


ORIGINAL ARTICLE

Persuasive Message Pretesting Using Non-Behavioral Outcomes: Differences in Attitudinal and Intention Effects as Diagnostic of Differences in Behavioral Effects

Daniel J. O'Keefe *

Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA

Persuasive message designers would like to be able to pretest messages to see which will be more effective in influencing behavioral outcomes, but pretesting using behavioral measures is commonly not practical. Examination of within-study effect size comparisons from 317 studies of 22 message variations suggests that persuasive messages' relative effectiveness is strikingly similar across attitudinal, intention, and behavioral outcomes—with messages' relative persuasiveness with respect to intention outcomes especially indicative of relative persuasiveness with respect to behavioral outcomes. Intention measures thus provide a convenient and accurate means of persuasive message pretesting.

Keywords: Formative research, Message pretesting, Persuasion

doi: 10.1093/joc/jqab017

For persuasive message design, formative research—research designed to shape subsequent campaign or intervention messages—can be valuable for many reasons. It can help identify how to reach target audiences, indicate what different messages might be needed for different audiences, and so forth (for general discussions, see [Atkin & Freimuth, 2013](#); [Rossmann, 2014](#)).

One specific purpose of such formative research can be to pretest alternative possible persuasive messages to identify those that are likely to be relatively more effective. But persuasive messages commonly seek behavioral effects, and it is often not practical to pretest messages by comparing their effects on behavioral outcomes. For example, a public health agency seeking to design messages aimed at

*Corresponding author: Daniel J. O'Keefe; e-mail: d-okeefe@northwestern.edu

encouraging flu shots is unlikely to be able to pretest candidate messages by seeing which ones are actually more effective at producing behavioral uptake.

The primary purpose of this article is to provide evidence concerning the possibility of using *non-behavioral* outcomes as a basis for identifying messages that will be more persuasive with respect to behavioral outcomes. The focus is specifically on attitude outcomes (general evaluations, e.g., consumer product liking) and intention outcomes (behavioral intentions, e.g., purchase intentions) as potential pretesting devices. The next section contextualizes this project against the backdrop of other message-pretesting methods.

Persuasive message pretesting methods

Because pretesting persuasive messages by obtaining behavioral-outcome data is commonly not practical, three other general kinds of pretesting methods have been suggested.

Checklist methods

One potential means of advance identification of relatively more persuasive messages is through appropriate content-analytic methods. In this approach, trained coders use a checklist to assess message properties thought to be relevant to persuasiveness. The supposition is that if message A ranks higher than message B with respect to such properties, then message A will likely be more persuasive than message B when behavioral outcomes are assessed. For example, for health education materials, [Paul, Redman, & Sanson-Fisher \(1997\)](#) developed a 62-item “checklist of the content and design characteristics of effective print materials . . . based on a critical review of the relevant research literature” (p. 153). [Armstrong \(2010\)](#) identified 194 “evidence-based principles” for persuasive advertising, with these organized into a coding scheme yielding a Persuasive Principles Index score for advertisements. (For examples of similar methods, see [Baumel, Faber, Mathur, Kane, & Muench, 2017](#); [Cole, Keller, Reynolds, Schaur, & Krause, 2016](#).)

To underwrite the message-pretesting use of such a method, the kind of research evidence needed would be evidence that those messages identified as relatively more persuasive using such a checklist are in fact ones relatively more effective when assessed using behavioral-outcome data. Unfortunately, such evidence is scant (see, e.g., [Hartnett, Greenacre, Kennedy, & Sharp, 2020](#); [Sharp & Hartnett, 2016](#)).

Perceived persuasiveness

A second pretesting method has pretest participants from the target audience rate or rank messages on perceived persuasiveness (perceived message effectiveness, PME). The supposition is that if message A is rated higher in perceived effectiveness than message B, then message A will likely be more persuasive than message B when behavioral outcomes (i.e., actual effectiveness outcomes) are assessed. For example, [Vaillancourt et al. \(2019\)](#) had participants rate messages on scales with

end-anchors such as *persuasive/not persuasive*, *effective/ineffective*, and *convincing/not convincing*; Drovandi, Teague, Glass, & Malau-Aduli (2019) had smokers rank-order cigarette warnings in terms of perceived effectiveness. (For some general discussions of such methods, see Baig et al., 2019; Dillard, Weber, & Vail, 2007.)

To underwrite the message-pretesting use of such a method, the kind of research evidence needed would be evidence that those messages identified as relatively more persuasive using such perceived-persuasiveness measures are in fact ones relatively more effective when assessed using behavioral-outcome data. Unfortunately, evidence underwriting the predictive diagnosticity of such methods is not strong (O'Keefe, 2018; for some discussion, see Cappella, 2018; Noar, Barker, & Yzer, 2018; O'Keefe, 2020).

Non-behavioral outcomes

A third possibility is to assess non-behavioral persuasive outcomes—specifically, attitude and intention outcomes—as pretest measures. The supposition is that relative message persuasiveness will be identical across attitude, intention, and behavioral outcomes, and hence if message A is more persuasive than message B when attitude or intention outcomes are assessed, then message A will likely be more persuasive than message B when behavioral outcomes are assessed.

To underwrite the message-pretesting use of such a method, the kind of research evidence needed would be evidence that those messages identified as relatively more persuasive using such non-behavioral outcomes are in fact ones relatively more effective when assessed using behavioral-outcome data. Such evidence is not in hand. Two other kinds of evidence might seem to be relevant here—but upon closer inspection, neither in fact underwrites the use of non-behavioral measures in message pretesting.

Positive attitude-behavior and intention-behavior correlations. One kind of evidence that would appear to support the use of non-behavioral measures in message pretesting is the existence of generally positive correlations between non-behavioral and behavioral measures. Individuals' attitudes and behaviors are commonly positively (if imperfectly) correlated (e.g., Glasman & Albarracín, 2006), as are individuals' intentions and behaviors (e.g., Sheeran, 2002). Thus, one might reason, attitudes and intentions will be good indicators of messages' relative behavioral effectiveness.

But this is not sound reasoning. (The following discussion is phrased in terms of the relationship of intentions and behavior, but the same analysis applies to the relationship of attitudes and behaviors.) The correlation between individuals' intention scores and behavior scores speaks to the question of whether one can predict individuals' behavior scores on the basis of their intention scores—more carefully, whether one can predict individuals' relative standing on behavioral outcomes on the basis of individuals' relative standing on intention outcomes. However, in message pretesting, one is trying to predict messages' relative standing on behavioral

outcomes, not individuals' relative standing; the interest is with messages' means, not individuals' scores.

The distinction is important. Even if individuals' scores on intention and behavior measures are positively correlated, that does not imply that the relative standing of messages' means on intention measures will match the relative standing of messages' means on behavior measures. This point can be illustrated with a hypothetical dataset based on three messages ($N = 12$): For message A, participants have the following (intention, behavior) pairs of scores: (52,48), (62,58), (72,68), and (82,78). For message B, the scores are (50,50), (60,60), (70,70), and (80,80). For message C, the scores are (78,82), (68,72), (58,62), and (48,52). The correlation between individuals' intention scores and behavior scores is strongly positive, $+0.96$. The ranking of messages on intention is A, B, C (means of 67.0, 65.0, and 63.0, respectively); the ranking of messages on behavior is C, B, A (means of 67.0, 65.0, and 63.0, respectively).¹

As this illustrates, it is possible for individuals' intention and behavior scores to be strongly positively correlated and yet for messages' relative standing on intention outcomes to be the opposite of those messages' relative standing on behavioral outcomes (even in the same dataset). Plainly, knowing that individuals' intention and behavior scores are generally positively correlated provides no guarantee that messages' relative standing on the two measures will generally match up. How messages' means sort themselves out on the two measures is something conceptually different from how individuals' scores on the two measures are related.²

Thus the existence of positive correlations between individuals' intention scores and behavior scores cannot possibly show that messages' relative standing on intention outcomes is a good indicator of messages' relative standing on behavioral outcomes. It's one thing to say "the relative standing of individuals on intention positively covaries with the relative standing of individuals on behavior" and something quite different to say "the relative standing of messages on intention positively covaries with the relative standing of messages on behavior." The first could be true while the second is false.

Similar mean effect sizes across outcomes. A second form of apparent evidence supporting the pretesting use of non-behavioral outcomes was offered by O'Keefe (2013), in the form of comparisons of meta-analytic averages of effect sizes for attitude, intention, and behavioral outcomes in studies of message-variable persuasive effects. Over a large number of different message variables, those meta-analytic averages were similar across the three different outcome variables. The conclusion was that for "practical persuasive-message pretesting" purposes, "campaign planners need not collect attitudinal, intentional, and behavioral outcome data. Any one of these three kinds of assessment will suffice to identify the most persuasive message" (p. 237).

But this is not sound reasoning, because the reported meta-analytic averages could potentially mask consequential within-study differences between effect sizes

for different outcomes. For example, imagine that some studies had a large positive effect size for intention (indicating message A was more effective than message B with respect to the intention outcome) and a small negative effect size for behavior (indicating message B was more effective than message A with respect to the behavioral outcome), while other studies had a small negative effect size for intention and a large positive effect size for behavior. When effect sizes were averaged across such studies, the mean intention effect size and the mean behavior effect size would look similar—even though in every individual study, the intention outcome would have misidentified which of the two messages was more effective in behavioral terms. To see whether attitude, intention, and behavioral outcomes are equivalent for purposes of pretesting messages, meta-analytic averages are not informative; examination of within-study comparisons, however, can provide relevant evidence.

Summary. In short, the use of non-behavioral measures in message pretesting is not justified—cannot possibly be justified—either by positive correlations between individuals' scores on behavioral and non-behavioral measures or by the similarity of meta-analytic mean effect sizes across behavioral and non-behavioral outcomes. Other evidence is needed.

The present project

Expressed narrowly, the present project is aimed at providing evidence bearing on the suggestion that non-behavioral persuasive outcomes—attitude and intention outcomes specifically—provide a sound basis for diagnosing differences between messages in behavioral outcomes. As just intimated, the evidence to be examined consists of within-study comparisons of effect sizes for behavioral and non-behavioral outcomes. If, within individual studies, messages' relative standing on non-behavioral outcomes is usually indicative of those messages' relative standing on behavioral outcomes, then formative researchers would have a straightforward message pretesting protocol: to see which message will be more persuasive with respect to behavioral outcomes, see which message is more persuasive with respect to non-behavioral (attitude or intention) outcomes.

Expressed more broadly, the question addressed here is whether messages' relative standing with respect to one persuasive outcome is indicative of their relative standing with respect to other outcomes. Hence, in addition to examining the relationship of non-behavioral (attitude and intention) effect sizes to behavioral effect sizes—the comparisons of interest for message pretesting questions—the relationship between attitude effect sizes and intention effect sizes is also considered.

Method

As an overview: Existing meta-analyses concerning the persuasive effects of message variations were used to identify primary research studies that had effect sizes (ESs)

for at least two outcomes of interest (attitude, intention, and behavior). Each such study yielded within-study comparisons of the effect sizes associated with the different outcomes. Comparisons were analyzed so as to assess whether messages' relative standing on one outcome was indicative of messages' relative standing on other outcomes.

Locating relevant studies

The data of interest are comparisons between message-variation persuasion ESs for different outcomes within a given study. In principle, any individual study (of the effects of a persuasive message variation) that contains assessments of at least two of the outcomes of interest would provide relevant evidence. Retrieving all such studies is not practical. However, many such studies can be efficiently located through examination of existing relevant meta-analyses. Hence this report is based on culling relevant within-study comparisons from meta-analyses of the persuasive effects of message variations.

The primary-research studies of interest were studies that compared the persuasiveness of two message forms (e.g., gain-framed vs. loss-framed) by assessing at least two of the three outcomes of interest (attitude, intention, behavior). Thus to be included, a meta-analysis had to have analyzed such studies. Meta-analyses of potential interest were identified by searches of the Web of Science, Medline, PsycINFO, PsycEXTRA, and ProQuest Dissertations and Theses databases, combining *meta-analysis* with such terms as *persuasion*, *message*, and *attitude*, through July 2020. Additional candidates were located through examination of reviews of relevant meta-analyses (e.g., [Eisend & Tarrahi, 2016](#); [Rains, Levine, & Weber, 2018](#)).

A meta-analysis was excluded if it examined the effects of non-message variations such as psychological states (e.g., [van Laer, de Ruyter, Visconti, & Wetzels, 2014](#); [Walter, Tukachinsky, Pelled, & Nabi, 2019](#); [Xu & Guo, 2018](#)) or the presence (vs. absence) of some preceding message (e.g., [Dillard, Hunter, & Burgoon, 1984](#); [Feeley, Anker, & Aloe, 2012](#)), compared a message form against a no-message control (e.g., [Braddock & Dillard, 2016](#); [Chan, Jones, Jamieson, & Albarracín, 2017](#)), included non-experimental studies or studies lacking a control condition (e.g., [Shen & Han, 2014](#); [Sun, Miu, Wong, Tucker, & Wong, 2018](#)), examined outcomes other than attitude, intention, or behavior (e.g., [Portnoy, Ferrer, Bergman, & Klein, 2014](#)), was restricted to studies examining only one kind of outcome of interest (e.g., [Burger & Caputo, 2015](#); [Carpenter, 2013](#); [Hamilton & Hunter, 1998](#); [Lee, Moon, & Feeley, 2016](#)), reported results only for composite persuasion outcomes (e.g., [Brugman, Burgers, & Vis, 2019](#); [Cheng & Yan, 2020](#); [Seo & Kim, 2018](#)), did not contain any studies yielding within-study comparisons (e.g., [Andrews, Carpenter, Shaw, & Boster, 2008](#); [Lee & Feeley, 2017](#)), analyzed multiple ESs for the same outcome based on a given message pair or set of participants (e.g., [Eisend, 2006](#)), or if necessary information about individual studies (ES, sample size, outcome variable) was not available even after correspondence with authors (e.g., [Edison, 2008](#); [Fischer &](#)

Huber, 2015; Freling, 2017; Freling, Vincent, & Henard, 2014; Grewal, Kavanoor, Fern, Costley, & Barnes, 1997; Lunt, 2016; Wirtz, Sparks, & Zimbres, 2018).

For most message variables, only one appropriate meta-analysis was available. In cases where more than one meta-analysis was available, the one with the largest number of contributing studies was included. If more than one meta-analysis of the same variable were to have been included, then individual studies would have been counted multiple times because the same study's results would have been included in multiple meta-analyses.

Data were obtained for a total of 22 message variables: gain-loss framing (data from the meta-analysis of O'Keefe & Jensen, 2006; 65 studies contributed comparisons), depicted threat severity (de Hoog, Stroebe, & de Wit, 2007; 41 studies), cultural value adaptation (Hornikx & O'Keefe, 2009; 33 studies), threat appeal strength (Witte & Allen, 2000; 28 studies), sidedness (O'Keefe, 1999, 22 studies), depicted threat vulnerability (De Hoog et al., 2007; 20 studies), humor (Walter, Cody, Xu, & Murphy, 2018; 20 studies), vividness (Blondé & Girandola 2016; 17 studies), depicted response efficacy (Witte & Allen, 2000; 12 studies), narrative vs. non-narrative messages (Shen, Sheer, & Li, 2015; 10 studies), depicted self-efficacy (Witte & Allen, 2000; 10 studies), sexual ad content (Lull & Bushman, 2015; 7 studies), recommendation specificity (O'Keefe, 2002; 6 studies), conclusion explicitness (O'Keefe, 2002; 5 studies), metaphor (Van Stee, 2018; 4 studies), pictorial vs. text warnings (Noar et al., 2016; 4 studies), negative political advertising (Lau, Sigelman, & Rovner, 2007; 3 studies), argument completeness (O'Keefe, 1998; 3 studies), guilt (O'Keefe, 2000; 3 studies), exemplars (Bigman, Bigsby, & Martinez Gonzalez, 2018; 2 studies), information-source citation (O'Keefe, 1998; 1 study), and narrative vs. statistical evidence (Zebregs, van den Putte, Neijens, & de Graaf, 2015; 1 study).

Extracting relevant comparisons

Each included meta-analysis yielded a set of effect sizes and associated sample sizes. The ESs were not adjusted, deleted, recomputed, or otherwise altered, except for converting all ESs to correlations (r s) for analysis.³

The unit of analysis was the comparison, within a given study, between the ES for one outcome and the ES for another outcome. A total of 317 studies reported results for at least two of the three outcomes of interest (attitude, intention, behavior) and so provided a basis for comparing effect sizes for different outcomes. Most studies (275, 87%) reported results for only two such outcomes and hence contributed only one comparison; 42 studies reported results for all three outcomes and hence contributed three comparisons. Thus a total of 401 comparisons were available.

Analyzing the comparisons

For each such comparison, several properties were of interest. One was whether the two effect sizes had the same direction of effect, that is, whether the message that

appeared more persuasive on one outcome was also the message that appeared more persuasive on the other outcome; comparisons were coded as having the same direction of effect if the signs of the two ESs were the same or if either of the ESs was zero. A second was the size of the difference between the two ESs, described by Cohen's q (the difference between z -transformed rs).⁴ A third was whether the two ESs were statistically significantly different, that is, whether q was statistically significant (two-tailed test, .05 alpha).

No one of these properties is necessarily informative with respect to assessment of message pretesting procedures. To concretize this point, consider the comparison between an intention ES and a behavior ES. Just because two such ESs are statistically significantly different in a given study does not necessarily suggest a weakness in using intention measures for message pretesting. Suppose that the intention ES (expressed as r) was $+0.15$, the behavior ES was $+0.25$, and these were significantly different. A pretest using intention outcomes would nevertheless have correctly identified the message that would be more persuasive with behavioral outcomes.

Similarly, just because two such ESs have different directions of effect does not necessarily suggest a weakness in using intention measures for message pretesting. Suppose that the intention ES was $+0.01$, the behavior ES was -0.01 , and these were not significantly different. The two ESs have different signs, but plainly the two messages did not differ much in persuasiveness and so relying on an intention assessment in pretesting would likely not make for a dramatically bad message choice.

Hence, in addition to examining those three properties individually, the number of *significant disjunctures* between ESs was also tallied. A comparison was coded as a significant disjuncture under three conditions: (1) for the comparison of any two ESs, if the two ESs had different signs and were significantly different; (2) for the comparison of an attitude ES and an intention ES, if one of the ESs was zero, the other ES was non-zero, and the ESs were significantly different; (3) for the comparison of a non-behavior (i.e., attitude or intention) ES and a behavior ES, if the non-behavior ES was zero, the behavior ES was non-zero, and the ESs were significantly different. Each of these three conditions represents a circumstance in which relying on the relative standing of messages on one outcome as a guide to relative standing on the other outcome would be a consequential error.

One additional useful way of describing the relationships of interest is provided by rank-order correlations. For a comparison in an individual study, the rank-order correlation (or, for that matter, the Pearson correlation) is $+1.00$ if the two messages' relative standing is identical on the two outcomes, -1.00 if the messages' relative standing is reversed, and $.00$ if either of the two ESs is zero (i.e., if the two messages are tied on an outcome).⁵ Thus across studies, a random-effects meta-analytic mean correlation (Borenstein & Rothstein, 2005) can provide an indication of the degree to which standing on one outcome is diagnostic of standing on the other.⁶

For analytic purposes data were combined across message variables. The present interest is not with relationships between different outcome measures for this or

that message variable in particular, but rather with whether in general these measures are equivalent indicators. And for any given variable, sometimes relatively few studies obtained data permitting comparison of effect sizes for different outcomes. For example, of the 50 studies in Lau et al.'s (2007) meta-analysis of negative political advertising, only three studies reported data on two outcomes of interest. Separate analysis of those cases would be uninformative.

Results

Table 1 provides results for each of the three kinds of comparisons of interest—between attitude ESs and intention ESs, between attitude ESs and behavior ESs, and between intention ESs and behavior ESs.⁷ Collapsed across the different kinds, a total of 401 ES comparisons were available. Of those, 346 (.863; 95% CI [.826, .893]) had the same direction of effect; 360 (.898; 95% CI [.864, .924]) were not statistically

Table 1 Comparison of Effect Sizes

	Attitude ES and intention ES	Attitude ES and behavior ES	Intention ES and behavior ES
Number of comparisons	232	60	109
Number with different directions of effect	37	11	7
Proportion	.159	.183	.064
95% CI	[.118, .212]	[.104, .301]	[.029, .129]
Number with significantly different ESs	24	7	10
Proportion	.103	.117	.092
95% CI	[.070, .150]	[.055, .225]	[.049, .162]
Number with significantly different ESs and different directions of effect	9	6	1
Proportion	.039	.100	.009
95% CI	[.019, .073]	[.043, .205]	[.000, .055]
Number of significant disjunctures	13	6	4
Proportion	.056	.100	.037
95% CI	[.032, .094]	[.043, .205]	[.011, .094]
Mean rank-order correlation	.996	.963	.994
95% CI	[.989, .998]	[.733, .995]	[.982, .998]
Mean q^a	.002	-.010	.006
Mean absolute q	.123	.127	.109

^a q (Cohen's q) is the difference between two z -transformed r s.

significantly different. In 16 cases (.040; 95% CI [.024, .064]) cases, the two ESs had different directions of effect and were statistically significantly different. In 23 cases (.057; 95% CI [.038, .085]) there was a significant disjuncture between the two ESs. The meta-analytic mean rank-order correlation was .994 (95% CI [.987, .997]). The mean difference between the ESs, expressed as Cohen's q , was .001; the mean of the absolute values of q was .120.

The proportion of comparisons in which the two ESs had different directions of effect was .159 for attitude-intention comparisons, .183 for attitude-behavior comparisons, and .064 for intention-behavior comparisons. The attitude-intention proportion was not significantly different (two-tailed test, .05 alpha) from the attitude-behavior proportion ($z = .44$, $p = .660$). The intention-behavior proportion was significantly smaller than both the attitude-intention proportion ($z = 2.45$, $p = .014$) and the attitude-behavior proportion ($z = 2.40$, $p = .016$).

The proportion of comparisons in which the two ESs differed in direction and were statistically significantly different from each other was .039 for attitude-intention comparisons, .100 for attitude-behavior comparisons, and .009 for intention-behavior comparisons. The attitude-intention proportion was not significantly different (two-tailed test, .05 alpha) from either the attitude-behavior proportion ($z = 1.91$, $p = .056$) or the intention-behavior proportion ($z = 1.51$, $p = .131$). The intention-behavior proportion was significantly smaller than the attitude-behavior proportion ($z = 2.84$, $p = .005$).

The proportion of comparisons in which there was a significant disjuncture between the two ESs was .056 for attitude-intention comparisons, .100 for attitude-behavior comparisons, and .037 for intention-behavior comparisons. No two of these proportions were significantly different (two-tailed test, .05 alpha). The attitude-intention proportion was not significantly different from either the attitude-behavior proportion ($z = 1.23$, $p = .219$) or the intention-behavior proportion ($z = .77$, $p = .441$); the intention-behavior proportion was not significantly different from the attitude-behavior proportion ($z = 1.67$, $p = .095$).

Discussion

Message pretesting

Similarity of effect size direction

The primary purpose of the present report is to provide evidence bearing on the suggestion that non-behavioral persuasive outcomes—attitude and intention outcomes specifically—can be a sound basis for diagnosing differences between messages in behavioral outcomes. These data indicate that these non-behavioral measures can indeed be effective ways of pretesting the relative behavioral impact of messages. When two messages have been compared for their persuasiveness in influencing both behavioral outcomes and these non-behavioral outcomes, the

relative standing of the two messages does not vary much between the two kinds of outcome.

Specifically, the direction of difference between messages on attitude outcomes commonly matches the direction of difference on behavioral outcomes (82% of comparisons). Only rarely do attitude and behavioral assessments differ significantly and have different directions of effect (10% of comparisons) or otherwise exhibit significant disjunctures (10% of comparisons). The meta-analytic mean rank-order correlation was .96.

Similarly, the direction of difference between messages on intention outcomes quite commonly matches the direction of difference on behavioral outcomes (94% of comparisons). Only exceptionally rarely do intention and behavioral assessments differ significantly and have different directions of effect (1% of comparisons) or otherwise exhibit significant disjunctures (4% of comparisons). The meta-analytic mean rank-order correlation was .99.

The straightforward implication is this: The message choices that would have been made in formative research using attitude or intention outcomes are generally the same as those that would have been made if behavioral outcomes had been examined. That is, the message that appears more effective when examining non-behavioral outcomes is generally the same as the message that is more effective when examining behavioral outcomes.

Similarity of effect size magnitudes

In addition to seeing whether the *direction* of ESs (the direction of difference in effectiveness between messages) is similar in behavioral and non-behavioral outcomes, the present data can also shed light on the degree to which the *magnitude* of ESs (the size of the difference in effectiveness between messages) is similar in behavioral and non-behavioral outcomes. Specifically, the present data can address two questions concerning ES magnitudes.

One question is whether the ES for one kind of outcome is generally larger or smaller than another. These data indicate that there are no systematic differences in ES magnitudes between behavioral and non-behavioral ESs; it's not the case that (say) behavioral ESs are generally smaller than non-behavioral ESs. On the contrary: the mean value of q was $-.01$ for attitude-behavior ESs and was $.01$ for intention-behavior ESs. That is, averaged across comparisons, the difference between attitude and behavior ESs or between intention and behavior ESs is functionally zero.

A second question concerns the degree to which behavioral-outcome ESs and non-behavioral-outcome ESs differ in magnitude in individual applications. The mean difference between the absolute values of q was $.13$ for comparisons between attitude and behavior ESs and was $.11$ for comparisons between intention and behavior ESs; these represent an average difference of roughly $.12$ between two correlation coefficients of the magnitudes seen for persuasion message-variation effects.⁸ The implication is that the magnitude of a non-behavioral ES might provide at least a general guide to the expected magnitude of the behavioral ES. For example, if in

pretesting using intention outcomes an ES (r) of .20 is observed, one might plausibly expect a behavior ES roughly between .08 and .32 (i.e., $.20 \pm .12$).

In short, behavioral and non-behavioral ESs characteristically do not differ dramatically—and behavioral ESs are neither generally larger nor generally smaller than non-behavioral ESs.

Comparing attitude and intention as pretest measures

As between attitude and intention measures as possible ways of pretesting relative message persuasiveness, these data give some reason to favor intention measures. The direction of behavioral-outcome effects (the direction of difference between messages) was significantly more likely to match the direction of intention outcomes (94% of comparisons matched) than the direction of attitude outcomes (82% of comparisons matched). And the proportion of comparisons in which the direction of effect did not match and the two ESs were statistically significantly different was significantly smaller for intention outcomes (1%) than for attitude outcomes (10%). There was no significant difference in the proportion of significant disjunctures observed with intention and attitude measures or in the meta-analytic mean rank-order correlations, though in each case the direction of difference also favored intention measures.

Moderating factors?

Identifying factors that moderate the diagnosticity of non-behavioral measures as message pretesting devices could be useful, because then better guidance could be given to formative researchers about the use of such message pretests. However, it should immediately be acknowledged that these data show considerable consistency across studies: it is exceptionally common that messages' relative standing on non-behavioral measures matches their relative standing on behavioral measures. There might well be moderators of the diagnosticity of non-behavioral measures as message pretesting devices, but these will be difficult to ferret out.

In looking for variables that might be moderators here, one might naturally initially be attracted to factors that moderate the correlation between individuals' intention and behavior scores. But that is a misleading path. In message pretesting one seeks to predict not individuals' scores on behavioral outcomes, but messages' scores; it's the relative standing of messages on the behavioral outcome that's of interest, not the relative standing of individuals. As discussed above, just because individuals' scores on intention and behavior are strongly positively correlated does not necessarily imply that messages' means will covary similarly. And because these are different relationships (the relationship between individuals' intention and behavior scores on the one hand, and the relationship between messages' mean scores on intention and behavior on the other), a variable that moderates one relationship will not necessarily moderate the other relationship.

So, for example: As the time interval between the (pretest) intention assessment and the behavioral assessment increases, the correlation between individuals'

intention and behavior scores will likely weaken (see, e.g., Sheeran & Orbell, 1998, pp. 234–235). Some people with initially positive intentions come, over time, to have negative intentions and hence at a delayed assessment have negative behavior scores; and some people who had negative intentions initially come, over time, to have positive intentions and hence at a delayed assessment have positive behavior scores. This reduces the predictability of individuals' relative standing on the behavioral outcome.

But in a message-pretesting setting, such changes over time will affect all participants, regardless of which message they saw in the pretest. Thus there is no reason to suppose that messages' means will be differentially affected by these processes, that is, no reason to suppose that messages' relative standing will be affected. With longer time intervals individuals' relative standing can be affected and hence the correlation between individuals' scores may weaken, but the relative standing of messages can be unaffected. The larger point here is: Just because a given variable moderates the relationship of individuals' intention and behavior scores does not necessarily imply that it will moderate the relationship of messages' means on intention and behavior. Identifying factors that moderate that latter relationship will require new avenues of approach.

Summary: non-behavioral measures as pretesting methods

These data clearly point to the usefulness of non-behavioral outcomes in persuasive message pretesting. Messages' relative standing on attitude and (especially) intention outcomes closely matches their relative standing on behavioral outcomes. Notably, both attitude assessments and intention assessments appear to be much better indicators of messages' relative standing on behavioral outcomes (meta-analytic mean rank-order correlations of .96 and .99, respectively, across 60 and 109 studies) compared to perceived-persuasiveness (PME) assessments as indicators of messages' relative standing on measures of actual effectiveness (AME); the mean PME-AME rank-order correlation was reported as $-.05$ across 35 studies (O'Keefe, 2018, p. 133).

At present, when a study reports differences between messages in persuasiveness based on non-behavioral outcomes, it is common for researchers to offer a disclaimer that their results do not speak to differences in behavioral effects.⁹ For example, using intention outcome measures, Ferrer, Klein, Zajac, Land, & Ling (2012, p. 459) found that affective boosters enhanced the persuasiveness of gain-framed appeals, but suggested that "research is necessary to determine whether affective boosters increase the persuasiveness of gain-framed messages for actual screening behavior."

But given the present results, such disclaimers are unnecessary—indeed, inappropriate. If two messages differ in persuasiveness with respect to attitude or intention outcomes, they are very likely to differ similarly with respect to behavioral outcomes. Thus these results suggest that non-behavioral outcome measures—

especially intention measures—provide formative researchers with a convenient and accurate tool for identifying relatively more effective persuasive messages.

However, this conclusion needs to be tempered in several ways. First, these non-behavioral measures are unlikely to be well-adapted to pretesting situations in which a very large number of messages are to be pretested and so a design is contemplated in which each pretest participant responds to multiple candidate messages (see Cappella & Kim, 2017). It is not clear that asking the same intention question after each message will yield results that could confidently be relied upon. It remains to be seen whether any pretesting method can be diagnostic in such a pretesting circumstance.

Second, if the messages being pretested differ only slightly on the non-behavioral pretest measure, the range of plausible differences in ES magnitudes between behavioral and non-behavioral outcomes suggests some caution. For example, if a non-behavioral pretest favors message A over message B but the ES corresponds to an r of .08, formative researchers should be prepared to see behavioral ESs in the range of $-.04$ to $.20$ (i.e., $.08 \pm .12$)—that is, including outcomes in which message B is superior in behavioral assessments. Such divergences appear to be rare (see Table 1), but formative researchers should be alert to the possibility.

Third, as with pretesting methods generally, if the candidate messages do not differ much in effectiveness, large pretest samples will be needed to reliably detect such differences. For example, to have 80% power (two-tailed test, .05 alpha) for detecting a population effect corresponding to $r = .10$, nearly 800 pretest participants will be required.

One final general caveat arises from the present analysis' acceptance of the underlying meta-analytic data as given. Those underlying data almost certainly contain error. In the present project, the decisions of the primary researchers (about the realization of experimental contrasts, outcome measurement, etc.) and the decisions of the meta-analysts (about inclusion criteria, how to compute effect sizes, etc.) are all embedded in the data set under analysis. And so all their errors, oversights, poor decisions, mistakes, etc., are inevitably part of these data.

However, there seems no reason to expect systematic error, that is, error that would bias the present results in some direction. And perhaps the diversity of the evidentiary base offers some reassurance on this point. Many different message variations, many different advocacy topics, many different kinds of assessments, many different primary researchers, many different meta-analysts are represented in these data. There surely is error in these data, but it seems likely to be haphazard.

Overall effects

The larger question addressed by these data is whether messages' relative standing with respect to one persuasive outcome is indicative of their relative standing with respect to other outcomes. As will be apparent, in general, attitude, intention, and behavior ESs do not systematically differ in the direction of effect. If message A is

more persuasive than message B on any one of these outcomes, it is likely to be more persuasive on the other two outcomes as well.

Moreover, in general, attitude, intention, and behavior ESs do not systematically differ in magnitude. As discussed above, the mean differences (expressed as Cohen's q) between attitude and behavior ESs (mean $q = -.01$) and between intention and behavior ESs (mean $q = .01$) were essentially zero. A similar result obtains for the mean difference between attitude and intention ESs (mean $q = .00$). The mean absolute values of q for those three comparisons were, respectively, .13, .11, and .12. So one might expect that the ES obtained for one of these three outcomes (expressed as r) might vary by roughly $\pm .12$ from either of the other two.

These results thus extend O'Keefe's (2013) conclusion that "the relative persuasiveness of *message types* will be substantively identical if compared using attitudinal, intention, or behavioral outcomes" (p. 244, emphasis added): the present results indicate that the relative persuasiveness of *individual messages* will also be substantively identical if compared using attitudinal, intention, or behavioral outcomes.

Relative vs. absolute persuasiveness

So: Where claims about relative message persuasiveness are concerned, attitudinal, intention, and behavioral outcomes are interchangeable; conclusions about relative persuasiveness will generally be the same no matter which of these outcomes is examined. This result has been framed here by focusing on message pretesting, because questions of relative message persuasiveness naturally arise in that enterprise.

But claims about relative message persuasiveness can appear in contexts other than formative research. In particular, such claims are common in persuasion research aimed at testing theory-based hypotheses, because those hypotheses characteristically involve claims about the relative persuasiveness of different message forms.

Some such hypotheses predict simple (main-effect) differences in persuasiveness between two message forms. For example, Meyerowitz and Chaiken (1987) used phenomena such as negativity bias and loss aversion to underwrite the hypothesis that loss-framed appeals will generally be more persuasive than gain-framed appeals. The hypothesis was not that loss-framed appeals will generally be highly persuasive in any absolute sense, only that such appeals will generally be more persuasive than gain-framed appeals.

Other hypotheses predict that the direction of difference in persuasiveness between two message forms will vary depending on some moderating condition. For example, Rothman and Salovey (1997) predicted that where the advocacy subject is disease detection behaviors, loss-framed appeals will generally be more persuasive than gain-framed appeals, but where the advocacy subject is disease prevention behaviors, gain-framed appeals will generally be more persuasive than loss-framed

appeals. The hypothesis concerned not the absolute persuasiveness of either message kind on a given subject, but rather the relative persuasiveness of the two kinds.

Still other hypotheses predict that the size (but not necessarily the direction) of difference in persuasiveness between two message forms will vary depending on some moderating condition. For example, the elaboration likelihood model predicts that the size of the difference in persuasiveness between strong-argument messages and weak-argument messages will be affected by the audience's level of involvement: as involvement increases, the size of the difference in persuasiveness between the two message types is predicted to increase (Petty & Cacioppo, 1986, p. 83). The hypothesis did not speak to the absolute persuasiveness of (say) strong-argument messages at any given level of involvement, but rather to the question of how the relative persuasiveness of strong-argument and weak-argument messages would be affected by involvement.

These various theoretically motivated hypotheses thus concern not the absolute persuasiveness of one message type, but rather the relative persuasiveness of two message types—and hence non-behavioral outcomes can appropriately be used in tests of such hypotheses. As the present data indicate, for assessing claims about *relative* message persuasiveness, data about non-behavioral outcomes can straightforwardly substitute for data about behavioral outcomes, because the relative persuasiveness of individual messages (or message kinds) is likely to be identical across these outcomes.

But for assessing claims or hypotheses about *absolute* message persuasiveness—that is, the absolute persuasiveness of a given message or message kind—behavioral outcome assessments are likely to be essential, because one cannot assume that non-behavioral-outcome data and behavioral-outcome data will yield identical conclusions where such claims are concerned.⁹ For example, imagine assessing the effectiveness of a given advertising campaign using post-campaign assessments of intention and behavior. The results for the two outcomes could be quite different: “90% of respondents reported positive intentions but only 20% performed the behavior,” say.¹⁰

Behavioral-outcome data may be desirable for other purposes as well. Consider, for example, causal claims embedded in frameworks such as reasoned action theory (Fishbein & Ajzen, 2010). The claims of interest here are ones suggesting that (e.g.) attitudes influence intentions, which in turn influence behaviors. For assessing such claims, one useful form of evidence would come from experimental studies with longitudinal data about the effects of interventions (e.g., different persuasive messages) on attitudinal, intention, and behavioral outcomes (see Weinstein, 2007). Such data could provide evidence about both relative and absolute persuasiveness with different outcomes at different points in time.

In short, even though behavioral-outcome data might not be needed for assessing claims about relative message persuasiveness, such data might be essential for other purposes. Researchers should not thoughtlessly abandon behavioral assessments of persuasive outcomes.

Conclusion

In formative research, one would like to be able to compare possible messages to see which will be relatively more effective in influencing behavioral outcomes, but pretesting using behavioral measures is often not practical. Examination of within-study effect size comparisons finds that messages’ relative standing is quite consistent across attitude, intention, and behavioral outcomes—and specifically indicates that messages’ relative standing on intention outcomes is indicative of their relative standing on behavioral outcomes. The practical implication is that in formative research, intention outcomes can confidently be used in persuasive message pretesting as a convenient means of identifying those messages that will be relatively more persuasive with respect to behavioral outcomes.

Supporting Information

Additional Supporting Information may be found in the online version of this article. Please note: Oxford University Press is not responsible for the content or functionality of any [supplementary materials](#) supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author.

Acknowledgments

Daniel J. O’Keefe is the Owen L. Coon Professor in the Department of Communication Studies at Northwestern University. ORCID: <https://orcid.org/0000-0003-1594-8892>. The author thanks Mike Allen, Natascha de Hoog, Rick Lau, Corinna Oschatz, Fuyuan Shen, Stephanie Van Stee, Nathan Walter, and Kim Witte for providing information about meta-analytic datasets, and Jon Guryan for helpful suggestions.

Notes

1. This dataset is taken from [O’Keefe \(2020, p. 28\)](#), where it was used to illustrate a similar point about PME and actual message effectiveness (AME): positive correlations between individuals’ PME and AME scores do not show that messages’ relative standing on PME will correspond with messages’ relative standing on AME.
2. Another way of expressing this point: The correlation between intention and behavior can vary depending on the unit of analysis. In that hypothetical dataset, individuals’ scores are positively correlated: when individual is the unit of analysis, the correlation between the intention measure and the behavior measure is $+0.96$. But messages’ scores (means) are negatively correlated: when message is the unit of analysis, the correlation between the intention measure and the behavior measure is -1.00 .
3. The dataset (effect sizes and sample sizes) is provided in the Supporting Information.
4. For the comparison of attitude and intention ESs, the difference (q) was the value for attitude (i.e., the z -transformed r for attitude) minus the value for intention (the z -transformed r for intention). For the comparison of attitude and behavior ESs, q was

the value for attitude minus the value for behavior. For the comparison of intention and behavior ESs, q was the value for intention minus the value for behavior. So, using the last as an example: positive values of q represent cases in which the intention ES was larger than the behavior ES; negative values of q represent cases in which the behavior ES was larger than the intention ES.

5. When there are only two cases, the Pearson r and the rank-order r are numerically identical, with only three possible values (+1.00, -1.00, .00). So these correlations might justifiably be described as Pearson correlations—but describing them as rank-order correlations is meant to make plain the connection to the usual message-pretesting protocol: choose whichever message ranks higher on the pretest assessment.
6. The N s used for these meta-analytic computations were the attitude N (for attitude-intention and attitude-behavior comparisons) and the intention N (for intention-behavior comparisons).
7. The CIs for reported proportions are 95% adjusted Wald (Agresti-Coull) CIs.
8. Cohen's q is the difference between two z-transformed r s (not between two r s). A q of .12 represents the difference between r s of .05 and .17, or .08 and .20, or .20 and .31, or .35 and .45, or .60 and .67.
9. Researchers have sometimes sought to justify such use of non-behavioral outcomes by pointing to positive correlations between individuals' scores on non-behavioral and behavioral measures (e.g., Cho & Choi, 2010, p. 310; Panozzo, Head, Kornides, Feemster, & Zimet, 2020, p. 260). As discussed above, such reasoning is defective. Even if individuals' scores on non-behavioral measures and behavioral measures are positively correlated, that does not show or imply that the relative standing of messages' means on non-behavioral measures will match the relative standing of messages' means on behavioral measures.
10. And just to make the contrast with relative persuasiveness explicit: Imagine that with intervention A 90% of respondents reported positive intentions and 20% performed the behavior, whereas with intervention B 80% of respondents reported positive intentions and 10% performed the behavior. The relative effectiveness of the two interventions is identical across the two kinds of outcome (with intervention A outperforming intervention B), but the absolute effectiveness of a given intervention differs dramatically depending on which outcome is examined.

References

- Andrews, K. R., Carpenter, C. J., Shaw, A. S., & Boster, F. J. (2008). The legitimization of paltory favors effect: A review and meta-analysis. *Communication Reports*, 21(2), 59–69. <https://doi.org/10.1080/08934210802305028>
- Armstrong, J. S. (2010). *Persuasive advertising: Evidence-based principles*. Palgrave Macmillan.
- Atkin, C. K., & Freimuth, V. (2013). Guidelines for formative evaluation research in campaign design. In R. E. Rice & C. K. Atkin (Eds.), *Public communication campaigns* (4th ed., pp. 53–68). Sage.
- Baig, S. A., Noar, S. M., Gottfredson, N. C., Boynton, M. H., Ribisl, K. M., & Brewer, N. T. (2019). UNC perceived message effectiveness: Validation of a brief scale. *Annals of Behavioral Medicine*, 53(8), 732–742. <https://doi.org/10.1093/abm/kay080>
- Baumel, A., Faber, K., Mathur, N., Kane, J. M., & Muench, F. (2017). Enlight: A comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth

- interventions. *Journal of Medical Internet Research*, 19(3), e82. <https://doi.org/10.2196/jmir.7270>
- Bigman, C. A., Bigsby, E., & Martinez Gonzalez, A. (2018). *A meta-analysis of the effects of exemplars on emotional, attitudinal, and behavioral intention responses to messages* [Paper presentation]. International Communication Association Annual Conference, Prague, Czech Republic.
- Blondé, J., & Girandola, F. (2016). Revealing the elusive effects of vividness: A meta-analysis of empirical evidences assessing the effect of vividness on persuasion. *Social Influence*, 11(2), 111–129. <https://doi.org/10.1080/15534510.2016.1157096>
- Borenstein, M., & Rothstein, H. (2005). Comprehensive meta-analysis (Version 2.2.023) [Computer software]. *Biostat*. Retrieved from <https://www.meta-analysis.com/>.
- Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, 83(4), 446–467. <https://doi.org/10.1080/03637751.2015.1128555>
- Brugman, B. C., Burgers, C., & Vis, B. (2019). Metaphorical framing in political discourse through words vs. concepts: A meta-analysis. *Language and Cognition*, 11, 41–65. <https://doi.org/10.1017/langcog.2019.5>
- Burger, J. M., & Caputo, D. (2015). The low-ball compliance procedure: A meta-analysis. *Social Influence*, 10(4), 214–220. <https://doi.org/10.1080/15534510.2015.1049203>
- Cappella, J. N. (2018). Perceived message effectiveness meets the requirements of a reliable, valid, and efficient measure of persuasiveness. *Journal of Communication*, 68(5), 994–997. <https://doi.org/10.1093/joc/jqy044>
- Cappella, J. N., & Kim, M. (2017). Media evaluation. In P. Rössler (Ed.), *International encyclopedia of media effects* (pp. 1020–1028). Wiley-Blackwell. <https://doi.org/10.1002/9781118783764.wbieme0020>.
- Carpenter, C. J. (2013). A meta-analysis of the effectiveness of the “but you are free” compliance-gaining technique. *Communication Studies*, 64(1), 6–17. <https://doi.org/10.1080/10510974.2012.727941>
- Chan, M.-P. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Cheng, Y., & Yan, X. (2020). *Effects of nostalgic messages on ad persuasiveness: A meta-analysis* [Paper presentation]. International Communication Association Annual Conference, Gold Coast, Australia.
- Cho, H., & Choi, J. (2010). Predictors and the role of attitude toward the message and perceived message quality in gain- and loss-frame antidrug persuasion of adolescents. *Health Communication*, 25(4), 303–311. <https://doi.org/10.1080/10410231003773326>
- Cole, G. E., Keller, P. A., Reynolds, J., Schaur, M., & Krause, D. (2016). CDC MessageWorks: Designing and validating a social marketing tool to craft and defend effective messages. *Social Marketing Quarterly*, 22(1), 3–18. <https://doi.org/10.1177/1524500415614817>
- De Hoog, N., Stroebe, W., & de Wit, J. (2007). The impact of vulnerability to and severity of a health risk on processing and acceptance of fear-arousing communications: A meta-analysis. *Review of General Psychology*, 11(3), 258–285. <https://doi.org/10.1037/1089-2680.11.3.258>

- Dillard, J. P., Hunter, J. E., & Burgoon, M. (1984). Sequential-request strategies: Meta-analysis of foot-in-the-door and door-in-the-face. *Human Communication Research, 10*(4), 461–488. <https://doi.org/10.1111/j.1468-2958.1984.tb00028.x>
- Dillard, J. P., Weber, K. M., & Vail, R. G. (2007). The relationship between the perceived and actual effectiveness of persuasive messages: A meta-analysis with implications for formative campaign research. *Journal of Communication, 57*(4), 613–631. <https://doi.org/10.1111/j.1460-2466.2007.00360.x>
- Drovandi, A., Teague, P. A., Glass, B., & Malau-Aduli, B. (2019). Smoker perceptions of health warnings on cigarette packaging and cigarette sticks: A four-country study. *Tobacco Induced Diseases, 17*(March), article (23). <https://doi.org/10.18332/tid/104753>
- Edison, A. S. (2008). *Does sex really sell? Research on sex in advertising: A meta-analysis*. (Publication No. 3385368). Doctoral dissertation, University of Alabama. ProQuest Dissertations and Theses Global.
- Eisend, M. (2006). Two-sided advertising: A meta-analysis. *International Journal of Research in Marketing, 23*, 187–198. <https://doi.org/10.1016/j.ijresmar.2005.11.001>
- Eisend, M., & Tarrahi, F. (2016). The effectiveness of advertising: A meta-meta-analysis of advertising inputs and outcomes. *Journal of Advertising, 45*(4), 519–531. <https://doi.org/10.1080/00913367.2016.1185981>
- Feeley, T. H., Anker, A. E., & Aloe, A. M. (2012). The door-in-the-face persuasive message strategy: A meta-analysis of the first 35 years. *Communication Monographs, 79*(3), 316–343. <https://doi.org/10.1080/03637751.2012.697631>
- Ferrer, R. A., Klein, W. M. P., Zajac, L. E., Land, S. R., & Ling, B. S. (2012). An affective booster moderates the effect of gain- and loss-framed messages on behavioral intentions for colorectal cancer screening. *Journal of Behavioral Medicine, 35*, 452–461. <https://doi.org/10.1007/s10865-011-9371-3>
- Fischer, M., & Huber, F. (2015). The picture communication effect: A meta-analysis. In A. K. Manrai & H. L. Meadow (Eds.), *Global perspectives in marketing for the 21st century: Proceedings of the 1999 World Marketing Congress* (pp. 65–68). Springer. https://doi.org/10.1007/978-3-319-17356-6_20
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press.
- Freling, R. E. (2017). *Understanding influences and biases in decision making: Three meta-analyses on information processing* (Publication No. 10759978). Doctoral dissertation, University of Texas at Dallas. ProQuest Dissertations and Theses Global.
- Freling, T. H., Vincent, L. H., & Henard, D. H. (2014). When not to accentuate the positive: Re-examining valence effects in attribute framing. *Organizational Behavior and Human Decision Processes, 124*, 95–109. <https://doi.org/10.1016/j.obhdp.2013.12.007>
- Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin, 132*(5), 778–822. <https://doi.org/10.1037/0033-2909.132.5.778>
- Grewal, D., Kavanoor, S., Fern, E. F., Costley, C., & Barnes, J. (1997). Comparative versus noncomparative advertising: A meta-analysis. *Journal of Marketing, 61*(4), 1–15. <https://doi.org/10.1177/002224299706100401>
- Hamilton, M. A., & Hunter, J. E. (1998). The effect of language intensity on receiver evaluations of message, source, and topic. In M. Allen & R. W. Preiss (Eds.), *Persuasion: Advances through meta-analysis* (pp. 99–138). Hampton Press.

- Hartnett, N., Greenacre, L., Kennedy, R., & Sharp, B. (2020). Extending validity testing of the Persuasion Principles Index. *European Journal of Marketing*, 54(9), 2245–2255. <https://doi.org/10.1108/EJM-12-2018-0865>
- Hornikx, J., & O'Keefe, D. J. (2009). Adapting consumer advertising appeals to cultural values: A meta-analytic review of effects on persuasiveness and ad liking. *Annals of the International Communication Association*, 33(1), 39–71. <https://doi.org/10.1080/23808985.2009.11679084>
- Lau, R. R., Sigelman, L., & Rovner, I. B. (2007). The effects of negative political campaigns: A meta-analytic reassessment. *The Journal of Politics*, 69(4), 1176–1209. <https://doi.org/10.1111/j.1468-2508.2007.00618.x>
- Lee, S., & Feeley, T. H. (2017). A meta-analysis of the pique technique of compliance. *Social Influence*, 12(1), 15–28. <https://doi.org/10.1080/15534510.2017.1305986>
- Lee, S., Moon, S.-I., & Feeley, T. H. (2016). A meta-analytic review of the legitimization of paltry favors compliance strategy. *Psychological Reports*, 118(3), 748–771. <https://doi.org/10.1177/0033294116647690>
- Lull, R. B., & Bushman, B. J. (2015). Do sex and violence sell? A meta-analytic review of the effects of sexual and violent media and ad content on memory, attitudes, and buying intentions. *Psychological Bulletin*, 141(5), 1022–1048. <https://doi.org/10.1037/bul0000018>
- Lunt, D. (2016). *Willing to help, but lacking discernment: The effects of victim group size on donation behaviors* (Publication No. 10302039). Doctoral dissertation, University of Texas at Arlington. ProQuest Dissertations and Theses Global
- Meyerowitz, B. E., & Chaiken, S. (1987). The effect of message framing on breast self-examination attitudes, intentions, and behavior. *Journal of Personality and Social Psychology*, 52(3), 500–510. <https://doi.org/10.1037/0022-3514.52.3.500>
- Noar, S. M., Barker, J., & Yzer, M. (2018). Measurement and design heterogeneity in perceived message effectiveness studies: A call for research. *Journal of Communication*, 68(5), 990–993. <https://doi.org/10.1093/joc/jqy047>
- Noar, S. M., Hall, M. G., Francis, D. B., Ribisl, K. M., Pepper, J. K., & Brewer, N. T. (2016). Pictorial cigarette pack warnings: A meta-analysis of experimental studies. *Tobacco Control*, 25, 341–354. <https://doi.org/10.1136/tobaccocontrol-2014-051978>
- O'Keefe, D. J. (1998). Justification explicitness and persuasive effect: A meta-analytic review of the effects of varying support articulation in persuasive messages. *Argumentation and Advocacy*, 35(2), 61–75. <https://doi.org/10.1080/00028533.1998.11951621>
- O'Keefe, D. J. (1999). How to handle opposing arguments in persuasive messages: A meta-analytic review of the effects of one-sided and two-sided messages. *Annals of the International Communication Association*, 22(1), 209–249. <https://doi.org/10.1080/23808985.1999.11678963>
- O'Keefe, D. J. (2000). Guilt and social influence. *Annals of the International Communication Association*, 23(1), 67–101. <https://doi.org/10.1080/23808985.2000.11678970>
- O'Keefe, D. J. (2002). The persuasive effects of variation in standpoint articulation. In F. H. van Eemeren (Ed.), *Advances in pragma-dialectics* (pp. 65–82). Sic Sat.
- O'Keefe, D. J. (2013). The relative persuasiveness of different message types does not vary as a function of the persuasive outcome assessed: Evidence from 29 meta-analyses of 2,062 effect sizes for 13 message variations. *Annals of the International Communication Association*, 37(1), 221–249. <https://doi.org/10.1080/23808985.2013.11679151>

- O'Keefe, D. J. (2018). Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication*, 68(1), 120–142. <https://doi.org/10.1093/joc/jqx009>
- O'Keefe, D. J. (2020). Message pretesting using perceived persuasiveness measures: Reconsidering the correlational evidence. *Communication Methods and Measures*, 14(1), 25–37. <https://doi.org/10.1080/19312458.2019.1620711>
- O'Keefe, D. J., & Jensen, J. D. (2006). The advantages of compliance or the disadvantages of noncompliance? A meta-analytic review of the relative persuasive effectiveness of gain-framed and loss-framed messages. *Annals of the International Communication Association*, 30(1), 1–43. <https://doi.org/10.1080/23808985.2006.11679054>
- Panozzo, C. A., Head, K. J., Kornides, M. L., Feemster, K. A., & Zimet, G. D. (2020). Tailored messages addressing human papillomavirus vaccination concerns improves behavioral intent among mothers: a randomized controlled trial. *Journal of Adolescent Health*, 67(2), 253–261. <https://doi.org/10.1016/j.jadohealth.2020.01.024>
- Paul, C. L., Redman, S., & Sanson-Fisher, R. W. (1997). The development of a checklist of content and design characteristics for printed health education materials. *Health Promotion Journal of Australia*, 7(3), 153–159.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.
- Portnoy, D. B., Ferrer, R. A., Bergman, H. E., & Klein, W. M. P. (2014). Changing deliberative and affective responses to health risk: A meta-analysis. *Health Psychology Review*, 8(3), 296–318. <https://doi.org/10.1080/17437199.2013.798829>
- Rains, S. A., Levine, T. R., & Weber, R. (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association*, 42(2), 105–124. <https://doi.org/10.1080/23808985.2018.1446350>
- Rossmann, C. (2014). Strategic health communication: Theory- and evidence-based campaign development. In D. Holtzhausen & A. Zerfass (Eds.), *The Routledge handbook of strategic communication* (pp. 409–423). Routledge. <https://doi.org/10.4324/9780203094440.ch27>.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: The role of message framing. *Psychological Bulletin*, 121(1), 3–19. <https://doi.org/10.1037/0033-2909.121.1.3>
- Seo, K., & Kim, N. Y. (2018). *Does adding images to texts influence persuasion? A meta-analysis of visual image effects on persuasive texts* [Paper presentation]. International Communication Association Annual Conference, Prague, Czech Republic.
- Sharp, B., & Hartnett, N. (2016). Generalisability of advertising persuasion principles. *European Journal of Marketing*, 50(1/2), 301–305. <https://doi.org/10.1108/EJM-12-2015-0842>
- Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, 12(1), 1–36. <https://doi.org/10.1080/14792772143000003>
- Sheeran, P., & Orbell, S. (1998). Do intentions predict condom use? Meta-analysis and examination of six moderator variables. *British Journal of Social Psychology*, 37(Pt 2), 231–250. <https://doi.org/10.1111/j.2044-8309.1998.tb01167.x>
- Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health communication: A meta-analysis. *Journal of Advertising*, 44, 105–113. <https://doi.org/10.1080/00913367.2015.1018467>

- Shen, F. Y., & Han, J. X. (2014). Effectiveness of entertainment education in communicating health information: A systematic review. *Asian Journal of Communication, 24*(6), 605–616. <https://doi.org/10.1080/01292986.2014.927895>
- Sun, W. H., Miu, H. Y. H., Wong, C. K. H., Tucker, J. D., & Wong, W. C. W. (2018). Assessing participation and effectiveness of the peer-led approach in youth sexual health education: Systematic review and meta-analysis in more developed countries. *Journal of Sex Research, 55*(1), 31–44. <https://doi.org/10.1080/00224499.2016.1247779>
- Vaillancourt, C., Bedard, A., Belanger-Gravel, A., Provencher, V., Begin, C., Desroches, S., & Lemieux, S. (2019). Promoting healthy eating in adults: An evaluation of pleasure-oriented versus health-oriented messages. *Current Developments in Nutrition, 3*(5), article no. nzz012. <https://doi.org/10.1093/cdn/nzz012>
- Van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The extended transportation-imagery model: A meta-analysis of the antecedents and consequences of consumers' narrative transportation. *Journal of Consumer Research, 40*(5), 797–817. <https://doi.org/10.1086/673383>
- Van Stee, S. K. (2018). Meta-analysis of the persuasive effects of metaphorical vs. literal messages. *Communication Studies, 69*(5), 545–566. <https://doi.org/10.1080/10510974.2018.1457553>
- Walter, N., Cody, M. J., Xu, L. Z., & Murphy, S. T. (2018). A priest, a rabbi, and a minister walk into a bar: A meta-analysis of humor effects on persuasion. *Human Communication Research, 44*(4), 343–373. <https://doi.org/10.1093/hcr/hqy005>
- Walter, N., Tukachinsky, R., Pelled, A., & Nabi, R. (2019). Meta-analysis of anger and persuasion: An empirical integration of four models. *Journal of Communication, 69*(1), 73–93. <https://doi.org/10.1093/joc/jqy054>
- Weinstein, N. D. (2007). Misleading tests of health behavior theories. *Annals of Behavioral Medicine, 33*(1), 1–10. https://doi.org/10.1207/s15324796abm3301_1
- Wirtz, J. G., Sparks, J. V., & Zimbres, T. M. (2018). The effect of exposure to sexual appeals in advertisements on memory, attitude, and purchase intention: A meta-analytic review. *International Journal of Advertising, 37*(2), 168–198. <https://doi.org/10.1080/02650487.2017.1334996>
- Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health programs. *Health Education & Behavior, 27*(5), 591–615. <https://doi.org/10.1177/109019810002700506>
- Xu, Z., & Guo, H. (2018). A meta-analysis of the effectiveness of guilt on health-related attitudes and intentions. *Health Communication, 33*(5), 519–525. <https://doi.org/10.1080/10410236.2017.1278633>
- Zebregs, S., van den Putte, B., Neijens, P., & de Graaf, A. (2015). The differential impact of statistical and narrative evidence on beliefs, attitude, and intention: A meta-analysis. *Health Communication, 30*(3), 282–289. <https://doi.org/10.1080/10410236.2013.842528>