Are All Hostility Scales Alike? Factor Structure and Covariation Among Measures of Hostility[1]

TODD Q. MILLER[2] AND C. DAVID JENKINS
Department of Preventive Medicine and Community Health
University of Texas Medical Branch
Galveston, Texas

GEORGE A. KAPLAN
Human Population Laboratory, Department of Health Services
Berkeley, California

JUKKA T. SALONEN
Research Institute of Public Health, University of Kuopio, Finland

Few studies have examined the correlational and structural relationships between measures of hostility that may predict coronary heart disease. Using data from a population-based study ($N = 2,682$), we examined the factor structure among the five subscales from Siegel's (1986) Multi-dimensional Anger Inventory, the three subscales from Spielberger et al.'s (1985) anger expression scales, a Cook-Medley (Cook & Madley, 1954) hostility subscale that measures cynicism, and items measuring hostility from four scales that assess Type A behavior. Eight separate factors were identified: hostile anger expression, perceived control over the expression of one's anger, frequency of anger, ease of anger provocation, brooding, hostile outlook, cynicism, and sullenness. These results suggest that the structure of hostility measures is more complex than previous factor analyses have suggested. Our findings point to the dangers in assuming that a factor structure can be generalized beyond the data on which it is based to provide theoretical statements about psychological structure. The new scales appear to be valid because they correlate with demographic variables and indicators of social support in the expected directions.

Although previous research strongly suggests that hostility is predictive of future coronary heart disease (CHD) and all-cause mortality (Booth-Kewley

1142

& Friedman, 1987; Matthews, 1988; Miller, Smith, Turner, Guijarro, & Hallet, in press; Smith, 1992), there is considerable conceptual disagreement over the assessment of hostility (Barefoot & Lipkus, 1994; Smith, 1994; Spielberger et al., 1985). Few studies that have examined the relationship between hostility and physical health have included more than one measure of hostility (Matthews, 1988). Therefore, researchers do not know the extent of uniqueness and covariation between the items on these different scales (Barefoot & Lipkus, 1994; Smith, 1992).

The goal of the current study was to seek a better understanding of the factor structure underlying hostility scales that are designed to predict physical health, and to create new scales with improved dimensionality and conceptual clarity. In addition, we examined the correlations among these newly restructured scales.

Previous Theory

Theory suggests a distinction between the behavioral, cognitive, and emotional components of hostility (Buss, 1961). The behavioral component of hostility pertains to either the physical or verbal expression of hostility toward others and has been referred to as aggression. The cognitive and emotional components of hostility relate to the experiential domain of hostility. The cognitive component of hostility includes hostile attitudes, ideas, or thoughts, such as resentment, the devaluation of others, suspicion, and cynicism (Barefoot & Lipkus, 1994; Smith, 1992). The emotional experience of hostility is the emotion of anger. This theory is supported by prior research (e.g., Bushman, Cooper, & Lemke, 1991; Costa, Zonderman, McCrae, & Williams, 1986; Musante, MacDougall, Dembroski, & Costa, 1989) that suggests that many previous scales have only two or three factors reflecting the behavioral, cognitive, and emotional components of hostility. However, most of the hostility scales used in prior research were originally designed to predict some other psychosocial outcome, and not physical health (e.g., violence, interpersonal conflict). Therefore, these scales may reflect aspects of hostility that are different from those identified in the CHD literature. Furthermore, many of these scales have a poorly defined internal structure. For example, a recent confirmatory factor analysis of the widely used Cook-Medley scale suggests that this scale lacks a coherent internal structure (Contrada & Jussim, 1992).

Several recently developed scales suggest that there may be additional dimensions of hostility. For example, Spielberger et al. (1985) suggests that the emotional experience of hostility can be divided into state anger (e.g., current feelings of anger that can be evoked by environmental stimuli) or trait anger that concerns habitual feelings of anger. In addition to identifying factors of

anger-in (i.e., the cognitive experience of hostility) and anger-out (i.e., the behavioral expression of hostility), Siegel's (1986) factor analysis found three additional factors: (a) an *anger-arousal* factor reflecting the frequency, intensity, and duration of the anger response; (b) a *range of anger-eliciting situations* factor that measures the likelihood that different situations will produce anger (e.g., blocking one's plans, being delayed); and (c) a *hostile-outlook* factor that appears to reflect the likelihood that various interpersonal situations will engender hostility. A study by Costa et al. (1986) suggests that there may be more than one dimension within the cognitive experiential domain of hostility. Their factor analysis of the Cook-Medley hostility scale (Cook & Medley, 1954) found two factors of suspiciousness and cynicism. These factors both appear to reflect different aspects of the cognitive experience of hostility.

Current Research

The current research attempted to ascertain the dimensionality of a broad range of hostility scales through a series of theory-driven confirmatory factor analyses (CFAs). We examineed the extent of intercorrelation and factor structure among hostility scales that were collected as part of the population-based ($N = 2,682$) Kuopio Ischemic Heart Disease (KIHD) study (Kaplan & Salonen, 1990; Salonen, 1988). The KIHD study includes the anger-out, anger-in, anger-arousal, range of anger eliciting situations, and hostile-outlook scales from Siegel's (1986) Multidimensional Anger Inventory (MAI) and Spielberger's Anger Expression Scales (Spielberger et al., 1985) that include anger-out, anger-in, and perceived anger control scales. Cynicism was assessed by a Cook-Medley hostility (Cook & Medley, 1954) subscale that has been validated in Finland (Greenglass & Julkunen, 1989). In addition, the KIHD study includes four self- report measures of Type A behavior that each contain several additional items that appear to assess hostility.

Method

Sample

The KIHD sample ($N = 2,682$) was drawn from the city of Kuopio in mid-Eastern Finland and surrounding communities that are not part of the central metropolitan area. Kuopio is the fifth largest city in Finland. Cohort 1 consisted of all men born in 1930-1931 who are permanent residents of the area. This cohort included 1,166 subjects, who represented 85.1% of eligible respondents enrolled between March 1984 and August 1986. All men were 54 years

old at the time of their examination. Cohort 2 included 1,526 men selected from a 33% stratified random sample of men aged 42, 48, 54, and 60 years old residing in the same areas. They were enrolled between August 1986 and December 1989 with a participation rate of 82.9%. Thus, at the time of their interviews, 12.5% of the sample were 42 years old, 13.3% were 48, 59.4% were 54, and 14.8% were 60. Compared with previous studies, the KIHD study has a more comprehensive assessment of psychosocial data (over 600 questionnaire items), sophisticated measurements of biological mechanisms (e.g., carotid artery ultrasonography and cardiovascular reactivity), and clinical cardiovascular endpoints. The occupational status of the subjects was represented by three categories: farmers (16.2%), blue collar jobs (44.3%), and white collar positions (39.5%). The educational status of the subjects was generally low: 10% had less than an elementary school education, 48.1% had an elementary school education, 35.1% had a middle school education, and 6.7% had a high school education. Most of the subjects were married (86.6%), although 6.6% were single and 6.9% were either divorced or widowed.

In any study which translates questionnaire items from one language to another, assuring equivalence of meaning is a critical, yet difficult task. English is widely taught in Finland and is generally read and spoken fluently by the well-educated professional. To assure equivalence of meaning, items that were originally in English were translated by consensus among several epidemiologists who were fluent in both languages and were translated back to English by epidemiologists who had not read the English version of the questionnaire items. Although there are other techniques, we believe back translation that utilizes translators who read and speak both languages well and who are highly conversant with the items and their conceptual and theoretical meaning is the preferred method.

Measures

Multidimensional Anger Inventory (MAI). The MAI (Siegel, 1986) includes several scales that were shortened for the current study. The Cronbach's alphas of these scales in Siegel's original study and in the KIHD were comparable: anger-arousal scale (original 8-item $\alpha = .83$; KIHD 4-item $\alpha = .71$), range of anger-eliciting situations scale (original 7-item $\alpha = .80$; KIHD 4-item $\alpha = .76$), hostile-outlook scale (original 4-item $\alpha = .70$; KIHD 4-item $\alpha = .71$), and anger-in scale (original 5-item $\alpha = .72$; KIHD 3-item $\alpha = .78$). The original two-item anger-out scale had a poor alpha in Siegel's original study ($\alpha = .51$). To improve the reliability of this scale for the KIHD study, three items were used from the item pool that the MAI scale was derived from which loaded highest on the anger-out dimension. Unfortunately, this procedure failed to

produce a more reliable scale (KIHD 3-item α = .46). All items were assessed on 5-point scales that asked subjects to rate how well each of the following statements described themselves: *very well, fairly well, I can't say, somewhat,* or *not at all.*

Anger expression scales. Spielberger et al. (1985) developed three scales that assess the ways in which people express their anger. These scales appear to be as valid and reliable in Finland as they do in the U.S. (Greenglass & Julkunen, 1989, 1991). The *anger-out* scale assesses the degree to which the respondent will do something hostile (e.g., slam doors, argue, say something nasty, lose one's temper) to a person who provokes their anger (KIHD α = .79). The *anger-in* scale assesses the likelihood that the respondent will conceal their anger toward others (KIHD α = .72). The *anger control* scale describes the extent of perceived ability to control one's expression of anger (KIHD α = .88). For all three scales, respondents were asked how frequently they express their anger on 4-point scales (*hardly ever, sometimes, often, almost always*).

Cynicism. Cynical attitudes were assessed by eight items from the 50-item Cook-Medley (Cook & Medley, 1954) hostility scale (KIHD α = .80). The 8-item subscale was found to correlate .77 with the full Cook-Medley scale in a Finnish sample and appears to be a reliable and valid scale in Finland (Greenglass & Julkunen, 1989, 1991). Respondents were asked to rate how accurately the statements described themselves on 4-point scales (*absolutely true for me, pretty much true, not really true,* and *absolutely not true*). A low score on this scale suggests that the respondent believes that people are dishonest and only care about themselves.

Hostility items from Type A scales. The KIHD study includes four Type A scales: (a) the Bortner Type A scale (Bortner, 1969), (b) the Jenkins Activity Survey (JAS; Jenkins, Zyzanski, & Rosenman, 1971), (c) the Framingham Type A scale (Haynes, Levine, Scotch, Feinleib, & Kannel, 1978), and (d) a Finnish scale of Type A behavior (Julkunen, Idänpään-Heikkilä, & Saarinen, 1992). The reliability and validity of these scales have been established in Finland (Greenglass & Julkunen, 1989, 1991; Julkunen et al., 1992). Although researchers have noted that these self-report measures of Type A behavior do not include many items that assess hostility, each of these scales includes a few items. The JAS and Bortner scales include questions that ask the respondent to assess one's current temper, and the JAS has an additional item concerned with one's temper when younger. In addition, the JAS includes a question about ease of irritability and several questions that assess aggressive behaviors such as putting words in the mouth of another person who is speaking slowly. The Finnish Type A scale includes questions about ease of irritation, hurrying others along, and peacefulness of one's personal character. A

Framingham scale item asks whether subjects would describe themselves as bossy or dominating.

Social support and related constructs. Indicators of social support have been shown to be predictive of ischemic heart disease (Kaplan et al., 1994). The KIHD study has a wide variety of measures of social support, including a 5-item scale of emotional support or affection from others (α = .79; e.g., Do you feel loved? How do you feel when you think of your friendships and quantity of love?), a 3-item scale on seeking social support in times of stress (α = .59 ; e.g., Do you talk to someone to find out more about the situation? Do you talk to someone who could do something concrete about the problem? Do you ask someone whom you respect for advice and follow it?), a 3-item scale assessing willingness to give social support to others (α = .76; items assess number of people who turn to them with their personal matters, borrow things from them) and a 17-item introversion scale (α = .79; items on this scale assess the extent of difficulty, confidence and satisfaction associated with meeting new people). We used these scales to speak to the construct validity of the hostility scales, based on the general hypothesis that hostile individuals would be less likely to give social support and would perceive less satisfaction and affection from others. In addition, some types of anger appear to be associated with introversion (Smith & Frohm, 1985), and these individuals may be less willing to seek social support in times of stress.

Statistical Analyses

Split-half replication. Of the 2,682 subjects who completed the questionnaire data, 2,653 (98.9%) responded to some hostility questions. For all subjects with at least some hostility data, two random subsamples of subjects were created for cross-validation analyses: Group 1 (N = 1,330) and Group 2 (N = 1,323).

Procedure. Our initial CFA models were specified to be identical with the factor structure that was found in the original exploratory factor analysis of the scale. After testing this initial CFA model, we tried modified CFA models to determine whether the goodness-of-fit of the initial models could be improved.

Model estimation. The questionnaire items used in our CFA models are ordinal. We used the recommended estimation procedure for ordinal variables, which is to use polychoric correlations with generalized weighted least squares (WLS) estimation (Babakus, Ferguson, & Jöreskog, 1987). We used LISREL VII software (Jöreskog & Sörbom, 1988a). We could not estimate all of the observed variables in one model because the number of variables is too large for the KIHD study's sample size. The sample size for WLS

estimation of polychoric correlations must be at least $1.5k(k + 1)$, where k is the number of variables (Jöreskog & Sörbom, 1988b). To reduce sample size requirements, we initially analyzed the anger inventories and cynicism scale separately.

Model evaluation. All CFA models we tested were rejected by a chi-square goodness-of-fit test at the $p < .0001$ level of significance. This finding is not surprising because the chi-square test is dependent on sample size. Therefore, the chi-square test will reject models that only depart trivially from the data in large samples (Jöreskog & Sörbom, 1988a), such as the KIHD study. To assess the relative goodness-of-fit of our models, we used several goodness-of-fit indices: the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), and the Tucker-Lewis coefficient (TLC). The GFI indices assess the relative amount of the variances and covariances that are predicted by the model. All of these indices usually have values between 1 and 0. Values greater than .90 indicate that the model fits the data reasonably well.

Model modification. We used an approach to model modification that has been suggested by Jöreskog (1993). After examining the goodness-of-fit of the initial CFA model, we attempted to improve the fit of the initial model by (a) removing items that appeared to have low reliability; (b) allowing questionnaire items to correlate with new latent variables, as suggested by modification indices; and (c) combining multiple factors that appeared to represent a single underlying dimension. As suggested by Jöreskog (1993), all items with *squared multiple correlations* (SMCs) that were less than .30 in either the initial or cross-validated sample were deleted from the modified model if modification indices did not suggest that the item may be more highly correlated with another latent variable. SMCs indicate the percentage of variance in the observed variable that is explained by all of the latent variables in the model (Jöreskog & Sörbom, 1988a). Therefore, the SMC is a lower-bound estimate of the reliability of the observed variable.

For initial models, all theta deltas were assumed to be equal to 0. Theta deltas (θ_δ) represent correlation between the error terms of the observed variables. Modification indices can suggest that goodness-of-fit can be improved by allowing theta deltas to be estimated. We decided, somewhat conservatively, that the theta deltas would be estimated in new models if the modification indices suggested that the theta deltas were greater than .25 in both split-half samples.

Using our theoretical perspective (e.g., cognitive, affective, and behavioral dimensions), we decided which model modifications were sufficiently theoretically plausible. Model modifications could lead to different sample sizes

because some items with missing data were dropped from modified models. To ensure that the magnitude of the goodness-of-fit tests were not influenced by changes in sample size, we used the subjects included in the initial model for all modified models.

Results

Confirmatory Factor Analyses

Table format. Tables 1 through 3 show the results of the initial and final CFA models for the cynicism scale and the two anger inventories. CFA with WLS estimation of the asymptotic correlation matrix requires listwise deletion. Therefore, the sample sizes for the two split-half samples vary for each analysis depending on the number of subjects who were deleted because they had missing data on one or more items. The item numbers in the tables indicate the questionnaire items. A brief description of the item follows. Each split-half sample has a column showing the lambda coefficients and SMCs for the final model. The final model is the model that we judged to be the best fitting model. For each latent variable, the lambda with the largest SMC was set equal to 1.00. Another lambda could have been chosen without influencing the results. The numbers in parentheses next to the lambda coefficients are the standard errors of the lambda coefficients. Next to the lambdas and SMCs for the final model, a brief description of each item is presented.

The goodness-of-fit indices for the initial and final models are presented below the lambdas and SMCs. The numbers in brackets are the goodness-of-fit indices for the initial model. The values of the goodness-of-fit indices that are not in brackets indicate the goodness-of-fit of the final modified model. Following the goodness-of-fit indices, Tables 1 through 3 display the correlations between the latent variables. In LISREL terminology, the correlations between latent variables are referred to as phi (ϕ) coefficients and represent an estimate of the unattenuated correlation between the two latent variables.

Cynicism scale. The initial model for the cynicism scale was a one factor model that was specified to test the goodness-of-fit of a model based on an exploratory factor analysis reported by Greenglass and Julkunen (1991). The GFI, AGFI, and TLC goodness-of-fit indices that appear in brackets in Table 1 are all higher than .95, suggesting that the initial model has an excellent goodness-of-fit. For the final model, only one change was made. We deleted Item 7 because in both split-half samples the SMCs were equal to .12. Item 7 appears to assess suspiciousness or mistrust while other items appear to assess a cynical view that other people are dishonest (Table 1). The SMCs for the final model are all moderate in size (.67 to .30).

MAI. As may occur with clearly inadequate models (Jöreskog & Sörbom, 1988a), an initial CFA using the MAI scales failed to converge. It appeared that the nonconvergence occurred because the inter-item correlations among the anger-out items were very low (rs = -.08, .12, and .33), and these items were not highly correlated with other MAI items.

Table 2 shows the goodness-of-fit indices of a second CFA model for the MAI that did not include the three anger-out items. These goodness-of-fit indices that appear in brackets in Table 2 suggest a moderate goodness-of-fit. For example, the TLC coefficient was .90 in both samples. Except for Items 4 and 9 (Sample 1, Item 4 = .29, Item 9 = .29; Sample 2, Item 4 = .35, Item 9 = .26), all SMCs were greater than .30.

Siegel (1986) started with a conceptual model that was based on a review of previous research on the performance of items similar to the ones used in her MAI scale. She was unable to test this factor structure explicitly with exploratory factor analysis. We used CFA to test Siegel's conceptual model and found that this model fit substantially better than did her empirically derived exploratory factor model or the modified model without the anger-out scale items. The TLC coefficient, chi-square test, GFI, AGFI, and SMCs were all substantially improved (Table 2).

For the final model, the name of the range of anger eliciting situations scale was changed to Siegel's original terminology of "ease of anger provocation." The only difference between the final model and the conceptual model is that one item from the anger-arousal scale (Item 4) and one item from the anger-in scale (Item 13) load on the brooding mode of expression scale.

The final model appears to be more cohesive than was the initial model. For example, items on the ease of anger provocation scale clearly assess how easily a person is angered by a variety of situations. The items on the final CFA brooding mode of expression scale appear to assess how long the respondent's anger lasts after an anger producing event (e.g., Item 14, "Even after I have expressed my anger").

In sum, Siegel's (1986) initial conceptual model, which was based on prior empirical research, fit our data better than did her empirical findings. The final scales assess (a) frequency of anger, (b) duration of anger, (c) anger-in, (d) hostile outlook, (e) range of anger-eliciting situations, and (f) brooding mode of expression. Similar to Siegel's original study, the anger-out items did not form a reliable scale.

Anger expression scales. For Spielberger et al.'s (1985) anger expression scales (Table 3), the goodness-of-fit indices for the initial model suggest a moderate to poor fit. In particular, the TLC coefficient was very low (Sample 1 TLC = .791; Sample 2 TLC = .813). For the initial model, one item on the

Table 1

Final Model for the Cynicism Scale.

Item number		Sample 1 (N = 1,089)		Sample 2 (N = 1,072)	
		λ	SMC	λ	SMC
1.	Believes people lie for their own benefit	.83 (.03)	.47	.85 (.03)	.46
2.	Believes people don't bother to help others	.79 (.03)	.42	.84 (.03)	.45
3.	Believes people only have friends to help themselves	.79 (.03)	.42	.79 (.03)	.40
4.	Believes its better not to trust anyone	.67 (.03)	.30	.77 (.03)	.38
5.	Believes people don't care what happens to others	.81 (.03)	.44	.89 (.03)	.50
6.	Believes honesty is only due to the fear of being caught	.79 (.03)	.42	.82 (.03)	.43
8.	Believes people are willing to be unscrupulous	1.00	.67	1.00	.64
	Item not used in final model				
7.	Thinks about real reasons why someone would help them				
	Indicators of goodness-of-fit for initial and final models				
	χ^2	74.5 [87.4]		89.5 [96.3]	
	df	14 [20]		14 [20]	
	GFI	.989 [.989]		.988 [.988]	
	AGFI	.979 [.979]		.975 [.978]	
	TLC	.971 [.971]		.966 [.969]	

Note. GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index; TLC = Tucker-Lewis Coefficient. All χ^2 and λ tests are significant at the $p < .001$ level of significance. Numbers in parentheses represent standard errors. Numbers in brackets represent goodness-of-fit indices for the initial model.

Table 2

Final Model for the Multidimensional Anger Inventory

Item number	Sample 1 (N=1,259)		Sample 2 (N=1,228)	
	λ	SMC	λ	SMC
Frequency of anger				
1. Tends to anger more than most people	.99 (.03)	.69	.99 (.03)	.67
2. Easy to make angry	1.00	.70	1.00	.68
3. Surprised at how often feels angry	.87 (.03)	.52	.91 (.03)	.57
Ease of anger provocation				
5. Gets angry when someone blocks plans	1.00	.54	.97 (.03)	.48
6. Gets angry when delayed	.87 (.03)	.41	.87 (.03)	.39
7. Gets angry when embarrassed	.90 (.03)	.44	.89 (.03)	.40
8. Gets angry when does something stupid	.81 (.03)	.36	.78 (.03)	.31
11. Gets angry taking orders from someone less capable (HO)	.99 (.03)	.53	.98 (.03)	.49
12. Gets angry when working with incompetent people (HO)	.92 (.03)	.45	1.00	.51
Hostile outlook				
9. Friends have habits that are very annoying	.81 (.03)	.34	.80 (.04)	.32
10. Bothered by some people just being around	1.00	.52	1.00	.49
Brooding				
4. Stays angry for hours (AA)	.69 (.02)	.38	.72 (.02)	.42
13. Harbors grudges without telling anyone	.69 (.02)	.38	.70 (.02)	.39
14. Has trouble forgetting anger	.96 (.03)	.73	.93 (.03)	.70
15. Thinks about anger for a long time	1.00	.79	1.00	.81

Items not used in final model

Anger-out

16. Doesn't forget anger soon if doesn't show it
17. Can't talk over problems without letting people know you are angry
18. Doesn't have trouble letting people know when angry

Indicators of goodness-of-fit for initial and final models

	Sample 1	Sample 2
χ^2	601.5 [957.2]	562.3 [912.9]
df	84 [84]	84 [84]
GFI	.970 [.952]	.970 [.952]
AGFI	.957 [.931]	.958 [.932]
TLC	.939 [.900]	.940 [.900]

Correlation matrix of relations between latent variables:

Sample 1

	Frequency	Ease	Brooding	Hostile
Frequency				
Ease	0.62 (.01)			
Brooding	0.43 (.02)	0.50 (.01)		
Hostile	0.63 (.02)	0.68 (.03)	0.65 (.03)	

Sample 2

	Frequency	Ease	Brooding	Hostile
Frequency				
Ease	0.57 (.01)			
Brooding	0.43 (.02)	0.52 (.01)		
Hostile	0.61 (.03)	0.69 (.03)	0.56 (.03)	

Note. GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index; TLC = Tucker-Lewis Coefficient; Frequency = Frequency of anger scale; Ease = Ease of anger provocation scale; Hostile = Hostile-outlook scale. The frequency of anger scales was originally referred to as the anger-arousal scale. The brooding scale was originally referred to as an anger-in scale. Items with an (HO) were originally part of the hostile-outlook scale. Item with an (AA) was originally part of the anger-arousal scale. All χ^2 and λ tests are significant at the $p < .001$ level of significance. Numbers in parentheses represent standard errors. Numbers in brackets represent goodness-of-fit indices for the initial model.

Table 3

Final Model for Spielberger et al.'s (1985) Anger Expression Scales

Item number	Sample 1 (N = 1,017)		Sample 2 (N = 993)	
	λ	SMC	λ	SMC
Anger-out				
2. When angered makes stinging remarks	.74 (.03)	.34	.78 (.03)	.39
3. When angered might slam doors	.88 (.03)	.48	.89 (.03)	.51
4. When angered starts arguments	.75 (.03)	.35	.73 (.03)	.34
5. When angered gets outraged at every annoying thing	.83 (.03)	.43	.88 (.03)	.50
6. When angered says something nasty	.82 (.03)	.42	.79 (.03)	.40
7. When angered loses temper	1.00	.62	1.00	.64
Sullenness				
10. When angered becomes sullen	.77 (.03)	.40	.77 (.03)	.41
13. When angered feels resentful without telling others	.74 (.03)	.38	.80 (.03)	.44
15. When angered becomes angrier than willing to admit	.87 (.03)	.52	.85 (.03)	.49
16. Frequently more irritated than others can guess	1.00	.69	1.00	.69

Anger control				
17. When angered can cool down	.85 (.02)	.56	.84 (.02)	.53
18. When angered is patient toward others	.85 (.02)	.57	.83 (.02)	.52
19. When angered remains calm	.91 (.02)	.65	.92 (.02)	.64
20. When angered can control oneself	.96 (.02)	.71	.97 (.02)	.70
21. When angered can control feelings	1.00	.78	1.00	.75
23. When angered tries to be patient and understanding	.88 (.02)	.61	.89 (.02)	.60
24. When angered can control feelings of anger	.92 (.02)	.66	.97 (.02)	.70
Theta delta				
Anger-out Items 2 and 6	.32 (.03)		.27 (.03)	

Items not used in final model

Anger-out
1. When angered expresses anger
8. When angered lets others know

Anger-in
9. When angered holds it inside
11. When angered withdraws to be alone
12. When angered won't show it
14. When angered criticizes others harshly inside oneself

Anger control
22. When angered calms down faster than others

(table continues)

Table 3 (Continued)

Item number	Sample 1 (N = 1,017)		Sample 2 (N = 993)	
	λ	SMC	λ	SMC

Indicators of goodness-of-fit for initial and final models

	Sample 1 (N = 1,017)		Sample 2 (N = 993)	
χ^2	284.7 [2,335.4]		318.4 [2,048.0]	
df	115 [249]		115 [249]	
GFI	.984 [.900]		.982 [.910]	
AGFI	.979 [.880]		.976 [.891]	
TLC	.979 [.791]		.974 [.813]	

Correlation matrix of latent variables

	Sample 1			Sample 2		
	Anger-out	Sullenness	Anger con	Anger-out	Sullenness	Anger con
Anger-out						
Sullenness	0.73 (.02)			0.70 (.02)		
Anger con	-0.79 (.01)	-0.48 (.01)		-0.74 (.01)	-0.43 (.01)	

Note. GFI = Goodness-of-Fit Index; AGFI = Adjusted Goodness-of-Fit Index; TLC = Tucker–Lewis Coefficient; Anger con = Anger control scale. All χ^2 and λ tests are significant at the $p < .001$ level of significance. Numbers in parentheses represent standard errors. Numbers in brackets represent goodness-of-fit indexes for initial models. The items on the sullenness scale were originally on a scale that was referred to as an anger-in scale.

anger-out scale (Item 8) had SMCs lower than .30, and there were four items on the anger-in scale (Items 9, 11, 12, & 14) that had SMCs less than .30.

Examination of the residuals and modification indices revealed further problems. Several items (1, 9, 11, & 12) were consistently estimated to be equally correlated with more than one latent variable, and the SMCs always remained lower than .30. Therefore, these items were dropped from the final model.

Only 17 of the original 24 items remained for the final factor model. The goodness-of-fit and SMCs for the final factor model are excellent. The error terms (theta deltas) on two similarly worded anger-out items (Items 2 & 6) were allowed to correlate.

Anger-out items deleted from the final model (Items 1 & 8) appear to deal with whether anger is expressed but do not indicate whether the expression is hostile or the specific form of the anger expression. These items do not distinguish between hostile and assertive anger expression. Therefore, the eliminated items may be too ambiguous or may measure assertiveness and not hostility. Anger-in items not included in the final model (Items 9, 11, 12, & 14) are concerned with whether the anger is held inside and not with whether the expression of the anger is hostile. Similar to the excluded anger-out items, these items may not directly assess hostile actions, thoughts, or feelings. Item 14 (critical of others) was the only item that was similar to included items and was excluded from our analysis. The SMC for item 14 was assessed at .28 in the first sample and .27 in the second sample, so this item was very close to our decision rule of .30. The only perceived anger control item that was not included in the final model is dissimilar to other included items in that it concerns the speed of calming down and does not explicitly concern control over anger expression. For example, a person could explode with anger and get the anger out of their system quickly. Therefore, the calming-down question is less directly related to perceived control than other items on the perceived anger control scale.

In sum, two items were excluded from the anger-out scale and four items were excluded from the anger-in scale. The perceived anger control scale replicated very well, and only one item was excluded. Most of the excluded items appear to assess hostility less directly than do included items.

Combined and higher-order factor models. For the modified anger expression, cynicism and MAI scales, there are 39 items. Using the $1.5k(k + 1)$ rule, the required sample size for a CFA with all these items is 2,340. Due to missing data, there was an insufficient sample size to estimate this model.

In order to determine whether two or more factors may be combined to represent a single latent variable, we examined subsets of items (between 28 and 32 items per analysis). These models indicated that all of our factors are

only moderately correlated with each other and that no factors could be combined. We also attempted exploratory maximum likelihood factor analyses, which confirmed our CFA findings and did not identify any additional factors or any higher-order factors.

Several higher-order factor analyses were attempted in which the latent variables from the original factor models were treated as observed variables for second-order latent variables. We attempted a model with two higher-order factors in which one higher-order latent variable represented hostile expression, and a second higher-order latent variable represented the experience of anger. In addition, a three factor higher-order model was attempted with separate cognitive and emotional factors. Perhaps because the hostility scales are only moderately correlated with each other, no satisfactory higher order latent variables could be identified. Therefore, we were left with eight scales having a reasonable degree of independence. The correlations among these scales are shown in Table 4. Only 6 of the 28 coefficients in Table 4 have an absolute value that is greater than .40.

Correlations Among New Scales, Deleted Items, Demographic Variables, and Social Support

Correlations between the new hostility scales and deleted items. Our next step was to examine the correlations among (a) the final scales, (b) the items from the hostility scales *not* used in the final scales, and (c) the hostility items from the Type A scales. No additional items were sufficiently highly correlated with a single scale to warrant inclusion in one of the final scales and the additional items were not highly intercorrelated. Therefore, we could not create any new additional scales.[3]

Reliability of the new hostility scales. The main diagonal of the correlation matrix in Table 4 shows the reliabilities of each scale. All of the reliabilities are

---

[3]Several items not used in the final factor models were highly correlated with more than one scale. For example, four items were correlated with both the frequency of anger and anger control scales: (a) a Bortner item that assessed whether or not the subjects perceived themselves to be short tempered, (b) a Finnish Type A item that asked whether they viewed themselves as easily irritated, and (c) a Jenkins Activity Survey (JAS) item and an anger-in item (Item 14) from the MAI both questioned subjects about their ease of irritation.

Other items (e.g., cynicism Item 7, anger-out Items 1 and 8, anger-in Items 11 and 12, and anger control Item 22) from the original hostility scales that were not used in the final models were most highly correlated with their original scale. In all of these cases, the correlation was not sufficiently high to warrant inclusion in the final scale. Several other items were correlated with another latent variable but the correlations were always low so that the SMCs would be less than our cutoff criterion of .30.

satisfactory, except for the hostile-outlook scale that has only two items. Tables 2 and 3 show the correlations between the latent variables. These correlations are corrected for unreliability, and therefore, represent what the correlation would be if the variables were measured without error. Therefore, these correlations are larger than the correlations presented in Table 4.

Correlations among the new hostility scales. Below each correlation coefficient in Table 4 is the sample size used to estimate the correlation.[4] Below the correlation matrix are the means, standard deviations, skewness, kurtosis, range of values and number of questionnaire items on each scale. The means are given as the items were originally coded (see Procedure section). To ease interpretation of the correlation matrix, the MAI and cynicism scales were reverse coded so that a high score indicates more hostility. For the anger control scale, a high score is interpreted as perceiving to have more control over one's anger. For the two anger-out and sullenness scales, a higher score is interpreted as more hostility.

The correlation matrix shows that the frequency of anger, ease of anger provocation, brooding and hostile outlook scales are moderately correlated with each other (.34 to .47). Similarly, Spielberger et al.'s (1985) anger expression scales are intercorrelated (-.35, -.61, .53). The cynicism scale has only small positive correlations with the other scales. The cynicism scale's highest correlations were with hostile outlook (.25), ease of anger provocation (.22), and brooding (.22). The perceived control over anger expression scale is most highly correlated with hostile anger-out (-.61), frequency of anger (-.38), and sullenness (-.35).

Validity of the new hostility scales. Due to the large sample size, correlations between the demographic variables and the hostility scales that were greater than .03 were statistically significant. However, most of these signifi-

---

[4]Table 4 presents the correlation matrices using pairwise deletion. In order to estimate the effects of the increased missing data for the listwide matrix, the differences between the pairwise and listwise *r*s were calculated for each correlation. The mean absolute difference was .020, with a standard deviation of .017, and a range from .057 to .000. The average correlation was only .007 smaller for the listwise procedure. The correlation between the pairwise and listwise *r*s was .997. Therefore, the pairwise and listwise correlation matrices produced similar results.

The scales were created by taking the sum of the standardized scores from the items that comprised a latent variable. We compared the standardized score correlation matrix with a correlation matrix based on scales that were a simple sum of the item scores. The correlations in the unstandardized matrix were always within .01 of the correlations based on the standardized scores.

All correlations were significantly different from zero at the $p < .0001$ level of significance, given the very large sample size of our study. Only the correlations in the total sample and not the split-half samples are reported because the correlations in the split-half samples never varied more than .04 from the total sample correlations.

Table 4

Pearson Correlations Between Restructured Scales, Demographic Variables, and Social Support

	Multidimensional anger scales				Anger expression scales			Cook-Medley subscale
	Frequency	Ease	Brooding	Hostile outlook	Anger-out	Sullen-ness	Anger control	Cynicism
Frequency	.80 2,584							
Ease	0.47 2,541	.79 2,560						
Brooding	0.34 2,558	0.41 2,537	.79 2,579					
Hostile outlook	0.40 2,574	0.46 2,555	0.41 2,573	.53 2,604				
Anger-out	0.34 2,126	0.23 2,110	0.12 2,121	0.17 2,141	.79 2,173			
Sullenness	0.22 2,162	0.24 2,147	0.35 2,158	0.26 2,179	0.53 2,130	.73 2,210		
Anger control	-0.38 2,124	-0.21 2,109	-0.08 2,117	-0.16 2,138	-0.61 2,091	-0.35 2,123	.89 2,172	
Cynicism	0.21 2,123	0.22 2,110	0.22 2,120	0.25 2,141	0.18 2,056	0.20 2,097	-0.13 2,055	.82 2,173

Variable	Frequency	Ease	Brooding	Anger-out	Sullenness	Anger control	Cynicism	Hostile outlook
Income (lowest to highest quintile)	-0.12 / *2,544*	-0.04 / *2,521*	-0.06 / *2,539*	-0.07 / *2,564*	0.03 / *2,145*	0.04 / *2,181*	0.04 / *2,144*	-0.17 / *2,142*
Education (lowest to highest grade)	-0.14 / *2,579*	-0.02 / *2,555*	-0.07 / *2,575*	-0.05 / *2,599*	0.04 / *2,169*	0.06 / *2,206*	0.03 / *2,168*	-0.21 / *2,169*
Age	0.11 / *2,579*	0.05 / *2,560*	0.05 / *2,579*	0.04 / *2,604*	-0.02 / *2,173*	-0.01 / *2,210*	0.01 / *2,172*	0.09 / *2,173*
Giving social support	-0.13 / *2,514*	-0.10 / *2,493*	-0.12 / *2,510*	-0.09 / *2,536*	-0.04 / *2,119*	-0.08 / *2,154*	0.08 / *2,118*	-0.18 / *2,123*
Feeling loved	-0.14 / *2,433*	-0.11 / *2,410*	-0.17 / *2,427*	-0.15 / *2,447*	-0.08 / *2,040*	-0.20 / *2,077*	0.09 / *2,036*	-0.19 / *2,048*

Variable	M	SD	Skewness	Kurtosis	Range of values	Number of items on scale
Frequency	7.54	2.72	-0.36	-0.26	0—12	3
Ease	11.66	4.68	0.17	-0.16	0—24	6
Brooding	8.48	3.88	-0.01	-0.68	0—16	4
Anger-out	10.60	2.70	1.03	2.17	6—24	6
Sullenness	7.59	2.21	0.65	0.46	4—16	4
Anger control	21.02	4.23	-0.46	-0.07	7—28	7
Cynicism	9.79	3.85	-0.05	-0.26	0—21	7
Hostile outlook	4.66	1.81	-0.15	-0.24	0—8	2

Note. Frequency = Frequency of anger scale; Ease = Ease of anger provocation scale. Sample sizes used to estimate correlations appear below correlations. Cronbach's alphas appear along the diagonal in italics.

cant associations were of very small magnitude (Table 4). In general, greater hostility was weakly associated with less income and education, and greater age. This finding is consistent with previous research (Smith, 1992). The cynicism scale was more closely associated with the demographic variables than were the other scales. A MANOVA with occupation (farmer, blue-collar worker, white collar worker) as the independent variable and the new hostility scales as the dependent variables showed that farmers were more hostile on all indicators except for sullenness. Similarly, a MANOVA based on marital status (married, divorced or widowed, never married) found that those who were never married had significantly lower levels of anger-out, sullenness, anger control, and cynicism. This finding is also consistent with previous research (Smith & Frohm, 1985). The small magnitude of these associations is also consistent with previous research.

We expected that people high on scales that reflect a generally negative view of all people would be (a) less willing to seek out support from others and (b) more introverted. In particular, we expected that people who were sullen, brooding, cynical or had a hostile outlook would be less willing to seek social support in times of stress and would be more introverted. The correlations between not seeking support, and sullenness, brooding, hostile outlook and cynicism were .16, .23, .08, and .09, respectively. The correlations between introversion and sullenness, brooding, hostile outlook, and cynicism were .32, .29, .30, and .27, respectively. As expected, the correlations between seeking social support and introversion were smaller for the other hostility scales that do not reflect a generally negative view of other people. As expected for all of the dimensions of hostility, more hostile individuals were (a) less likely to give social support to others and (b) more likely to perceive less affection from others (Table 4).

Discussion

Several researchers (Barefoot & Lipkus, 1994; Siegel, 1986; Spielberger et al., 1985) have noted that there has been considerable confusion in the measurement of hostility scales that may be linked to physical health. There appears to be many dimensions of hostility (Costa et al., 1986; Siegel, 1986; Spielberger et al., 1985), and the relationships between dimensions require further delineation (Smith, 1994).

We attempted to describe the underlying structure of two anger inventories containing eight anger subscales, a measure of cynicism, and hostility items taken from well-known measures of Type A behavior. In general, our results in a Finnish sample replicated the results of studies conducted in English speaking populations (Musante et al., 1989; Siegel, 1986; Spielberger et al., 1985). For

example, the MAI scales appear to replicate well with Siegel's (1986) initial conceptual model that was based on a review of previous empirical research in English-speaking countries. Similarly, most of the items on Spielberger et al.'s anger-out and anger control scales replicated well. In addition, all of the new hostility scales appear to have reasonable validity because they correlated modestly in the expected directions with a variety of demographic characteristics and different aspects of social support. Research suggests that a lack of social support is a risk factor for CHD (Kaplan et al., 1988). The correlations with social support are consistent with research suggesting that hostile individuals engender less social support which, in turn, increases their vulnerability to CHD (Miller, Markides, Chiriboga, & Ray, in press).

Previous factor analyses (e.g., Bushman et al., 1991; Musante et al., 1989) may not have identified the additional factors found in our analysis and in other recent factor analyses (e.g., Costa et al., 1986; Siegel, 1986). One potential explanation is that some items used in previous scales appear to tap more than one factor. For example, we found that several items on Type A scales appear to tap more than one dimension of hostility. The use of such items may blur distinctions between factors and reduce the number of factors that are identified.

We also found that some hostility items did not correlate with any latent variables. One explanation is that most of the items that were not part of the final factor model did not assess hostility. Most of the deleted items appear to assess the likelihood that a person will not display anger toward others (e.g., a lack of assertiveness). Therefore, these items may be too ambiguous or may assess assertiveness and not hostility.

Cognitive Dimensions of the Experience of Hostility

Costa et al.'s (1986) factor analysis of the Cook-Medley cynical hostility scale (Cook & Medley, 1954) found two highly correlated factors. One factor relates to suspiciousness—the belief that others intend one harm (e.g., paranoia, suspicion, fear of threat to self). Suspiciousness may produce anger through the perception that one is being harmed by others (Shaver, Schwartz, Kirson, & O'Connor, 1987).

The second factor, cynicism, reflects a negative view of human nature as untrustworthy and selfish. The scale used in our study appears to reflect this second aspect of cynicism without the implication of oneself as the intended victim of others. Cynicism appears to be relatively independent of other indicators of hostility. In our study, it had small correlations ($rs = .13$ to $.25$) with all of the other scales. It was most closely correlated with the hostile outlook scale ($r = .25$). Consistent with previous research (Smith, 1992), cynicism was found to correlate with lower socioeconomic status, not giving social support,

less affection from others, greater introversion, and not seeking support from others.

Emotional Dimensions of the Experience of Hostility

Frequency of anger. The "frequency of anger" MAI subscale is best described as an assessment of how often one experiences the emotion of anger. Not surprisingly, this scale was most highly correlated with ease of anger provocation, lack of perceived anger control, and a hostile outlook. The frequency of anger measure may reflect a hostile environment because frequency of anger may, in part, be the result of a hostile environment.

Ease of anger provocation and hostile outlook. The ease of anger provocation and hostile outlook scales in our study indicate "how short a fuse" a person has for experiencing anger in several typical situations that can make a person angry. The hostile-outlook scale specifically asks about typical situations that evoke anger toward other persons. Thus, these scales may be closely related to the construct of trait anger (Spielberger et al., 1985). The hostile outlook and ease of anger provocation scales were moderately intercorrelated ($r = .46$). Both scales were also moderately correlated with frequency of anger. Hostile outlook was also correlated with brooding, not seeking support in times of stress and introversion.

Behavioral Dimensions of Hostility

Anger-out. Anger-out scales assess the expression of hostile behaviors that represent the behavioral component of hostility. Prior research has found modest correlations between anger-out and the experiential aspects of hostility (Smith, 1992). Thus, it is not surprising that anger-out was not highly correlated with the MAI scales or cynicism, which emphasize the experience of anger.

Perceived anger control. Little is known about the construct of perceived anger control (Spielberger et al., 1985). In our study, anger control was strongly negatively correlated with anger-out ($r = -.61$) which is consistent with the view that people who direct their anger toward others perceive themselves as having less control over their anger. Lack of anger control was also moderately correlated with sullenness ($r = -.35$). Data from the KIHD study has found that perceived anger control is predictive of future carotid artery disease (Julkunen, Salonen, Kaplan, Chesney, & Salonen, 1994).

Styles of Expressing and Experiencing Hostility

Brooding. The brooding scale measures the average duration of the experience of anger and appears to tap dimensions of both the emotional and

cognitive experience of hostility. The brooding scale may assess a particular hostile style that involves behavioral withdrawal and cognitive aspects, such as rumination and harboring grudges. In support of this hypothesis, brooding was moderately correlated with other cognitive experiential dimensions of hostility such as frequency of anger, ease of anger provocation, and hostile outlook. In addition, this scale was correlated with introversion and not seeking social support in times of stress, which suggests a tendency to not reveal one's emotions to others.

Sullenness. The sullenness scale is based on four items from Spielberger et al.'s (1985) anger-in scale. The sullenness scale appears to reflect a particular mode of both the experience and expression of hostility. A high score on the sullenness scale suggests that the respondent suppresses anger and tries not to express continuing resentment. Although sullenness may not reflect direct anger expression, the sullenness scale correlated more highly with anger-out than with any other scale in our study and was correlated with not seeking social support in times of stress. Therefore, the sullenness scale may, in part, reflect aspects of a passive-aggressive behavioral dimension of hostility ($r = .53$). In addition, items on the sullenness scale assess resentment and irritation that appear to reflect the experience of hostility. Sullenness was also moderately correlated with brooding and introversion. Thus, the brooding and sullenness scales may reflect a particular dysfunctional style of coping with hostility that has implications for both the behavioral and experiential aspects of hostility.

The sullenness and brooding scales appear to relate most closely to the concept of anger-in that has been widely discussed in previous research (Averill, 1982). Anger-in scales typically include a range of items reflecting brooding and sullenness, as well as guilt, cynicism, and covert hostility (Bushman et al., 1991; Musante et al., 1989; Spielberger et al., 1985).

Directions for Future Research

Our research suggests that there may be more dimensions of anger and hostility than have been identified in past factor analyses. Miller et al. (in press) report that 63 different measures of hostility have been used in previous research on hostility and physical health. Previous factor analyses may have failed to find some dimensions because some types of items were under-represented in their sample. Because of such limits on generality, researchers must be very cautious in interpreting how the structure found among one group of items may generalize to other samples. Similarly, the battery of this study had only a few items representing aggression, so the number of anger-out factors identified in our study may be incomplete.

Most research has been conducted on male-only or male-dominated

Caucasian samples. The current study is no exception and more work is required on females and on other ethnic groups. Previous research has varied in the ages of participants. Siegel (1986) validated her scale on college students and middle-aged factory workers. Spielberger et al. (1985) developed their scale on a high-school student sample. Costa et al.'s (1986) factor analysis of the Cook-Medley scale was based on middle-aged angiography patients. Our sample represented subjects between the ages of 42 and 60 because this is the age range that is at highest risk for ischemic heart disease. Therefore, the factor structure we found may not generalize to younger subjects.

References

Averill, J. R. (1982). *Anger and aggression: An essay on emotion*. New York, NY: Springer-Verlag.

Babakus, E., Ferguson, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research*, **24**, 222-228.

Barefoot, J. C., & Lipkus, I. M. (1994). The assessment of anger and hostility. In A. W. Siegman & T. W. Smith (Eds.), *Anger, hostility and the heart* (pp. 43-66). Hillsdale, NJ: Lawrence Erlbaum.

Booth-Kewley, S., & Friedman, H. S. (1987). Psychological predictors of heart disease: A quantitative review. *Psychological Bulletin*, **101**, 343-362.

Bortner, R. W. (1969). A short rating scale as a potential measure of Pattern A behavior. *Journal of Chronic Diseases*, **22**, 87-91.

Bushman, B. J., Cooper, H. M., & Lemke, K. M. (1991). Meta-analysis of factor analyses: An illustration using the Buss-Durkee hostility inventory. *Personality and Social Psychology Bulletin*, **17**, 344-349.

Buss, A. H. (1961). *The psychology of aggression*. New York, NY: John Wiley & Sons.

Contrada, R. J., & Jussim, L. (1992). What does the Cook-Medley Hostility Scale measure? In search of an adequate measurement model. *Journal of Applied Social Psychology*, **22**, 615-627.

Cook, W., & Medley, D. (1954). Proposed hostility and pharisaic-virtue scales for the MMPI. *Journal of Applied Psychology*, **38**, 414-418.

Costa, P. T., Jr., Zonderman, A. B., McCrae, R. R., & Williams, R. B., Jr. (1986). Cynicism and paranoid alienation in the Cook and Medley Ho Scale. *Psychosomatic Medicine*, **48**, 283-285.

Greenglass, E. R., & Julkunen, J. (1989). Construct validity and sex differences in Cook-Medley hostility. *Personality and Individual Differences*, **10**, 209-218.