NIH Public Access
Author Manuscript
Commun Methods Meas. Author manuscript; available in PMC 2015 January 05.

Published in final edited form as:
Commun Methods Meas. 2011 March 4; 5(1): 48–75. doi:10.1080/19312458.2010.547822.

A Measure of Perceived Argument Strength: Reliability and Validity

Xiaoquan Zhao,
George Mason University

Andrew Strasser,
University of Pennsylvania

Joseph N. Cappella,
University of Pennsylvania

Caryn Lerman, and
University of Pennsylvania

Martin Fishbein
University of Pennsylvania

Abstract

Studies of the content of persuasive messages in which the central arguments of the message are scrutinized have traditionally relied on the technique of thought-listing to assess argument strength. Although the validity of the thought-listing procedure is well documented, its utility can be limited in situations involving non-adult populations and sensitive topics. In this paper we present a self-reported scale that can be used to assess perceived argument strength in contexts where thought-listing may be less appropriate. This scale taps into perceived argument strength from multiple points of view, including but also extending beyond the potential of the argument to elicit positive and negative thoughts. Reliability and validity of this scale were assessed in health communication contexts involving anti-drug PSAs directed at adolescents and anti-smoking PSAs targeting adults. Evidence of convergence between this scale and the thought-listing technique was also obtained using the classical comprehensive exam arguments.

Argument strength appears to be the most commonly manipulated message feature in persuasion research (Johnson, Maio, Smith-McLallen, 2005). Despite variable results depending on topic and audience (Petty & Cacioppo, 1986), there is relatively consistent evidence that, when processed carefully, strong arguments produce more attitude and belief change than do weak arguments (Johnson, Smith-McLallen, Killeya-Jones, & Levin, 2004; Johnson & Eagly, 1989; Wood & Quinn, 2003). Thought-listing, a procedure that assesses audience's cognitive responses to persuasive messages, is often used to measure argument strength. When audiences are processing information in an elaborated fashion, arguments that elicit predominantly favorable thoughts are considered to be strong, whereas arguments eliciting predominantly unfavorable thoughts are considered to be weak.

The validity of cognitive responses as indicants of argument strength (as well as the thought-listing procedure) has been considered to be empirically well-established. (e.g., Cacioppo, Harkins, & Petty, 1981; Cacioppo & Sandman, 1981; Petty & Cacioppo, 1979). The research we present in this article contributes to this long tradition by proposing a simpler, closed-ended measure of cognitive response to arguments as an alternative to the thought-listing approach under conditions where thought-listing might be difficult, inefficient or unrepresentative (Cacioppo, von Hippel, & Ernst, 1997; Shapiro, 1994; Stephenson & Palmgreen, 2001). This scale assesses perceived argument strength from multiple points of view, including but also extending beyond the potential of the argument to elicit positive and negative thoughts. Like thought-listing, this scale can be used as a pretest of message quality, a check of strength manipulation, or an outcome of message exposure in persuasion research. The reliability and validity of the scale were evaluated in health communication contexts involving anti-drug PSAs directed at adolescents and anti-smoking PSAs targeting adults. Convergence between this new measure and the thought-listing procedure was also assessed using the classical comprehensive exam arguments (Petty, Harkins, & Williams, 1980).

Argument Strength and Thought-Listing

Researchers have noted two different approaches to argument strength, one focusing on intrinsic message features, the other message receivers' cognitive responses (O'Keefe, 2003; O'Keefe & Jackson, 1995). Identifying independent message features that may influence persuasion represents a direct approach to argument strength. However, because persuasion is often a situational event, highly generalizable findings from this line of research are still limited (Fishbein & Ajzen, 1975; Petty and Wegener 1998). The cognitive response approach represents an indirect approach to argument strength. It focuses on indicators of argument strength in receivers' reactions to the persuasive message.

The cognitive response approach has been widely adopted through the influence of the Elaboration Likelihood Model (ELM, Petty & Cacioppo, 1986). The ELM posits that people devote different amounts of cognitive resources to the processing of persuasive messages depending on their motivation and ability. Testing of the ELM often involves manipulating argument strength, such that messages advocating the same position can be differentially persuasive. The ELM relies on thought-listing to determine argument strength. Typically, potential arguments are presented to members of the appropriate subject population, who are instructed to read the arguments carefully and then list all their thoughts. These thoughts are then content analyzed and sorted into different categories along certain dimensions. Although the potential dimensions on which the thoughts can be categorized are many, traditionally researchers have focused on valence. That is, thoughts are most often grouped in terms of whether they are favorable, unfavorable, or neutral toward the advocated position. After coding the thoughts, a profile is created for each argument in terms of the numbers of favorable and unfavorable responses it has engendered. A "strong" argument is then defined as one that has elicited predominantly favorable or positive thoughts. A "weak" argument, by contrast, is defined as one that has elicited predominantly unfavorable or negative thoughts.

Like the ELM, our research also approaches argument strength from the perspective of recipients' cognitive responses. To make this perspective clear, we will use *perceived argument strength* to refer to the construct our measurement scale aims to assess. Formally defined, perceived argument strength refers to audience members' perceptions of the quality, strength and persuasiveness of the arguments employed in a persuasive communication. Perceived argument strength is an important concept in persuasion. Many theories consider message-evoked reactions a critical mediating variable in the persuasion process. These include not only theories in the cognitive response tradition (Brock, 1967; Petty & Cacioppo, 1986; Petty & Wegener, 1998), but also theories evolving from other perspectives (e.g., advertising research linking ad liking to brand attitude, Shimp, 1981). Moreover, recent efforts to reconceptualize the input variables in experimental media effects research have also accentuated the importance of message-evoked psychological states as a mediating mechanism between message attributes and message effects (Tao & Bucy, 2007; also see Dillard & Peck, 2000; O'Keefe, 2003). All of these theoretical perspectives call for careful attention to perceived argument strength.

On the practical side, there is clear evidence that audience perceptions of the merits of persuasive messages are a strong and reliable predictor of actual persuasion effects such as attitude change (Dillard, Weber, & Vail, 2007). Perceived argument strength, thus, offers persuasion researchers and practitioners an alternative footing (and sometimes the only footing) on which to gauge the effectiveness of persuasive communication. A case in point would be formative research prior to large-scale information campaigns (Dillard et al.). Focusing on perceived argument strength instead of actual effects in formative research can often help campaign designers overcome constraints of time and resources that could otherwise prove insurmountable.

Limitations of Thought-Listing

While thought-listing has rightfully enjoyed wide application in persuasion research, it does have several potential limitations (Cacioppo, von Hippel, & Ernst, 1997; Shapiro, 1994; Stephenson & Palmgreen, 2001). First, people sometimes may not want to report their thoughts accurately, for example when they believe that their thoughts are socially undesirable. Second, members of certain population segments, such as young children, may not have the ability to report their thoughts accurately. Third, since thought-listing is essentially a memory-based reconstructive process, systematic bias may be introduced by this procedure. Fourth, thought-listing is often an inefficient method, particularly when the listed thoughts have to be analyzed and scored by independent judges. Fifth, certain types of message features (e.g., the visual image of a beautiful woman in an ad for a beauty product) can contribute to message effectiveness through affective or experiential processes that may not be well captured by thought-listing. Finally, lack of literacy skills, low motivation to complete the task, and heavy demand from multiple messages may all negatively affect the quality of data generated by the thought-listing procedure.

In addition to these inherent limitations of thought-listing, much of the research using the thought-listing technique has also relied on a relatively simplistic coding scheme for categorizing thoughts (Eagly & Chaiken, 1993). As mentioned earlier, the vast majority of

the studies using thought-listing have focused on thought valence, i.e., whether the thoughts are favorable or unfavorable to the message-advocated position. Only coding for thought valence ignores other useful information provided by thought-listing, such as the relative amount of central vs. peripheral processing generated by a message. By focusing only on valence of thoughts both thought-listing and perceived argument strength are limited. The measure of perceived argument strength, while having certain advantages, does not allow for qualitative features of arguments provided by thought-listing.

Previous Close-Ended Measures

Although open-ended thought-listing has been the dominant approach to the measurement of argument strength, various close-ended measures have also appeared in previous research. Many of these measures are introduced for the purpose of manipulation check. Some of them are single-item measures (e.g., Darke & Chaiken, 2005; DeBono & McDermott, 1994; Ziegler & Dieh, 2001). Others are multi-item scales, typically including items such as strong/weak, convincing/unconvincing, persuasive/unpersuasive, and so forth (e.g., Andrews & Shimp, 1990; Lavine & Snyder, 1996; Munch & Swasy, 1988). Although many of these scales share similar items, the lack of consistency or continuity is apparent in the literature. Almost every study uses a somewhat different scale, and very rarely is there a clear rationale for the selection of scale items. Moreover, none of these existing scales have been submitted to rigorous validation tests beyond internal consistency checks. Clearly, a well-validated and widely applicable scale of perceived argument strength is still absent.

Development of the Current Scale

The scale presented in this article is intended to serve as an alternative to thought listing incorporating both the strengths of the traditional thought-listing procedure and at the same time overcoming some of its limitations. Relevant research from both within and outside the ELM paradigm has been consulted to inform item selection. Overall, our goal was to develop a theoretically and empirically motivated set of components that stand the best chance of accurately capturing perceived argument strength.

Despite the wide use of thought-listing in persuasion research, seldom have the existing scales of perceived argument strength included items that directly assess the relative valence of message-evoked thoughts. In a rare exception, Stephenson and his colleagues used a close-ended measure to gauge the amount and valence of adolescents' cognitive responses to anti-drug PSAs (Stephenson & Palmgreen, 2001; Stephenson, 2003). This close-ended measure was moderately correlated with the traditional open-ended thought-listing measure (r = .33 and .24 for amount and valence respectively, Stephenson & Palmgreen, 2001) and exhibited good internal consistency across different studies (Stephenson & Palmgreen, 2001; Stephenson, 2003). Based on this work, we decided to include two thought-related items in our scale of perceived argument strength, one assessing favorable thoughts engendered by the argument, the other unfavorable thoughts. We also included an item assessing the likelihood that an argument will elicit agreement from the audience. This item is consistent with the thought-listing approach which essentially profiles the audience's overall agreement with the argument by calculating the ratio of positive (agreeing) over negative

(disagreeing) cognitive responses. Some previous research has also included agreement in their measures of perceived argument strength (e.g., Munch & Swasy, 1988).

In addition to the ELM (and the thought-listing methodology), two other theoretically based, effects-oriented approaches have appeared in the persuasion literature to address argument strength. The first approach is represented by Morley's work (Morley, 1987). Building upon the earlier work by Toulmin (1958) on practical arguments and Bayesian inference making, Morley's research sought to explain persuasion by looking into the "cognitive operations" receivers perform on the arguments presented by persuasive messages. Morley posited that, for an argument to be persuasive, the information it carries has to be perceived by the receivers as plausible, important, and novel. He presented some empirical evidence to show that these three subjective perceptions were necessary for belief change to occur.

The second approach to argument strength was first proposed by Areni and Lutz (1988; also see Areni, 2002) and recently revitalized by other researchers (Johnson et al., 2004; van Enschot-van Dijk, Hustinx, & Hoeken, 2003). This approach viewed argument strength from an expectancy value perspective that underlies the Theory of Reasoned Action (TRA, Fishbein & Ajzen, 1975) and the Theory of Planned Behavior (TPB, Ajzen, 1991). In these expectancy value theories, the acceptability of an outcome belief is considered a function of both the subjective likelihood and the perceived desirability of the outcome. Following this logic, researchers have argued that subjective likelihood and desirability are also essential features of an argument's acceptability and hence important determinants of its strength for the audience.

Amongst their differences, the two approaches reviewed above both see an important role for the perceived probability that the outcome featured in an argument is real. Based on these conceptualizations, we decided that our scale of perceived argument strength should include items to assess the truth value of the argument as perceived by the target audience. We also decided to include Morley's importance component in the scale because other research has produced corroborative evidence that arguments considered important by the audience are likely to be more influential (e.g., Igou & Bless, 2003). Evidence for the centrality of novelty to argument strength is scarce beyond Morley's work. Nevertheless, we included a novelty component in our scale, hoping to further evaluate its relevance to argument strength using our data.

Areni and other's work suggests that the desirability of the outcome featured by an argument should factor in perceived argument strength (Areni & Lutz, 1988; Johnson et al., 2004; van Enschot-van Dijk et al., 2003). This dimension is captured in part by the thought items in our scale which ask whether the argument put favorable (if the outcome is desirable) or unfavorable (if the outcome is undesirable) thoughts in one's mind. We did not include items to explicitly assess outcome desirability, however, because these items (e.g. *[an outcome]* is good/bad; pleasant/unpleasant) would be too close to the typical measures used to assess attitudes (e.g., *smoking* is good/bad; pleasant/unpleasant). Including such items, thus, would likely introduce confound between perceived argument strength and attitude and inflate the predictive success of the perceived argument strength scale.

An important implication of Areni and his followers' work is that arguments should be assessed in light of the major behavioral predictors as identified by theories such as the TRA and TPB. So far this line of research has primarily focused on attitudinal beliefs. In the TPB, however, a sense of confidence in the ability to undertake a behavior is also deemed an important determinant of intention to act and/or actual behavioral performance. An argument's ability to enhance such confidence, therefore, could be an important indicator of the strength of the argument as well. Considering this, we decided to include an additional component in our scale to assess the level of confidence created in the audience by the argument.

Finally, while the components described above focus on specific aspects of perceived argument strength, we also included a component in our scale to directly assess the audience's overall perceptions of the argument's strength and quality. These types of items are commonly used in previous research and appear to have good face validity (e.g., Andrew & Shimp, 1990; Garlick & Mongeau, 1992; Limon & Boster, 2001).

In sum, the perceived argument strength scale includes items to tap into the following components: positive and negative thoughts, agreement, plausibility, importance, novelty, confidence, and overall quality. Together, these components appear to have covered the full range of argument perceptions that existing theory and research suggest should matter. Mindful of thought-listing's limitations, we intended to develop a relatively short scale that can be used with both adult and non-adult populations. For this reason, each of the theoretical components is represented by only 1 or 2 items in our scale. We envision the scale to have a single factor structure because we see no compelling reason to predict *a priori* that any of the components would cluster to form subdimensions.

Overview of Studies

The perceived argument strength scale was empirically evaluated in three studies. Study 1 applied the scale to anti-drug PSAs targeting adolescents and assessed its reliability and validity with an adolescent sample. Study 2 further tested the reliability and validity of the scale by applying it to anti-smoking PSAs for adults. Finally, study 3 assessed the convergence between the proposed scale and the traditional thought-listing procedure using the classical comprehensive exam arguments (Petty, Harkins, & Williams, 1980).

Study 1

In study 1 adolescent participants were presented with multiple arguments extracted from anti-drug PSAs. They evaluated each of the arguments using the perceived argument strength scale. We first assessed the internal consistency of the scale for each argument. Then we conducted a confirmatory factor analysis to test the hypothesized single factor structure underlying the scale. Finally, we examined the construct validity of the scale by assessing the relationship between participants' individual-level perceived argument strength ratings on the one hand and their marijuana use intentions and related cognitions on the other.

Method

Sample—The sample included 322 adolescents, 49.7% of whom were male. About two thirds (66.8%) of the sample were Caucasians, 22.4% African Americans, and 10.8% from other ethnic or racial groups. Age ranged from 12 to 18, with a mean of 15.4 (*SD* = 1.95). Participants were recruited through mall intercept solicitations at 15 locations across the country. Both parental consent and youth assent were obtained before participation. Participants were each paid $5 as compensation.

Arguments—Arguments used in this study were derived from 69 televised anti-drug PSAs. Most of these ads have been used in national or state anti-drug campaigns, and many were produced by the Partnership for a Drug-Free America. We undertook a careful procedure to derive a verbal representation of the central argument made by each of the PSAs.[1] Examples of these arguments are listed in Appendix A.

Procedure—Participants completed the study using touch screen computers in individual sessions. The study began with standard demographic questions and a brief set of questions on sensation seeking. Then 10 randomly selected arguments against marijuana use were presented. After reading each argument, participants evaluated the argument in terms of its perceived strength. After finishing the evaluation task, participants completed a set of measures assessing their marijuana-related cognitions and behavioral intentions.

Measures

Perceived argument strength: We developed a 10-item instrument to measure perceived argument strength. First, participants were asked to indicate on a 5-point scale (1 *strongly disagree*, 5 *strongly agree*) the extent to which they think that: The statement 1) is a reason for staying away from marijuana that is *believable*, 2) is a reason for staying away from marijuana that is *new*, 3) is a reason for staying away from marijuana that is *unconvincing*, 4) gives a reason for staying away from marijuana that is *important* to me, 5) *helped me feel confident* about how best to stay away from marijuana, 6) would *help my friends* stay away from using marijuana, if they were offered, 7) put *thoughts* in my mind *about staying away from marijuana*, and 8) put *thoughts* in my mind *about wanting to try marijuana*. Then they were asked: 9) Overall, how much do you *agree or disagree with the statement*? Finally, they were asked: 10) Is the reason the statement gave for staying away from marijuana a *strong or weak reason*? This last item was scored on a 5-point scale ranging from 1 *very weak* to 5 *very strong*.

The negatively phrased item (*unconvincing*) was reverse coded before analysis. Following traditional practices in handling thought-listing data, we subtracted the unfavorable thoughts score (*thoughts about wanting to try marijuana*) from the favorable thoughts score (*thoughts about staying away from marijuana*) to generate an index of favorability in participants'

[1]First, undergraduate and graduate research assistants familiar with anti-drug ads viewed the ads and crafted a verbal description of the argument made by each ad. This description included relevant information from both the visual and verbal components of the ad. Second, faculty members on the research team did the same thing and produced another set of verbal descriptions. These two sets of descriptions were then compared and a final version of the central argument made by each ad was developed by the authors to (a) be comprehensive, (b) be clear and coherent, and (c) be as consistent as possible with the ad's intent.

cognitive responses to arguments. This index had a potential range from −4 to +4. To convert it into a 5-point scale (so that it would have equal weight as other scale items), we divided the index by 2 and then added a constant of 3 to the resulting scores.

Intention: Participants were first asked how likely it is that they would try marijuana even once or twice over the next 12 months using a 4-point scale ranging from *definitely will not* to *definitely will*. Those who gave answers other than *definitely will not* were then asked, using the same 4-point scale, how likely it is that they would use marijuana nearly every month for the next 12 months. Their answers to these two questions were combined to create a 4-point regular use intention measure with the following options: 1 *definitely will not try*, 2 *definitely will not use regularly*, 3 *probably will not use regularly*, 4 *probably or definitely will use regularly* (*M* = 1.61, *SD* = 1.04).

Attitude: Attitude was measured by asking whether using marijuana regularly would be *bad/good*, *dumb/smart*, *unenjoyable/enjoyable*, and *unpleasant/pleasant*. Answers were indicated on a 7-point scale (−3 to +3) and later averaged to produce an overall attitude score (α = .94, *M* = −1.87, *SD* = 1.61).

Subjective norm: Subjective norm was measured by a single 5-point scale asking how "most people important to you" would feel about one's using marijuana regularly (−2 *strongly disapprove* to +2 *strongly approve*) (*M* = −1.30, *SD* = 1.14).

Self-efficacy: Self-efficacy was measured by asking participants how sure they were that they could say no to marijuana, if they really wanted to. Answering options ranged from 1 (*completely sure)* to 4 (*not at all sure*) (*M* = 1.37, *SD* = .72).

Risk: A risk measure was constructed to index participants' initial propensity for regular marijuana use (*M* = −3.35, *SD* = 1.63). The predictive model underlying this measure was derived from a previous independent survey of 600 adolescents (Fishbein et al., 2002). It included 4 predictors: age, sensation seeking, number of friends who used marijuana even once or twice in the past 12 months, and number of times participants were offered marijuana in the last 30 days.[2] While age, number of friends using, and number of offers to self were each measured with a single item, sensation seeking was measured using a 4-item scale (α = .76) adapted from previous research (Hoyle, Stephenson, Palmgreen, Lorch, & Donohew, 2002).

Results

Internal Consistency—As mentioned earlier, participants in this study were each presented with 10 randomly selected arguments from a pool of 69. Under this design, the number of participants evaluating any given argument varied from 30 to 60, with a mean of 44. We inspected the internal consistency of the scale for each argument individually. The scale overall exhibited a satisfactory level of internal consistency (mean α = .85, range = .75 to .92). However, we also noted that, *for every single argument*, coefficient α would

[2]The equation used is as follows: Use (Past 12 months) = −9.34 + .66 Number of offers + .62 Number of User Friends + .11 Sensation Seeking + .19 Age.

increase if *convincingness* and *novelty* were deleted (mean α = .90, range = .81 to .95). On the other hand, coefficient α would decrease if any other item was deleted. This suggests that *convincingness* and *novelty* were either invalid or weak indicators of perceived argument strength and their inclusion in the scale needs to be reconsidered.

Confirmatory Factor Analysis—To further examine the inter-item reliability of the perceived argument strength scale and to investigate its factorial validity, we submitted the scale to a confirmatory factor analysis (CFA). Although we could conduct a CFA for each argument separately, as we did with the internal consistency analyses, the small sample sizes for individual arguments made it difficult for such tests to produce stable results. To get around this problem, we randomly selected an argument for each participant and conducted a CFA with the entire sample (N = 322). Because the argument for each participant was randomly chosen, we assumed that the idiosyncrasies of the arguments would be a random factor and cause no systematic bias in the analysis. The descriptive statistics and inter-correlations of scale items are presented in Table 1.

The basic measurement model for the perceived argument strength scale was a simple one-level factor structure with one latent variable, nine indicators (*convincingness* and *novelty* included), and zero error correlations. Model estimation was conducted using Amos 7.0. We used Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) to evaluate model fit. Generally, reasonable model fit is indicated by a CFI value higher than .95, and a RMSEA value lower than .06 (Holbert & Stephenson, 2008). Chi-squares are also reported following traditions in the literature.

Initial estimation of the measurement model indicated an inadequate fit, χ^2 (27, N = 322) = 142.4, p < .001, CFI = .92, RMSEA = .12, 90% confidence interval (CI) of RMSEA = .10 – .13. Inspection of the standardized regression weights revealed that both *convincingness* (.19) and *novelty* (.38) had low loadings on the latent factor, while the loadings for the other variables were all relatively high (ranging from .61 to .84). This is consistent with the earlier finding that *convincingness* and *novelty* tended to compromise the internal consistency of the perceived argument strength scale. In light of these findings, we decided to remove *convincingness* and *novelty* from the hypothesized measurement model. After removing these two items, model fit improved significantly, but still did not reach an acceptable level, χ^2 (14, N = 322) = 57.49, p < .001, CFI = .97, RMSEA = .10, 90% CI of RMSEA = .07 – .13.

Modification indices suggested that the model could be further improved by allowing for correlations between two pairs of error terms (*help myself feel confident about staying away from marijuana* and *help friends stay away from marijuana; agreement with the argument* and *argument as a strong or weak reason*). Methodologists have advised that error correlations in CFAs should only be allowed when they can be meaningfully interpreted (Byrne, 2001, 2005; Joreskog & Sorbom, 1993). Both error correlations in our CFA model appear to be explainable. The *confidence* and *friends* items had similar syntactic structure (which could lead to common method variance), while the *agreement* and *reason* items both tapped into the overall effectiveness of the argument. In light of this, we decided to relax the constraints on these error correlations.[3] The new model provided an excellent fit to the data,

χ^2 (12, N = 322) = 16.75, p = .16, CFI = 1.0, RMSEA = .04, 90% CI of RMSEA = 0 – .07. Also importantly, the free estimation of the two error correlations did not bring appreciable change to factor loadings, which remained high (ranging from .62 to .83). This suggests that the hypothesized factor structure was reasonable, and the latent factor was able to account for the vast majority of the shared variance among the items. The factor loadings of the 7 items are presented in Table 2.

Construct Validity—After participants finished evaluating the 10 arguments, they were asked to report their intentions and other cognitions regarding future marijuana use (hereafter referred to as marijuana expectations). This allowed us to inspect the construct validity of the perceived argument strength scale. Specifically, if the scale was valid, we should expect greater perceived argument strength to be associated with more anti-drug expectations. This relationship could arise through two potential mechanisms. First, participants exposed to strong arguments might be more likely to shift their marijuana expectations in an anti-drug direction compared to those exposed to weak arguments. Second, participants holding more anti-drug positions might be more likely to evaluate anti-drug arguments as strong compared to those holding more pro-drug positions. The cross-sectional nature of our data makes it difficult to determine which of the two mechanisms was at work. But both mechanisms should lead to the same type of association between perceived argument strength and marijuana expectations.

To test this hypothesis, we first constructed an overall perceived argument strength score for each participant by averaging their ratings of the 10 arguments (M = 3.78, SD = .72).[4] We then assessed the partial correlations between this overall perceived argument strength measure and the marijuana expectation measures while controlling for risk. As expected, participants who perceived the arguments they read as relatively strong reported lower intentions to use (r = −.35, p < .001), more negative attitudes (r = −.47, p < .001), less favorable subjective norms (r = −.16, p < .01), and stronger self-efficacy to stay away from marijuana (r = .14, p < .05). Our hypothesis was supported.

Discussion

Data from study 1 provided strong support for the reliability of the modified scale for perceived argument strength. This scale exhibited excellent internal consistency for virtually every argument to which it was applied. However, it should be noted that this scale was not the original scale we presented. Two of the original items, *convincingness* and *novelty*, had to be deleted in order for the scale to achieve the highest level of inter-item reliability. We suspect that the *convincingness* item worked to lower reliability in part because of the

[3]In CFA, items with correlated errors may sometimes represent subdimensions of the latent construct. In this case, for example, one may argue that *confidence* and *friends* together represent the efficacy subdimension of perceived argument strength. This interpretation is consistent with the theoretical basis of the scale. We may therefore alternatively specify a two-level factorial model, with efficacy as one of the first-order factors that in turn load on the higher-level factor of perceived argument strength. However, because there were only two items in each of the subdimensions, specifying a two-level factor structure would add much complexity to the measurement model while at the same time subjecting the model to the risk of underidentification. In view of this, we have decided to retain the one-level factor structure in the model while allowing for the corresponding error terms to correlate.

[4]We have alternatively tried to construct the overall argument strength measure using average ratings from the entire sample instead of participants' individual ratings. The construction procedure was otherwise the same. The measure based on average ratings had limited variation and did not predict marijuana expectations. For the sake of simplicity, analyses involving this alternative measure are not reported here.

negative wording. In our unpublished past work with adolescents using the need for cognition scale (Cacioppo, Petty, & Kao, 1984), negatively and positively worded items tended to break into separate clusters regardless of the other semantic characteristics of the stem. Also in past ad effectiveness research, *convincingness* tended to function reliably with other items when worded positively (Fishbein, Jamieson, Zimmer, von Haeften, & Nabi, 2002). Further test of our original scale with a positively phrased *convincingness* item is thus needed.

The weak performance of *novelty* in our scale is inconsistent with Morley's research (1987). However, as mentioned earlier, the importance of *novelty* as a dimension of argument strength is rarely documented beyond Morley's work. In a way, our finding in study 1 was reflective of this general lack of corroborative evidence for the relevance of novelty to argument strength. We will continue to examine the performance of the *novelty* item in study 2.

The modified scale has appeared to be a valid instrument. The CFA showed that the hypothesized single factor structure underlying this scale fit the data well. The loadings for the scale items were all very high. Our data also provided evidence for the construct validity of the scale. We predicted greater perceived argument strength to be associated with more anti-drug expectations about future marijuana use. This prediction was confirmed. Importantly, we observed the predicted relationship after controlling for risk. In other words, participants discriminated between strong and weak arguments regardless of their risk status. This suggests that our scale was indeed assessing perceived argument strength, not just low risk participants' tendency to agree and high risk participants' tendency to disagree with anti-drug arguments.

Study 2

Study 2 continued to evaluate the perceived argument strength scale by applying it to a different type of persuasive message (anti-smoking PSAs) targeting a different population (adults). In this study the *convincingness* item was positively phrased and its relationship with the other items in the scale was expected to become stronger. The role of *novelty* (or the lack thereof) in perceived argument strength was also a focus of investigation. Furthermore, we tested the discriminant validity of the scale by examining its relationship with need for cognition, an individual difference variable with which perceived argument strength is not expected to correlate.

Method

Sample—Three hundred current smokers were recruited via shopping mall intercept. Recruitment sites were distributed nationwide and participant eligibility criteria included being between 18 and 65 years of age, currently smoking every day, smoking a minimum of 5 cigarettes daily, and having smoked at least 100 cigarettes in their lifetime. The sample included 221 Whites (73.7%), 56 African Americans (18.7%), and 23 people from other ethnic or racial groups (7.6%). Approximately half of the sample was male (50.3%). The mean age of the sample was 36.8 (*SD* = 12.7).

Arguments—A total of 99 Anti-smoking public service announcements (PSAs) were collected from sources such as the Centers for Disease Control and Prevention and several state Departments of Health. The central arguments of these PSAs were extracted using a procedure similar to that used in study 1.

Procedure—Study 2 was completed using touch screen computers in individual sessions. Participants first completed demographic and smoking history-related questions. They were then presented with 12 randomly selected arguments. Participants completed the perceived argument strength scale after reading each argument. Finally, participants completed questions related to their perceived vulnerability to the health risks of smoking and intentions to quit.

Measures

Perceived Argument strength: Since the target behavior of the arguments evaluated in this study was quitting smoking, all 10 items in the perceived argument strength scale were modified accordingly to reflect this focus (e.g., The statement is a reason for quitting smoking that is *believable*). Additionally, the *convincingness* item was now positively phrased. All items were scored on a 5-point scale where 1 represented *strongly disagree* or *very weak* and 5 represented *strongly agree* or *very strong*. As in study 1, the positive and negative cognitive response items were combined into a thought favorability index which was then treated as a single item in the instrument. The final perceived argument strength score for each argument was obtained by averaging across scale items.

Perceived vulnerability: Perceived vulnerability to the health risks of smoking was assessed by asking the following: 1) How much do you think you can smoke without harming your health?; 2) To what extent do you feel your overall health has been affected by smoking?; and 3) How much do you think that quitting smoking could help your health? Answering options for the first item were 1 *none*, 2 *an occasional cigarette*, 3 *a few cigarettes per day*, 4 *half a pack a day*, and 5 *1 or more packs a day*. Answering options for the last two items were 1 *not at all*, 2 *a little*, 3 *somewhat*, 4 *quite a bit*, and 5 *very much*. Responses to item 1) were reverse scored and the responses to all three items were then averaged to produce an index of the degree to which participants believed that they were vulnerable to the health effects of smoking (α =.56, M = 3.80, SD = .89).

Intention to quit: Participants were asked two questions: 1) How likely is it that you will try to quit smoking completely and permanently in the next three months? 2) How likely is it that you will quit smoking completely and permanently in the next three months? Possible responses for each question included 1 *I definitely will not*, 2 *I probably will not*, 3 *I probably will*, and 4 *I definitely will*. The intention score was obtained by averaging the two items (r =.77, M = 2.65, SD = .81).

Interest in quitting: Interest in quitting was assessed by asking participants: On a scale of 0 to 10 (0 = *not interested at all*, 10 = *very interested*), how interested are you in quitting smoking? (M = 7.96, SD = 2.92).

Need for cognition: Need for cognition was measured with 5 items adopted from the established instrument developed by Cacioppo and his colleagues (Cacioppo et al., 1984). The items were scored on a 5-point scale from 1 *not at all like me* to 5 *a lot like me* (α =.65, M = 2.29, SD = .82).

Results

Internal Consistency—Arguments in this study were each evaluated by a minimum of 36 participants. Internal consistency of the full perceived argument strength scale ranged from . 76 to .94, with a mean of .88. Unlike in study 1, the alpha coefficients tended to decrease now when the positively worded *convincingness* item was dropped from the scale. However, when *novelty* was deleted from the scale, the alpha coefficients still tended to increase (mean α = .90, range = .82 – .96).

Confirmatory Factor Analysis—As in study 1, we randomly selected one argument for each participant and performed a confirmatory factor analysis of the perceived argument strength scale (see Table 1 for descriptive statistics and correlations). We again used CFI and RMSEA to assess model fit. Initial estimation of the basic measurement model did not reveal an adequate fit to the data, χ^2 (27, N = 300) = 248.05, p < .001, CFI = .86, RMSEA = .17, 90% CI of RMSEA = .15 – .19. Estimation results showed that *novelty* had a very low loading on the latent factor (.23) while all other items had fairly strong loadings (.59 – .88). In view of this and the earlier finding regarding internal consistency, we decided to remove *novelty* from the scale. Removing this item improved CFI, but slightly increased RMSEA. Overall the model still did not fit the data at an acceptable level, χ^2 (20, N = 300) = 204.17, p < .001, CFI = .88, RMSEA = .18, 90% CI of RMSEA = .15 – .20.

Modification indices suggested that three error correlations could be freely estimated to further improve model fit (see Byrne, 2001, 2005; Joreskog & Sorbom, 1993; also see note 3). Two of the error correlations were identical to those added to the basic model in study 1 (between *agreement* and *reason*, and between *confidence* and *friends*). The third added error correlation was between *believability* and *convincingness* (note that *convincingness* was not retained in the model in study 1).[5] This error correlation appeared to make sense because *believability* and *convincingness* both tap into the perceived plausibility of the argument. After freeing these three correlations, the model produced an excellent fit to the data, χ^2 (17, N = 300) = 18.71, p = .35, CFI = 1.0, RMSEA = .02, 90% CI of RMSEA = 0 – .06. Furthermore, as was the case in study 1, the free estimation of the three error correlations did not cause appreciable change to factor loadings, which were all relatively high (ranging from .55 to .85, see Table 2). These results suggest that the specified factor structure was reasonable and that the latent factor was able to account for substantial variance in the observed items.

Construct Validity—We examined the construct validity of the perceived argument strength scale by first examining the relationship between the overall perceived strength of

[5]A model without this third error correlation did not fit the data adequately, χ^2 (18, N = 300) = 97.65, p < .001, CFI = .95, RMSEA = . 12, 90% CI of RMSEA = .10 – .15.

the arguments participants were exposed to and their smoking-related perceptions and expectations. It was hypothesized that greater perceived argument strength would be associated with more anti-smoking perceptions and expectations. Once again, this hypothesis was derived based on two potential mechanisms. First, exposure to strong arguments might move participants toward more anti-smoking positions. Second, participants holding more anti-smoking positions might be more likely to evaluate anti-smoking arguments favorably.

The overall perceived argument strength scores were obtained by averaging the perceived argument strength scores of the 12 arguments each participant rated ($M = 3.71$, $SD = .62$). We then assessed the partial correlations between overall perceived argument strength and perceived vulnerability, intention to quit, and interest in quitting, while controlling for current risk status (i.e., number of cigarettes smoked last week). As expected, participants reporting greater perceived argument strength indicated greater perceived vulnerability to the health risks of smoking ($r = .51$, $p < .001$), greater intentions to quit ($r = .44$, $p < .001$), and greater interest in quitting ($r = .54$, $p < .001$). Our hypothesis was supported.

In study 2 we also examined the discriminant validity of the perceived argument strength scale by testing its relationship with need for cognition. If the scale was a valid assessment of perceptions of argument quality, it should not be influenced by individuals' propensity for cognitive activity. In other words, need for cognition may influence how much people think about an argument, but it should not influence how favorably people think about an argument. The correlation between need for cognition and the overall perceived argument strength scores was consistent with this reasoning ($r = -.03$, $p = .58$).

Discussion

Results from study 2 provided additional evidence for the reliability and validity of the perceived argument strength scale minus the *novelty* item. The scale exhibited excellent internal consistency across arguments. The hypothesized factor structure was supported by the CFA. Evidence for the construct validity of the scale was also obtained. Participants' argument ratings were found to be positively correlated with quitting intentions and other anti-smoking cognitions but unrelated to need for cognition.

Study 2 replicated study 1's finding regarding *novelty*. In both studies, the *novelty* item lowered the scale's internal consistency and failed to load strongly on the latent construct. In view of its weak performance across contexts, we decided to permanently remove *novelty* from the perceived argument strength scale.

If novelty is not an important component of perceived argument strength, what role does it play in persuasion? This question is beyond the scope of this research. But it is worth noting that some researchers have considered novelty to be a dimension of message sensation value (Palmgreen, Stephenson, Everett, Baseheart, & Francies, 2002). From that perspective, the real impact of novelty lies not in its contribution to argument strength, but in its ability to attract attention. Other researchers have also explored the possibility that novelty might work as a moderator in the relationship between perceived message quality and actual

message effectiveness. This research, however, has turned out only inconclusive evidence at this point (Dillard et al., 2007).

A difference between study 2 and study 1 is the wording of the *convincingness* item. When positively phrased in study 2, *convincingness* emerged as a reliable indicator of the perceived argument strength construct. It served to increase, rather than decrease, the internal consistency of the scale. *Convincingness* also had the second highest factor loading among all scale items, showing that it is an important indicator of perceived argument strength. In view of these findings, we recommend using positive wording for the *convincingness* item in future application of the perceived argument strength scale. The final form of the scale is presented in Appendix C.

Study 3

Studies 1 and 2 have produced consistent evidence for the reliability and validity of the perceived argument strength scale. However, direct evidence for the convergence between the traditional thought-listing procedure and the proposed scale is still lacking. To test the convergent validity of our scale, we conducted a third study to assess the consistency between thought-listing and the proposed scale using existing arguments advocating senior comprehensive exams (Petty, Harkins, & Williams, 1980). Strong and weak arguments in support of senior comprehensive exams were presented to a group of college students. Thought-listing responses and scale ratings were obtained for each argument. Consistency between these two sets of data was then assessed using both the individual and the argument as the units of analysis.

Method

Sample—Sixteen (16) male and 43 female undergraduate students from George Mason University participated in this study. The average age of the sample was 20.7 (SD = 2.81) and 68% were White. Participants were recruited via extra credit from introductory level communication classes.

Design and Procedure—The study had a mixed design with argument strength (strong vs. weak) as a between-subject factor and measurement technique (thought-listing vs. scale) as a within-subject factor.

Participants were randomly assigned to argument strength conditions. In each condition, participants were presented with 8 arguments in support of instituting senior comprehensive exams in their university. The arguments used in the two conditions were shown by previous data to be either strong or weak in quality (Petty et al., 1980).

The 8 arguments appropriate to condition were presented in random order for each participant. The participant responded to the first 4 arguments using the thought-listing procedure and then rated the remaining 4 arguments using the perceived argument strength scale. This design feature ensured that each participant would provide only one type of data for each argument (either thought-listing or scale rating), yet each argument would have both types of data coming from different participants. Also of note is that the order of the

two measurement methods was fixed for all participants in this study (with thought-listing always preceding scale rating). This was considered necessary because exposure to scale items first could potentially influence participants' cognitive responses to ensuing arguments.

After evaluating the arguments, participants were asked to indicate their attitudes toward instituting comprehensive exams in their university. Demographic questions were asked last. Participants completed the entire study using laptop computers in small group sessions.

Measures

Thought-listing: After reading each of the first 4 arguments, participants were asked to write down all of their thoughts while reading the argument. The instructions for the thought-listing procedure were copied from an example offered by Cacioppo and Petty (1981) and were minimally modified to adapt to the computerized format. The recorded thoughts were coded by two independent coders into three categories: favorable, unfavorable, or neutral to the advocated policy (Cohen's *kappa* = .82; differences resolved through discussion). An index of thought favorability for each argument was then obtained by subtracting the number of unfavorable thoughts from the number of favorable thoughts in response to that argument.

Perceived argument strength: The items of the final perceived argument strength scale (see Appendix C) were adjusted to reflect the focus on the comprehensive exam policy (e.g., The statement gives a reason for instituting comprehensive exams for seniors at George Mason University that is *important* to me). As in study 2, all items were scored on a 5-point scale (1 to 5) and the thought items were combined before calculating the average perceived argument strength score for each argument (mean α = .92, range = .84 to .98).

Attitude: Attitude toward instituting comprehensive exams was assessed with three semantic differentials (*bad/good, unfavorable/favorable, foolish/wise*). Participants indicated their positions on a 7-point scale (1 to 7). The three items were averaged to produce an overall attitude score (α = .91, M = 2.96, SD = 1.65).

Results

To replicate previous research (Petty et al. 1980), we first tested whether participants exposed to strong vs. weak arguments differed in their thought favorability and post-exposure attitudes. Because each participant responded to 4 arguments using thought-listing, we constructed an overall favorability index by averaging the 4 thought favorability scores. If the strong and weak arguments identified by Petty et al. (1980) were indeed different in strength, we should see higher favorability scores and more positive attitudes from the strong condition than from the weak condition. These expectations were supported. Participants in the strong condition (M = −.63, SD = 1.31) reported relatively more favorable thoughts than participants in the weak condition (M = −1.51, SD = 1.10), t = 2.77, df = 57, p = .008. Participants in the strong condition (M = 3.37, SD = 1.73) also reported more positive attitudes toward senior comprehensive exams than participants in the weak condition (M = 2.42, SD = 1.40), t = 2.27, df = 57, p = .03.

We next examined whether a similar difference existed in perceived argument strength ratings. Again, an overall rating index was constructed for each participant by averaging across the 4 rated arguments. If our scale was indeed a valid measure of perceived argument strength, we should see higher ratings from the strong condition than from the weak condition. Participants in the strong condition rated the arguments significantly higher (M = 2.78, SD = .81) than participants in the weak condition (M = 1.83, SD = .77), t = 4.54, df = 57, $p < .001$.

We also assessed the correlations among overall thought favorability, overall perceived argument strength, and attitude on the individual level. Although each participant used thought-listing and the perceived argument strength scale for different arguments, these arguments have been shown by previous research to be uniformly strong or weak. Thus, if the perceived argument strength scale was an adequate substitute for thought-listing, we should expect the overall strength index to correlate with attitude at roughly the same level as the overall favorability index. As it turned out, overall perceived argument strength correlated with attitude more strongly (r = .84, $p < .001$) than overall thought favorability (r = .50, $p < .001$). The strength and favorability indexes were also strongly correlated with one another (r = .51, $p < .001$).

Finally, we examined the relationship between thought favorability and perceived argument strength using the argument as the unit of analysis. By virtue of our design, each argument in this study was evaluated by two independent subgroups of participants using thought-listing and the perceived argument strength scale respectively (mean group n = 15.3). Using group means as data points, we were able to assess the correlation between thought favorability and perceived argument strength on the argument level. This correlation was substantial in size, r = .77, N = 16, p = .001.[6]

Discussion

Study 3 provided strong evidence for the convergent validity of the perceived argument strength scale. Arguments identified as strong and weak by previous research were successfully differentiated by both thought-listing and the perceived argument strength scale. On the individual level, perceived argument strength ratings correlated strongly with thought favorability and attitude. When individual arguments were used as the unit of analysis, perceived argument strength and thought favorability were also highly correlated. This latter finding was particularly impressive because individual differences were averaged out in that analysis. We have intended the perceived argument strength scale to serve as an alternative to thought-listing. Results from study 3 provided strong justification for that intention.

Conclusion

This paper presents a perceived argument strength scale as a complement to the conventional thought-listing measure of argument strength. The thought-listing measure,

[6]The perceived argument strength scale contained two items specifically assessing positive and negative thoughts in response to an argument. This scale component (indexed by their difference score) also correlated strongly with thought favorability, r = .64, N = 16, p = .008.

while well established, has some inherent limitations that can compromise its usefulness in some contexts, such as those involving sensitive topics and non-adult populations (Cacioppo, von Hippel, & Ernst, 1997; Shapiro, 1994; Stephenson & Palmgreen, 2001). The close-ended measurement scale we developed can overcome some of these limitations – it lessens the threat of sensitivity and is much less labor-intensive. At the same time it also extends the measurement of perceived argument strength beyond just thought favorability. The items on the scale are all theoretically motivated and represent the most current understanding of the factors underlying perceived argument strength. The empirical studies reported in this paper have provided clear and consistent evidence for the scale's reliability and validity. The wide applicability of the scale was also demonstrated in three different persuasive contexts, two of which involved non-student populations processing realistic health communication messages. Taken as a whole, the evidence we have garnered so far suggests that this scale can be a useful tool for persuasion researchers in a variety of contexts.

To be sure, the perceived argument strength scale is not without its own limitations. As a self-report measure, it is not immune to the influence of social desirability biases. As a multiple-item instrument, the scale could also fall victim to response set, particularly under circumstances of low motivation or fatigue. Finally, although the scale is an adequate substitute for thought-listing as a measure of perceived argument strength, it cannot replace thought-listing for its other important uses, e.g., as a way to assess the relative amount of central vs. peripheral processing generated by a message or eliciting the content of positive and negative thoughts. Researchers need to be mindful of these limitations when adopting this scale for use in their own research.

An important context in which we envision the perceived argument strength scale to be useful is research dealing with non-adult populations. For this reason, we have prioritized simplicity in scale construction. Our final scale has only 9 items, although we have considered a number of different theories in item development. The relative brevity of the scale has limited our ability to fully explore the dimensionality of the perceived argument strength construct. We specified and found support for a single-factor model in our CFAs. However, as our studies showed, some of the items on our scale seemed to share additional variance beyond that directly determined by the latent factor. Whether such shared variance represented sub-dimensions in the latent construct was left an open question in our research. Future research might consider expanding the current set of items to enable clearer understanding of scale dimensionality. This effort could also result in a more elaborate form of the current scale to be used with adult populations.

Future research should continue to test the validity of the perceived argument strength scale. Some new research pertinent to predictive validity has been and is being undertaken. For example, one study showed that perceived argument strength interacted with message sensation value (MSV) in predicting message effectiveness of anti-drug messages among adolescents (Kang, Cappella & Fishbein, 2006). In another study, anti-smoking ads with smoking cues created different levels of smoking urge depending on the perceived argument strength of the ad (Kang, Cappella, Strasser, & Lerman, 2009). In still another study, anti-smoking ads with strong arguments elicited stronger physiological reaction in adult smokers

than those with weaker arguments (Strasser et al., 2009). Research in progress has employed perceived argument strength as a way to differentiate anti-smoking arguments varying in objective (linguistic) features. Other studies are underway to evaluate the brain's responses to PSAs with strong and weak arguments. Results from this and other research will provide evidence to further assess the applicability and utility of this self-reported measure.

Acknowledgments

The research reported here was supported by grant from the National Cancer Institute to Robert Hornik (CECCR I: P50 CA095856-01) and from the National Institute on Drug Abuse to Martin Fishbein (NIDA: R01 DA12356-01). The views expressed are those of the authors alone.

Appendix A

Examples of arguments with highest and lowest perceived argument strength ratings in study 1

Ranking	Argument	Average Perceived Argument Strength Rating
1	Don't do things you wouldn't want your younger siblings to do because they will try to be like you. Drugs don't just affect you, they affect those around you too.	4.16
2	Doing drugs can cause lots of bad things to happen, like make you: depressed and anxious, lose your girl (or boyfriend), flunk out of school, act clumsy, smell bad, and do stupid things. It can also get you in trouble with the law.	4.12
3	Smoking marijuana costs you, your family and your home. Marijuana won't help you fit in and it won't help with your problems, it'll only add to them. Marijuana can lead you to harder drugs.	4.03
	…	…
67	Lots of different types of people say no to marijuana. So, when you are offered marijuana, say what's on your mind. Confident kids say no to marijuana.	3.49
68	Kids who are drug-free do fun things like sports and jumping out of an airplane. Lots of kids are drug-free and doing cool things.	3.47
69	If you smoke pot once, it probably won't affect you. But, if you keep smoking it, you will get dumber and dumber.	3.42

Appendix B

Examples of arguments with highest and lowest perceived argument strength ratings in study 2

Ranking	Argument	Average Perceived Argument Strength Rating
1	Every cigarette you smoke causes damage to your lungs. All you need are damaged cells to develop lung cancer. Quitting today may save your life.	4.01
2	When you die from smoking you leave behind many sad family members who miss you.	3.99
3	Smoking causes emphysema. Emphysema has no cure, only a long painful life filled with taking many medicines, and eventually death.	3.98

Ranking	Argument	Average Perceived Argument Strength Rating
	…	
97	People don't want to be around smokers. Women do not find it attractive. Quit to avoid public disapproval.	3.07
98	Smoking can interfere with your social life and cause embarrassing moments. If this hasn't happened yet, it will.	2.72
99	People have many varying preferences in types of people they enjoy spending time with. All people prefer non-smokers.	2.56

Appendix C

Perceived Argument Strength Scale (Final Form)

1. The statement is a reason for ___ that is believable.
2. The statement is a reason for ___ that is convincing.
3. The statement gives a reason for ___ that is important to me.
4. The statement helped me feel confident about how best to ___.
5. The statement would help my friends ___.
6. The statement put thoughts in my mind about wanting to___.
7. The statement put thoughts in my mind about not wanting to___.
8. Overall, how much do you agree or disagree with the statement?
9. Is the reason the statement gave for ____ a strong or weak reason?

Instructions: Fill in the blanks with the target behavior for the persuasive argument. Use a 5-point Likert scale (*strongly disagree* to *strongly agree*) to score items 1–8. Use a 5-point Likert type scale (*very weak* to *very strong*) to score item 9. Subtract item 7) from item 6) to create a single thought favorability item and then convert the new item to a 5-point scale by dividing it by 2 and then adding a constant of 3.

References

Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes. 1991; 50:179–211.

Andrews JC, Shimp TA. Effects of involvement, argument strength, and source characteristics on central and peripheral processing of advertising. Psychology and Marketing. 1990; 7:195–214.

Areni CS. The proposition-probability model of argument structure and message acceptance. Journal of Consumer Research. 2002; 29:168–187.

Areni CS, Lutz RJ. The role of argument quality in the elaboration likelihood model. Advances in Consumer Research. 1988; 15:197–203.

Brock TC. Communication discrepancy and intent to persuade as determinants of counterarguments. Journal of Experimental Social Psychology. 1967; 3:296–309.

Byrne, BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Mahwah, NJ: Lawrence Erlbaum; 2001.

Byrne BM. Factor analytical models: Viewing the structure of an assessment instrument from three perspectives. Journal of Personality Assessment. 2005; 85:17–32. [PubMed: 16083381]

Cacioppo, JT.; Sandman, CA. Psychophysiological functioning, cognitive responding, and attitudes. In: Petty, RE.; Ostrom, TM.; Brock, CT., editors. Cognitive responses in persuasion. Hillsdale, NJ: Lawrence Erlbaum; 1981. p. 81-103.

Cacioppo, JT.; Harkins, SG.; Petty, RE. The nature of attitudes and cognitive responses and their relationships to behavior. In: Petty, RE.; Ostrom, TM.; Brock, TC., editors. Cognitive responses in persuasion. Hillsdale, NJ: Lawrence Erlbaum; 1981. p. 32-54.

Cacioppo JT, Petty RE, Kao C. The efficient assessment of need for cognition. Journal of Personality Assessment. 1984; 482:306–307. [PubMed: 16367530]

Cacioppo JT, von Hippel W, Ernst JM. Mapping cognitive structures and processes through verbal content: The thought-listing technique. Journal of Consulting and Clinical Psychology. 1997; 65:928–940. [PubMed: 9420354]

Darke PR, Chaiken S. The pursuit of self-interest: Self-interest bias in attitude judgment and persuasion. Journal of Personality and Social Psychology. 2005; 89:864–883. [PubMed: 16393021]

DeBono KG, McDermott JB. Trait anxiety and persuasion: Individual differences in information processing strategies. Journal of Research in Personality. 1994; 28:395–407.

Dillard JP, Peck E. Affect and persuasion: Emotional responses to public service announcements. Communication Research. 2000; 27:461–495.

Dillard JP, Weber KM, Vail RG. The relationship between the perceived and actual effectiveness of persuasive messages: A meta-analysis with implications for formative campaign research. Journal of Communication. 2007; 57:613–631.

Eagly, AH.; Chaiken, S. The psychology of attitudes. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers; 1993.

Fishbein, M.; Ajzen, I. Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley; 1975.

Fishbein, M.; Cappella, JN.; Hornik, R.; Sayeed, S.; Yzer, MC.; Ahern, RK. The Role of Theory in Developing Effective Antidrug Public Service Announcements. In: Crano, WD.; Burgoon, M., editors. Mass media and drug prevention: Classic and contemporary theories and research. Mahwah, NJ: Lawrence Erlbaum; 2002. p. 89-117.

Fishbein M, Jamieson KH, Zimmer E, von Haeften I, Nabi R. Avoiding the boomerang: Testing the relative effectiveness of antidrug public service announcements before a national campaign. American Journal of Public Health. 2002; 92:238–245. [PubMed: 11818299]

Garlick R, Mongeau P. Majority/minority size and the evaluations of persuasive arguments. Communication Research Reports. 1992; 9:43–53.

Holbert, RL.; Stephenson, MT. Commentary on the uses and misuses of structural equation modeling in communication research. In: Hayes, AF.; Slater, MD.; Snyder, LB., editors. The Sage sourcebook of advanced data analysis methods for communication research. Thousand Oaks, CA: Sage; 2008. p. 185-218.

Hoyle RH, Stephenson MT, Palmgreen P, Lorch EP, Donohew RL. Reliability and validity of a brief measure of sensation seeking. Personality and Individual Differences. 2002; 32:401–414.

Igou ER, Bless H. Inferring the importance of arguments: Order effects and conversational rules. Journal of Experimental Social Psychology. 2003; 39:91–99.

Johnson BT, Eagly AH. Effects of involvement on persuasion: A meta-analysis. Psychological Bulletin. 1989; 106:290–314.

Johnson, BT.; Maio, GR.; Smith-McLallen, A. Communication and attitude change: Causes, processes, and effects. In: Albarracín, D.; Johnson, BT.; Zanna, MP., editors. Handbook of attitudes. Mahwah, NJ: Erlbaum; 2005.

Johnson, BT.; Smith-McLallen, A.; Killeya, LA.; Levin, KD. Truth or consequences: Overcoming resistance to persuasion with positive thinking. In: Knowles, ES.; Linn, J., editors. Resistance and persuasion. Mahwah, NJ: Lawrence Erlbaum; 2004. p. 215-233.

Jöreskog, KG.; Sörbom, D. LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software; 1993.

Kang Y, Cappella JN, Fishbein M. The attentional mechanism of message sensation value: Interaction between message sensation value and argument quality on message effectiveness. Communication Monograph. 2006; 73:351–378.

Kang Y, Cappella JN, Strasser A, Lerman C. The effect of smoking cues in antismoking advertisements on smoking urge and psychophysiological reactions. Nicotine and Tobacco Research. 2009; 11:254–261. [PubMed: 19251767]

Lavine H, Snyder M. Cognitive processing and the functional matching effect in persuasion: The mediating role of subjective perceptions of message quality. Journal of Experimental Social Psychology. 1996; 32:580–604. [PubMed: 8979935]

Limon MS, Boster FJ. The impact of varying argument quality and minority size on influencing the majority and perceptions of the minority. Communication Quarterly. 2001; 49:350–365.

Morley DD. Subjective message constructs: A theory of persuasion. Communication Monographs. 1987; 54:183–203.

Munch JM, Swasy JL. Rhetorical question, summarization frequency, and argument strength effects on recall. Journal of Consumer Research. 1988; 15:69–76.

O'Keefe DJ. Message properties, mediating states, and manipulation checks: Claims, evidence, and data analysis in experimental persuasive message effects research. Communication Theory. 2003; 13:251–274.

O'Keefe, DJ.; Jackson, S. Argument quality and persuasive effects: A review of current approaches. In: Jackson, S., editor. Argumentation and values: Proceedings of the ninth Alta conference on argumentation. Annandale, VA: Speech Communication Association; 1995. p. 88-92.

Palmgreen P, Donohew L, Lorch E, Rogus M, Helm D, Grant N. Sensation seeking, message sensation value, and drug use as mediators of PSA effectiveness. Health Communication. 1991; 3:217–227.

Palmgreen P, Stephenson MT, Everett MW, Baseheart JR, Francis R. Perceived message sensation value (PMSV) and the dimensions and validation of a PMSV scale. Health Communication. 2002; 14:403–428. [PubMed: 12375769]

Petty RE, Cacioppo JT. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. Journal of Personality and Social Psychology. 1979; 37:1915–1926.

Petty, RE.; Cacioppo, JT. Communication and persuasion: Central and peripheral routes to attitude change. New York: Springer-Verlag; 1986.

Petty, RE.; Wegener, DT. Attitude change: Multiple roles for persuasion variables. In: Gilbert, D.; Fiske, S.; Lindzey, G., editors. The Handbook of Social Psychology. 4th ed.. New York: McGraw-Hill; 1998. p. 323-390.

Petty RE, Harkins SG, Williams KD. The effects of group diffusion of cognitive effort on attitudes: An information processing view. Journal of Personality and Social Psychology. 1980; 38:81–92.

Shapiro, MA. Think-aloud and though-list procedures in investigating mental processes. In: Lang, A., editor. Measuring psychological responses to media messages. Hillsdale, NJ: Lawrence Erlbaum; 1994. p. 1-14.

Shimp TA. Attitude toward the ad as a mediator of consumer brand choice. Journal of Advertising. 1981; 10(2):9–15.

Stephenson MT. Examining adolescents' responses to antimarijuana PSAs. Human Communication Research. 2003; 29:343–369.

Stephenson MT, Palmgreen P. Sensation seeking, message sensation value, personal involvement, and processing of antidrug PSAs. Communication Monographs. 2001; 68:49–71.

Strasser AA, Cappella JN, Jepson C, Fishbein M, Tang KZ, Han E, Lerman C. Experimental evaluation of anti-tobacco PSAs: Effects of message content and format on physiological and behavioral outcomes. Nicotine and Tobacco Research. 2009; 11:293–302. [PubMed: 19246628]

Tao C, Bucy EP. Conceptualizing media stimuli in experimental research: Psychological versus attribute-based definitions. Human Communication Research. 2007; 33:397–426.

Toulmin, SE. The uses of argument. Cambridge: Cambridge University Press; 1958.

van Enschot-van Dijk, R.; Hustinx, L.; Hoeken, H. The concept of argument quality in the elaboration likelihood model: A normative and empirical approach to Petty and Cacioppo's "strong" and "weak" arguments. In: van Eemeren, FH.; Blair, JA.; Willard, CA.; Snoeck Henkemans, AF.,

editors. Anyone who has a view: Theoretical contributions to the study of argumentation. Amsterdam: Kluwer; 2003. p. 319-335.

Wood W, Quinn JM. Forewarned and forearmed? Two meta-analysis syntheses of forewarnings of influence appeals. Psychological Bulletin. 2003; 129:119–138. [PubMed: 12555796]

Ziegler R, Diehl M. The effect of multiple source information on message scrutiny: The case of source expertise and likeability. Swiss Journal of Psychology. 2001; 60:253–263.

Table 1

Descriptive statistics and inter-correlations of the perceived argument strength scale (studies 1 & 2)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Correlations					Study 2				
Believability (1)	-	.16	.86	.67	.53	.45	.43	.59	.60
*Novelty (2)	.36	-	.19	.18	.33	.21	−.11	.14	.14
Convincingness (3)	.17	−.23	-	.70	.58	.47	.44	.62	.65
Importance (4)	.68	.38	.15	-	.58	.49	.50	.65	.66
Confidence (5)	.67	.33	.14	.66	-	.59	.32	.58	.53
Friend (6)	.61	.35	.06[a]	.63	.71	-	.25	.50	.48
Thoughts (7)	.52	.06[a]	.27	.48	.55	.50	-	.47	.43
Agreement (8)	.57	.20	.18	.56	.55	.48	.46	-	.76
Reason (9)	.64	.28	.17	.61	.65	.56	.41	.66	-
					Study 1				
Mean (Study 1)	3.78	3.18	3.19	3.79	3.70	3.58	1.56	3.86	3.85
SD (Study 1)	1.08	1.25	1.29	1.15	1.13	1.20	1.81	1.20	1.13
Mean (Study2)	3.91	2.90	3.92	3.90	3.53	3.47	3.60	3.94	3.87
SD (Study 2)	1.05	1.30	1.02	1.04	1.08	1.07	.79	1.09	1.17

Note.

*Item deleted from final scale. For study 1, $N = 322$; Correlations presented below diagonal;

[a]Not significant at the .05 level; All other correlations significant at the .01 level. For study 2, $N = 300$; Correlation presented above diagonal: All correlations significant at .05 level.

Table 2

Factor loadings for the perceived argument strength scale

	Loading	
Item	**Study 1**	**Study 2**
Believability	.83	.77
Importance	.81	.85
Convincingness	-	.82
Friends	.75	.58
Confidence	.82	.69
Thoughts	.62	.55
Agreement	.68	.78
Reason	.76	.78