

The Anger Self Report: A Psychometrically Sound (30 Item) Version

Nicola S. Reynolds
Department of Justice, Wellington

Frank H. Walkey* and Dianne E. Green
Victoria University of Wellington

Most of the existing measures of anger lack adequate psychometric documentation and there is a need to develop soundly based research instruments in the area. In the present study a systematic psychometric evaluation and revision of the Anger Self Report Questionnaire (ASR) was undertaken. This evaluation revealed adequate reliability for the full 89 item scale but moderate to low reliability for the four subscales. A 30 item single scale ASR questionnaire which measured a general anger factor was developed, using items from the original 89 item ASR. This shorter questionnaire has high reliability and as a relatively brief scale, appears particularly suited for exploration of anger in other New Zealand samples. Norms have been included for the 30 item scale, derived from the responses of 101 male and 100 female students.

Introduction

The current lack of research into anger appears to be caused by factors, including the conceptual confusion between anger and aggression, the difficulty of objectively researching emotional states (Biaggio, 1980, 1987a, 1987b, 1989a, 1989b; Johnson & Wilborn, 1991), and establishing suitable research techniques (Novaco, 1975).

Consequently, there is much confusion about the relevant assessment measures for anger. Biaggio (1980) and Unverzagt and Schilt (1989), in their assessment of the Novaco Anger Inventory (Novaco, 1985) and the Spielberger Trait Anger Scale (Spielberger, Jacobs, Russell & Crane, 1983), have also pointed out that many of the measures lack adequate psychometric documentation, and that there is still a need to develop and to evaluate adequate research instruments in the area. One promising instrument which has been subjected to tentative evaluation procedures, is the Anger Self Report (ASR) described by Zelin, Adler, and Myerson (1972).

The original ASR questionnaire was an 89 item Likert type self response measure, providing six response options ranging from strongly agree to strongly disagree. Zelin et al. (1972), reduced the 89 item ASR to 64 items by item analysis, using

inter-item correlations between items and total scale scores.

The authors of the ASR indicated that the questionnaire consisted of five scales, Awareness of Anger, Total Expression of Anger, Guilt, Condemnation of Anger and Mistrust or Suspicion. The second scale, Total Expression of Anger, is further divided into three subscales, General Expression, Physical Expression and Verbal Expression of Anger. The number of items in each scale or subscale ranges from five to 13.

Items for the ASR were taken from the Buss-Durkee Hostility Inventory (BDHI) (Buss & Durkee, 1957), the MMPI (Dahlstrom & Welsh, 1960), and the Masochism Questionnaire (Shore, Clifton, Zelin & Myerson, 1971). Other items were written specifically for inclusion in the ASR. Zelin et al. (1972) found that there was no item overlap in the scales.

In addition to the original work by Zelin et al. (1972), other studies have attempted to examine the psychometric properties of the ASR questionnaire. Biaggio (1980) included the ASR in a study of four measures of anger and hostility. Interscale relationships were evaluated and a subscale factor analysis (as contrasted with item factor analysis) of the four measures, was performed. The analysis produced five factors which Biaggio (1980) identified as a willingness to experience and express anger (58.9% of variance), physical

*Address for correspondence: Victoria University of Wellington, PO Box 600, Wellington, New Zealand.

and verbal expression of anger (16.8% of variance), an attitudinal set of resentments, mistrust and guilt (10.9% of variance), anger provoking incidents (7.8% of variance) and negativism (6.0% of variance). The ASR loaded on three of these factors: a willingness to experience and express anger, physical and verbal expression of anger and an attitudinal set of resentment, mistrust and guilt.

Biaggio (1980) also computed correlations between all four of the measures examined in her study and between their subscales. Significant correlations were obtained between the BDHI Total Hostility Scale and the ASR Total Expression Scale ($r = .64$, $p < .01$). A number of hypothesised correlations for BDHI and ASR subscales were found to be significant, although the magnitude of the correlation coefficients was not high, varying between .28 and .78. Biaggio (1980) concluded that these correlations provided additional evidence for the convergent and discriminant validity of the ASR and BDHI, but needed to be interpreted conservatively as the predictive validity of neither scale had been firmly established.

In a further study, Biaggio, Supplee, and Curtis (1981) attempted to assess the test-retest reliability and predictive validity of the same four measures. Generally these studies provided equivocal evidence for the reliability and validity of the ASR. Zelin et al. (1972) had obtained adequate split half reliability for most subscales. Although reliability estimates of the Physical Expression Scale ($r = .64$), and the General Expression Scale ($r = .66$), were only moderate, they decided to retain the scales on the grounds that they contain 'valid information'. Biaggio et al. (1981), however, found test-retest reliabilities, ranging from a relatively low .28 $p < .05$ (Guilt) up to a maximum of .76 $p < .01$ (Condemnation of Anger). They suggested that both the Verbal Expression ($r = .35$), and Guilt Scales ($r = .28$), were particularly questionable. Zelin et al. (1972) claimed good discriminant and convergent validity of the ASR from their studies on students and psychiatric patients.

Biaggio et al. (1981) also suggested that the validity studies of Zelin et al. (1972) provide some evidence of predictive validity for all subscales except Mistrust. Their own study showed good evidence of predictive validity for subscales of Awareness of Anger, General Expression of Anger, Physical Expression of Anger, Condemnation of Anger and Total Expression. However, predictive validities for Verbal Expression, Mistrust and Gilt were considered questionable.

Schill, Ramanaiah, and Conn (1990) administered seven covert and 14 overt hostility scales from the BDHI, together with the ASR, to 65 college students. From the responses of this small student group to the Questionnaires, they found a correlation of $r = .60$ for males and $r = .66$ for females, with both correlations significant at the $p < .05$ level.

The present study set out to evaluate the capacity of the ASR to distinguish between Awareness of Anger, Expression of Anger and the amount of Guilt and Mistrust as claimed by Zelin et al. (1972), and to show whether any distinction found is sufficiently clear to warrant the scoring of separate scales. After reviewing the results of these psychometric evaluations it was decided to attempt to develop a brief but reliable, unidimensional measure of anger by selecting 30 of the 89 items contained in the original ASR.

Evaluation of the ASR

Method

Subjects

The original 89 item questionnaire was administered to 246 Victoria University students (127 males, 119 females), whose ages ranged from 16 to 47 years. Volunteer subjects were recruited from library users, and from first year Geography and Psychology students.

Procedure

Pearson Product Moment Correlation Coefficients were calculated to obtain item to total test correlations, item to subtest correlations, subtest to subtest correlations and subtest to total test correlations.

Kuder-Richardson 20 (KR20) reliability was calculated for each subtest and for the whole (89 item) test.

Factor analyses were conducted using a strategy suggested by Walkey (1983), who argued that the number of factors to be rotated should be derived from an assessment of the developers' intended subscale structure rather than by using some mathematically based criterion such as extracting factors with eigenvalues greater than one. For situations where such information is not available or has proved to be doubtful, Walkey and McCormick (1985) have developed a method to identify an optimal number of factors and to verify such a structure using a criterion of replicability, by comparing the factor analytic solutions derived from a number of respondent groups. This method uses the s index described by Cattell, Balcar, Horn, and Nesselroade (1969), which has been incorporated into a procedure—FACTOREP, which has been described in detail in earlier papers by Siegert, McCormick, Taylor, and Walkey (1987); Green, Walkey, Taylor, and McCormick (1989); and Walkey and Green (1992).

Subjects were divided into three groups, each with

82 members. Factor analyses of items were conducted using the Statistical Analysis Systems package (SAS Institute Inc. 1985), and solutions derived from the responses of each of the three groups of subjects were compared using the FACTOREP procedure.

Zelin et al. (1972) had divided the questionnaire into seven scales, three of which combined to make up a Total Expression of Anger Scale. Two analyses which extracted five and seven factors were therefore conducted on the original 89 item questionnaire to ascertain whether the test developers' intended structure was confirmed.

Factor analyses were also undertaken using the 64 items of Zelin et al.'s (1972) final version of the questionnaire since it was considered that the 25 filler items in the original questionnaire could be confounding the results. To investigate whether any replicable factor structure was to be found in the 64 item version of the ASR, seven, five, four, three and two factors were rotated, and the results derived from the three groups were compared, again using the FACTOREP procedure.

Results

Item to Total Test Correlations

Thirty-nine of the total 89 items correlated significantly ($p < .001$) and positively with total test scores, suggesting a variable underlying a substantial number, but not all of the items. Overall correlations ranged from $r = .28$ to $.66$ ($p < .001$).

Subtest Correlations

Twenty-four of the 36 subtest to subtest correlations were significant while all subtest scores correlate significantly with total test scores (See Table 1), again suggesting the presence of a single pervasive underlying variable.

Reliabilities

A moderate total test (KR20) reliability of .85 was obtained for the 89 item test using the sample of 246 subjects. Subtest reliabilities ranged from low, General Expression (KR20 = .48), and Mistrust (KR20 = .57), to moderate, Verbal Expression of Anger (KR20 = .70) and Awareness of Anger (KR20 = .79).

Factor Analyses

Factor Analyses were conducted on the 89 item questionnaire to compare the seven factor and five factor solutions using analyses obtained from three samples of 82 subjects. Three criterion hyperplane levels were employed (.30, .40 and .50) for the FACTOREP procedure. At none of these criterion levels was a replicable seven or five factor structure evident.

Factor analyses were then undertaken using the 64 items that had made up Zelin et al.'s (1972) final version of the ASR. Three and two factor solutions were compared across the three subject groups using the same three criterion levels. None of these analyses revealed a clear, replicable factor structure. The subgroups were then combined, and analyses of seven, five, four, three and two factors were undertaken, which showed a total lack of comparability with Zelin et al.'s (1972) subscale structure.

Although these results did not provide any support for the notion of a replicable factor structure, the presence of a large number of positive and significant subtest to subtest correlations, subtest to total test correlations, and item to total test correlations suggested that a general anger factor was being measured by the ASR.

Table 1: *Correlations of Subsets with One Another and with Total Test*

| | Awareness | General Expression | Physical Expression | Verbal Expression | Total Expression | Guilt | Condemnation | Mistrust |
|---------------------|-----------|--------------------|---------------------|-------------------|------------------|-------|--------------|----------|
| Total Test | .85* | .71* | .57* | .60* | .77* | .37* | .38* | .46* |
| Mistrust | .29* | .20 | .24 | -.03 | .15 | .34* | -.13 | |
| Condemnation | .34* | .35* | .07 | -.25* | -.28* | .13 | | |
| Guilt | .17 | .00 | -.01 | -.14 | -.07 | | | |
| Total expression | .70* | .80* | .71* | .86* | | | | |
| Verbal expression | .51* | .57* | .35* | | | | | |
| Physical expression | .54* | .43* | | | | | | |
| General expression | .65* | | | | | | | |

* significant at $p < .001$

General Factor Identification

The unrotated loadings obtained for the original 89 item questionnaire were therefore subsequently re-examined in an attempt to determine whether a general factor was being measured by the ASR. The first factor of the unrotated factor matrix accounted for 10.06% of the variance, Factor Two accounted for 5.84% of the variance, Factor Three 3.51%, Factor Four 3.51% and Factor Five 3.06%.

Examination of the unrotated factor loadings showed that the mean loading on the first factor was .24, considerably greater than the mean loadings on the other factors, none of which was greater than .07. These results supported the previously described indicators of a substantial general factor, possibly attenuated by the presence of a number of items with low loadings on it, forming a unidimensional scale with high inter-item correlations for which more elaborate underlying factor structure would be claimed or assumed.

Development and Evaluation of a Brief Single Scale Questionnaire

Procedure

A homogeneous single scale questionnaire was constructed, using the procedure suggested by Nunnally and Wilson (1975), who indicated that when undertaking the construction of a homogeneous tests, items which have poorest item to test correlation should be deleted, and recalculation undertaken on the reduced item pool. This procedure continues with the remaining items until a test of desired length is found.

Item to total test correlations were calculated

for the original 89 items of the ASR. The 10 items with the lowest item to total test correlations were removed, and the item to total test correlations recalculated. This procedure was repeated until 39 items remained. On the final analysis, nine items were removed. The remaining 30 items then formed the final version of the questionnaire (See Appendix 1). A principal components analysis was conducted on these items to demonstrate the presence of the general factor in the unrotated first factor loadings.

Finally, some preliminary norms were derived from the responses to the 30 item scale by groups of male ($N = 101$), and female students ($N = 100$).

Results

Twenty-nine of the 30 items in the final version of the questionnaire were those that loaded most highly on the first unrotated factor in the original five factor analysis, suggesting a substantial level of support for the item selection procedure. Item to total test correlations of the 30 selected items were all positive and significant beyond the level $p < .001$.

A principal component analysis of responses to the 30 item scale showed clear evidence for a general factor in the unrotated first factor loadings. These loadings had a mean value of .48, compared with .006 for the second factor and .005 for the third factor. Standard deviations of these loadings were .08, .27 and .27 respectively. Factor loadings on the first factor were therefore high and closely clustered, while those on the second and third clustered relatively loosely around zero. As

Table 2: Final 30 Item Single Scale Questionnaire, Item to Total Correlations and Unrotated first factor Loadings

| Question | Item to total test correlations | Unrotated first factor loadings | Question | Item to total test correlations | Unrotated first factor loadings |
|----------|---------------------------------|---------------------------------|----------|---------------------------------|---------------------------------|
| 1 | .52 | .54 | 53 | .46 | .47 |
| 5 | .45 | .44 | 60 | .38 | .38 |
| 9 | .48 | .47 | 61 | .61 | .63 |
| 19 | .54 | .56 | 62 | .70 | .65 |
| 20 | .47 | .49 | 67 | .64 | .66 |
| 21 | .51 | .50 | 69 | .49 | .51 |
| 22 | .38 | .38 | 70 | .45 | .44 |
| 27 | .49 | .49 | 73 | .44 | .46 |
| 29 | .56 | .57 | 76 | .48 | .48 |
| 31 | .51 | .51 | 77 | .45 | .45 |
| 32 | .39 | .39 | 78 | .52 | .51 |
| 33 | .49 | .49 | 79 | .44 | .43 |
| 37 | .40 | .39 | 80 | .38 | .39 |
| 43 | .40 | .37 | 83 | .39 | .39 |
| 44 | .56 | .56 | 86 | .53 | .54 |

Table 3: *Percentile Ranks of ASR30 Scores for Male and Female University Students*

| Percentiles | Raw Scores | |
|-------------|------------|--------|
| | Male | Female |
| 95 | 146 | 144 |
| 90 | 137 | 136 |
| 80 | 127 | 126 |
| 70 | 120 | 118 |
| 60 | 114 | 112 |
| 50 | 108 | 107 |
| 40 | 102 | 100 |
| 30 | 96 | 95 |
| 20 | 89 | 87 |
| 10 | 78 | 79 |
| 5 | 71 | 69 |
| Mean | 109 | 107 |
| N | 101 | 100 |

the loadings, which reflect the correlations between the items and the general factor, are almost identical with the item to total scale correlations, ($r = .98$), it would appear that the general factor is an almost perfect representation of the total scale score. Item to total correlations and unrotated first factor loadings are set out in Table 2.

Internal consistency of the shortened 30 item scale was more than satisfactory with $KR20 = .89$, compared with $.85$ for the original 89 item scale.

Preliminary Norms

In order to provide some preliminary information for users of the 30 item scale, norms were derived from the responses of 101 male and 100 female students. These norms are included in Table 3.

Discussion

Evaluation of the ASR

$KR20$ reliabilities obtained in the present evaluation of the original version of the ASR were generally lower than those obtained by Zelin et al. (1972). Although the results as a whole show only moderate reliabilities, those for the General Expression of Anger ($KR20 = .48$), and Mistrust ($KR20 = .57$), scales are particularly questionable.

Previous studies provide only modest evidence for discriminant, convergent and predictive validity of the ASR, although the evidence is particularly questionable for some subscales. No attempt was made in the present study however to examine its validity. Factor analyses provided little support for previous attempts to show construct validity. Zelin et al. (1972) reported only subtest

to subtest correlations, which ranged from $.29$ (Total Expression with Condemnation and Condemnation with General Expression) to $.91$ (Total Expression with Verbal Expression). They claimed that the subscales of the test were measuring different components of anger on the basis of these correlations and made no attempt to further examine their subscales. Neither factor analyses nor correlations obtained for the subscales of the ASR support the contention that the subscales represent replicable common factors, so it must be assumed that some other characteristic of the scale accounts for the high correlations. The present study therefore offers no support to Zelin et al.'s (1972) division of the ASR into seven separate subscales, three of which they suggested could be combined to make up a Total Expression scale. The present study also failed to replicate Biaggio's (1980) findings and revealed only a very tenuous three factor structure. It was not possible to interpret these factors which each included items from a number of Zelin et al.'s (1972) subscales. However, the present results did show that many items of the ASR load significantly on a general factor and it appears most probable that it is the pervasive presence of this general factor rather than the presence of stable, replicable specific factors that led to the moderate level of internal consistency found *within* the subscales and to the significant correlations which have been found *between* many of them.

Indeed, the present study offers considerable support for the existence of a general anger factor being measured by the ASR. This is seen in the presence of a large first factor accounting for 10.06% of the variance in the unrotated factor matrix of the original 89 item inventory and the mean loading of $.24$ obtained for the first factor (unrotated) in that analysis. Evidence is also provided by the pattern of item to total test correlations, item to subtest correlations and subtest to total correlations. Biaggio's (1980) first factor which accounted for 58.9% of total common variance over four anger measures also provides strong evidence for the existence of a general factor.

The Single Scale Questionnaire

On the basis of the foregoing results, it was decided to construct a single scale questionnaire measuring the general anger factor using existing items of the ASR. This was done by calculating item to total test correlations on the original 89

item questionnaire and removing those items with negative and lowest positive correlations in groups of nine or ten.

The final test consisted of 30 items from the original ASR. Items are included from all scales of the ASR except Guilt and Mistrust/Suspicion. Two items were filler items in the original test, ten items are from the Awareness subscale, five from Total Expression, four from Physical Expression, six from Verbal Expression and two from Condemnation of Anger. All but one of these items were those that loaded most highly on the general factor in the unrotated factor pattern of the 89 item version of the ASR. Factor analysis of responses to the 30 items showed clear evidence for a general factor in the unrotated factor loading matrix, with the mean loading for the first factor of .48 and for subsequent factors close to zero. KR20 reliability of this test was good at .89.

Preliminary norms were derived for the 30 item unidimensional scale, based on the responses of 101 male students and 100 female students. Items for this 30 item version of the ASR were selected in a New Zealand environment. It is a short, highly reliable, homogeneous measure of anger, and is apparently a significant psychometric improvement over the original ASR, in that though claiming to measure considerably less, it would appear to be measuring substantially better. It has proved to be suitable for use in New Zealand and its brevity, compared with the previous 89 item ASR, may be potentially useful for situations where the concentration span of subjects may be limited.

References

- Biaggio, M. K. (1980). Assessment of anger arousal. *Journal of Personality Assessment*, 44, 289-298.
- Biaggio, M. K. (1987a). Clinical dimensions of anger management. *American Journal of Psychotherapy*, Vol. XLI, 417-427.
- Biaggio, M. K. (1987b). Therapeutic management of anger. *Clinical Psychology Review*, 7, 663-665.
- Biaggio, M. K. (1989). Sex differences in behavioral reactions to provocation of anger. *Psychological Reports*, 64, 23-26.
- Biaggio, M. K., Supplee, K., & Curtis, N. (1981). Reliability and validity of four anger scales. *Journal of Personality Assessment*, 45, 639-648.
- Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, 21, 343-349.
- Cattell, R. B., Balcar, K. R., Horn, J. L., & Nesselroade, J. R. (1969). Factor matching procedure: An improvement of the *s* index; with tables. *Educational and Psychological Measurement*, 29, 781-792.
- Dahlstrom, W. J., & Welsh, G. S. (1960). *An MMPI handbook: A guide to use in clinical practice and research*. Minneapolis: University of Minnesota Press.
- Green, D. E., Walkey, F. H., Taylor, A. J. W., & McCormick, I. A. (1989). Replication of the factor structure of the Hopkins Symptom Checklist with New Zealand and United States Respondents. *New Zealand Journal of Psychology*, 18, 60-64.
- Johnson, W., & Wilborn, B. (1991). Group counselling as an intervention in anger expression and depression in older adults. *Journal for Specialists in Group Work*, 16, 133-142.
- McCann, J. T., & Biaggio, M. K. (1989). Narcissistic personality features and self-reported anger. *Psychological Reports*, 64, 55-58.
- Novaco, R. W. (1975). *Anger Control: The development and evaluation of an experimental treatment*. Massachusetts: Lexington Press.
- Nunnally, J. C., & Wilson, W. H. (1975). Method and theory for developing measures in evaluation research. In E. L. Struening & M. Guttentag (Eds.). *Handbook of evaluation research*. Beverly Hills: Sage.
- SAS Institute Inc. (1985). *SAS users guide*. Cary, N. C.: Author.
- Schill, T., Ramanaiah, N. V., & Conn, S. R. (1990). Development of covert and overt hostility scales from the Buss-Durkee Inventory. *Psychological Reports*, 67, 671-674.
- Shore, M. F., Clifton, A., Zelin, M., & Meyerson, P. G. (1971). Patterns of masochism: An empirical analysis. *British Journal of Medical Psychology*, 44, 59-66.
- Siebert, R. J., McCormick, I. A., Taylor, A. J. W., & Walkey, F. H. (1987). An examination of reported factor structures of the General Health Questionnaire and the identification of a stable replicable structure. *Australian Journal of Psychology*, 39, 89-100.
- Spielberger, C. D., Jacobs, G., Russell, S., & Crane, R. S. (1983). Assessment of anger: The State-Trait Anger Scale. In J. N. Butaker & C. D. Spielberger (Eds.), *Advances in Personality Assessment Vol. 2*. Hillsdale, NJ: Erlbaum.
- Unverzagt, F. W., & Schill, T. (1989). Anger assessment and its relation to self-report aggressive behavior. *Psychological Reports*, 65, 585-586.
- Walkey, F. H. (1983). Simple versus complex factor analyses of responses to multiple scale questionnaires. *Multivariate Behavioral Research*, 18, 401-421.
- Walkey, F. H., & Green, D. E. (1992). An exhaustive examination of the replicable factor structure of the Maslach Burnout Inventory. *Educational and Psychological Measurement*, 52, 309-323.
- Walkey, F. H., & McCormick, I. A. (1985). FACTOREP: A Pascal program to examine factor replication. *Educational and Psychological Measurement*, 45, 145-150.
- Zelin, M. T., Adler, G., & Meyerson, P. G. (1972). Anger self report: An objective questionnaire for the measurement of aggression. *Journal of Consulting and Clinical Psychology*, 39, 340.

Appendix

The 30 Item Anger Self Report Questionnaire

We would like you to consider carefully each of the following statements and indicate as accurately as you can how it applies to you. There are no right or wrong answers, we just want to know how you feel. Please mark next to each statement according to the amount of your agreement or disagreement by using the following scale:

1. Strong disagreement
2. Moderate disagreement
3. Slight disagreement
4. Slight agreement
5. Moderate agreement
6. Strong agreement

Mark all statements. If a statement is unclear to you please place an 'X' next to it in the margin but mark it anyway. If a statement does not apply to you, place a '?' next to it in the margin but mark it anyway.

Please begin.

1. I get mad easily.
2. I seldom strike back, even if someone hits me first.
3. I never feel hate towards members of my family.
4. Even when my anger is aroused, I don't use strong language.
5. If I am mad, I really let people know it.
6. Sometimes I feel that I could injure someone.
7. I will criticize someone to their face if they deserve it.
8. I find that I cannot express anger at someone until they have really hurt me badly.
9. Even when people yell at me, I don't yell back.
10. At times I have a strong urge to do something harmful or shocking.
11. I have many quarrels with members of my family.
12. I don't feel guilty when I swear under my breath.
13. Feeling angry is terrible.
14. I have physically hurt someone in a fight.
15. At times I feel like smashing things.
16. I find it easy to express anger at people.
17. My conscience would punish me if I tried to exploit someone else.
18. I hardly ever feel like swearing.
19. I couldn't hit anyone if I were extremely angry.
20. I hardly ever get angry.
21. I find it hard to think badly of anyone.
22. I can think of no good reason for ever hitting anyone.
23. I am rarely cross and grouchy.
24. In spite of how my parents treated me, I didn't get angry.
25. I could not put someone in their place even if they needed it.
26. When I really lose my temper, I am capable of slapping someone.
27. It's easy for me not to fight with those I love.
28. If someone annoys me, I am apt to tell them what I think of them.
29. It's useless to get angry.
30. If someone crosses me, I tend to get back at them.