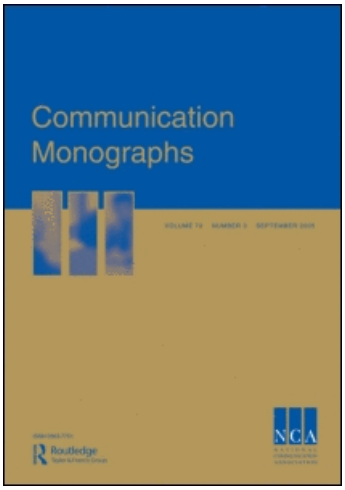


Communication Monographs
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713695619

A Multitrait-Multimethod Validity Assessment of the Verbal Aggressiveness and Argumentativeness Scales

Michael R. Kotowski; Timothy R. Levine; Colin R. Baker; Jeffrey M. Bolt

Online publication date: 01 December 2009

To cite this Article Kotowski, Michael R., Levine, Timothy R., Baker, Colin R. and Bolt, Jeffrey M.(2009) 'A Multitrait-Multimethod Validity Assessment of the Verbal Aggressiveness and Argumentativeness Scales', Communication Monographs, 76: 4, 443 — 462
To link to this Article: DOI: 10.1080/03637750903300247
URL: http://dx.doi.org/10.1080/03637750903300247

Communication Monographs
Vol. 76, No. 4, December 2009, pp. 443–462

Routledge
Taylor & Francis Group

A Multitrait–Multimethod Validity Assessment of the Verbal Aggressiveness and Argumentativeness Scales

Michael R. Kotowski, Timothy R. Levine, Colin R. Baker & Jeffrey M. Bolt

The construct validity of Infante and Wigley's verbal aggressiveness scale and Infante and Rancer's argumentativeness scale are assessed with confirmatory factor analysis and multitrait–multimethod analysis. The factor analytic data replicate previous findings that the verbal aggressiveness scale measures two constructs, verbal aggressiveness and verbal benevolence communication style, and that the argumentativeness scale is unidimensional with some poor items. The multimethod data, however, show near zero correlations between self-reports and observed behavior and evidence of method variance. These data indicate a discrepancy between conceptual definitions and behaviors. Rather than measuring behavioral dispositions to communicate argumentatively or aggressively, the scales may function as attitude or self-concept scales.

Keywords: Argumentativeness; Multitrait–Multimethod; Verbal Aggressiveness

A substantial literature on aggressive and argumentative communication exists (cf. Infante & Rancer, 1996; Rancer & Avtgis, 2006) and much of this research employs Infante and Wigley's (1986) verbal aggressiveness scale (VAS), Infante and Rancer's (1982) argumentativeness scale (ARG), or both. These two scales were designed to measure personality traits predisposing people to communicate in specific ways. Verbal aggressiveness is conceptually defined as "a personality trait that predisposes

Michael R. Kotowski (PhD, Michigan State University) is an Assistant Professor in the School of Communication Studies at the University of Tennessee. Timothy R. Levine (PhD, Michigan State University) is a Professor in the Department of Communication at Michigan State University. Colin R. Baker (PhD, Michigan State University) is an Assistant Professor in the Department of Communication at Wayne State University. Jeffrey M. Bolt (MA, Michigan State University) is a Doctoral Candidate in the Department of Communication at Kent State University. Correspondence to: Michael R. Kotowski, School of Communication Studies, University of Tennessee, Knoxville, TN 37996, USA. E-mail: mkotowsk@utk.edu

ISSN 0363-7751 (print)/ISSN 1479-5787 (online)
DOI: 10.1080/03637750903300247

people to attack the self-concepts of others" (Infante & Wigley, 1986, p. 61), whereas argumentativeness is a "trait that predisposes people to advocate positions on controversial issues while attacking verbally the positions which other people take on these issues" (Infante & Rancer, 1982, p. 72). These conceptual definitions specify, explicitly, behavioral manifestations as the key defining feature of the respective constructs. Consequently, construct-valid measures of these traits must be predictive of relevant behaviors.

A considerable quantity of supportive validation research exists for the VAS and the ARG scale, and they are widely used and accepted. The primary focus of the previous validation research has been on establishing the dimensionality of the scales. Consequently, the factor structures of both scales are well known and documented. Previous construct-validation studies have also found that the two scales correlate generally as expected with other self-report measures. Notably absent from the validity portfolios of the two scales, however, is observational research linking self-reported scores with behaviors. This omission is crucial because these constructs were, and continue to be, conceptualized as behavioral predispositions.

Therefore, the goal of this study is to assess the extent to which scores on the VAS and the ARG scale are associated with observations of relevant behaviors. Constructs are crossed with the methods to provide the first multitrait–multimethod (MTMM) validation study of these two important communication scales. This method allows for the test of behavioral indicators and the control of method variance artifacts.

Validity of the VAS and the ARG Scale

The construct validity of a measure generally refers to the extent to which scores on a measure are indicative of the construct that it was designed to assess, and only that construct, as specified by an a priori conceptual definition existing within a specified theoretical network. Although measures are often said to be valid or invalid, validity indeed is a continuous variable. Furthermore, the stronger and the more consistent the validity evidence, the more confidence that can be placed in claims of a measure's validity. As a result, making a strong case for construct validity requires several forms of evidence (Campbell & Fiske, 1959; Nunnally & Bernstein, 1994).

One common form of validity evidence comes from factor analysis. The dimensionality of the measure needs to match the dimensionality of the construct, with each construct being a distinct dimension. Perhaps the best method for assessing the dimensionality of measures is confirmatory factor analysis (CFA). Factor analysis (especially CFA) along with a face validity assessment is informative about how many dimensions are assessed by a measure, but additional evidence is required to determine the substantive content of the measure in the absence of information indicating how the measure relates to other known entities.

The factor structures (structural validity) of the VAS and the ARG scale are documented well in the literature. The VAS was designed as a 20-item unidimensional measure, but initial tests of the measure indicated two factors, with all the aggressively worded items on one dimension and all the reflected, benevolently

worded items on a second dimension (Infante & Wigley, 1986). Infante and Wigley (1986) attributed the second factor to a methodological artifact, but Beatty, Rudd, and Valencic (1999) and Levine et al. (2004) concluded that the second factor is a distinct construct based upon the face validity of the items. For example, items such as *I try to make people feel good about themselves even when their ideas are stupid* describe behaviors involving active esteem supportiveness through acts of verbal benevolence rather than just a lack of aggression. In any case, the two-factor solution is a replicable finding (Beatty et al., 1999; Infante & Wigley, 1986; Levine et al., 2004; Suzuki & Rancer, 1994).

The intended factor structure of the ARG scale is unclear. On the one hand, Infante and Rancer (1982) posited that tendencies for argument approach and argument avoidance are separate constructs. On the other hand, they present a unidimensional conceptual definition for argumentativeness and suggest scoring the 20-item measure as if it were unidimensional. Subsequent studies that have assessed the face and structural validity of the ARG scale find that a unidimensional model fits after dropping a number of poor items (e.g., Boster, Kotowski, & Andrews, 2006; Boster & Levine, 1988; Boster, Levine, & Kazoleas, 1993). Meta-analytic evidence lends support to this finding (Hamilton & Mineo, 2002). Thus, the available evidence indicates that the ARG scale is unidimensional with some problematic items.

A second form of validity evidence derives from Cronbach and Meehl's (1955) nomological network approach to validation. This approach involves hypothesizing a theoretically predicted network of correlations among a set of measures of different constructs. To the extent that the correlations match theoretical predictions, construct validity evidence is obtained.

The nomological network approach has generated evidence consistent with the validity of the VAS and the ARG scale. Both scales correlate predictably with a variety of measures (e.g., Avtgis & Rancer, 1997; Beatty, Zelley, Dobos, & Rudd, 1994; Infante & Rancer, 1982; Infante & Wigley, 1986; Levine et al., 2004; Onyekwere, Rubin, & Infante, 1991). Consistent with the two-factor model of the VAS, Levine et al. (2004) found that when controlling for the other subscale, the verbal aggressiveness items and the verbal benevolence items differentially predicted aggressive and prosocial communication. Virtually all of the evidence, however, comes from various types of self-report methods.

The nomological network approach has limitations, however. First, it is possible that a construct other than the one purported to be indicated by the focal measure is associated in the same ways with the other constructs measured in the network. Second, the observed associations between constructs could be inflated by method variance or obscured by other method-construct confounds. For these reasons, the MTMM approach provides a stronger test of construct validity than the nomological network approach.

Meta-analytic procedures can also present strong validity evidence for a measure (e.g., Hamilton & Mineo, 2002; Hunter & Schmidt, 2004). Although powerful, meta-analyses are limited by the presence of data in the extant literature. In the case of the VAS and the ARG scale, there is a lack of research examining the relationship between

self-reports on the scales and behaviors. To the authors' knowledge, no previous study has examined the link between scores on the VAS and verbally aggressive behavior. For the ARG scale, at least three previous studies exist (i.e., Infante, 1981; Levine & Boster, 1996; Semic & Canary, 1997). Infante (1981) found that scores on the ARG scale were associated with several dimensions of behavior in actual arguments. Statistically significant associations with behaviors ranged in size from $r = .11$ to $r = .33$. The effects for will-to-argue and argumentative skill were $r = .20$ and $r = .26$, respectively. These findings, however, were not replicated in studies by Levine and Boster (1996) and Semic and Canary (1997). These later studies reported no statistically significant relationships between self-reported argumentativeness and argumentative behavior. Due to the dearth of research examining the relationship between behavior and self-report scores on the VAS and the ARG scale, the links between the VAS, the ARG scale, and communication behavior remain important open questions. Consequently, meta-analysis only provides a partial picture of these scales' construct validity. Therefore, this study employed a MTMM approach utilizing behavioral observation.

Multitrait–Multimethod Validation

Campbell and Fiske's (1959) MTMM approach uses triangulation to establish construct validity evidence. As a solution to construct-method confounds, an MTMM validation crosses a minimum of two measurement methodologies with at least two constructs. This procedure produces four kinds of correlations. These correlations include monotrait–monomethod, monotrait–heteromethod, heterotrait–monomethod, and heterotrait–heteromethod correlations. The strength of the MTMM approach rests in the fact that each correlation provides unique construct validity evidence.

Conceptually, a reliability coefficient (e.g., Cronbach's α) is the proportion of variance shared between an indicator, or set of indicators, and the construct, with the remainder being attributable to measurement error assuming the indicators are valid. The monotrait–monomethod correlations represent this concept. Substantial reliability coefficients are consistent with, yet not sufficient for, construct validity claims. Although rules of thumb are of limited utility, a coefficient of .70 or better is indicative of a measure with relatively little measurement error. Previous research has found using Cronbach's α that reliabilities of the VAS and the ARG scale fluctuate around .80 (Infante & Rancer, 1982; Infante & Wigley, 1986; Levine et al., 2004).

A second type of correlation produced by the MTMM approach is the monotrait–heteromethod correlation. These correlations are measures of association between different measurement methodologies used to measure the same construct and are sometimes referred to as validity coefficients. These correlations will also be substantial, although lower than the corresponding reliabilities, to the extent there is evidence for convergent validity. According to the consensus view, positive monotrait–heteromethod correlations exceeding approximately .50 can be considered consistent with validity. Further, monotrait–heteromethod correlations must be

larger than heterotrait correlations consistently to claim evidence of validity. Substantial and positive monotrait–heteromethod correlations that are systematically larger than the heterotrait correlations of the same row and column provide evidence of convergent validity because alternative measures of the same construct will demonstrate greater concomitant variation than alternative measures of different constructs.

The heterotrait–heteromethod correlations are associations between different measurement methodologies used to measure different constructs. The only specific prediction for these correlations is that they be consistent with theory-based hypotheses. To the extent that these correlations are consistent with theoretical predictions there is evidence for convergent and discriminant validity.

Finally, heterotrait–monomethod correlations are measures of the association between different constructs as measured using a common methodology. Evidence for convergent validity exists to the extent that the heterotrait–monomethod correlations are smaller than the monotrait–heteromethod correlations and to the extent that the associations among different constructs are consistent with theory-based hypotheses. When the heterotrait–monomethod correlations are systematically larger than the monotrait–heteromethod correlations, evidence for method effects and a lack of discriminant validity exists.

This study investigates the extent to which the VAS and the ARG scale are construct-valid measures of verbal aggressiveness, verbal benevolence, and argumentativeness employing the MTMM framework. Specifically, the MTMM approach is used to cross these three constructs with four measurement methodologies: self-report, message-selection, coded message-generation, and coded behavioral observation.

Method

Participants

This study's sample consisted of 103 undergraduates from a large Midwestern university. Statistical power for the validity coefficients exceeds .99 and the power to detect medium effects of $r = .24$ is greater than .80 for a sample of this size. The sample's sex composition was 31% male and 69% female. The average age of the sample was $M = 19.32$, Mdn = 19.00, $SD = 1.32$. Although ethnicity demographics for the sample were not measured, undergraduates at this university were Caucasian in the main, with 8% African American, 5% Asian/Pacific Islander, 3% Hispanic, and 1% Native American. All participants were compensated for their time with class research credit.

Procedure

Data collection followed an IRB approved protocol, in a multiroom laboratory, with two participants scheduled for each time slot in a study titled, *Debate on Current Issues*. Following a consent procedure, the experimenter administered a current issues questionnaire used to assign the topic of debate. The topics were the war in Iraq and the use of the phrase "under God" in the Pledge of Allegiance. The assigned topic of

debate was the issue about which the participant reported feeling more strongly. Following the issue rating task, participants first completed either a packet of measures or debated with a research confederate. All participants completed both tasks in a counterbalanced random order.

To start the debate, the experimenter led one participant into a second room where a research confederate, having the appearance of being another participant, was sitting in one of two chairs. A video camera made an audio and video recording of the debate for later coding. The camera was out of sight behind a two-way mirror and the angle was constant. The experimenter instructed the participant to sit in the chair across from the confederate at which point the participant and confederate were instructed to debate the assigned topic. The experimenter then left the room and turned on the video camera. The confederate was trained to debate the opposing side of the issue chosen by the participant using arguments identified in pilot testing. After 10 minutes, the experimenter turned off the video camera and reentered the room. The experimenter escorted the participant back to the first room at which time the packet of measures was administered.

The timing of the procedures was such that the participant who completed the packet of measures first had time to complete the packet by the conclusion of the first debate. Therefore, at the time the participant who completed the debate first was escorted back to the first room, the experimenter collected the completed packet from, and repeated the debate procedures with, the participant who debated second. Following the second debate, all participants were debriefed.

Measures

The packet of measures included Infante and Wigley's (1986) VAS, Infante and Rancer's (1982) ARG scale, a message-selection task (Appendix 1), a message-generation task (Appendix 2), and basic demographic items. These self-report measures, along with the behavioral observations, provided the basis for crossing traits and methods.

The VAS and the ARG scale. The VAS is a 20-item Likert-type measure using 5-point response scales. When scored as designed by reverse coding the benevolent items before averaging (Infante & Wigley, 1986), scores from the participants in this study were distributed approximately normally (Skew = 0.13, $\text{Skew}_{SE} = 0.24$), $M = 2.44$, $SD = 0.48$ with a standardized item α (SI α) = .79. These data matched closely the descriptives reported by Infante and Wigley (1986), $M = 2.47$ and $\alpha = .81$.

Given the evidence that the VAS consists of two dimensions, verbal aggressiveness and verbal benevolence (Beatty et al., 1999; Levine et al., 2004), confirmatory factor analysis (CFA) was employed in this study to test the internal consistency and parallelism of the measurement model (Hunter & Gerbing, 1982). The CFA method used estimated factor loadings with a centroid procedure using communalities on the diagonal and standardized matrices (Hunter & Hamilton, 1992). Factor loadings and model specifications were used to generate a predicted interitem correlation matrix, which was subtracted from the observed interitem correlation matrix. The resulting residual matrix was examined to assess model fit. To the extent that the residuals in

the matrix were within sampling error of zero, there was evidence that the model fit the data. This method was employed for all CFAs.[1]

CFA found that the size of almost one-half of the factor loadings for a one-factor verbal aggressiveness measurement model were weak (45% were below .40), the internal consistency of the model displayed 10 residuals that were greater than zero by more than sampling error and these residuals were accumulated on several items rather than distributed across the entire matrix. Thus, the evidence indicated that a one-factor measurement model did not fit the data well, $\chi^2(df = 170,\ N = 103) = 272.77$, Root Mean Squared Error (*RMSE*) = .11, Goodness of Fit Index (*GFI*) = .77.

Consistent with previous research (Beatty et al., 1999; Infante & Wigley, 1986; Levine et al., 2004), the two-factor model provided a better fit, $\chi^2(df = 151,\ N = 103) = 174.89$, *RMSE* = .09, *GFI* = .85. Factor loadings were ample (100% were above .40) and both the verbal aggressiveness factor (*RMSE* = .07) and the verbal benevolence factor (*RMSE* = .08) were internally consistent. Furthermore, there was a marginally acceptable level of parallelism between the two factors, *RMSE* = .11.[2] Finally, a nested model comparison provided evidence that the two-factor model fit the data better than the one-factor model, $\chi^2(df = 1,\ N = 103) = 87.56,\ p < .05$. Therefore, the two factors were scored separately.

Averaged responses on the 10-item verbal aggressiveness measure were distributed approximately normally (Skew = 0.11, $\text{Skew}_{SE} = 0.24$), $M = 2.45$, $SD = 0.62$, and SI $\alpha = .77$. The measure's reliability adjusted by the Spearman-Brown Prophecy Formula (SI α_{SB}) to estimate a 20-item measure was SI $\alpha_{SB} = 0.87$. Averaged responses on the 10-item verbal benevolence measure were also distributed approximately normally (Skew = −0.50, $\text{Skew}_{SE} = 0.24$), $M = 3.56$, $SD = 0.54$, and SI $\alpha = .72$. SI $\alpha_{SB} = 0.84$ estimating a 20-item measure. Because a subsequent CFA in this study indicated parallelism problems between item 11 from the verbal aggressiveness measure and several items from the ARG scale the item was dropped. Averaged responses on the reduced set of nine verbal aggressiveness items were distributed approximately normally (Skew = 0.11, $\text{Skew}_{SE} = 0.24$), $M = 2.42$, $SD = 0.63$, and SI $\alpha = .76$. SI $\alpha_{SB} = 0.88$ after adjusting the nine-item measure's reliability estimate to allow a comparison to the 20-item VAS.

The ARG scale consists of 20 Likert-type items with five-point response formats. Although designed as two dimensions with approach and avoidance tendencies as separate constructs, the measure is conventionally scored unidimensionally (cf. Infante & Rancer, 1982). Averaged responses on the 20-item ARG scale as designed were distributed approximately normally (Skew = −0.04, $\text{Skew}_{SE} = 0.24$), $M = 3.18$, $SD = 0.52$, SI $\alpha = .87$.

The CFA for the ARG scale tested a hypothesized three-factor measurement model consisting of the 20-item ARG scale, the 10-item verbal aggressiveness measure, and the 10-item verbal benevolence measure so as to test parallelism. Analysis revealed several items that were not internally consistent, parallel, or both. Twelve argumentativeness items (1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 19, and 20) and one verbal aggressiveness item (11) had disproportionate accumulations of statistically significant residuals. A face validity assessment of these items identified wording

ambiguities. For example, argumentativeness item 14, "I prefer being with people who rarely agree with me," had a weak average interitem correlation with the other argumentativeness items, $M = -0.03$.

Removal of the poor items resulted in a unidimensional set of argumentativeness items, $\chi^2(df = 20,\ N = 103) = 34.76$, $RMSE = .06$, $GFI = .92$. Factor loadings were ample (all greater than .34 and 75% greater than .50), all residuals in the internal consistency matrix were within sampling error of zero, $RMSE = .07$, and the eight ARG scale items had acceptable parallelism characteristics with the verbal aggressiveness and verbal benevolence items ($RMSE = .11$). This finding is consistent with previous meta-analytic research (Hamilton & Mineo, 2002). Averaged responses across the remaining eight items remaining on the ARG scale were distributed approximately normally (Skew $= -0.24$, $\text{Skew}_{SE} = 0.24$), $M = 3.25$, $SD = 0.60$, SI $\alpha = .80$. Adjusting the eight-item reliability to estimate a 20-item measure yielded SI $\alpha_{SB} = .91$.

Message-selection. Participants read a vignette for which they imagined themselves in dispute with the campus parking office about a questionable parking ticket that they received. Participants were instructed to report the likelihood that they would use each of 12 messages in response to the situation. Four messages were verbally aggressive, four messages were verbally benevolent, and four messages were argumentative. Likelihood of use was measured by a five-point Likert-type scale, with 5 corresponding to "very likely."

The intended three-factor measurement model of the 12 message-selection items was tested using CFA, $\chi^2(df = 51,\ N = 103) = 100.04$, $RMSE = .12$, $GFI = .86$. All factor loadings were greater than .40. The internal consistency of each set of four items was acceptable and all residuals were attributable to sampling error (verbal aggressiveness $RMSE = .05$, verbal benevolence $RMSE = .05$, and argumentativeness $RMSE = .06$). Parallelism, however, was less than ideal ($RMSE = .14$ to .16). Because the problematic parallelism residuals were not attributable to any particular items, all items were retained and each set of four items was scored as a unidimensional measure.

The distribution of averaged verbal aggressiveness item responses was skewed positively (Skew $= 1.80$, $\text{Skew}_{SE} = 0.24$), $M = 1.53$, $SD = 0.67$, and SI $\alpha = .80$.[3] Averaging across the verbally benevolent items resulted in a distribution having a slight negative skew (Skew $= -1.08$, $\text{Skew}_{SE} = 0.24$), $M = 3.74$, $SD = 0.79$, and SI $\alpha = .67$. Finally, averaging across the argumentativeness items formed a distribution with a slight negative skew (Skew $= -0.67$, $\text{Skew}_{SE} = 0.24$), $M = 3.90$, $SD = 0.68$, SI $\alpha = .57$.

Message generation. Participants read a second vignette describing a locksmith demanding payment for attempting to retrieve keys from inside of a locked car, even though the locksmith failed to do so after an hour of trying. Participants imagined that it was their car and were instructed to write down what they would say to the locksmith in open-ended format. Four coders later assessed the level of verbal aggressiveness, verbal benevolence, and argumentativeness in each participant's response. The coders were trained to assess the extent to which each participant's response was exemplary of each construct's conceptual definition (Appendix 3). The

coders accomplished this assessment by employing a five-point response scale with 5 being scored as high. To create verbal aggressiveness, verbal benevolence, and argumentativeness measures, each coder was considered an item on a four-item measure. Because each coder measured three constructs, this procedure was hypothesized to form three unidimensional four-item measures.

The extent to which this hypothesized three-factor measurement model fit the data was tested using CFA. All factor loadings were greater than .65, and all but two exceeded .80. The residual matrix provided evidence that the three measures had acceptable levels of internal consistency, and all but one residuals were within sampling error of zero (verbal aggressiveness $RMSE = .09$, verbal benevolence $RMSE = .04$, argumentativeness $RMSE = .00$). All residuals were within sampling error of zero in the verbal aggressiveness parallelism residual matrices as well (verbal aggressiveness and verbal benevolence $RMSE = .05$ and verbal aggressiveness and argumentativeness $RMSE = .06$). The verbal benevolence and argumentativeness measures did not, however, exhibit acceptable parallelism, with problems attributable to Coder 3's verbal benevolence measure. The removal of Coder 3's verbal benevolence measure improved parallelism across the measurement model; all residuals were within sampling error of zero (verbal aggressiveness and verbal benevolence $RMSE = .04$, verbal aggressiveness and argumentativeness $RMSE = .06$, verbal benevolence and argumentativeness $RMSE = .08$). Considered together, these analyses present evidence for the three-factor measurement model's fit with the data, $\chi^2(df = 41,\ N = 103) = 85.52$, $RMSE = .05$, $GFI = .87$.

When averaged across coders according to the CFA, the verbal aggressiveness data were distributed as follows: Skew = 1.67, $\text{Skew}_{SE} = 0.24$, $M = 1.59$, $SD = 0.80$, SI $\alpha = .86$. The verbal benevolence data were distributed as follows: $Skew = 0.70$, $Skew_{SE} = 0.24$, $M = 2.14$, $SD = 1.20$, SI $\alpha = .92$, SI $\alpha_{SB} = 0.94$ adjusting the reliability estimate to allow a comparison to a four coder measure. Finally, the argumentativeness data were distributed as follows: $Skew = -1.07$, $Skew_{SE} = 0.24$, $M = 3.77$, $SD = 1.25$, and SI $\alpha = .95$.

Behavioral observation. The videotapes of the interactions between the participants and confederate were coded by the same four coders of the message-generation task to ensure conceptual consistency across measures. To avoid carryover, the message-generation responses and behavioral observations were coded at different times in different participant orders. The coders watched the videotaped interactions and assessed the degree to which each participant's behavior during the debate was exemplary of the conceptual definitions of verbal aggressiveness, verbal benevolence, and argumentativeness behaviors on a five-point response scale with 5 scored as high (Appendix 4). Like the message-generation task, each coder was considered an item on a four-item measure creating a hypothesized three-factor measurement model.

The fit of the measurement model with the data was assessed using CFA. All factor loadings were larger than .70. Examination of the residual matrix revealed that the internal consistency of the measures was acceptable (verbal aggressiveness $RMSE = .05$, verbal benevolence $RMSE = .04$, and argumentativeness $RMSE = .03$); all residuals were within sampling error of zero. The verbal aggressiveness measure also exhibited parallelism with verbal benevolence ($RMSE = .05$) and

argumentativeness (*RMSE* = .05); there were no residuals larger than what would be expected due to chance. Parallelism between the verbal benevolence and argumentativeness measures was, however, problematic (*RMSE* = .17) and verbal benevolence Coder 4 was removed from further analyses because the coder produced the largest residuals out of all the items in the matrix. The adjustment improved the parallelism matrices across all the three measures (verbal aggressiveness and verbal benevolence *RMSE* = .04, verbal aggressiveness and argumentativeness measure *RMSE* = .05, and verbal benevolence and argumentativeness *RMSE* = .13). Therefore, the evidence is consistent with the claims that the fit of the three-factor measurement model to the data is acceptable, $\chi^2(df = 41, N = 103) = 81.67$, *RMSE* = .08, *GFI* = .88, and that the behavioral measures have structurally validity.

When averaged across coders according to the CFA, the verbal aggressiveness distribution had the following characteristics: Skew = 3.42, $\text{Skew}_{SE} = 0.24$, $M = 1.19$, $SD = 0.46$, and SI $\alpha = .90$. The verbal benevolence distribution had the following characteristics: Skew = 0.31, $\text{Skew}_{SE} = 0.24$, $M = 2.44$, $SD = 0.92$, and SI $\alpha = .83$, SI $\alpha_{SB} = 0.87$ after adjusting the reliability estimate to allow a comparison to the complete four-coder measure. Finally, the argumentativeness had the following distributional characteristics: Skew = -0.88, $\text{Skew}_{SE} = 0.24$, $M = 3.78$, $SD = 0.95$, and SI $\alpha = .89$. Therefore, in addition to having structurally validity, the behavioral measures are highly reliable.

Results

With the exception of the monotrait—monomethod diagonal which contains SI α reliabilities, the MTMM matrices were constructed with ordinary Pearson product-moment correlation coefficients between all the measures collected, organized by trait and method. The MTMM correlation matrix using optimized measures based on the CFA is presented in Table 1 and the matrix based on measures as designed is presented in Table 2. Both tables include correlations corrected (above the diagonal) and uncorrected (below the diagonal) for attenuation due to measurement error. The results are straightforward, and similar conclusions are drawn from each of the matrix variations.

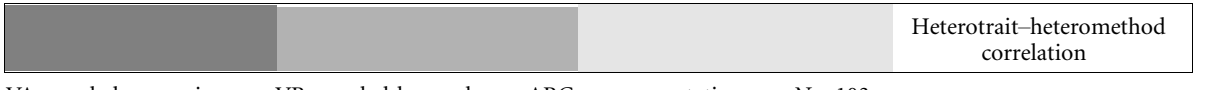
Monotrait–Monomethod

To review, the monotrait–monomethod diagonal contains reliabilities, which are estimates of the proportion of a measure's variability attributable to variability in the construct and not measurement error, assuming the measure is valid. Although rules of thumb are best used with caution, values near .80 are expected from previous research employing the VAS and the ARG scale, and values below .70 cause reason for concern. In the main, the monotrait–monomethod correlations were consistent with the claim that each of the measures was a reasonably reliable measure of its underlying construct. The message-selection method was an exception, however, with low reliability when measuring verbal benevolence (SI $\alpha = .67$) and argumentativeness (SI $\alpha = .57$). Thus, there was evidence for differential reliability across the

Table 1 CFA Optimized Multitrait–Multimethod Correlation Matrix

	Verbal aggressiveness				Verbal benevolence				Argumentativeness			
Instruments	Self-report	Msg. select	Msg. gen.	Behave	Self-report	Msg. select	Msg. gen.	Behave	Self-report	Msg. select	Msg. gen.	Behave
VA self-report	.76	.33	.35	−.10	−.32	−.08	−.28	−.04	.18	.16	.18	−.07
VA msg. select	.43	.80	.29	−.14	−.21	−.37	.01	−.05	.06	.04	−.02	−.12
VA msg. gen.	.43	.35	.86	−.01	−.32	−.27	−.54	−.07	.08	−.05	.53	.11
VA behave	−.12	−.17	−.01	.90	−.13	.13	−.10	−.46	.08	−.02	.16	.40
VB self-report	−.43	−.28	−.40	−.16	.72	.18	.18	.20	−.17	.04	−.24	−.20
VB msg. select	−.11	−.50	−.35	.17	.26	.67	.10	.01	.04	.31	−.06	.04
VB msg. gen.	−.33	.01	−.61	−.11	.23	.13	.92	.09	−.11	−.06	−.81	−.03
VB behave	−.05	−.06	−.08	−.53	.25	.01	.10	.83	−.06	.00	−.14	−.14
ARG self-report	.23	.08	.09	.09	−.22	.05	−.13	−.08	.80	.29	.18	.10
ARG msg. select	.24	.07	−.07	−.03	.05	.50	−.08	.00	.42	.57	−.01	.01
ARG msg. gen.	.21	−.02	.58	.17	−.29	−.08	−.86	−.15	.20	−.01	.95	.12
ARG behave	−.08	−.14	.12	.44	−.25	.05	−.03	−.16	.12	.01	.13	.89

Color coding key:



VA = verbal aggressiveness, VB = verbal benevolence, ARG = argumentativeness. $N = 103$.

Table 2 As Designed Multitrait–Multimethod Correlation Matrix

	Verbal aggressiveness				Verbal benevolence				Argumentativeness			
Instruments	Self-report	Msg. select	Msg. gen.	Behave	Self-report	Msg. select	Msg. gen.	Behave	Self-report	Msg. select	Msg. gen.	Behave
VA self-report	.77	.34	.35	−.08	−.33	−.08	−.26	.03	.19	.18	.20	−.07
VA msg. select	.44	.80	.29	−.14	−.21	−.37	.01	−.04	.08	.04	−.02	−.12
VA msg. gen.	.43	.35	.86	−.01	−.32	−.27	−.55	−.04	.06	−.05	.53	.11
VA behave	−.10	−.17	−.01	.90	−.13	.13	−.12	−.45	.08	−.02	.16	.40
VB self-report	−.44	−.28	−.40	−.16	.72	.18	.19	.20	−.21	.04	−.24	−.20
VB msg. select	−.10	−.50	−.35	.17	.26	.67	.09	−.01	.05	.31	−.06	.04
VB msg. gen.	−.30	.02	−.61	−.13	.24	.12	.93	.12	−.11	−.04	−.87	−.03
VB behave	.03	−.05	−.05	−.53	.27	−.02	.14	.78	−.10	.04	−.11	−.07
ARG self-report	.23	.09	.07	.09	−.27	.06	−.12	−.13	.87	.24	.20	.15
ARG msg. select	.27	.07	−.07	−.03	.05	.50	−.06	.06	.35	.57	−.01	.01
ARG msg. gen.	.24	−.02	.58	.17	−.29	−.08	−.93	−.13	.22	−.01	.95	.12
ARG behave	−.08	−.14	.12	.44	−.25	.05	−.03	−.08	.17	.01	.13	.89

Color coding key:

Monotrait–monomethod correlation	Monotrait–heteromethod correlation	Heterotrait–monomethod correlation	Heterotrait–heteromethod correlation

VA = verbal aggressiveness, VB = verbal benevolence, ARG = argumentativeness. $N = 103$.

measures. For this reason, both corrected and uncorrected correlations are reported in the tables. Consistent conclusions, however, are drawn from both the corrected and uncorrected correlations. Consequently, a differential pattern of measurement error is not a viable explanation for the MTMM results.

Monotrait–Heteromethod

The extent to which different measurement methodologies used to measure the same constructs are correlated positively is indicative of convergent validity because scores on valid alternative measures of the same construct are caused by the same construct. Across all three constructs, the self-report, message-selection, and message-generation measures demonstrated some convergent validity (ranging from $r = .43$ to $r = .20$ for the scales with selection and generation), although below the criterion level of $r = .50$ proposed by Campbell and Fiske (1959). Most noteworthy, however, is the finding that the scale–behavior correlations for all three constructs were within sampling error of zero. The statistical power for these tests was in excess of .995. Put differently, for the constructs of verbal aggressiveness, verbal benevolence, and argumentativeness, paper-and-pencil measures (self-report, message-selection, and message-generation) of the same construct exhibited some convergence, with the behavioral measure approximating orthogonality. These findings are inconsistent with the validity claims for the VAS and the ARG scale because the constructs are conceptualized as behavioral traits. Validity coefficients of the minimum magnitude recommended by Campbell and Fiske are outside the 95% and 99% confidence intervals.

Heterotrait–Heteromethod

Heterotrait–heteromethod correlations are indicators of association between different measurement methodologies used to assess different constructs. To the extent that that the correlations between measures are consistent with theory-based associations between constructs purportedly indicated by the measures there is evidence consistent with convergent and discriminant validity. Consistent with previous research (Beatty et al., 1999; Infante & Wigley, 1986; Levine et al., 2004), the verbal benevolence measures generally correlated negatively with verbal aggressiveness and argumentativeness. Furthermore, the verbal aggressiveness measures correlated positively, albeit within sampling error of zero, with the argumentativeness measures. Although this association is not predicted by theory, it is consistent with meta-analytic research (Hamilton & Mineo, 2002).

Heterotrait–Monomethod

Heterotrait–monomethod correlations are measures of the association between different constructs as measured by a common measurement methodology. Examination of these correlations revealed problems with the measures' discriminant validity because there was a tendency for these correlations to be larger than the monotrait–heteromethod correlations, indicating method effects. That is, measures of different constructs employing the same measurement methodology correlated

stronger than measures of the same constructs using different measurement methodology. The message-generation measure, for example, produced the largest method effects ($r_{va,vb} = -.61$, $r_{va,arg} = .58$, $r_{vb,arg} = -.86$) indicating that the common method introduces concomitant variation into the measurement of different constructs.

Additional Analyses

The reported results were based on correlations between the measures as optimized using CFA (Table 1). Structurally invalid items were removed from several measures. Although excluding these items provided a stronger test of the measures, some may argue that dropping the items damaged the measures' integrity and therefore, obscured the test of the measures' validity as designed. Therefore, Table 2 presents a MTMM matrix of correlations between the measures as designed, including all items. A visual comparison of Table 1 and Table 2 demonstrated that all but one difference between the two matrices were within sampling error of zero.[4] Therefore, the dropping of poor items did not change the findings of the MTMM analyses, although it did improve many of the measures' structural validity.

Discussion

This paper examined the construct validity of Infante and Wigley's (1986) VAS and Infante and Rancer's (1982) ARG scale. Both measures are used frequently in communication research and are believed, in the main, to be both valid and reliable. This study, however, reveals reasons to reconsider beliefs about the validity of the measures. Whereas conclusions about these scales' factor structures, interrelationships, and reliabilities replicated previous findings (Hamilton & Mineo, 2002; Levine et al., 2004), the MTMM results were inconsistent with claims of construct validity. The lack of correlation between scores on the scales and behavior coupled with the indication of method-variance precludes claims of construct validity.

The CFA results were consistent with other recent findings (Beatty et al., 1999; Hamilton & Mineo, 2002; Levine et al., 2004). First, the VAS measures two constructs. Therefore, the VAS is best thought of as two 10-item measures rather than one 20-item measure. Second, the ARG scale contains a number of poor items (most ambiguously worded) that lack internal consistency amongst themselves, parallelism with other measures, or both. Furthermore, there was not evidence in the face validity of the items or the residual matrix indicating that the poor items measured an additional construct. Dropping the poor items did, however, result in a set of items consistent with the unidimensional conceptual definition proposed by Infante and Rancer (1982).

It was the MTMM findings that challenge the construct validity of the VAS and the ARG scale. The 10-item verbal aggressiveness measure correlated moderately with the selection and generation measures but not with observations of aggressive behavior. The 10-item verbal benevolence measure correlated with selection, generation, and observation measures, but those correlations were relatively small. The ARG scale

correlated moderately with the selection measure, less strongly with the generation measure, and did not correlate with behavior. None of the cross-method correlations could be considered as strong validity coefficients. Thus, this study has found a situation where measures of verbal aggressiveness and argumentativeness do not correlate with behaviors. At least two other studies report similar near-zero scale–behavior correlations indicating that the current validity findings are not isolated (Levine & Boster, 1996; Semic & Canary, 1997). Because both constructs are defined as behavioral predispositions, this study documents an important construct–indicator mismatch.

The concomitant method variance observed across the heterotrait–monomethod correlations is also problematic for these measures. These correlations were as large if not larger than the correlations among measures of the same construct using different methodologies for verbal aggressiveness, verbal benevolence, and argumentativeness. This pattern of correlations provides evidence consistent with the claim that these measures have difficulty measuring the constructs without confounding the trait with the measure itself.

Perhaps the most important and most obvious question raised by the current findings is how to interpret findings from past research using the VAS and the ARG scale. Whereas the findings of this study indicate that the VAS and the ARG scale lack construct validity as measures of trait verbal aggressiveness and trait argumentativeness, it does not follow that scores on these measures are useless. The results of this study and prior research point to the likelihood that these scales do measure something. For example, the measures' factor structures and correlations with some outside measures are replicable.

Given that scores on the VAS and the ARG scales correlate with message selection and message generation, but not with behaviors, it is plausible that rather than behavioral dispositions, the VAS and the ARG scale reflect attitudes or self-concept constructs. As measures of attitudes for example, the VAS and the ARG scale may indicate a motivation to act positively or negatively toward verbally aggressive or argumentative communication. Alternatively, as self-concept measures the VAS and the ARG scale may measure the tendency to see one's self as verbally aggressive or argumentative. As measures of affective or cognitive constructs in the context of the data observed in this study, the VAS and the ARG scales' larger correlations with message selection and message generation and the lack of substantial positive correlations with behaviors are less problematic. This study was not designed, however, to test these alternatives.

A finding tangential to the goal of this study was the relatively small monotrait–heteromethod correlations between message selection and message generation. Both the verbal benevolence and argumentativeness correlations were within sampling error of zero and the verbal aggressiveness correlation was modest, $r = .35$. Although both measures were reports of intend responses, the intentions reported in response to the structured choice of the message-selection task did not covary with the intentions reported in response to the free choice of the message-generation task. Notably, neither selection nor generation correlated with observed behavior. This

finding is reminiscent of debates in the compliance-gaining message production literature over the relative validity of selection and generation tasks. This finding is consistent with the conclusion that paper-and-pencil reports have limited predictive validity for at least some behaviors. Studies interested in behavioral outcomes need to observe behavior.

As with all research, the conclusions of this study are not without limitation. In spite of the fact that verbal aggressiveness and argumentativeness are, by definition, behavioral traits, it could be argued that because participants were in a controlled laboratory setting and a particular sample of their behavior was observed, they were unlikely to engage in verbally aggressive and argumentative communication because of situation-specific social forces inhibiting the behavior. Although the verbally aggressive and argumentative behaviors expressed by participants in this study did occur less frequently than what would be predicted by responses on the paper-and-pencil measures, the behaviors nevertheless occurred. The CFA also demonstrated that the measures of these behaviors had structural validity and were highly reliable. The consistency of the current findings with the few previous behavioral studies that have been conducted (e.g., Levine & Boster, 1996; Semic & Canary, 1997) lends confidence to the current finding that verbally aggressive and argumentative behaviors occurred approximately orthogonal to responses on the paper-and-pencil measures.

A second potential criticism concerns the sample size, specifically, the concern of whether the sample size affects the findings meaningfully. Whereas larger sample sizes are always preferred to smaller samples, a larger sample functions to reduce sampling error, shrinking confidence intervals, resulting in more stable population estimates. Considering that the sample size is just over 100 and the MTMM conclusions are based on zero-order correlations, for population correlations of $\rho = \text{zero}$, the 95% confidence intervals are approximately $r \pm .20$. Further, as ρ departs from zero with any fixed sample size, the confidence intervals shrink further. Therefore, the present analyses have a precision of $\pm .20$ or better. Variations of this range do not change the substantive conclusions of this study. For example, consider the correlations between the VAS, the ARG scale, and their respective behaviors. These are $r = -.12$, $-.10$, $+.12$, $+.17$, etc. Even modest validity coefficients are well outside the 95% confidence intervals for these correlations. Consequently, this fact along with the fact that the CFA results replicate previous research (e.g., Hamilton & Mineo, 2002; Levine et al., 2004) minimizes the likelihood that a larger sample would produce findings that would change the substantive conclusions of this study.

Finally, some readers might wonder why a structural equation modeling (SEM) approach to analyzing the MTMM correlation matrix was not employed (e.g., Kenny & Kashy, 1992). It is true that various SEM-based approaches can offer a useful analytic strategy for MTMM data (e.g., testing the monotrait–heteromethod correlations as a unidimensional set of alternate indicators of a construct). In this study's data, however, simple zero-order correlations were informative enough to unambiguously answer to the questions guiding the research.

Conclusion

In summary, the results here demonstrate data inconsistent with the validity of the VAS and the ARG scale as measures of the behavioral predispositions toward verbally aggressive or argumentative communication. Scores on the scale fail to correlate significantly or substantially with behaviors. If measures of behavioral predispositions to verbally aggress or argue are desired, alternative measures appear to be needed. Even so, scores on the scales may plausibly measure attitudes or self-perceived communication tendencies toward verbal aggression and argumentativeness. Future research is necessary to determine if these alternative conceptualizations have merit.

Notes

[1] Hunter and Gerbing's (1982) centroid method does not calculate GFI. Therefore, AMOS was used in a parallel manner to obtain GFI for the CFA. It is worth noting that the Hunter and Gerbing CFA estimates were highly consistent with the AMOS CFA estimates.

[2] Whereas the internal consistency of a set of indicators purported to measure the same construct is the extent to which each indicator correlates predictably with other indicators in that set, the parallelism between two or more sets of indicators is the extent to which indicators purported to measure the same construct correlate predictably with indicators purported to measure a different construct.

[3] None of the skewed distributions showed evidence of range restriction. Furthermore, Havlicek and Peterson's (1977) Monte Carlo study found that Pearson's r is robust to considerable violations of the normality assumption. Therefore, although the MTMM matrix relied on zero-order correlations, some of which contained variables with skewed distributions, the matrix can be interpreted meaningfully.

[4] The only correlation pair that had a difference greater than what could be attributable to sampling error was the message-generation correlation between verbal benevolence and argumentativeness. The correlation between the measures as designed ($r = -.93$) was outside of the 95% confidence interval for the correlation between the optimized measures, $r = -.86$, $P(-.91 \leq \rho \leq -.81) = .95$.

References

Avtgis, T., & Rancer, A. (1997). Argumentativeness and verbal aggressiveness as function of locus of control. *Communication Research Reports*, *14*, 441–450.

Beatty, M. J., Rudd, J. E., & Valencic, K. M. (1999). A re-examination of the verbal aggressiveness scale: One factor or two? *Communication Research Reports*, *16*, 10–17.

Beatty, M., Zelley, J., Dobos, J., & Rudd, J. (1994). Fathers' trait verbal aggressiveness and argumentativeness as predictors of adult sons' perceptions of fathers' sarcasm, criticism, and verbal aggressiveness. *Communication Quarterly*, *42*, 407–415.

Boster, F. J., Kotowski, M. R., & Andrews, K. R. (2006, November). *Identifying influentials: Development of the connector, persuader, and social maven scales*. Paper presented at the annual meeting of the National Communication Association, San Antonio, TX.

Boster, F. J., & Levine, T. R. (1988). Individual differences and compliance gaining message selection: The effects of verbal aggressiveness, argumentativeness, dogmatism, and negativism. *Communication Research Reports*, *5*, 114–119.

Boster, F. J., Levine, T. R., & Kazoleas, D. (1993). The impact of argumentativeness and verbal aggressiveness on strategic diversity and persistence in compliance-gaining behavior. *Communication Quarterly*, *41*, 405–414.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–301.

Hamilton, M. A., & Mineo, P. J. (2002). Argumentativeness and its effect on verbal aggressiveness: A meta-analytic review. In M. Allen, R. W. Preiss, B. M. Gayle, & N. Burrell (Eds.), *Interpersonal communication research: Advances through meta-analysis* (pp. 281–314). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Havlicek, L. L., & Peterson, N. L. (1977). Effect of the violation of assumptions upon significance levels of the Pearson r. *Psychological Bulletin*, *84*, 373–377.

Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second order factor analysis, and causal models. *Research in Organizational Behavior*, *4*, 267–320.

Hunter, J. E., & Hamilton, M. A. (1992). *CFA: A program to do confirmatory factor analysis*. East Lansing, MI: Michigan State University.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Infante, D. (1981). Trait argumentativeness as a predictor of communicative behavior in situations requiring argument. *Central States Speech Journal*, *32*, 265–272.

Infante, D. A., & Rancer, A. S. (1982). A conceptualization and measure of argumentativeness. *Journal of Personality Assessment*, *46*, 72–80.

Infante, D. A., & Rancer, A. S. (1996). Argumentativeness and verbal aggressiveness: A review of recent theory and research. In B. R. Burleson (Ed.), *Communication yearbook 19* (pp. 319–351). Thousand Oaks, CA: Sage.

Infante, D. A., & Wigley, C. J. (1986). Verbal aggressiveness: An interpersonal model and measure. *Communication Monographs*, *53*, 61–69.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112*, 165–172.

Levine, T. R., Beatty, M. J., Limon, S., Hamilton, M. A., Buck, R., & Chory-Assad, R. M. (2004). The dimensionality of the verbal aggressiveness scale. *Communication Monographs*, *71*, 245–268.

Levine, T. R., & Boster, F. J. (1996). The impact of self and other's argumentativeness on talk about controversial issues. *Communication Quarterly*, *44*, 345–358.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Onyekwere, E., Rubin, R., & Infante, D. (1991). Interpersonal perception and communication satisfaction as a function of argumentativeness and ego-involvement. *Communication Quarterly*, *39*, 35–47.

Rancer, A. S., & Avtgis, T. A. (2006). *Argumentative and aggressive communication*. Thousand Oaks, CA: Sage.

Semic, B. A., & Canary, D. J. (1997). Trait argumentativeness, verbal aggressiveness, and minimally rational argument: An observational analysis of friendship discussions. *Communication Quarterly*, *45*, 355–379.

Suzuki, S., & Rancer, A. S. (1994). Argumentativeness and verbal aggressiveness: Testing for conceptual and measurement equivalence across cultures. *Communication Monographs*, *61*, 256–279.

Appendix 1. Message-Selection Task

Instructions: *Read the hypothetical situation below carefully and imagine that you are in that situation. Following the situation are several statements that some people may say in response.*

Imagine that while on a trip to the library you happen to find a parking spot in the parking lot behind the library. You notice that a sign above the space reads, "Reserved for university

vehicles only: Mon–Fri 7a.m.–6p.m." It being a Friday evening after 6p.m. you do not think much of the sign and figure it is okay to park your car there and go into the library. Later, upon returning to your car you notice a parking ticket on your windshield. Since you figure it must have been placed there by mistake, you decide to appeal the ticket. On Monday, you go to the police station. Although you explain the situation, the clerk at the parking violation desk repeatedly asks you to provide evidence of the situation as you explain it. At that point you leave the police station and return to the spot you received the parking ticket at and take a picture of the sign for evidence. After this you return to the police station and speak to the same clerk as before. You show the clerk the picture and explain your situation again. The clerk responds by saying that you still must pay the $15.00 since there is no way to know whether or not the picture you took is of the spot you were parked in.
What would you say in response?

(Please read each statement and using the scale provided indicate how likely you would be to use the statement as a response.)

1. I understand where you are coming from, what you are saying is reasonable. You must hear things like this everyday, but the ticket really is an error.[VB1]
2. You self-righteous jerk, I cannot believe you are calling me a liar. Do you get off on screwing people over?[VA1]
3. You told me to give evidence, implying that if I took a picture of the spot, you would believe me. If you thought I was a liar, why did you have me provide evidence in the first place?[ARG1]
4. I know it's your job and that you're just following the rules and that's fair, but please think about my point for a minute.[VB2]
5. I parked there legally. I wouldn't go looking around campus looking for some random sign that supports my case. I have better things to do with my time, that's how you know I'm not lying.[ARG2]
6. I'm sorry if you're unhappy with your position in life as a clerk but don't take it out on me. I took a picture and went through the trouble to please your perverted power fetish and you know as well as I, I don't deserve the ticket.[VA2]
7. Stop giving me the run around and acting like you have some almighty power. You're a clerk and I'll be your boss someday so you'd better watch it. Now waive my fine.[VA3]
8. Since by your reasoning there is no way for me to prove I was parked there, isn't it at least fair to argue that it is possible that the parking attendant made a mistake giving me the ticket?[ARG3]
9. If I were in your shoes, I'd be acting the same. I must sound like every other person that comes in here. You're not being unreasonable, but take a look at it from my perspective.[VB3]
10. It is obvious that you take your job seriously and I think that is great. I feel that I was parked legally and I am wondering what other proof you think is necessary to prove my innocence.[VB4]
11. You are only a clerk so this must be hard for you to understand. It appears you are too incompetent to figure this out for yourself so I want to speak with someone who can; perhaps your supervisor.[VA4]
12. I'm sure if you check the ticket, you'll see that the location reported is the same as the one in the picture.[ARG4]

Appendix 2. Message-Generation Task

Instructions: *Read the hypothetical situation below carefully and imagine that you are in that situation.*

Imagine you have just parked your car. After getting out of your car you notice that you have locked your keys in the car and you don't have a spare. Since you have no other option, you call a locksmith. The locksmith arrives and for almost an hour the locksmith tries every trick he knows attempting to get your keys out. However, in the end he has no success and gives up trying. As the locksmith is leaving he says that even though he did not get the keys out, he needs money for the hour he was trying since he is on a service call.
What would you say in response?

(Please write exactly what you would say, or describe how you would respond.)

Appendix 3. Message-Generation Coding Instructions

Instructions: *After reading the participant's open-ended response, read and respond to each of the several statements below. Please indicate how much you agree or disagree with each statement by circling the item on the scale provided that most accurately represents how you feel. Use your overall impression of the participant's response to guide your response.*

1. The participant's response seemed to take efforts to appear supportive of and react positively to the position that the locksmith took, regardless of whether or not the participant agreed with them.
2. The participant's response seemed to confront and counter the position that the locksmith took without inhibition.
3. The participant's response seemed to attack the self-concept of the locksmith, instead of or in addition to attacking the locksmith's position.

Appendix 4. Behavioral Observation Coding Instructions

Instructions: *After viewing the videotaped interaction, read and respond to each of the several statements below. Please indicate how much you agree or disagree with each statement by circling the item on the scale provided that most accurately represents how you feel. Use your overall impression of the participant during the interaction to guide your response.*

1. The participant seemed to take efforts to appear supportive of and react positively to the position that the other person took on issues, regardless of whether or not the participant agreed with them.
2. The participant seemed to confront and counter the position that the other person took on issues without inhibition.
3. The participant seemed to attack the self-concept of the other person, instead of or in addition to attacking the other person's position on issues.