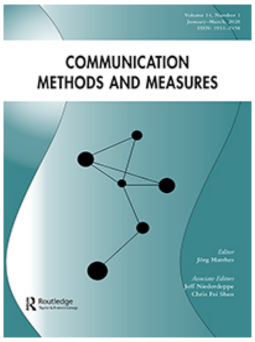

Communication Methods and Measures

ISSN: 1931-2458 (Print) 1931-2466 (Online) Journal homepage: https://www.tandfonline.com/loi/hcms20

Message Pretesting Using Perceived Persuasiveness Measures: Reconsidering the Correlational Evidence

Daniel J. O'Keefe

To cite this article: Daniel J. O'Keefe (2020) Message Pretesting Using Perceived Persuasiveness Measures: Reconsidering the Correlational Evidence, Communication Methods and Measures, 14:1, 25-37, DOI: 10.1080/19312458.2019.1620711

To link to this article: https://doi.org/10.1080/19312458.2019.1620711

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.

Published online: 30 May 2019.

Submit your article to this journal

Article views: 501

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hcms20

COMMUNICATION METHODS AND MEASURES
2020, VOL. 14, NO. 1, 25–37
https://doi.org/10.1080/19312458.2019.1620711

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Message Pretesting Using Perceived Persuasiveness Measures: Reconsidering the Correlational Evidence

Daniel J. O'Keefe

Department of Communication Studies, Northwestern University, Evanston, USA

ABSTRACT

Perceived message effectiveness (PME) measures have been used in persuasive message pretesting to diagnose differences between messages in actual message effectiveness (AME). This practice has been underwritten by pointing to positive correlations between individuals' PME and AME scores. This essay argues that such correlations do not – and cannot – show that messages' relative standing on PME matches their relative standing on AME, and thus correlations between *individuals'* PME and AME scores are irrelevant to the question of whether PME assessments accurately identify relatively more effective messages. Only correlations between *messages'* PME and AME scores can indicate whether PME measures provide successful prediction of messages' relative AME standing. But positive correlations between individuals' PME and AME scores do suggest a possible repurposing of PME measures in formative research.

Persuasive message pretesting is a valuable element of formative research. By pretesting alternative possible messages, a message designer can choose ones likely to be relatively more effective. Measures of perceived message persuasiveness are widely used in this enterprise. The message-pretesting use of these measures has often been underwritten by pointing to evidence showing that such measures are positively correlated with measures of persuasive outcomes such as attitude, intention, and behavior.

The purpose of this paper is to re-examine that correlational evidence, with an eye to suggesting that it is less compelling than might be supposed. In what follows, the use of perceived-persuasiveness measures in message pretesting is given some further specification, and the correlational evidence is then examined more closely. The argument will be that the commonly reported correlations between measures of perceived persuasiveness and measures of actual persuasiveness – correlations between individuals' scores on perceived and actual persuasiveness – are not relevant to the question of whether perceived-persuasiveness measures are diagnostic of differences in messages' actual effectiveness. A concluding section identifies appropriate criteria for assessing potential message-pretesting procedures and suggests a possible new role for perceived-persuasiveness measures in formative research.

Message Pretesting and Perceived Persuasiveness Measures

Formative research can have many purposes, but one common purpose is that of pretesting potential messages so as to identify the most effective ones. In the simplest case, two messages will be under consideration for use, and the question is: Which message will be more effective?[1]

CONTACT Daniel J. O'Keefe d-okeefe@northwestern.edu Department of Communication Studies, Northwestern University, 2240 Campus Drive, Evanston, IL 60208

Author Note
Daniel J. O'Keefe is the Owen L. Coon Professor in the Department of Communication Studies at Northwestern University.
Thanks to Nancy Grant Harrington, Andy King, Bruce Lambert, Nathan Walter, and the journal's reviewers for helpful commentary.

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Measures of perceived or expected persuasiveness (PME, perceived message effectiveness, or PE, perceived effectiveness) have commonly been used to address this question.[2] The data usually come from members of the target audience, though sometimes from experts in the subject matter of advocacy. PME measures can take various forms. For example, participants have been asked to respond to questionnaire items asking for ratings of messages on persuasiveness, convincingness, effectiveness, and the like (e.g., Mackert et al., 2014; Popova, Neilands, & Ling, 2014), to rank-order messages in terms of persuasiveness (e.g., Mouneyrac, Le Floch, Lemercier, Py, & Roumegue, 2017; Pollard et al., 2016), or to engage in focus-group discussions concerning relative message persuasiveness (e.g., Mowbray, Marcu, Godinho, Michie, & Yardley, 2016; Record, Harrington, Helme, & Savage, 2018). For some general discussions of such measures, see Choi and Cho (2016), Noar, Bell, Kelley, Barker, and Yzer (2018), and Yzer, LoRusso, and Nagler (2015).

Regardless of the particular form of PME data, the formative-research decision procedure is straightforward: whichever message is rated higher in perceived persuasiveness is the one presumed to be more effective and hence is selected for deployment. If message A is rated higher in PME than message B, then message A is chosen.[3]

Recently, questions have been raised about how successfully this procedure identifies the more effective message. In O'Keefe's (2018a) review of 151 message-pair comparisons, PME assessments correctly identified the more effective message in only 58% of cases, not significantly different from chance (for commentary and discussion, see Cappella, 2018; Davis & Duke, 2018; Noar, Barker, & Yzer, 2018; O'Keefe, 2018b).

However, a number of researchers have pointed to other evidence that seems to support the diagnostic use of PME assessments, namely, positive correlations between PME measures and measures of actual persuasiveness (AME, actual message effectiveness) such as assessments of attitude, intention, and behavior. The next section discusses such evidence.

PME-AME Correlations as Evidence

Abstractly put, the formative-research problem is one of predicting messages' relative standing on AME. The specific question of interest is whether PME assessments provide accurate predictions of that outcome, that is, whether messages' relative PME standing is diagnostic of their relative AME standing.

Naturally, then, one form of evidence that has often been adduced in support of PME diagnosticity is the ability to predict AME outcomes (e.g., subsequent behavior) on the basis of PME assessments. The evidence of this predictability takes the form of positive correlations between individuals' PME and AME scores (henceforth "individual-level" PME-AME correlations).

For example, Dillard, Weber, and Vail's (2007, p. 613) meta-analysis of individual-level PME-AME correlations found a mean correlation of .41 across 40 cases, concluding that "the results empirically demonstrate the value of PE judgments in formative research." In a more focused meta-analysis, Noar, Barker, Bell, and Yzer (in press, p. 1) reviewed six longitudinal studies of tobacco education messages, reporting that individuals' PME assessments predicted their quitting intentions (mean individual-level r = .26) and cessation behaviors (mean individual-level r = .20); these results were interpreted as suggesting that "PME provides some predictive value as to the likely effectiveness of messages."

Similarly, single studies reporting positive individual-level PME-AME correlations have been taken as underwriting the use of PME measures for message pretesting. For example, Brennan, Durkin, Wakefield, and Kashima (2014, p. 412) found that individuals' scores on a measure of PME were positively correlated with subsequent intention and behavior change, saying that such findings "increase confidence in the use of PE measures to pretest and evaluate antismoking television advertisements." Davis, Nonnemaker, Duke, and Farrelly (2013, p. 461) reported that individuals' PME assessments of smoking-cessation television ads predicted subsequent quitting intentions, arguing that "our findings support the use of PE in quantitative ad pretesting as part of a standard regimen of formative research for cessation television campaigns." Davis, Uhrig, Bann, Rupert, and Fraze (2011, p. 39) found individuals' scores on PME measures to predict subsequent HIV testing intentions, concluding that "measures of perceived ad effectiveness should be used to quantitatively pretest future HIV/AIDS-focused social marketing campaign messages." (See, similarly, Choi & Cho, 2016; Noar et al., 2016.)

The argument to be advanced here is that these individual-level PME-AME correlations are in fact not relevant to whether PME measures are a useful way of pretesting messages to assess their relative effectiveness. As a preview of the eventual key point: Individual-level PME-AME correlations speak to the question of whether one can predict *individuals'* relative standing on AME using individuals' relative standing on PME. But for message-pretesting purposes, the question of interest is whether one can predict *messages'* relative standing on AME using messages' relative standing on PME. As will be seen, these are different questions – with independent answers.

Two Kinds of PME-AME Correlations

As noted above, the positive PME-AME correlations that are commonly invoked as underwriting the message-pretesting use of PME measures are correlations between individuals' scores on PME and AME measures. A positive individual-level PME-AME correlation shows that individuals' relative standing on PME is indicative of individuals' relative standing on AME.

But where PME and AME data are available for two (or more) messages, it is also possible to compute a second kind of PME-AME correlation, a "message-level" correlation in which the unit of analysis is messages, not individuals. In the relevant sort of dataset, each message has a mean PME value (providing a basis for ranking messages on PME) and a mean AME value (providing a basis for ranking messages on AME) and hence a rank-order correlation can be computed. A positive message-level PME-AME correlation indicates that messages' relative standing on PME is indicative of messages' relative standing on AME.

Independence of the Two Correlations

One might think that these two kinds of PME-AME correlations must somehow be closely related. In particular, it might be tempting to think that if the individual-level PME-AME correlation is positive, then the message-level PME-AME correlation must also be positive.

But in fact, there is no necessary relationship between the signs of the two correlations. The absence of that relationship is not a consequence of the specific variables of interest here (PME and AME) but rather reflects the properties of correlation coefficients generally and so can be expressed in appropriately abstract terms, as follows.

The two kinds of PME-AME correlations distinguished here (individual-level and message-level) are instantiations of a more abstract general contrast between individual-level and group-level correlations. So imagine a dataset with two variables, X and Y, where the data come from two groups, A and B. One can compute the familiar individual-level correlation between X and Y (indicating the degree to which individuals' relative standing on X is indicative of their relative standing on Y); one can also compute a group-level correlation between X and Y, that is, the rank-order correlation of the two groups' X and Y scores (indicating the degree to which the relative standing of groups on X is indicative of the relative standing of groups on Y).

But individual-level and group-level correlations are largely independent. As a place to start in seeing that, consider that the correlation coefficient is invariant under conditions of linear transformations of the variables.[4] Imagine, for example, that in a given dataset the correlation between variable X and variable Y is .6578. Suppose one adds 29.3 to each case's score on X (thereby changing the mean of X) and divides each case's score on Y by 3.76 (thereby changing the mean of Y). The recomputed correlation will still be .6578.

This invariance implies that knowing the correlation coefficient cannot possibly provide a basis for inference about the value of the mean on either variable, because the exact same correlation could have been obtained with a great many numerically different means. Similarly, knowing both the correlation coefficient and the mean value for one variable does not provide a basis for inference about the mean value for the other variable; no matter what the mean happens to be for one variable, the exact same correlation could have been obtained for any number of numerically different means for the other variable.[5]

So now consider the circumstance in which the data come from two (or more) groups or conditions. Just as it is not possible to draw inferences about overall variable means on the basis of the individual-level correlation between two variables, so it is also not possible to draw inferences about the relationship of group means (i.e., the group-level correlation) on the basis of the individual-level correlation between two variables. For example, the occurrence of a positive individual-level correlation is consistent with positive, negative, or zero group-level correlations. This is easily illustrated with hypothetical data.

Imagine a data set in which data for variables X and Y are available for two groups, with n = 5 for each group. For group A, the participants' (X, Y) pairs of scores are (90, 92), (80,81), (70,70), (60,59), and (50,48). For group B, the scores are (85,75), (75,70), (65,65), (55,60), and (45,55). The individual-level correlation between X and Y is quite positive, .94. Group A has a higher mean on X (70.0) than does group B (65.0), and group A also has a higher mean on Y (70.0) than does group B (65.0). Thus, the group-level correlation is positive (rank-order correlation = +1.00), indicating that the relative standing of the two groups on X matches their relative standing on Y.

But a strongly positive individual-level correlation is also consistent with a different relationship of group means, as can be illustrated with a different data set. For group A, the participants' (X,Y) pairs of scores are (90, 85), (80,75), (70,65), (60,55), and (50,45). For group B, the scores are (85,90), (75,80), (65,70), (55,60), and (45,50). The individual-level correlation is again .94. Group A has a higher mean on X (70.0) than does group B (65.0) – but group B has a higher mean on Y (70.0) than does group A (65.0). That is, even though *individuals'* scores on X and Y are strongly positively correlated, the *groups'* scores are negatively correlated (rank-order correlation = −1.00), that is, the relative standing of the two groups is different on the two variables.

So, expressed abstractly: The sign of the individual-level correlation between two variables cannot possibly indicate the sign of the group-level correlation (that is, the relationship between groups' means on those variables).[6] And this is a general point about the properties of individual-level correlation coefficients, a point unaffected by the substance of the particular variables involved. No matter what the variables are, a positive individual-level correlation between two variables cannot possibly justify a conclusion about the group-level correlation. Individual-level correlations do not contain information about means – not about overall variable means, and a fortiori not about group means.

Application to PME-AME Correlations

As perhaps is apparent, this general reasoning about the properties of individual-level and group-level correlations applies straightforwardly to individual-level and message-level PME-AME correlations. A positive individual-level correlation between PME and AME cannot possibly justify drawing a conclusion about the message-level correlation. The ability to predict individuals' relative standing on AME from knowing individuals' relative standing on PME (a positive individual-level PME-AME correlation) does not imply an ability to predict messages' relative standing on AME from knowing messages' relative standing on PME (a positive message-level PME-AME correlation).

For example, even if the correlation between individuals' scores on PME and AME is strongly positive, the correlation between messages' scores on PME and AME can be strongly negative. As an illustrative hypothetical dataset involving three messages (N = 12): For message A, participants have the following (PME, AME) 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 individual-level PME-AME correlation is .96. The rank-order of messages on PME is A, B, C (means of 67.0, 65.0, and 63.0, respectively); the rank-order on AME is C, B, A (means of 67.0, 65.0, and 63.0, respectively). That is, the messages' ranking on PME is the opposite of their ranking on AME; the message-level PME-AME correlation is negative (−1.00). So despite a strongly positive individual-level PME-AME correlation, PME gives exactly the wrong answer about the relative actual effectiveness of these messages.

And notice: The possibility of such divergence between individual-level and message-level PME-AME correlations has nothing to do with the substantive characteristics of PME or AME as variables, or the way in which those variables are measured, or whether some third variable might be influencing both PME and AME, or whether the PME–AME relationship is causal, or whether the correlations are based on cross-sectional or longitudinal data, or whether moderators might influence the strength of the correlation, etc. The divergence is possible simply because of the abstract properties of individual-level and group-level correlations, that is, the ways in which correlations can be influenced by the unit of analysis.

The Upshot

The independence of individual-level and message-level PME-AME correlations has a straightforward implication: Individual-level PME-AME correlations do not – cannot – provide evidence that is relevant to the question of whether PME measures are diagnostic of differences between messages in AME. Whether PME measures are a good way of pretesting messages for relative effectiveness cannot be deduced from examination of individual-level PME-AME correlations.

To concretize the consequences of this conclusion, recall Dillard et al.'s (2007, p. 613) meta-analysis; the reported mean individual-level PME-AME correlation of .41 was said to "empirically demonstrate the value of PE judgments in formative research." But those individual-level correlations are actually not relevant to the question of whether PME assessments are useful in formative research aimed at diagnosing differences in messages' effectiveness. Similarly, Noar et al.'s (in press, p. 1) claim that positive individual-level PME-AME correlations indicate that "PME provides some predictive value as to the likely effectiveness of messages" is incorrect. Those individual-level correlations cannot possibly show that PME provides guidance either about the likely absolute effectiveness of any one message or about the likely relative effectiveness of several messages.

Unfortunately, such reports of positive individual-level PME-AME correlations have often been relied on in ways that have led to unwarranted inferences. Many researchers have incorrectly reasoned that because individuals' scores on PME and AME are generally positively correlated, one can draw inferences about messages' relative AME on the basis of knowing messages' relative PME (e.g., Dailey, McCracken, & Romo, 2011; Davis & Duke, 2018; Durkin, Bayly, Cotter, Mullin, & Wakefield, 2013; Grummon, Hall, Taillie, & Brewer, 2019; Niederdeppe, Farrelly, Nonnemaker, Davis, & Wagner, 2011; Noar, Palmgreen, Zimmerman, Lustria, & Lu, 2010; Ramsay, Yzer, Luciana, Vohs, & MacDonald, 2013; Yzer et al., 2018). A particularly clear illustration of this mistaken reasoning is provided by O'Keefe (2018a):

> When PME and AME data are collected concerning two messages, the PME–AME correlation computed for data combined across messages does provide evidence relevant to the question of the diagnosticity of PME data. Strong positive across-message PME–AME correlations are an indication that messages' relative standing on PME will be diagnostic of their relative standing on AME; weak or negative correlations are a sign of poor diagnosticity. (O'Keefe, 2018a, p. 124)

Such reasoning is simply wrong. No matter what their magnitude or direction, individual-level PME-AME correlations do not – cannot – speak to the question of whether messages' relative standing on PME assessments is a good guide to those messages' relative standing on AME assessments. Even if individuals' PME and AME scores are strongly positively correlated, messages' relative AME standing can be the opposite of their relative PME standing.

Perhaps one source of confusion here has been the ambiguity of referring to "PME-AME correlations" without specifying the unit of analysis. If one is inattentive to that ambiguity, then reasoning of the following sort might seem correct: "Message A had a higher mean on PME than did message B. And PME and AME are positively correlated. Therefore, message A is likely to have a higher mean on AME than does message B."

But notice how specifying the unit of analysis can spotlight the mistake in reasoning: "Message A had a higher mean on PME than did message B. And individuals' scores on PME and AME are positively correlated; that is, individuals' relative ranking on PME is indicative of individuals' relative ranking on AME. Therefore, message A is likely to have a higher mean on AME than does message B." As perhaps is obvious, the second premise is not the required sort, because it's the wrong unit of analysis. Whether *individuals'* scores on PME and AME are positively correlated is not what's relevant here. The question is whether *messages'* scores on PME and AME are positively correlated.

Moving Forward

How to Underwrite a Message Pretesting Procedure

Positive individual-level PME-AME correlations cannot be used to justify the use of PME assessments as a means of diagnosing relative message persuasiveness, because the direction of individual-level PME-AME correlations is independent of whether messages' relative standing on PME matches their relative standing on AME (that is, independent of the direction of message-level PME-AME correlations). And that independence reflects a general abstract truth about individual-level and message-level correlations, not something specific to PME or AME.

Thus this observation has implications beyond PME assessments specifically. For *any* proposed message pretesting measure, correlations between individuals' scores on that measure and individuals' AME scores will not be relevant to the question of whether that measure accurately diagnoses messages' differences in effectiveness. For any given message-pretesting variable P, knowing that individuals' scores on P are correlated with individuals' AME scores provides no information about whether P can accurately distinguish relatively more and less persuasive messages.

Instead, relevant evidence concerning a possible pretest measure will consist of *message-level* correlations, that is, correlations between messages' scores on the pretest variable and on AME. To show that a given message-pretesting procedure is diagnostic of differences in messages' AME, one needs evidence of positive message-level correlations between the pretest variable and AME.

In order to produce such evidence, one will need to have data that permits one to compare different messages' standing on AME. Expressed differently: If a study does not have data concerning the rank-ordering of messages in terms of relative actual persuasiveness, then that study does not have data that can speak to the question of how to *predict* the rank-ordering of messages in terms of relative actual persuasiveness. In general, if a study does not contain an assessment of X, that study cannot speak to the question of how to predict X.

That point might seem too obvious to be worth mentioning, but consider: Some of the studies that have commonly been cited as supporting the use of PME assessments in message pretesting have *not* had data that allowed rank-ordering of messages in terms of relative AME. For example, some studies exposed all participants to the same message or message materials, as in Davis et al.'s (2011) design; all participants saw the same HIV-testing campaign materials (three ads and a booklet), and hence post-exposure AME assessments did not permit comparison of different messages' standing on AME. In Bigsby, Cappella, and Seitz's (2013) design, each participant saw four antismoking ads (randomly chosen from a set of 100) and then reported quitting intentions; different participants saw different ads, but because each participant saw four ads before reporting intentions, it was not possible to compare the relative actual effectiveness of individual ads. In Alvaro et al.'s (2013) research, participants were exposed to (and provided PME assessments of) various different anti-marijuana television ads; post-exposure AME measures were obtained, but because participants may have seen multiple ads, it was not possible to compare the effectiveness of different ads.[7]

It might seem surprising that anyone would offer a study that did not have data about different messages' standing on AME as somehow being relevant to the question of how to predict different messages' standing on AME. But this misstep perhaps becomes more understandable if one remembers that the data were commonly described as "PME-AME correlations," that is, without explicit characterization of the unit of analysis.

The ambiguous language of "PME-AME correlations" might have easily misled researchers. For example, the (implicit) reasoning might have been something like this: "PME and AME are positively correlated, so one can use PME to predict AME. And our goal is to predict messages' AME – so the evidence supports using PME to identify relatively more effective messages."

But, as explained above, this reasoning is flawed because of inattention to the unit of analysis. If individuals' scores on PME and AME are correlated, then one can use individuals' PME scores to predict individuals' relative standing on AME. But evidence of the ability to predict individuals' relative AME standing is not evidence of the ability to predict messages' relative AME standing. Individual-level correlations and message-level correlations are different things – and only message-level correlations speak to the ability to predict messages' relative AME standing.

No matter what the reasoning was that led researchers to think that individual-level PME-AME correlations provided evidence about whether PME assessments are diagnostic of differences in messages' mean AME scores, that reasoning was plainly defective. One cannot justify the message-pretesting use of PME assessments by invoking positive individual-level PME-AME correlations, because the direction of individual-level PME-AME correlations is independent of whether messages' rankings on PME match their rankings on AME.

Instead, in future research aimed at identifying dependably diagnostic message-pretesting procedures, the appropriate criterion to use is whether the pretesting procedure accurately predicts messages' relative actual persuasiveness. When the purpose of message pretesting is to identify the relatively more effective messages, the ability to do that – identify the relatively more effective messages – is the appropriate thing to demand of a message pretesting procedure.

And with respect to PME measures specifically: O'Keefe (2018a, p. 133) reported that the average message-level rank-order PME-AME correlation, across 35 studies, was effectively zero (mean message rank-order $r = -.053$). That is, even though individual-level PME-AME correlations are characteristically positive (e.g., Dillard et al., 2007), message-level PME-AME correlations are not. Individuals' relative standing on AME can generally be predicted by individuals' relative standing on PME, but messages' relative standing on AME cannot generally be predicted by messages' relative standing on PME.

Repurposing PME Assessments in Formative Research?

Against the backdrop of the distinction between individual-level and message-level PME-AME correlations, the fact that individual-level PME-AME correlations are commonly positive raises the possibility that PME data might say more about people than about messages. Specifically, variation between individuals in PME ratings might reflect differences in individuals' readiness for (or resistance to) change, such that respondents who are on the cusp of behavior change (those who are initially more favorably inclined toward the advocated view) rate messages as more persuasive than do respondents who are more resistant to change (for an earlier discussion of such a hypothesis, see Zhao, Strasser, Cappella, Lerman, & Fishbein, 2011, p. 60).

This picture would explain why individuals' PME and AME scores are generally positively correlated, because the individuals with higher PME scores would precisely be the individuals most likely to exhibit subsequent change. To concretize that idea, imagine showing a set of antismoking ads to a sample of smokers, obtaining their PME assessments of each ad, and creating a total PME score for each respondent; six months later, AME data are collected on the participants (e.g., number of quit attempts, whether the respondent had successfully quit, etc.). Suppose that individuals' PME and AME scores were found to be positively correlated: participants who rated the ads as highly persuasive were more likely to quit (or to try to quit, etc.).[8]

Such a pattern of results would not necessarily have to reflect any persuasive effects of the antismoking ads to which participants were exposed. In fact, such a set of results could arise even if the ads had no effect whatever.

The alternative process that could give rise to such results would be one in which smokers who were already predisposed to change – those already favorably inclined toward quitting and so already more likely to try to quit in the future – would generally perceive the ads as persuasive, whereas those smokers who were remote from change would generally see the ads as unpersuasive. The smoker who had already started to taper down, or who had set a quit date, or who had already been thinking about nicotine replacement – that smoker, upon seeing the ads, has the reaction "yes, those are good ads, those arguments are persuasive, those are convincing ads." But the committed smoker, the resolute smoker – that smoker, upon seeing those same ads, has the reaction "those arguments are stupid, they're not going to convince anybody, those ads are not persuasive." Unsurprisingly, at the delayed assessment, the former sort of smoker shows behavioral change but the latter does not – even if neither was influenced by the ads to which they were exposed.

This picture of the underlying process is also consistent with research findings of individual-level associations between PME assessments and several relevant *pre-exposure* states. Such associations have been reported in many tobacco-related studies. For example, in Guillaumier et al.'s (2017) research, smokers with more positive pre-exposure cessation cognitions rated various antismoking messages significantly higher in PME than did smokers with less positive cognitions. In Thrasher et al.'s (2012) study of antismoking pictorial health warning labels, daily smokers perceived the warnings to be less effective (and lower in credibility) than did nondaily smokers. Biener, Ji, Gilpin, and Albers (2004) found that adolescents who owned a tobacco promotional item, compared to those who did not, rated antismoking ads as significantly less effective. In a number of studies, smokers with higher PME scores for antismoking ads, compared to those with lower PME scores, have been found to be significantly more likely to have made a previous quit attempt, to smoke fewer cigarettes, to have a greater desire or readiness to quit, to already have plans to quit, and so on (Bigsby et al., 2013; Brennan et al., 2014; Davis et al., 2017; Davis, Nonnemaker, Farrelly, & Niederdeppe, 2011; Hitchman et al., 2012; Van Dessel, Smith, & De Houwer, 2018; Willemsen, 2005). And several studies have found antismoking messages to be rated higher in perceived effectiveness by nonsmokers than by smokers (Biener, McCallum-Keeler, & Nyman, 2000; Lingwall et al., 2018; Sidhu, 2011; Sobani, Nizami, Raza, Ul Ain Baloch, & Khan, 2010).

Similar individual-level associations have been reported in domains other than tobacco. For example, resolute nonusers of marijuana were found to evaluate antimarijuana ads significantly higher on PME than did users (Alvaro et al., 2013). In O'Donnell and Willoughby's (2017) research, participants with more positive pre-exposure attitudes toward condom use rated various condom-use Instagram posts significantly higher in PME than did participants with less positive attitudes. People who had received the HPV vaccine evaluated HPV vaccine messages significantly more positively than did those who had not received the vaccine (Dillard, 2013), and people who had received an HIV test in the last 12 months rated messages advocating HIV testing as higher in perceived effectiveness than did those who had never been tested (Davis et al., 2011). In Santa and Cochrane's (2008) study, anti-DUI (driving under the influence) PSAs were rated as more effective by participants who (pre-exposure) perceived DUI as relatively high in dangerousness than by participants with relatively low dangerousness perceptions. Chen, Grube, Bersamin, Waiters, and Keefe (2005) found that greater frequency of alcohol use was associated with greater perceived effectiveness of beer advertisements. Noar et al. (2010) reported that more consistent condom users rated safer-sex PSAs as more effective than did less consistent users. Weber, Westcott-Baker, and Anderson (2013) found that high-drug-risk respondents rated all manner of different antidrug messages as low in perceived effectiveness.

All these various findings are consistent with the idea that individuals' PME scores might be indicators of readiness for change, favorable predispositions toward the advocacy, susceptibility (or resistance) to influence, and the like.[9] Expressed in transtheoretical model terms (Noar, 2017;

Prochaska & Velicer, 1997), PME assessments might reflect the respondent's stage of change, such that individuals who have higher PME scores are more likely to be at more advanced stages. Indeed, in two studies that assessed stage of change, participants at more advanced stages were found to generally give more positive message assessments (Donovan, Leivers, & Hannaby, 1999; Santa & Cochran, 2008).

Briefly, then: It appears that individuals' PME scores are positively correlated with AME or AME-like assessments (attitude, intention, behavior) – positively correlated both with *post-exposure* AME measures (the usual sort of individual-level PME-AME correlation) and with *pre-exposure* AME-like measures (as in the studies just reviewed). Parsimony would look for one explanation of those two relationships – and the readiness-for-change interpretation of PME scores fits the bill.

Such an interpretation would point to a possible repurposing of PME measures in formative research. PME measures might turn out to be useful in identifying which individuals might be predisposed to change – and which might be expected to be relatively resistant. Where correlates of PME could be identified (e.g., demographic characteristics), those correlates might be used as markers of likely susceptibility (or resistance) to change. Such information might be used to guide decisions about how to design messages for different audiences, how to best target messages to ensure appropriate exposure, and so forth.

To place all this in a larger context: For present purposes, it's less important whether this specific hypothesis is sound (for some discussion, see Bigsby et al., 2013; Zhao et al., 2011) than that this line of thinking illustrates the ways in which the existence of generally positive individual-level PME-AME correlations might be re-interpreted. Those correlations cannot properly be interpreted as evidence that PME assessments accurately diagnose differences in message effectiveness, but they might be understood in other ways that point to novel uses for PME assessments in formative research.

Summary

Individuals' scores on measures of perceived message persuasiveness (PME) and actual persuasiveness (AME) are commonly positively correlated, but those individual-level correlations do not – and cannot – show that messages' relative standing on PME assessments matches their relative standing on AME assessments. The implication is that positive individual-level PME-AME correlations do not provide evidence supporting the message-pretesting practice of using PME assessments to identify relatively more effective messages.

The appropriate kind of evidence for underwriting the use of a message pretesting measure (whether PME assessments or some other method) is the ability of that measure to correctly identify relatively more effective messages. What is wanted to be predicted is not individuals' relative standing on AME but messages' relative standing on AME, and hence research on message pretesting measures should focus on whether messages' pretest scores are positively correlated with messages' AME scores, that is, whether those pretest measures provide accurate prediction of messages' relative AME standing.

However, even if PME assessments are not helpful in identifying which messages will be relatively more persuasive, they might turn out to have other uses in formative research. For example, PME measures might assess the degree to which individuals are likely to be disposed or resistant to change – as evidenced by the common positive correlations between individuals' PME and AME scores.

Notes

1. This is a simplified version of formative-research circumstances. Often more than two messages are under consideration, sometimes different sets of messages are being considered for different target audiences, and so forth. But the key issues can be laid out plainly by considering simpler situations.

2. Although the focus of interest here is message *persuasiveness* specifically, for reasons of convenience and familiarity, the acronym PME (perceived message effectiveness) will be used.
3. Again, this is a simplified version of the circumstance faced in formative research. For example, sometimes the decision involves not a choice between just two messages but rather selecting some subset of messages from a larger set (as when, say, 30 different antismoking arguments are under consideration, with an interest in identifying 10 to be deployed).
4. More carefully expressed: The correlation coefficient is invariant under transformations of the form $aX + b$ and $cY + d$, where a, b, c, and d are constants and a and c are positive (for some discussion, see Duncan & Litwiller, 2004).
5. If a concrete illustration is needed, consider these two hypothetical datasets. The mean for X is 70.0 in both sets. The mean for Y is 75.0 in one set and 25.0 in the other – but in each set, the correlation is .8693. The (X,Y) pairs in set #1 are (80,95), (75,75), (70,70), (65,69), and (60,66); in set #2 the pairs are (80,45), (75,25), (70,20), (65,19), and (60,16).
6. Related phenomena have been discussed as the "ecological fallacy" (Firebaugh, 2015; Robinson, 1950), which paradigmatically refers to the fallacy of supposing that the properties of individuals (e.g., individual-level correlations between variables) can be deduced from the properties of groups (e.g., group-level correlations between those variables). The present circumstance may represent the ecological fallacy in reverse, that is, the fallacy of supposing that the properties of messages (e.g., message-level correlations between variables) can be deduced from the properties of individuals (e.g., individual-level correlations between those variables).
7. Even when data were available that allowed comparing messages on AME, those data have sometimes been ignored. Brennan et al.'s (2014, p. 413) design assessed AME after exposing participants to one of two messages – but the analysis collapsed the data across message conditions.
8. This is not entirely suppositional. See, for example, Davis et al. (2017).
9. Although the effects of interest here concern judgments of perceived effectiveness (PME), parallel effects have been observed for ad liking ("attitude-toward-the-ad"): persons with more positive pre-exposure views about the advertised item have more positive assessments of corresponding ads (see, e.g., Chen et al., 2005; Edell & Burke, 1987, Study 1; Lienemann et al., 2019; Messmer, 1979; Unger, Johnson, & Rohrbach, 1995; cf. Wyllie, Zhang, & Casswell, 1998).

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Daniel J. O'Keefe http://orcid.org/0000-0003-1594-8892

References

Alvaro, E. M., Crano, W. D., Siegel, J. T., Hohman, Z., Johnson, I., & Nakawaki, B. (2013). Adolescents' attitudes toward antimarijuana ads, usage intentions, and actual marijuana usage. *Psychology of Addictive Behaviors*, *27*, 1027–1035. doi:10.1037/a0031960

Biener, L., Ji, M., Gilpin, E. A., & Albers, A. B. (2004). The impact of emotional tone, message, and broadcast parameters in youth anti-smoking advertisements. *Journal of Health Communication*, *9*, 259–274. doi:10.l080/10810730490447084

Biener, L., McCallum-Keeler, G., & Nyman, A. L. (2000). Adults' response to Massachusetts anti-tobacco television advertisements: Impact of viewer and advertisement characteristics. *Tobacco Control*, *9*, 401–407. doi:10.1136/tc.9.4.401

Bigsby, E., Cappella, J. N., & Seitz, H. H. (2013). Efficiently and effectively evaluating public service announcements: Additional evidence for the utility of perceived effectiveness. *Communication Monographs*, *80*, 1–23. doi:10.1080/03637751.2012.739706

Brennan, E., Durkin, S. J., Wakefield, M. A., & Kashima, Y. (2014). Assessing the effectiveness of antismoking television advertisements: Do audience ratings of perceived effectiveness predict changes in quitting intentions and smoking behaviours? *Tobacco Control*, *23*, 412–418. doi:10.1136/tobaccocontrol-2012-050949

Cappella, J. N. (2018). Perceived message effectiveness meets the requirements of a reliable, valid, and efficient measure of persuasiveness. *Journal of Communication*, *68*, 994–997. doi:10.1093/joc/jqy044

Chen, M.-J., Grube, J. W., Bersamin, M., Waiters, E., & Keefe, D. B. (2005). Alcohol advertising: What makes it attractive to youth? *Journal of Health Communication*, *10*, 553–565. doi:10.1080/10810730500228904

Choi, J., & Cho, H. (2016). Perceived message effectiveness, attitude toward messages, and perceived realism. In D. K. Kim & J. W. Dearing (Eds.), *Health communication research measures* (pp. 139–152). New York, NY: Peter Lang.
Dailey, R. M., McCracken, A. A., & Romo, L. K. (2011). Confirmation and weight management: Predicting effective levels of acceptance and challenge in weight management messages. *Communication Monographs*, *78*, 185–211. doi:10.1080/03637751.2011.564638
Davis, K. C., Duke, J., Shafer, P., Patel, D., Rodes, R., & Beistle, D. (2017). Perceived effectiveness of antismoking ads and association with quit attempts among smokers: Evidence from the tips from former smokers campaign. *Health Communication*, *32*, 931–938. doi:10.1080/10410236.2016.1196413
Davis, K. C., & Duke, J. C. (2018). Evidence of the real-world effectiveness of public health media campaigns reinforces the value of perceived message effectiveness in campaign planning. *Journal of Communication*, *68*, 998–1000. doi:10.1093/joc/jqy045
Davis, K. C., Nonnemaker, J., Duke, J., & Farrelly, M. C. (2013). Perceived effectiveness of cessation advertisements: The importance of audience reactions and practical implications for media campaign planning. *Health Communication*, *28*, 461–472. doi:10.1080/10410236.2012.69653
Davis, K. C., Nonnemaker, J. M., Farrelly, M. C., & Niederdeppe, J. (2011). Exploring differences in smokers' perceptions of the effectiveness of cessation media messages. *Tobacco Control*, *20*, 26–33. doi:10.1136/tc.2009.035568
Davis, K. C., Uhrig, J., Bann, C., Rupert, D., & Fraze, J. (2011). Exploring African American women's perceptions of a social marketing campaign to promote HIV testing. *Social Marketing Quarterly*, *17*(3), 39–60. doi:10.1080/15245004.2011.595536
Dillard, J. P. (2013). The effects of prior behavior on judgments of perceived message effectiveness: Evaluating HPV vaccine messages. *Journal of Vaccines and Vaccination*, *4*, 193. doi:10.4172/2157-7560.1000193
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*, 613–631. doi:10.1111/j.1460-2466.2007.00360.x
Donovan, R. J., Leivers, S., & Hannaby, L. (1999). Smokers' responses to anti-smoking advertisements by stage of change. *Social Marketing Quarterly*, *5*(2), 56–65. doi:10.1080/15245004.1999.9961046
Duncan, D. R., & Litwiller, B. H. (2004). The constancy of the correlation coefficient with respect to linear transformations. *Alabama Journal of Mathematics*, *28*, 9–13.
Durkin, S., Bayly, M., Cotter, T., Mullin, S., & Wakefield, M. (2013). Potential effectiveness of anti-smoking advertisement types in ten low and middle income countries: Do demographics, smoking characteristics and cultural differences matter? *Social Science & Medicine (1982)*, *98*, 204–213. doi:10.1016/j.socscimed.2013.09.022
Edell, J. A., & Burke, M. C. (1987). The power of feelings in understanding advertising effects. *Journal of Consumer Research*, *14*, 421–433. doi:10.1086/209124
Firebaugh, G. (2015). Ecological fallacy, statistics of. *International encyclopedia of the social & behavioral sciences* (Vol. 6, 2nd ed., pp. 865–867). doi:10.1016/B978-0-08-097086-8.44017-1
Grummon, A. H., Hall, M. G., Taillie, L. S., & Brewer, N. T. (2019). How should sugar-sweetened beverage health warnings be designed? A randomized experiment. *Preventive Medicine*, *121*, 158–166. doi:10.1016/j.ypmed.2019.02.010
Guillaumier, A., Bonevski, B., Paul, C., D'Este, C., Durkin, S., & Doran, C. M. (2017). Which type of antismoking advertisement is perceived as more effective? An experimental study with a sample of Australian socially disadvantaged welfare recipients. *American Journal of Health Promotion*, *31*, 209–216. doi:10.4278/ajhp.141125-QUAN-593
Hitchman, S. C., Mons, U., Nagelhout, G. E., Guignard, R., Mcneill, A., Willemsen, M. C., … Fong, G. T. (2012). Effectiveness of the European Union text-only cigarette health warnings: Findings from four countries. *European Journal of Public Health*, *22*, 693–699. doi:10.1093/eurpub/ckr099
Lienemann, B. A., Rose, S. W., Unger, J. B., Meissner, H. I., Byron, M. J., Baezconde-Garbanati, L., … Cruz, T. B. (2019). Tobacco advertisement liking, vulnerability factors, and tobacco use among young adults. *Nicotine & Tobacco Research*, *21*, 300–308. doi:10.1093/ntr/nty220
Lingwall, C., Nehl, E., Topuridze, M., Sturua, L., Maglakelidze, N., & Berg, C. J. (2018). Perceived effectiveness of differing health warning label messaging strategies among adults in the Republic of Georgia: One size does not fit all. *International Journal of Environmental Research and Public Health*, *15*, 2221. doi:10.3390/ijerph15102221
Mackert, M., Lazard, A., Champlin, S., Liang, M. C., Mabry, A., Stroever, S., … Watkins, L. (2014). "Take time. save lives. clean hands protect": A comparison of two hand hygiene health promotion posters. *American Journal of Infection Control*, *42*, 530–532. doi:10.1016/j.ajic.2014.01.017
Messmer, D. J. (1979). Repetition and attitudinal discrepancy effects on the affective response to television advertising. *Journal of Business Research*, *7*, 75–93. doi:10.1016/0148-2963(79)90027-4
Mouneyrac, A., Le Floch, V., Lemercier, C., Py, J., & Roumegue, M. (2017). Promoting responsible gambling via prevention messages: Insights from the evaluation of actual European messages. *International Gambling Studies*, *17*, 426–441. doi:10.1080/14459795.2017.1350198

Mowbray, F., Marcu, A., Godinho, C. A., Michie, S., & Yardley, L. (2016). Communicating to increase public uptake of pandemic flu vaccination in the UK: Which messages work? *Vaccine*, *34*, 3268–3274. doi:10.1016/j.vaccine.2016.05.006

Niederdeppe, J., Farrelly, M. C., Nonnemaker, J., Davis, K. C., & Wagner, L. (2011). Socioeconomic variation in recall and perceived effectiveness of campaign advertisements to promote smoking cessation. *Social Science & Medicine*, *72*, 773–780. doi:10.1016/j.socscimed.2010.12.025

Noar, S. M. (2017). Transtheoretical model and stages of change in health and risk messaging. *Oxford Research Encyclopedia of Communication*. doi:10.1093/acrefore/9780190228613.013.324

Noar, S. M., Barker, J., Bell, T., & Yzer, M. C. (in press). Does perceived message effectiveness predict the actual effectiveness of tobacco education messages? A systematic review and meta-analysis. *Health Communication*. doi:10.1080/10410236.2018.1547675

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*, 990–993. doi:10.1093/joc/jqy047

Noar, S. M., Bell, T., Kelley, D., Barker, J., & Yzer, M. C. (2018). Perceived message effectiveness measures in tobacco education campaigns: A systematic review. *Communication Methods and Measures*, *12*, 295–313. doi:10.1080/19312458.2018.1483017

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. doi:10.1136/tobaccocontrol-2014-051978

Noar, S. M., Palmgreen, P., Zimmerman, R. S., Lustria, M. L. A., & Lu, H.-Y. (2010). Assessing the relationship between perceived message sensation value and perceived message effectiveness: Analysis of PSAs from an effective campaign. *Communication Studies*, *61*, 21–45. doi:10.1080/10510970903396477

O'Donnell, N. H., & Willoughby, J. F. (2017). Photo-sharing social media for eHealth: Analysing perceived message effectiveness of sexual health information on Instagram. *Journal of Visual Communication in Medicine*, *40*, 149–159. doi:10.1080/17453054.2017.1384995

O'Keefe, D. J. (2018a). Message pretesting using assessments of expected or perceived persuasiveness: Evidence about diagnosticity of relative actual persuasiveness. *Journal of Communication*, *68*, 120–142. doi:10.1093/joc/jqx009

O'Keefe, D. J. (2018b). Whistling past the graveyard: Response to commentaries. *Journal of Communication*, *68*, 1001–1005. doi:10.1093/joc/jqy046

Pollard, C. M., Howat, P. A., Pratt, I. S., Boushey, C. J., Delp, E. J., & Kerr, D. A. (2016). Preferred tone of nutrition text messages for young adults: Focus group testing. *JMIR Mhealth and Uhealth*, *4*, 252–265. doi:10.2196/mhealth.4764

Popova, L., Neilands, T. B., & Ling, P. M. (2014). Testing messages to reduce smokers' openness to using novel smokeless tobacco products. *Tobacco Control*, *23*, 313–321. doi:10.1136/tobaccocontrol-2012-050723

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*, 38–48. doi:10.4278/0890-1171-12.1.38

Ramsay, I. S., Yzer, M. C., Luciana, M., Vohs, K. D., & MacDonald, A. W., III. (2013). Affective and executive network processing associated with persuasive antidrug messages. *Journal of Cognitive Neuroscience*, *25*, 1136–1147. doi:10.1162/jocn_a_00391

Record, R. A., Harrington, N. G., Helme, D. W., & Savage, M. W. (2018). Using the theory of planned behavior to guide focus group development of messages aimed at increasing compliance with a tobacco-free policy. *American Journal of Health Promotion*, *32*, 143–152. doi:10.1177/0890117116687467

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, *15*, 351–357. doi:10.2307/2087176

Santa, A. F., & Cochran, B. N. (2008). Does the impact of anti-drinking and driving Public Service Announcements differ based on message type and viewer characteristics? *Journal of Drug Education*, *38*, 109–129. doi:10.2190/DE.38.2.b

Sidhu, A. N. (2011). *Affective responses towards anti-tobacco advertisement concepts* (master's thesis). Retrieved from ProQuest Dissertations and Theses Global (ProQuest no. 1507182).

Sobani, Z., Nizami, S., Raza, E., Ul Ain Baloch, N., & Khan, J. A. (2010). Graphic tobacco health warnings: Which genre to choose? *International Journal of Tuberculosis and Lung Disease*, *14*, 356–361.

Thrasher, J. F., Arillo-Santillán, E., Villalobos, V., Pérez-Hernández, R., Hammond, D., Carter, J., … Regalado-Piñeda, J. (2012). Can pictorial warning labels on cigarette packages address smoking-related health disparities? Field experiments in Mexico to assess pictorial warning label content. *Cancer Causes and Control*, *23*(S1), 69–80. doi:10.1007/s10552-012-9899-8

Unger, J. A., Johnson, C. A., & Rohrbach, L. A. (1995). Recognition and liking of tobacco and alcohol advertisements among adolescents: Relationships with susceptibility to substance use. *Preventive Medicine*, *24*, 461–466. doi:10.1006/pmed.1995.1074

Van Dessel, P., Smith, C. T., & De Houwer, J. (2018). Graphic cigarette pack warnings do not produce more negative implicit evaluations of smoking compared to text-only warnings. *PLoS ONE*, *13*, e0194627. article no. e0194627. doi:10.1371/journal.pone.0194627

Weber, R., Westcott-Baker, A., & Anderson, G. L. (2013). A multilevel analysis of antimarijuana public service announcement effectiveness. *Communication Monographs*, *80*, 302–330. doi:10.1080/03637751.2013.788254

Willemsen, M. C. (2005). The new EU cigarette health warnings benefit smokers who want to quit the habit: Results from the dutch continuous survey of smoking habits. *European Journal of Public Health*, *15*, 389–392. doi:10.1093/eurpub/cki061

Wyllie, A., Zhang, J. F., & Casswell, S. (1998). Positive responses to televised beer advertisements associated with drinking and problems reported by 18 to 29-year-olds. *Addiction*, *93*, 749–760. doi:10.1046/j.1360-0443.1998.93574911.x

Yzer, M., LoRusso, S., & Nagler, R. H. (2015). On the conceptual ambiguity surrounding perceived message effectiveness. *Health Communication*, *30*, 125–134. doi:10.1080/10410236.2014.974131

Yzer, M., Rhodes, K., McCann, M., Harjo, J., Nagler, R. H., LoRusso, S. M., & Gollust, S. E. (2018). Effects of cultural cues on perceptions of HPV vaccination messages among parents and guardians of American Indian youth. *Preventive Medicine*, *115*, 104–109. doi:10.1016/j.ypmed.2018.08.021

Zhao, X., Strasser, A., Cappella, J. N., Lerman, C., & Fishbein, M. (2011). A measure of perceived argument strength: Reliability and validity. *Communication Methods and Measures*, *5*, 48–75. doi:10.1080/19312458.2010.547822