FORUM

# Whistling Past the Graveyard: Response to Commentaries

Daniel J. O'Keefe

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

All parties to this conversation want dependably diagnostic means of pretesting persuasive messages, but we vary in how optimistic we are about the prospects for the use of perceived-persuasiveness (PME) assessments in this enterprise. Three questions organize the following remarks.

The first is: Where is the burden of proof for the use of perceived-persuasiveness assessments in message pretesting? I believe that the current evidence sets presumption firmly against any general message pretesting use of PME measures. Across 35 studies, the mean rank-order correlation between messages' ranks on perceived and actual persuasiveness was almost literally zero (−.05; O'Keefe, 2018). To my eye, this means that one can no longer comfortably assume that a given PME assessment will be useful for message design decisions; justification is required.

In this regard, the heterogeneity in the reviewed PME assessments that Noar, Barker, and Yzer (this issue) find so troubling actually *strengthens* the case against the general diagnosticity of PME measures. (If all the PME measures had been identical, no general conclusions could have been underwritten.) Moreover, the moderator-variable analyses did not provide any reason to think that (at least some) more homogeneous sets of PME measures (e.g., ones in which the referent was always the self) would yield dramatically better diagnosticity.

And evidence has continued to accumulate. Bruneau, Kteily, and Falk (2018, p. 445) reported that lay predictions of the effectiveness of anti-prejudice videos were inaccurate, describing their results as "extending the literature on individuals' poor forecasting of effective interventions." Erlandsson, Nilsson, and Västfjäll's (in press, p. 23) study of charity appeals found that "people at least to some extent mistake themselves regarding the relative effectiveness of different appeals in that they incorrectly expect that positive appeals will … be more effective in increasing

---

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

donations than negative appeals." Thijsen et al. (2017) studied messages designed to encourage blood donors to adopt behaviors that would prevent undesirable side effects, but donors' pretest ratings did not identify the messages that were actually most effective. In Hoover, Johnson, Boghrati, Graham, and Dehghani's (2018, p. 13) research, "while people reported thinking that care and loyalty frames would increase donation behavior (Studies 2 and 3), we find no evidence for any such effects when people are actually exposed to these frames."

In short, it's just not plausible to believe that most perceived-persuasiveness measures are generally diagnostic of differences in actual persuasiveness.

However, even if presumption is set against this use of PME measures, their use in specific applications might still be justifiable. The presumption of innocence for U.S. criminal defendants is no bar to showing guilt in specific cases, but the prosecution must satisfy its burden of proof. Similarly here: any message-pretesting use of a PME measure requires justification, because presumption lies against it. To satisfy that burden of proof, I believe that what's required is evidence of the diagnosticity of relative actual message persuasiveness: evidence that when using a given PME assessment in a specific application, the usual design protocol—choosing the message rated higher on PME—will identify the message higher in AME.

The second question is: How might more diagnostic, application-specific perceived-persuasiveness measures be developed? Even if perceived-persuasiveness assessments are thought not to generally be diagnostic of differences in relative actual persuasiveness, measures of perceived persuasiveness might be shown to be diagnostic under a specified set of conditions—but which conditions?

To make progress on this front, Noar et al. (this issue) suggest that "PME studies should include a set of messages or message types with moderate or more expected variability, perhaps including control messages, rather than a narrow set of messages that may all score similarly on PME." If one's goal is to make PME measures look good, creating such artificial variability might be attractive. But if the goal is finding diagnostic pretesting procedures that are useful when trying to detect small but meaningful differences in actual effectiveness, that approach is not obviously helpful.[1]

Some might think that locating conditions that promote diagnosticity can be found in the phrasing of perceived-persuasiveness assessments, such as whether the assessment specifies the behavior of interest. I would look elsewhere: to the intersection of lay theories and the messages presented for assessment.

Consider the parallel with lay deception–detection research. Suppose people's naive beliefs are that behaviors A, B, and C are signs of deception. And suppose that, in fact, behavior A is indicative of deception, but B and C are not. Well, then, sometimes lay research participants would be good at deception detection (when behavior A was what varied in the stimulus materials), but sometimes they would perform badly (when B and C were varied).

Similarly, lay assessments of likely persuasiveness are similar. Those assessments reflect (implicit or explicit) lay theories of persuasion, and hence the diagnosticity of

those assessments will vary depending on the accuracy of the relevant lay beliefs in a given application. When lay perceptions of the persuasiveness of a set of messages are driven by (or correlated with) the same factors that influence the relative actual persuasiveness of those messages, then PME measures will be diagnostic, because the lay beliefs being relied upon for judgments of perceived persuasiveness are, in that application, accurate. But when lay perceptions of persuasiveness are driven by factors unrelated to those that influence relative actual persuasiveness, then PME measures will not be diagnostic, because the lay beliefs being relied upon for judgments of perceived persuasiveness are, in that application, inaccurate.

On this view, PME assessments should be very useful for assessing certain sorts of message variations but unhelpful for other message variations, depending on the accuracy of the lay theories concerning the variations in question. For example, if laypeople believe that adding vivid graphics to messages enhances persuasiveness, then their PME assessments of text-only and text-plus-graphics cigarette package warnings will vary accordingly. And if that lay theory is correct in that application, then perceived-persuasiveness assessments will be diagnostic of relative actual persuasiveness. But observing diagnosticity in such a case would not be evidence for any general diagnosticity of perceived-persuasiveness measures. Each application must be underwritten by evidence of its diagnosticity for the particular message variations under investigation.

The third question is: What evidence of diagnosticity is offered in the commentaries? No commentary offers evidence of the general diagnosticity of perceived-persuasiveness measures. Davis and Duke (this issue) argue that for "hard-hitting and graphic" antismoking messages, PME assessments are diagnostic of AME differences. Cappella (this issue) suggests that for certain antitobacco messages, PME measures predict various biobehavioral and behavioral outcomes.[2]

Even taken at face value, then, the commentaries do not justify any continued general use of perceived-persuasiveness assessments for message pretesting. There's only pleading on behalf of specific applications, as indeed is appropriate. But nothing here should embolden message designers generally to unthinkingly continue using perceived-persuasiveness measures as they have been.

Moreover, the application-specific evidence the commentaries do offer isn't always that compelling. Consider, for example, Cappella's invocation of Bigsby, Cappella, and Seitz's (2013) results. That evidence is imperfect, if only because each participant saw multiple public service announcements, and hence the effectiveness of any individual message could not be determined straightforwardly. And Cappella (this issue) is that "the degree of aggregate PME predicted intentions to quit and to reduce consumption of tobacco"—but the correlations were only 0.09 and 0.13, respectively, (Bigsby et al. 2013, p. 10, Table 1). Yes, the correlations were positive and statistically significant, but this is not strikingly impressive predictability. (The observed overall 58% diagnosticity is numerically larger than 50%, but it's not impressive either.)

Similarly, Davis and Duke point to (inter alia) Alvaro et al.'s (2013) research in support of a claim of PME-AME associations. But Alvaro et al.'s design did not provide ad-specific AME assessments (i.e., it was not possible to compare the

effectiveness of different ads in that study), which limits the extent to which those data can support claims about whether PME assessments accurately diagnose differences in message effectiveness.

## Summary

I continue to believe that the general use of perceived-persuasiveness assessments in message pretesting is not justified by the evidence in hand. But one might reasonably hope to see specific applications in which such assessments are demonstrably diagnostic.

## Notes

1 Astronomer #1: "This telescope can't distinguish different stars in this part of the sky; we need better resolution." Astronomer #2: "OK, let's point the telescope at a region where the stars are farther apart."
2 Note, however, that PME could predict intention, behavior, and other outcomes without necessarily being diagnostic of messages' relative AME. For example, if variation between respondents in PME (i.e., respondents' relative PME) reflects differences in their readiness for behavior change (with respondents on the cusp of change generally rating messages as more persuasive), PME and subsequent behavioral outcomes would be correlated, even if messages' relative PME was not diagnostic of messages' relative AME.

## 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

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

Bruneau, E., Kteily, N., & Falk, E. (2018). Interventions highlighting hypocrisy reduce collective blame of Muslims for individual acts of violence and assuage anti-Muslim hostility. *Personality and Social Psychology Bulletin, 44*, 430–448. doi:10.1177/0146167217744197

Cappella, J. N. (this issue). Perceived message effectiveness meets the requirements of a reliable, valid, and efficient measure of persuasiveness. *Journal of Communication*.

Davis, K. C., & Duke, J. C. (this issue). Evidence of the real-world effectiveness of public health media campaigns reinforces the value of perceived message effectiveness in campaign planning. *Journal of Communication*.

Erlandsson, A., Nilsson, A., & Västfjäll, D. (in press). Attitudes and donation behavior when reading positive and negative charity appeals. *Journal of Nonprofit & Public Sector Marketing*. doi:10.1080/10495142.2018.1452828

Hoover, J., Johnson, K., Boghrati, R., Graham, J., & Dehghani, M. (2018). Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology, 4*, 9. doi:10.1525/collabra.129

Noar, S. M., Barker, J., & Yzer, M. (this issue). Measurement and design heterogeneity in perceived message effectiveness studies: A call for research. *Journal of Communication*.

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*, 120–142. doi:10.1093/joc/jqx00

Thijsen, A., Fisher, J., Gemelli, C. N., Bell, B., Davison, T. E., & Masser, B. M. (2017). Facilitating donor compliance with strategies to prevent vasovagal reactions: Comparison of web-based and in-center approaches. *Transfusion, 57*, 2449–2457. doi:10.1111/trf.14243