.

Analysis 2

Mood induction. Hypothesis 2 was supported. There was a significant difference between mean anger intensity ratings for Scale A items ($F_{(127,1)}=181.96$, $p \leq .001$). Table 9 presents a summary of the repeated measures analysis of variance for Scale A items.

Table 9

Summary table of repeated measures analysis of variance of Scale A Items

Source	Sums of Squares	*df*	Mean Square	*F*
Within Cells	2100.98	127	16.54	181.96*
Constant	3010.23	1	3010.23	

*$p \leq .0001$.

Further, orthonormalized polynomial contrasts between the quantitative responses to Scale A items revealed a linear trend component ($t = -4.42$, $df = 126$, $p \leq .0001$). Table 10 presents the results of this analysis of trend.

Table 10

Analysis of trend for Scale A responses

Source	Sums of Squares	*df*	Mean Square	*F*
Within Cells	913.84	508	1.80	10.97*
Constant	78.96	4	19.74	

$*p \leq .0001$.

Table F-2 (see Appendix F) presents mean anger ratings for each item from Scale A.

Analysis 3

Review. Results from analysis 3 supported, in part, hypothesis 3. There was an adequate three-factor solution to the correlation matrix comprised of the 15 AEAI variables. However, the results obtained through the extraction of four factors seemed to be more adequate.

Similarly, the results from analysis 3 supported, in part, hypothesis 4. There was an adequate one-factor solution to Scale A variables. However, a two-factor solution proved to be more adequate.

Hypothesis 5 was, in part, supported. Results indicated that a three-factor solution described well the relationship between the variables in Scale E.

However, results from analysis 3 did not support hypotheses 6 and 7. There was not an adequate one-factor solution for Scale I. A two-factor solution proved to be more adequate. Nor was there an adequate two-factor solution for Section I variable inter-correlations. The three-factor solution presented below was preferable.

Maximum-likelihood (ML) factor analysis of the AEAI: three factor solution

Maximum-likelihood extraction with a three-factor criteria limit accounted for nearly 62 percent of the total variance in the 15 variable set. Table F-3 (see Appendix F) presents the unreduced correlation matrix for the original variables.

Analysis of the sample distribution and sample characteristics indicated that the sample was drawn from a multivariate, normal distribution and that factor analysis was appropriate for this data. A significant Bartlett chi-square transformation test of sphericity (X^2=2034.69, $p \leq .001$) indicated that there were significant correlations between all variables. Small magnitude partial correlations between the unique factors were assessed by the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (KMO statistic = 0.80).

This high ratio indicated that the correlations between the variables could be explained by the other variables and could not be accounted for by correlations between unique variance components.

Varimax rotation of the three factors revealed a latent structure consistent with the three AEAI scales. Using factor loadings of .30 or greater as inclusion criteria, all but one variable loaded exclusively on one factor. Table 11 illustrates the rotated factor matrix for the 15 AEAI items. Using the cut-off criteria suggested by Thorndike (1978), factor loadings of 0.40 would be used as inclusion criteria. Table 11 illustrates the results which are obtained through the use this cut-off criteria which accommodates the reduced correlation matrix. Communality may be used as a measure of the total variance accounted for by the factors extracted: that is, for orthogonal factors communality squared equals total variance accounted for (Thorndike, 1978). The three factors were able to account for large portions of the variance for most items except for items 1, 2, 4, and, 11 (see Table 11, Communality column). The three -factor solution could not account for any of the variance in item 2 and could only account for about two percent of the variance for items 1 and 4. There were 25 residuals (23 percent) that were greater in magnitude than .05. Table 12 illustrates which items would be included as components of each of the three factors extracted, using the two-factor loading cut-off criteria.

Table 11

Varimax rotated maximum-likelihood factor matrix for 15 AEAI variables

Variable	Factor 1	Factor 2	Factor 3	Communality
Item 1	0.31066*	0.20704	0.03026	0.14029***
Item 2	0.13946	0.13062	0.11368	0.04934***
Item 3	0.68408**	0.21951	0.01385	0.51634
Item 4	0.05902	0.14179	0.36876*	0.15957***
Item 5	0.84973**	0.19465	0.06132	0.76369
Item 6	0.07243	0.07359	0.73322**	0.54828
Item 7	0.94950**	0.18739	0.16047	0.96538
Item 8	0.09757	0.15296	0.98290**	0.99900
Item 9	0.92345**	0.18434	0.20326	0.92806
Item 10	0.11106	0.14325	0.87896**	0.80542
Item 11	0.22676	0.32991*	0.08152	0.16691***
Item 12	0.24414	0.60325*	0.08851	0.43135
Item 13	0.22584	0.89846**	0.17623	0.88928
Item 14	0.19876	0.94868**	0.20490	0.98148
Item 15	0.17834	0.90142**	0.21549	0.89080

*Loadings of ≥0.30 were considered significant. **Thorndike's criteria:loadings of ≥0.40 considered significant. ***Less than 25% of the item variance accounted for by this solution.

Table 12

Variables included in the three-factor varimax rotated maximum likelihood solution

Factor 1		Factor 2		Factor 3	
Minimum factor loading for inclusion[a]					
.30	.40	.30	.40	.30	.40
Item 1	------	Item 11	------	Item 4	------
Item 3	Item 3	Item 12	Item 12	Item 6	Item 6
Item 5	Item 5	Item 13	Item 13	Item 8	Item 8
Item 7	Item 7	Item 14	Item 14	Item 10	Item 10
Item 9	Item 9	Item 15	Item 15	------	------

[a] Refers to factor loading cut-off criteria that was used to determine significance.

Maximun-likelihood factor analysis of the AEAI: four-factor solution.

Table F-4 presents a forced four-factor solution matrix to the correlations between the 15 AEAI variables. Lower residual values (only 6% of the residuals are >.05) make this solution preferable to the three-factor solution presented above. As reflected in the larger variable communalities, this solution was also able to account for a larger proportion of the variance in the total variable set (i.e., 71 percent). Table 13 presents the significant variable loadings for each factor, using Thorndike's cut-off criteria.

Table 13

Significant variable loadings with the four-factor solution[a]

Factor 1	Factor 2	Factor 3	Factor 4
Item 3* (.6232)[b]	Item 12 (.4814)	Item 4 (.3711)	Item 1 (.6413)
Item 5 (.8287)	Item 13 (.8636)	Item 6 (.7327)	Item 3* (.4970)
Item 7 (.9435)	Item 14 (.9397)	Item 8 (.9778)	------
Item 9 (.9204)	Item 15 (.8997)	Item 10 (.8744)	------

[a] Using Thorndike's cut-off criteria .39. [b] Factor loadings.
* Variables that loaded significantly on more than one factor.

ML factor analysis of Scale A items.

Maximum-likelihood extraction with a one-factor criteria limit accounted for 65 percent of the variance of Scale A items.

Analysis of the sample distribution and sample characteristics indicated that the sample was drawn from a multivariate, normal distribution and that factor analysis was appropriate for this data. Bartlett's test of sphericity for identity matrices in the population was negative (X^2=665.79, $p \leq .001$). The KMO measure of sampling adequacy was 0.79, which indicated that variables in Scale A share common factors.

All variables loaded significantly on the extracted factor. However, there were two residuals (20%) that were greater than .05. It will be recalled that the magnitude of the residua's coefficients between the original correlation matrix and the reproduced correlation matrix is one indicator of the adequacy of a factor solution. Table 14 presents the unrotated factor loadings for Scale A items.

Table 14

Maximum-likelihood factor matrix for Scale A variables

Variable	Factor 1	Communality
Item 1	0.34136*	0.25006***
Item 3	0.70067**	0.59215
Item 5	0.86656**	0.76072
Item 7	0.98464**	0.91686
Item 9	0.96160**	0.90359

* Loadings of >0.30 considered significant. **Thorndike's criteria: loadings of ≥0.40 considered significant. ***Less than 25% of the variance accounted for by this solution.

Although the solution was not ideal, the one-factor solution was able to account for an adequate proportion (more than 25 percent) of most item variance (see Table 14 Communalities). Nevertheless, Table 14 also illustrates the more conservative results obtained in utilizing Thorndike's cut-off criteria.

ML factor analysis of Scale A items: two factor solution

Table A-2 (see Appendix) presents a forced two factor solution matrix to the correlations between the 5 AEAI Scale A variables. Lower residual values (none of the residuals are >.05) make this solution preferable to the one-factor solution presented above. Also, as reflected in the much larger variable communalities, this solution was able to account for a greater proportion of the variance in the total variable set (i.e., 78%). Table 15 presents the significant variable loadings for each factor, using Thorndike's cut-off criteria.

Table 15

Significant variable loadings on each factor for Scale A[a]

Factor 1	Factor 2
Item 5* (.7439)[b]	Item 1 (.4370)
Item 7 (.9264)	Item 3 (.9234)
Item 9 (.8901)	Item 5* (.4628)

[a] Thorndike's cut-off criteria = .40. [b] Factor loadings.
* Variables that loaded on more than one factor.

ML factor analysis of Scale E items: three factor solution

Maximum-likelihood extraction with a three-factor criteria limit accounted for 69 percent of the total variance of Scale E items.

Analysis of the sample distribution and sample characteristics indicated that the sample was drawn from a multivariate, normal distribution. Bartlett's test for sphericity was negative

(X^2=392.66, $p \leq .001$). The KMO measure of sampling adequacy for factor analysis was 0.70, which indicated that the variables in Scale E share common factors.

Table 16 presents the rotated factor matrix for the three-factor solution to Scale E variables. None of the residual correlation values were greater than .05 in magnitude.

Table 17 presents the significant variable loadings for each of the three factors using Thorndike's cut-off criteria.

Table 16

Varimax rotated maximum-likelihood factor matrix for AEAI Scale E variables: three-factor solution

Variable	Factor 1	Factor 2	Factor 3	Communality
Item 2	.03288	.06655	.54720**	.30494***
Item 4	.27851	.15386	.52064**	.37231***
Item 6	.44413**	.83445**	.23728	.94986
Item 8	.88500**	.36738*	.16650	.94592
Item 10	.87656**	.23727	.20283	.86580

*Loadings of ≥0.30 were considered significant. **Thorndike's criteria: loadings of ≥0.37 would be considered significant. ***Less than 25% of the variance accounted for by this solution.

Table 17

Significant variable loadings with the three-factor maximim-likelihood solution for Scale E variables[a]

Factor 1	Factor 2	Factor 3
Item 6* (.4441)[b]	Item 6* (.8345)	Item 2 (.5472)
Item 8* (.8850)	Item 8* (.3674)**	Item 4 (..5206)
Item 10 (.8757)	------	------

[a] Using Thorndike's cut-off criteria .37. [b] Factor loadings.
* Variables that loaded significantly on more than one factor. **Slightly under cut-off criteria.

ML Factor Analysis of Scale E: one factor solution

Table 18 presents a forced unrotated one-factor solution matrix to the correlations between the five Scale E items. Higher residual values (three residuals [30%] are >.05) make this solution less preferable than the three-factor solution presented above. Also, as reflected in the lower communalities, this solution was not able to account for as great a proportion of the variance in the total variable set (i.e., 51%). Only four of the five variables of this scale loaded on the extracted factor. Furthermore, Factor 1 was only able to account for more than 25 percent of the variance in three items (see Table 18 Communalities).

Table 18

Maximum-likelihood factor matrix for Scale E variables: one-factor solution

Variable	Factor 1	Communality
Item 2	0.14511	0.10568**
Item 4	0.39023*	0.23002**
Item 6	0.73942*	0.56573
Item 8	0.99950*	0.85292
Item 10	0.89712**	0.80828

*Thorndike's criteria; loadings of ≥0.35 considered significant. **Less than 25% of the variance accounted for by this solution.

ML factor analysis of Scale I items

Maximum-likelihood extraction with a one-factor criteria limit accounted for 66 percent of the variance of Scale I items. Analysis of the sample distribution and sample characteristics indicated that the sample was drawn from a multivariate, normal distribution and that factor analysis was appropriate for this data. Bartlett's test of sphericity for identity matrices in the population was negative (X^2=777.95, $p \leq .001$). The KMO measure of sampling adequacy was 0.75, which indicated that variables in Scale I share common factors.

All variables loaded significantly on the extracted factor. However, there were 2 residuals (10%) that were greater than .05. Table 19 presents the unrotated factor loadings for Scale I items.

Table 19

Maximum-likelihood factor matrix for Scale I variables

Variable	Factor 1	Communality
Item 11	0.37470*	0.14040***
Item 12	0.64041**	0.41012
Item 13	0.94050**	0.88453
Item 14	0.99260**	0.98525
Item 15	0.94276**	0.88880

* Loadings of >0.30 considered significant. **Thorndike's criteria: loadings of ≥0.39 considered significant. ***Less than 25% of the variance accounted for by this solution.

Although the solution was not ideal, the one factor was able to account for an adequate proportion (more than 25 percent) of most item variance (see Table 19 Communalities). Nevertheless, Table 19 also illustrates the more conservative results obtained in utilizing Thorndike's cut-off criteria.

ML factor analysis of Scale I items: two factor solution

Table F-6 presents a forced two-factor solution matrix to the correlations between the five AEAI Scale I variables. Lower residual values (none of the residuals are > .05) make this solution preferable to the one-factor solution presented above. Also, as reflected in the much larger variable communalities, this solution was able to account for a greater proportion of the variance in the total variable set (i.e., 84%). Table 20 presents the significant variable loadings for each factor, using Thorndike's cut-off criteria.

Table 20

Significant variable loadings on each factor for Scale I[a]

Factor 1	Factor 2
Item 12* (.3730)[b]	Item 11 (.6342)
Item 13* (.8558)	Item 12* (.9273)
Item 14 (.9573)	Item 13* (.3975)
Item 15 (.8941)	------

[a] Thorndike's cut-off criteria = .37. [b] Factor loadings.
* Variables that loaded on more than one factor.

ML factor analysis of Section I items.

Maximum-likelihood extraction with a two-factor criteria limit accounted for 58 percent of the variance of Section I items (items 1 to 10). Analysis of the sample distribution and sample characteristics indicated that the sample was drawn from a multivariate, normal distribution. Bartlett's test for identity matrices was negative (X^2=1108.86, $p \leq .001$). The KMO measure of sampling adequacy of .76 indicated that there were only small magnitude partial correlations between unique variance factors for Section I items.

All variables except item 2 loaded significantly on the extracted factors. However, the two factors did not account for significant proportions of the variance in items 1, 2, and 4 (see Table 21 Communalities). There were thirteen (28%) residuals that were greater than .05 in magnitude. Table 21 presents the rotated factor matrix for Section I items and the results obtained through the use of Thorndike's cut-off criteria.

Table 21

Varimax rotated maximum-likelihood factor matrix for AEAI Section I variables

Variable	Factor 1	Factor 2	Communality
Item 1	0.33854*	0.05282	0.11740***
Item 2	0.15769	0.12758	0.04114***
Item 3	0.70889**	0.03186	0.50355
Item 4	0.07902	0.38355*	0.15336***
Item 5	0.86938**	0.07248	0.76108
Item 6	0.07837	0.73527**	0.54677
Item 7	0.96712**	0.17679	0.96658
Item 8	0.11620	0.99272**	0.99900
Item 9	0.94020**	0.21035	0.92823
Item 10	0.12873	0.88818**	0.80544

*Loadings of ≥0.30 were considered significant. **Thorndike's criteria: loadings of ≥0.39 would be considered significant. ***Less than 25% of the variance accounted for by this solution.

ML factor analysis of Section I: three factor solution

Table F-7 presents a forced three-factor solution matrix to the ten AEAI Section I variables (Scales A and E combined). The low residual values (only 13% are greater than .05) make this solution more accurate description of the original correlation matrix than the two-factor solution presented above. This solution accounted for 65 percent of the variance in Section I subtotals. Table 22 presents the significant factor loadings for each factor using Thorndike's cut-off criteria.

Table 22

Significant variable loadings with the three-factor solution to Section I variables[a]

Factor 1	Factor 2	Factor 3
Item 3* (.6588)[b]	Item 4* (.3536)**	Item 1 (.4715)
Item 5 (.8506)	Item 6 (.7246)	Item 2 (..4266)
Item 7 (.9591)	Item 8 (.9885)	Item 3* (.4666)
Item 9 (.9447)	Item 10 (.8833)	Item 4* (.4465)

[a] Using Thorndike's criteria .36. [b] Factor loadings.
* Variables that loaded significantly on more than one factor.
** Slightly under cut-off criteria.

Analysis 4

Results from analysis 4 could not support hypothesis 8 but did support hypothesis 9.

Scale A. There were no significant differences between the means for any of the five orders of presentation of Scale A items ($F_{(4,35)}=1.16$, $p \geq .36$). Table 23 presents the summary statistics for the analysis of variance of Scale A items between the five order-modified administrations of the AEAI. Table 24 presents descriptive statistics for each of the five cell means.

Table 23

Analysis of variance summary table for analysis 4: Scale A

Source	*df*	Sum of Squares	Mean Square	*F*
Between groups	4	93.00	23.25	1.16*
Within groups	35	701.38	20.04	
Total	39	794.38		

$^{*}p \geq .35$.

Table 24

Descriptive statistics for Scale A items

Form	*n*	Mean	*S.D.*
2	8	5.88	4.79
3	8	8.88	4.79
4	8	8.13	2.95
5	8	6.88	4.32
6	8	4.63	5.18

Scale E. There were no significant differences between the means for any of the five orders of presentation of Scale E items ($F_{(4,35)}=1.03$, $p \geq .41$). Table 25 presents the summary statistics for the analysis of variance of Scale E items between the five order-modified administrations of the AEAI. Table 26 presents descriptive statistics for each of the five cell means.

Table 25

Analysis of variance summary table for analysis 4: Scale E

Source	*df*	Sum of Squares	Mean Square	*F*
Between groups	4	34.75	8.69	1.03*
Within groups	35	295.63	8.45	
Total	39	330.38		

*$p \geq .41$.

Table 26

Descriptive statistics for Scale E items

Form	*n*	Mean	*S.D.*
2	8	9.88	3.60
3	8	7.75	3.33
4	8	10.25	2.32
5	8	8.38	2.56
6	8	9.38	2.50

Scale I. There were no significant differences between the means for any of the five orders of presentation of Scale I items ($F_{(4,35)}$=1.16, $p \geq .35$). Table 27 presents the summary statistics for the analysis of variance of Scale I items between the five order-modified administrations of the AEAI. Table 28 presents descriptive statistics for each of the five cell means.

Table 27

Analysis of variance summary table for analysis 4: Scale I

Source	*df*	Sum of Squares	Mean Square	*F*
Between groups	4	106.35	26.59	1.16*
Within groups	35	802.75	22.95	
Total	39	909.10		

*$p \geq .41$.

Table 28

Descriptive statistics for Scale I items

Form	*n*	Mean	*S.D.*
2	8	10.88	4.55
3	8	10.88	6.23
4	8	9.75	3.62
5	8	8.75	4.74
6	8	6.50	4.34

Numbers of days. There were no significant differences between the mean number of days transpired for any of the five orders of presentation of the AEAI scales ($F_{(4,35)}=0.85$, $p \geq .50$). Table 29 presents the summary statistics for the analysis of variance for the mean number of days transpired for each of the five orders of administration of the AEAI. Table 30 presents descriptive statistics for each of the five order-modified forms.

Table 29

Analysis of variance summary table for analysis 4: number of days

Source	*df.*	Sum of Squares	Mean Square	*F*
Between groups	4	3.40	0.85	0.85*
Within groups	35	35.00	1.00	
Total	39	38.40		

*$p \geq .50$.

Table 30

Descriptive statistics for number of days transpired

Form	*n*	Mean	*S.D.*
2	8	3.50	0.93
3	8	3.75	1.28
4	8	3.00	0.93
5	8	3.00	0.93
6	8	3.25	0.87

Analysis 5

Descriptive Statistics. Table 31 presents mean scores and other descriptive statistics for each of the measures administered to Group 1 subjects. There was wide distribution of scores for all measures. Also, the distribution of scores on convergent and discriminant measures was consistent with those reported by other authors on similar samples (Novaco, 1976; Biaggio, 1981; Siegel, 1983).

Relationship between tests administered and demographic variables

There were no significant correlations between the Awareness and Expression of Anger Indicator (AEAI) and any measured demographic variable.

There were no significant correlations between the the Novaco Anger Inventory (NAI) and any measured demographic variable.

There was a significant correlation between the Multidimensional Anger Inventory (MAI) total scores and the number of years of schooling subjects had obtained ($r = -.15$, $p \leq .05$, $N = 125$). These correlations were reflected in the relationship between MAI frequency and anger-out subscales and the number of years of schooling subjects had obtained ($r = -.21$, $p \leq .01$ & $r = -.15$, $p \leq .05$ respectively).

Table 31

Descriptive statistics of total test scores for convergent and discriminant measures

	AEAI[a]	NAI[b]	MAI[c]	MC[d]	CAS[e]
Mean	40.19	154.49	103.79	15.10	13.87
S.D.	14.16	20.77	14.31	9.76	2.93
Range	65.00	126.00	73.00	28.00	15.00
Minimum	0.00	95.00	71.00	1.00	4.00
Maximum	65.00	221.00	144.00	29.00	19.00

[a]Awareness and Expression of Anger Indicator. [b]Novaco Anger Inventory. [c]Multidimensional Anger Inventory. [d]Marlowe-Crowne Social Desirability Scale. [e]Cognitive Attitude Scale.

The Marlowe-Crowne Social Desirability Scale (M-C). There was a significant correlation between subjects' age and their scores on the M-C ($r = -.15$, $p \leq .05$, $N = 127$).

The Cognitive Attitudes Scale (CAS). There was no significant correlation between the CAS and any demographic variable.

Relationships between convergent and discriminant measures

There were significant correlations between the M-C and the NAI ($r = -.25$, $p \leq .005$), the MAI ($r = -.43$, $p \leq .001$), and the CAS ($r = .17$, $p \leq .05$). The M-C correlated significantly with all but two subscales of the MAI.

There was a significant correlation between NAI total scores and MAI total scores ($r = .27$, $p \leq .001$). Also, NAI scores correlated significantly with MAI subscales: Anger-In ($r = .27$, $p \leq .001$), Brooding ($r = .27$, $p \leq .001$), Discuss ($r = .20$, $p \leq .01$), Hostility ($r = .18$, $p \leq .02$), and Range ($r = .42$, $p \leq .001$).

Finally, there was a significant correlation between CAS total scores and MAI total scores ($r = -.15$, $p \leq .05$). Moreover, CAS total scores correlated significantly with MAI subscales: Frequency ($r = -.29$, $p \leq .001$), Magnitude ($r = -.26$, $p \leq .002$), and Brooding ($r = -.19$, $p \leq .05$).

<u>Relationship between the AEAI and convergent and discriminant measures: Total scores</u>

Table 32 presents correlations between AEAI total test scores and convergent and discriminant scores.

Convergent validity of the AEAI. Hypothesis 10 was, in part, supported. There were significant positive correlations between AEAI total scores and NAI total scores ($r = .23$, $p \leq .005$, $N = 127$). However, there were no significant correlations between AEAI total scores and MAI total scores.

Discriminant validity of the AEAI. Hypothesis 12 was supported. There were no significant correlations between AEAI total scores and M-C scores.

However, hypothesis 13 was not supported. There were significant positive correlations between AEAI total scores and CAS scores ($r = .22$, $p \leq .01$; $N = 120$).

Table 32

Correlations[a] between AEAI scale and total scores and convergent and discriminant measures

	Scale A	Scale E	Scale I	Total
NAI[b]	.20*	.02	.32**	.23**
MAI[c]	-.04	-.05	.18*	.00
Frequency	.00	-.13	.15*	-.02
Duration	.05	.07	.10	.06
Magnitude	-.07	-.09	.15*	-.03
Anger-In	.00	-.02	.20*	.02
Anger-out	-.02	.03	.06	.00
Guilt	-.08	-.11	.09	-.07
Brooding	-.08	.02	.16*	-.04
Discuss	-.08	.00	-.14*	-.10
Hostility	-.03	.00	.21*	.03
Range	.09	.05	.20*	.12
M-C[d]	.08	.13	-.09	.06
CAS[e]	.21*	.21*	.05	.22**

[a] Pearson Product-Moment Correlation Coefficients. [b] Novaco Anger Inventory. [c] Multidimensional Anger Inventory. [d] Marlowe-Crowne Social Desirability Scale. [e] Cognitive Attitude Scale.
* $p \leq .05$. ** $p \leq .01$.

Multitrait-Multimethod matrices

Convergent validity: awareness of anger. Hypothesis 11b was generally supported. There was some evidence of convergent validity between Scale A and the NAI. However, there was no evidence of convergent validity between Scale A measures of the awareness of anger and MAI measures of the same trait.[1]

Table 33 presents the completed multitrait-multimethod matrix using original scales of all tests. AEAI measures of awareness of anger correlated positively with the NAI as a measure of the awareness of anger ($r = .20$, $p \leq .01$, $N = 127$), but not with MAI measures of the same trait ($r = .03$, NS). However, there was a significant positive correlation between the two convergent measures of the awareness of anger, the NAI and the MAI ($r = .36$, $p \leq .01$).

Convergent validity: expression of anger. There was no evidence of convergence between AEAI measures of the expression of anger and the convergent measures of expression used in this study. There were no significant relationships between AEAI measures of the expression of anger and any of the convergent measures of the expression of anger.

Convergent validity: induced awareness of anger. Hypothesis 11a was supported. There was evidence of convergent validity between Scale I measures of the awareness of anger and the two convergent measures of the awareness of anger.

There was a significant positive correlation between AEAI measures of the induced awareness of anger and the two convergent measures of the awareness of anger. AEAI Scale I correlated significantly with both the NAI ($r = .32$, $p \leq .01$) and MAI measures of the

[1] It should be noted that Larzelere and Mulaik (1977) and others (e.g., Hays, 1973) have reported that multiple tests of correlation on data from a single sample increases the likelihood of Type I error as the number of tests of significance increases. They recommend a multistage Bonferroni correction procedure be employed to maintain an acceptable familywise Type I error rate. This procedure was not presented for the following reasons: First, the miltistage Bonferroni procedure is usually employed when a large number of correlations are being examined simultaneously. In this multitrait-multimethod matrix only 15 of the 28 correlations are interpreted as having direct bearing on the hypotheses. Second, of those 15 applicable correlations, only five are at risk for Type I error (i.e., concerned with convergence of the AEAI). Finally, the use of the correlation coefficient in multitrait-multimethod matrices is one of description and not particularly one of inference: coefficients are employed to describe the general relationship between scores, not as proof of causality or independence.

awareness of anger ($r = .25$, $p \leq .01$).

Discriminant validity: awareness of anger. Hypothesis 12 was supported. There was evidence of discriminant validity for AEAI measures of awareness of anger (Scale A).

There was no significant correlation between AEAI Scale A scores and a subject's M-C scores (see Table 33). There was also no significant correlation between AEAI awareness measures and MAI expression measures (see Table 33). However, Hypothesis 13 was not supported. There was a significant correlation between AEAI awareness measures and CAS scores ($r = .21$, $p \leq .01$).

Discriminant validity: Expression of anger. Hypothesis 12 was again supported. There was evidence of discriminant validity for AEAI measures of the expression of anger.

There were no significant correlations between Scale E scores and M-C scores, MAI awareness scores, or, NAI awareness scores (see Table 33). However, hypothesis 13 was, once again not supported. There was a significant positive correlation between Scale E scores and CAS scores ($r = .21$, $p \leq .01$).

Discriminant validity: Induced awareness of anger. Hypotheses 12 and 13 were supported with respect to Scale I. There was evidence of discriminant validity between AEAI measures of the induced awareness of anger and all discriminant measures.

There were no significant correlations between Scale I scores and MAI expression scores, M-C scores, or CAS scores (see Table 33).

Table 33

Multitrait-Multimethod Matrix for AEAI Scales and Convergent and Discriminant Measures[a]

	1	2	3	4	5	6	7	8
AEAI								
1.Awareness	---							
2.Expression	.31*	---						
3.I. Awareness	.37*	-.06	---					
MAI								
4.Awareness	**.03**	-.03	**.25***	---				
5.Expression	-.05	**.02**	.00	.33*	---			
NAI								
6.Awareness	**.20***	.03	**.32***	**.36***	.06	---		
OTHER								
7.Soc. Des.[b]	.07	.13	-.09	-.39*	-.21*	-.25*	---	
8.Beliefs[c]	.21*	.21*	.05	-.20*	.03	-.03	.16**	---

[a] Bold face values are validity coefficients. [b] Based on Marlowe-Crowne Social Desirability scores. [c] Based on Cognitive Attitudes Scale scores.
**p* ≤.01. ***p* ≤.05.

Discriminant validity: other considerations

Table 34 presents a comparison between monotrait-heteromethod (M-H), heterotrait-monomethod (H-M), and, heterotrait-heteromethod (H-H) correlation coefficients from the above multitrait-multimethod correlation matrix. Four of the six (67%) M-H coefficients are significant (see column M-H, Table 34). However, three of the four H-M coefficients are also significant (see colum H-M, Table 34). Finally, the magnitude of the validity coefficients (M-H column, Table 34), are generally larger than the H-H coefficients.

Table 34

Discriminant Validity of the AEAI: Comparison of Monotrait-Heteromethod, Heterotrait-Monomethod, and Heterotrait-Heteromethod Correlations[a]

M-H[b]	H-M[c]	H-H[d]	
.03	.31	.05	.07
.25	.37	.03	.13
.02	.06	.00	.09
.20	.33	.03	.21*
.32	---	.07	.05
.36	---	.21*	---

[a] Absolute values used. [b] Monotrait-Heteromethod. [c] Heterotrait-Monomethod.
[d] Heterotrait-Heteromethod.
*Correlations with Cognitive Attitude Scale.

Multitrait-Multimethod Matrix using MAI Factor Scales

Table 35 presents an additional analysis of the AEAI against MAI factor-derived scales of the awareness and expression of anger. There are no significant changes in the direction of the relationships between measures described in the preceeding section. However, several of correlations did increase in magnitude. Also, the use of MAI factor-derived scales resulted in: a significant positive correlation between AEAI measures of the induced awareness of anger and MAI measures of the expression of anger ($r = .21$, $p \leq .01$), and, a significant positive correlation between NAI awareness measures and MAI expression measures ($r = .25$, $p \leq .01$). Finally, the use of factor scales resulted in a significant correlation between MAI expression scores and subjects' beliefs about the consequences of expressing their anger ($r = -.15, p \leq .01$).

Table 35

Multitrait-Multimethod Matrix for AEAI Scales and Convergent and Discriminant Measures Using MAI Factors[a]

	1	2	3	4	5	6	7	8
AEAI								
1.Awareness	---							
2.Expression	.31*	---						
3.I. Awareness	.37*	-.06	---					
MAI								
4.Awareness[b]	**.05**	.01	**.24***	---				
5.Expression[c]	-.05	**.04**	.21*	.30*	---			
NAI								
6.Awareness	**.20***	.03	**.32***	**.36***	.25*	---		
OTHER								
7.Soc. Des.[d]	.07	.13	-.09	-.44*	-.31*	-.25*	---	
8.Beliefs[e]	.21*	.21*	.05	-.17*	-.15*	.03	.16**	---

[a] Bold face values are validity coefficients. [b] Comprised of MAI Factors 1, 2, and 3. [c] Comprised of MAI Factors 4 and 5. [d] Based on M-C scores. [e] Based on CAS scores. *$p \leq .01$. **$p \leq .05$.

DISCUSSION

The purpose of this study was to examine specific aspects of internal and external validity of the AEAI.

It was suggested that data was needed on the internal consistency and homogeneity of the three subscales of the AEAI. Factor analytic studies were thought to be most able to confirm or refute the structural validity and independence of the three dimensions measured by the these subscales. Last, the veracity of previous research conclusions based on the use of the AEAI was challenged by the absence of data on the effect of order of presentation of the scale items.

The need for further data on external validity was discussed as being necessary to support the constructs of non-induced and induced awareness of anger and expression.

A research design was implemented which addressed these issues. The results supported for the most part the research hypotheses and predictions. However, several of the findings were unexpected. A discussion of these sometimes ambiguous, sometimes illuminating results ensues.

Internal validity

Interpreting the extracted factors: the AEAI. The non-planned four-factor solution was chosen as more adequate. The lower residual values (none were >.05), the higher communalities, and the fact that the four-factor solution unmasked the effect of the weighting system on scale scores contributed to the selection of this solution.

Factor 1 was comprised entirely of items from Scale A. But does this mean it was an awareness factor? It does not seem so. Factor 4 was also comprised entirely of items from Scale A. However, only items 1 and 3 loaded on this factor. From what was revealed about the frequency of actual responses and because items 1 and 3 were barely affected by the weighting system, it is concluded that Factor 4, more than Factor 1, is the purer non-induced awareness of anger factor. Where does that leave Factor 1? Because the loadings on this factor increased sequentially with the items (as did the effects of the weighting system), it seems that Factor 1, to a large extent, is an AEAI weighting system factor. Factor 2 is labelled Induced awareness of anger, and Factor 3 is labelled Expression of anger. Again, both these factors were comprised exclusively of items from the respectively named scales. These two factors also appear to be contaminated by the weighting system. The concurrently increasing factor loadings with each sequential item corresponds to the sequentially increasing effects of the weighting system. This conclusion was corroborated by the individual scale analyses which are discussed next.

Scale A : non-induced awareness of anger. For the same reasons presented earlier, the two-factor solution is selected. In this analysis all factors were comprised entirely from Scale A items. Item 5 loaded on both factors. Likely this reflects the high frequency of both actual and weighted responses for this item.

Factor 1 was labelled the weighting factor, and Factor 2 was labelled the Non-induced awareness of anger factor. Isolated analysis of Scale A, as expected, was not inconsistent with the conclusions presented above. The ramifications of this interpretation is that the last two items on this scale were of limited value as behavioral indicators. It appears that for the large part, item 7 and 9 scores were inferred. The implications of this will be discussed in a later section.

Scale E: expression of anger. As predicted, the three factor solution was selected as most adequate for the analysis of Scale E. It would be tempting to label these factors as reflecting solely the hypothesized processes of identification, confrontation and resolution. Closer inspection of the factor loadings contradicted this interpretation.

Factor 1 is labelled a modification point (weighting system) factor. Factors 2 and 3 are labelled second-order Expression factors. It appears that Factors 2 and 3 were comprised of fairly pure verbal responses. The variables load exclusively on these factors and not on Factor 1. An examination of the actual responses given by the subjects, supports the differentiation of these factors into: identification and confrontation/resolution. Therefore, Factor 2 is labelled Confrontation/Resolution and Factor 3 is labelled Identification. Clearly, these labels do not validate these processes as tangible entities. Rather, the responses that were given seem to be consistent with, or can be described as, what has been operationally defined as Identification, Confrontation and Resolution.

Scale I : non-Induced awareness of anger. For the same reasons presented earlier (i.e., residual values, communalities, and interpretation), the two factor solution is selected. In this analysis all factors were comprised entirely from Scale I items. Items 12 and 13 loaded on both factors. Again, this reflects the frequency of actual and non-scored responses for these items.

Factor 1 was labelled the modification point factor, and Factor 2 was labelled the Induced awareness of anger factor. Isolated analyses of Scale I, as expected, was not inconsistent with the conclusions presented in the main analysis. The ramifications of this interpretation is that the last two items on this scale were also of limited value as behavioral indicators. It appears that for the large part, items 14 and 15 contribute very little. This was the intent of the test developers. Scale I was originally designed to test the hypothesis that the wording of the question (or method) could alter the meaning of the response. Scale I was not originally intended to contribute any clinically relevant information. Ironically, results from this study (i.e., convergence of Scale I with NAI & MAI responses) indicate that Scale I may be of equal utility in assessing other dimensions of the anger experience. Perhaps part of processing anger involves an ability to acknowledge or receive environmental cues about how one should be feeling.

Internal consistency of the AEAI. The pattern of inter-item correlations presented in analysis 1 may be interpreted in at least two ways. Perhaps these patterns reflect the true nature of the relationship between the items; there may be some type of effect occurring whereby a subject's final responses are somewhat different from his/her early responses. Or, perhaps some aspect of the test design was responsible.

Several factors support the second argument over the first. First, order effects did not exert any systematic bias in test scores. Second, this pattern reasserted itself in the factor solutions. Third, the fact that variables which had a low frequency of actual subject responses loaded highly on the weighting factors and did not tend to load on any of the other factors, suggests that these patterns were at least in part due to the AEAI method of weighting item scores. Last, it seems unlikely that such a consistent pattern could be attributed to the random effects of chance. Regardless, it seems plausible to conclude that coefficients of reliability for Scales A and E, were somehow artificially increased.

Scale homogeneity. A question remains: what, if any, significance is there in the fact that Scale I did not appear as homogeneous as Scales A and E?

In brief, Scales A and E were affected by the modification point system: Scale I was not. As we have seen, the first few items in each scale had a higher frequency of actual responses than did the last two items of each scale. However, while Scale I was unaffected by the weighting system, its coefficients of reliability were also biased. Following the last item of each scale that had an actual subject response (called the last morning), items from Scales A and E received modification points. Items following the last morning in Scale I received zero points. This also produced a skewed measure of homogeneity, in this case, artificially decreased alpha coefficients. Given these two facts, it may be reasonable to conclude that a more accurate index of AEAI scale homogeneity is somewhere between the artificially repressed alpha coeffients for Scale I and the artificially inflated alpha coefficients for Scales A and E. In any case, this value would be sufficiently large (i.e., .43< alpha > .89) to be acceptable.

The recurring scenario as a mood induction technique. The recurring scenario seemed to have been an effective anger-provoking stimulus. Most subjects reported some degree of anger in response to Scale A questions. However, the weighting system seems to have concealed, rather than enhanced the linear trend. By assigning the same score to all absent responses, the weighting system (Braha & Catchlove, 1984) flattened out the higher end of the reported anger intensity curve. It seems safe to conclude that in the absence of the weighting system, either the linear trend would be stronger, or it would become more quadratic. This further supports the hypothesis that the AEAI recurring scenario acts as a mood induction device. However, the latter conclusion is limited by several factors. Subjects may have learned the appropriate response from interviewer statements. Reinforcement of affective responses was built into the test design. Subjects who reported affect were asked to rate the intensity of their feelings. These prompts may have acted as reinforcers. Conversely, not all subjects who received prompts continued to report affect. Also, because not all affective responses were scored, the same factors which may have enhanced the findings of analysis 2, served to suppress the trend. This would have occurred when non-angry affective responses were reported.

It would be interesting to determine the frequency of other, non-anger, affective responses to Scale A questions. This unexplored domain may reveal further potential for the AEAI. Judging from the number of non-anger affective responses it seems the AEAI is more than an anger test. It may have some potential in assessing subjects' ability to identify or label general arousal with very few contextual cues. Many subjects responded initially to Scale A questions with undifferentiated affective responses: for example, subjects would report feelings such as frustrated, upset, funny, and put-off. By the second or third hypothetical morning most of these subjects would report anger. The question is, were these subjects aware of angry feelings even though they did not report them, or did they need the extra provocation or contextual cues derived from the recurring scenario to label their earlier arousal as anger or to actually provoke anger. Unfortunately, the inclusion of Scale I does not entirely answer this question. Many subjects' responses to Scale A questions (i.e., How would you feel?) were not congruent with their responses to inducing Scale I questions (i.e., Would you feel angry?) for the same hypothetical morning. Often subjects reported no anger in Scale A, but, responded affirmatively to the inducing type questions of Scale I. The absence of order effects (analysis 4) eliminates the possibility of conditioning effects accounting for these incongruencies. It may be that these subjects' have a tendency to acquiesce in response to inducing type questions from an interviewer. This supports earlier findings which suggested that there was a difference between what Scale A and Scale I items were assessing (Catchlove & Braha, 1985). To be certain, separate analyses of these cases would need to be conducted.

- *Order effects*. Hypothesis 9 was not supported. Contrary to what was originally believed, there was no difference in scores when the order of presentation of scale items was varied. This is an intruiging finding. The implications seem clear. The presentation of inducing, suggestive questions (e.g., Would you feel angry?), had no overall effect on subjects' responses to non-inducing, neutral questions (e.g., How would you feel?) about the same situation -- regardless of subjects' responses to the inducing question, and even when the question were presented consecutively. It would be interesting to examine this effect with regards to subjects' locus of control. Davis and Mettee (1971) have discussed some of the implications of internal versus external locus of control and the labelling of emotions from environmental cues. However, this line of research does not seem to have been pursued.

Conclusion: analysis of the internal validity of the AEAI

Statistically, the AEAI demonstrated acceptable intra-subscale homogeneity and internal reliability.

The AEAI seems to be assessing at least three dimensions of anger: non-induced awareness, expression and induced awareness of anger. However, AEAI scores are not awareness, expression and induced awareness of anger. However, AEAI scores are not entirely pure indicators of subject responses. Scores also reflect an inferential weighting system. Factor analysis was able to isolate the weighting component from individual item scores and still extract factors which confirmed the above dimensions. An analysis of the frequency of subject responses revealed that at least 75 percent of all scores up to the fourth item in each scale were based on actual subject responses. Hence, the effects of the weighting system were most apparent on the last two items of each scale. Because of this, it seems likely that the results found with the use of this data would have been found had no weighting system been used.

Separate analysis of Scale E variables produced second-order factors which provide support for the constructs of identification and confrontation/resolution.

Final comment on internal validity. The effects of the weighting system are largely theoretical. Some may find the use of this weighting system a practical, valid solution to an unorthodox test design. Others may find the weighting system too speculative. Either way the origin of the problem lies in the test design. By way of explanation, subjects who resolve the situation in only a few hypothetical days may score fewer points than subjects who never resolve the hypothetical situation but give scorable responses to all 15 items. This would not be a problem if the scoring method were consistent with that of most personality measures (e.g., low and high scores representing statistical deviance, mid-range scores representing average responses). Rather, the AEAI is scored in the same way as many achievement and aptitude tests. With these tests, as with the AEAI, more points are presumed to be indicative of more skill in the test domain. Hence, the problem: how to avoid the situation where subjects who never resolve the situation end up with more points than subjects who resolve the situation early. The modification point system attempts to compensate for this deficit in the test design. The test developers attempted several solutions to this problem.

Appendix F

Table F-4

Maximum-likelihood four-factor, varimax rotated matrix for 15 AEAI variables (*N* =147)

Item	Factor 1	Factor 2	Factor 3	Factor 4
1	.22	.09	.04	.64
2	.09	.06	.12	.34
3	.62	.13	.01	.50
4	.03	.10	.37	.25
5	.83	.16	.05	-.23
6	.07	.07	.73	.10
7	.94	.18	.15	.15
8	.11	.17	.98	.03
9	.92	.18	.18	.13
10	.12	.16	.87	.05
11	.12	.18	.09	.78
12	.16	.48	.09	.68
13	.20	.86	.16	.28
14	.19	.94	.18	.19
15	.17	.89	.19	.16

Appendix F

Table F-5

Maximum-likelihood, two-factor, varimax rotated factor matrix for Scale A (N =147)

Item	Factor 1	Factor 2
1	.19	.44
3	.38	.92
5	.74	.46
7	.93	.35
9	.89	.35

Appendix F

Table F-6

Maximum-likelihood, two-factor, varimax rotated factor matrix for Scale I (*N* =147)

Item	Factor 1	Factor 2
11	.19	.63
12	.37	.93
13	.86	.40
14	.96	.29
15	.89	.28

Appendix F

Table F-7

Maximum-likelihood, three-factor, varimax rotated factor matrix for Section I (N=147)

Item	Factor 1	Factor 2	Factor 3
1	.28	.02	.47
2	.10	.10	.43
3	.66	.00	.47
4	.02	.35	.45
5	.84	.05	.30
6	.07	.72	.17
7	.96	.16	.12
8	.12	.99	.08
9	.94	.20	.08
10	.13	.88	.09

Appendix G

Table G-1

Frequency of responses for each item of the CAS

Response Choice	Item 1	Item 2	Item 3	Item 4
1)	37(25.2)[a]	8(5.4)	52(35.4)	12(8.2)
2)	9(6.1)	5(3.4)	9(6.1)	3(2.0)
3)	30(20.4)	11(7.5)	23(15.6)	1(0.7)
4)	40(27.2)	18(12.2)	19(12.9)	15(10.2)
5)	3(2.0)	78(53.1)	17(11.6)	89(60.5)

[a] Percent of total responses.