FORUM

Measurement and Design Heterogeneity in Perceived Message Effectiveness Studies: A Call for Research

Seth M. Noar[1,2], Joshua Barker[1], & Marco Yzer[3]

1 School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
2 Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
3 Hubbard School of Journalism and Mass Communication, University of Minnesota, Minneapolis, MN 55455, USA

Ratings of perceived message effectiveness (PME) are commonly used during message testing and selection, operating under the assumption that messages scoring higher on PME are more likely to affect actual message effectiveness (AME)—for instance, intentions and behaviors. Such a practice has clear utility, particularly when selecting from a large pool of messages.

Recently, O'Keefe (2018) argued against the validity of PME as a basis for message selection. He conducted a meta-analysis of mean ratings of PME and AME, testing how often two messages that differ on PME similarly differ on AME, as tested in separate samples. Comparing 151 message pairs derived from 35 studies, he found that use of PME would only result in choosing a more effective message 58% of the time, which is little better than chance. On that basis, O'Keefe concluded that "message designers might dispense with questions about expected or perceived persuasiveness (PME), and instead pretest messages for actual effectiveness" (p. 135). We do not believe that the meta-analysis supports this conclusion, given the measurement and design issues in the set of studies O'Keefe analyzed.

Measurement issues

One of the most vexing issues in the PME literature has been a lack of clear conceptualization of what PME is and how to measure it (Yzer, LoRusso, & Nagler, 2015). To examine this, we recently conducted a systematic review of the PME measures used in tobacco education campaigns (Noar, Bell, Kelley, Barker, & Yzer, 2018).

Corresponding author: Seth M. Noar; e-mail: noar@unc.edu.

Across 75 studies, we found substantial heterogeneity in PME measures, including (a) use of 16+ persuasive constructs; (b) assessment of message perceptions or effects perceptions; (c) inclusion of a target referent; and (d) referencing of behavior. In essence, our results indicate little consensus on how to measure PME, even in a literature focused only on anti-smoking media campaigns. When constructs are poorly measured, prediction suffers. For instance, poor measurement of attitudes and risk perceptions clouded the association between these constructs and behavior, but when measurement improved, so did prediction (Brewer et al., 2007; Fishbein & Ajzen, 2010).

In O'Keefe's (2018) set of studies—spanning at least 18 different topics—there is similar heterogeneity. Some studies used a single, generic PME item that asked participants to rate how "effective" a message was—an item with unknown validity. Other single items or scales asked about message preference, strong or weak reasons, expected compliance, willingness to engage in the behavior, predicted purchase likelihood, message satisfaction, importance, or motivation. While several scales included an item such as "persuasive," some measures included items with less relevance to effectiveness (e.g., visibility, understandability, usefulness). Many measures did not specify a target referent, potentially biasing PME ratings (Dillard & Ye, 2008). Thus, the PME measures in O'Keefe's (2018) analysis differ so much that we cannot say—with any confidence—that they are measuring the same construct. We should thus interpret the analyses testing the diagnosticity of PME with caution.

Design issues

Several studies in the meta-analysis showed a lack of PME-AME correspondence in measurement, samples, or design. For instance, several PME measures asked participants to rate how a message would affect *others*, while the AME assessment was concerned with message effects on the *participant*. In addition, while the PME and AME studies used the same messages, the samples were sometimes drawn from different sources, creating non-comparability though the use of different recruitment methods (e.g., MTURK for PME vs. community sample for AME) or entirely different populations (experts for PME vs. target audience for AME). The studies were also not designed to test the diagnosticity of PME, and many of the PME studies were preliminary, with sample sizes of 40 or less. Further—and perhaps most importantly—in many studies, only tiny differences in PME means were observed. For instance, when a message pair scored similarly on PME ($M = 3.07$ and 3.09), but with corresponding AME means that differed in direction from PME ($M = 3.14$ and 3.08, respectively; Pettigrew et al., 2016), this was treated as a failure of PME in the meta-analysis. This strikes us as an overly conservative test, as PME is much more likely to provide useful guidance when differences are larger. In fact, O'Keefe (2018) found that when differences in PME were statistically significant, the diagnostic rate improved from 53% to 67%. This illustrates a very important point: namely, that PME studies should include a set of messages or message types with moderate or large

expected variability, perhaps including control messages, rather than a narrow set of messages that may all score similarly on PME.

Where we go from here

While we disagree with O'Keefe's dismissal of PME, we believe he has done the field a service by highlighting the need for more conceptual, measurement, and validation work on PME. To date, some PME measures appear to lack the rigorous psychometric work that is needed. More research is needed to understand the role of underlying persuasive constructs, message perceptions versus effects perceptions, influence of target referents, and referencing of behavior in PME measures (Noar et al., 2018). We also need additional, rigorously-designed validation studies of PME. Such studies will advance our understanding of the role of PME ratings in message selection, an area that has broad applicability across the communication field. In the end, improved measures will increase the likelihood of selecting more effective messages, better realizing the potential of communication to do good.

Acknowledgments

References

Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, *26*, 136–145. doi:10.1037/0278-6133.26.2.136

Dillard, J. P., & Ye, S. (2008). The perceived effectiveness of persuasive messages: Questions of structure, referent, and bias. *Journal of Health Communication*, *13*, 149–168. doi:10.1080/10810730701854060

Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. New York, NY: Psychology Press.

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

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. doi:10.1093/joc/jqx009

Pettigrew, S., Jongenelis, M. I., Glance, D., Chikritzhs, T., Pratt, I. S., Slevin, T., … Wakefield, M. (2016). The effect of cancer warning statements on alcohol consumption intentions. *Health Education Research*, *31*, 60–69. doi:10.1093/her/cyv067

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