EXAMINING PERCEIVED MESSAGE EFFECTIVENESS AS A MARKER FOR THE IMPACT OF BRIEF HEALTH BEHAVIOR INTERVENTIONS

Sabeeh A. Baig

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Health Behavior in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:
Noel T. Brewer
Nisha C. Gottfredson
Allison J. Lazard
Seth M. Noar
Kurt M. Ribisl
ABSTRACT

Sabeeh A. Baig: Examining Perceived Message Effectiveness as a Marker for the Impact of Brief Health Behavior Interventions
(Under the direction of Noel T. Brewer)

Introduction. Interventionists often select health communication messages based on audience ratings of perceived message effectiveness (PME). I sought to examine the roles of message perceptions (persuasive potential) and effects perceptions (potential for behavior change), two types of PME, in message selection using anti-smoking messages about cigarette smoke chemicals (chemical messages) as a case study.

Methods. In three papers, I examined several aspects of the validity of the UNC PME Scale focusing on effects perceptions. The first paper used data from three national samples of adults (n = 999 and n = 1,692) and adolescents (n = 869). The second and third papers used data from a three-week trial of the impact of chemical messages among 703 U.S. adult smokers. At the final visit, a survey assessed message perceptions, effects perceptions, quitting behaviors and quitting antecedents drawn from the UNC Tobacco Warnings Model (TWM).

Results. In the first paper, the UNC PME Scale demonstrated strong psychometric properties in diverse populations and across varied chemical messages. In the second paper, message perceptions demonstrated predictive validity only with an early behavioral antecedent from the TWM (attention to the message). However, effects perceptions demonstrated predictive validity with four later antecedents from the TWM: negative affect; thinking about the chemicals in cigarette smoke, or harms of smoking; and quit intentions. Effects perceptions also demonstrated predictive validity with butting out or forgoing a cigarette and quit attempts. In the
third paper, effects perceptions, but not message perceptions, mediated the impact of chemical messages on these three quitting behaviors, although the corresponding effect sizes were small to medium.

**Conclusions.** Effects perceptions, but not message perceptions, were a proxy for chemical messages’ impact on three quitting behaviors. This finding supports the diagnostic value of effects perceptions in formative research on messages seeking to change smoking and, potentially, other behaviors. The distinct patterns of predictive validity further suggest that effects perceptions are more relevant to behavior change than message perceptions.
ACKNOWLEDGMENTS

I would like to express immense gratitude to my dissertation committee for guiding me through this process. Their thoughtful comments greatly improved the quality of my work. I especially thank Dr. Brewer, my advisor and committee chair, for his support throughout my training and for always pushing me to clarify my thinking and presenting it in a way that is accessible to everyone. Noel, you have been an incredible advocate and friend, and I look forward to working together on other topics in the future.

To my parents, I am incredibly indebted to you for showing me the importance of service to others, which has always been my motivation for working in public health. I also appreciate that I can count on my hands the number of days in the last four or so years that I did not talk to either one of you, if only for a few minutes. Thank you for enabling all my successes—they are yours more than mine.

Finally, I would like to thank the open source community for developing the software tools that enabled all aspects of my research. I hope to eventually make some contributions to those tools so that other users can continue to benefit from them as I have.
# TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................. ix

LIST OF FIGURES .............................................................................................................. x

CHAPTER 1: OVERVIEW ..................................................................................................... 1

CHAPTER 2: CONCEPTUAL REVIEW OF PERCEIVED MESSAGE EFFECTIVENESS ........... 5

Conceptual criteria for perceived message effectiveness measures ................................. 6

Sensitive to differences in messages ................................................................................. 7

Predicts behavior ............................................................................................................... 8

Focuses on behavior ......................................................................................................... 13

About the respondent ....................................................................................................... 16

Psychometric criteria for perceived message effectiveness measures ............................ 17

Construct validity ............................................................................................................. 17

Minimal measurement error ......................................................................................... 18

Brief in length .................................................................................................................. 20

Discussion ....................................................................................................................... 21

CHAPTER 3: VALIDATION OF THE UNC PERCEIVED MESSAGE EFFECTIVENESS SCALE .... 23

Methods ............................................................................................................................. 26

Participants ....................................................................................................................... 26

Procedures ....................................................................................................................... 28

Measures .......................................................................................................................... 28

Data analysis ..................................................................................................................... 29

Results ............................................................................................................................... 32

Psychometric properties ................................................................................................. 32
Differential item and test functioning ................................................................. 33
Construct validity ............................................................................................. 34
Discussion ......................................................................................................... 35
Strengths and limitations ............................................................................... 36
Conclusions ..................................................................................................... 37

CHAPTER 4: RESEARCH DESIGN AND METHODS ............................................ 43
Overview of parent study ............................................................................... 43
Data preparation .............................................................................................. 44
Analysis plan for Aim 2 .................................................................................. 45
Analysis plan for Aim 3 .................................................................................. 55
Power calculations .......................................................................................... 57

CHAPTER 5: PREDICTIVE VALIDITY OF MESSAGE PERCEPTIONS AND
EFFECTS PERCEPTIONS IN THE CONTEXT OF ANTI-SMOKING MESSAGES ........ 61
Methods ........................................................................................................... 64
Participants and procedures ......................................................................... 64
Measures .......................................................................................................... 65
Data analysis .................................................................................................... 66
Results ............................................................................................................. 68
Dimensionality ............................................................................................... 68
Predictive validity .......................................................................................... 70
Discussion ......................................................................................................... 71
Conclusions ..................................................................................................... 74

CHAPTER 6: MESSAGE PERCEPTIONS AND EFFECTS PERCEPTIONS AS
PROXIES FOR IMPACT ON BEHAVIOR ............................................................ 80
Methods ........................................................................................................... 83
Participants and procedures ......................................................................... 83
Measures ............................................................................................................................... 84
Data analysis .......................................................................................................................... 85
Results .................................................................................................................................. 87
Mediation by message perceptions ....................................................................................... 87
Sequential mediation by message perceptions through effects perceptions ...................... 88
Mediation by effects perceptions ......................................................................................... 88
Discussion ............................................................................................................................ 89
Conclusion ............................................................................................................................ 91

CHAPTER 7: FUTURE DIRECTIONS FOR OPTIMIZING USE OF PERCEIVED
MESSAGE EFFECTIVENESS IN FORMATIVE RESEARCH ................................................. .98
Message perceptions as a moderator .................................................................................... .98
Effects perceptions and health promoting behaviors .......................................................... 100
Effects perceptions item pool ............................................................................................ 101
Strengths and limitations ..................................................................................................... 102
Conclusions .......................................................................................................................... 105

APPENDIX: SURVEY MEASURES FOR PROPOSED ANALYSES IN AIMS 2 AND 3 .......... 106
REFERENCES ......................................................................................................................... 109
LIST OF TABLES

Table 3.1. Messages about the chemicals in cigarette smoke .................................................. 38

Table 3.2. Participant characteristics .......................................................................................... 39

Table 3.3. Factor loadings for the UNC Perceived Message Effectiveness Scale ....................... 40

Table 4.1. Construct validators categorized according to relevance in message processing based on the Message Impact Framework and UNC Tobacco Warnings Model .......................................................................................................................... 58

Table 5.1. Measures for all construct validators with ordinal reliability coefficients for multi-item scales ........................................................................................................................................ 75

Table 5.2. Participant characteristics .......................................................................................... 77

Table 6.1. Measures for all outcomes in multiple mediation analyses with ordinal reliability coefficients for multi-item scales ......................................................................................................................... 94

Table 6.2 Participant characteristics .......................................................................................... 95
LIST OF FIGURES

Figure 3.1. Information curves for the UNC Perceived Message Effectiveness Scale ................................................................. 41

Figure 3.2. Effect sizes for differential item functioning and differential test functioning exhibited by the UNC Perceived Message Effectiveness Scale .............................................................................. 42

Figure 4.1. Cigarette pack labels placed on smokers’ cigarette packs in intervention and control arms ............................................................. 59

Figure 4.3. Complex mediation model of the impact of chemical messages on micro-quitting through effects perceptions ........................................ 60

Figure 5.1. Confirmatory factor analytic model of message perceptions and effects perceptions .............................................................................. 78

Figure 5.2. Predictive validity of message perceptions and effects perceptions with behavioral antecedents from the UNC Tobacco Warnings Model, quitting behavior, and related constructs ............................................................. 79

Figure 6.1. Messages placed on smokers’ cigarette packs in the intervention and control arms .............................................................................. 93

Figure 6.2. Path diagram for structural equation model with mediated effects of chemical messages on a given outcome through message perceptions alone, message perceptions and effects perceptions sequentially, and effects perceptions alone .............................................................................. 96

Figure 6.3. Standardized indirect effects of chemical messages through message perceptions and effects perceptions .............................................................................. 97
CHAPTER 1: OVERVIEW

Brief health behavior interventions often rely on high-impact health messages, but developing such messages pose unique methodological challenges. First, it is difficult to identify health messages that are promising in the early stages of development without longitudinal data on their behavioral impact that can only be collected after the messages have been disseminated in the field. Second, brief exposure to a health message, as in an online or phone survey experiment, usually does not lead to meaningful changes in cognitive mediators of behavior from psychological theories that an effective message should change. Third, evaluation studies that use extensive self-report measures and ask participants to rate many brief messages may impose a high cognitive burden on participants.

In order to overcome the aforementioned methodological challenges, some researchers have assessed perceived message effectiveness (PME) as an early indicator of the importance of a health message in changing behavior specifically. As PME is applicable to brief messages in general, a common definition is the "extent to which a message recipient believes that a health message will affect him or her personally in terms of the particular message objectives." The theoretical literature on PME references two operationalizations of the construct, one that focuses on message perceptions and the other on effects perceptions. The former is rooted in communication theories of persuasion and focuses on message features that would lead to further processing of a health message and subsequent attitude changes. The latter is rooted in research on intentions, risk perceptions, and other potential mediators of behavior and focuses on the likelihood of a health message actually changing a referenced behavior among message
recipients. Many existing measures of PME combine aspects of both operationalizations. A major barrier to theoretically informed use of PME is the lack of scientific consensus on any relationship between PME and behavioral impact. PME may mediate the impact of brief messages on behavior; or behavior may mediate the impact of messages on PME. Alternatively, PME and behavior may be on separate causal pathways while being strongly correlated. Another major barrier to appropriately using PME is the dearth of data on the validity of the distinction between the two types of PME. Without this data, it is difficult to determine whether brief messages should be evaluated on both types of PME and, if so, how the types of assessments correspond to stages of message development; or only one type and, if so, which.

In my dissertation, I propose to examine PME as a psychological marker of the behavioral impact of health messages specifically in the context of brief messages about the chemicals in cigarette smoke (chemical messages). Chemical messages are an important case study because tobacco use is the leading cause of preventable deaths in the US, with cigarette smoking accounting for the majority of lung cancer deaths among men and women. The main source of harm from cigarette smoking is the toxic chemicals in cigarette smoke, many of which are known carcinogens. The public is interested in learning more about the chemicals in cigarette smoke, and awareness about these chemicals is associated with greater discouragement from wanting to smoke.

This research will involve three large national samples of smokers and non-smokers: an online convenience sample of 1,035 adults (Study 1), an online probability sample of 1,758 adults (Study 2), and an online probability sample of 879 adolescents (Study 3). Participants rated at least two chemical messages in each of the three studies. This research will also examine data from the parent randomized controlled trial (Study 4) in which 719 U.S. adult smokers rated
three novel anti-smoking messages either about chemicals or littering cigarette butts, rotated on their cigarette packs. The specific aims of the dissertation are to:

Aim 1. Establish the psychometric properties of a PME scale that focuses on effects perceptions. Activities to address this aim:

1. Conduct an item analysis of the UNC PME scale to assess its reliability and construct validity in Study 1.
2. Examine differential item functioning and differential test functioning in the UNC PME scale by key grouping variables related to smoking disparities in Study 1.
3. Assess the construct validity of the UNC PME Scale using multitrait-multimethod confirmatory factor analysis in Study 1.
4. Replicate the results from Study 1 in Studies 2 and 3.

Aim 2. Examine the construct validity of two PME scales, one that focuses on effects perceptions and the other on message perceptions. Activities to address this aim:

1. Examine the dimensionality of items from the UNC PME Scale and an established message perceptions scale\(^4\) using multiple-group confirmatory factor analysis in Study 4.
2. Compare the strength of the correlations between both scales with hypothesized constructs (drawn primarily from the UNC Tobacco Warning Model and the Message Impact Framework) using structural equation modeling in Study 4.

Aim 3. Identify whether PME focusing on effects or message perceptions is a better proxy for behavioral outcomes. Activities to address this aim:

1. Empirically test a structural model for the impact of chemical messages on micro-quitting simultaneously through both types of PME as measured by the UNC PME Scale and the established message perceptions scale in Study 4.
2. Replicate the results for three additional outcomes in Study 4: quit intentions, smoking-related social interactions, and chemical information seeking.

The proposed research will advance the theory of PME so that researchers can confidently use PME to develop brief messages that optimally impact behavior. First, it will clarify the psychometric properties, including the construct validity, of two types of PME that researchers have used interchangeably in message development. Second, it will provide evidence for the roles that both types of PME may have in message processing and impact. The main results of the proposed research should be generalizable to the use of PME in other contexts besides cigarette smoking, such as breast cancer screening and human papillomavirus vaccination.

This dissertation has six additional chapters. Chapter 2 is a conceptual review of the empirical and theoretical literature on PME and develops the conceptual framework for the proposed research. Chapter 3 is a validation study of the UNC PME Scale completed in fulfillment of Aim 1, and Chapter 4 lays out the analytic framework for Aims 2 and 3. Chapters 5 and 6 are two comparative validation studies involving effects and message perceptions completed in fulfillment of Aims 2 and 3, respectively. Finally, Chapter 7 provides additional insights about PME by synthesizing across all three aims.
CHAPTER 2: CONCEPTUAL REVIEW OF PERCEIVED MESSAGE EFFECTIVENESS

Interventions in health psychology and health communication often rely on high-impact health messages. Developing brief messages that can change risky health behaviors is a multi-stage process, often beginning with a pool of candidate messages. It is critical to identify the most promising messages during the early stages of development. However, assessing the behavioral impact of a health message is exceedingly resource and time-intensive and requires longitudinal data that can usually only be collected after the message has been disseminated. In behavioral contexts, *perceived message effectiveness* (PME) refers to the extent that a message recipient thinks that a message will affect their own behavior, potentially through effects on related attitudes, beliefs, and intentions. Researchers have used PME to select important messages in many contexts, including cigarette smoking and marijuana use. The theoretical literature on PME robustly shows that it is correlated with attitudes and behavioral intentions in a variety of substantive contexts (e.g., nutrition and dental hygiene). Thus, researchers have treated selecting important messages based on PME ratings as a sufficient strategy for winnowing down a pool of candidate messages during formative research.

The theoretical literature on PME does not have a consensus on how to define the construct, posits different psychological mechanisms for how PME may be related to behavior, and is largely silent on the relevant measurement challenges. In this review, I identify conceptual and psychometric criteria for the efficient use of PME in behavioral contexts and summarize key studies corresponding to each. I also make recommendations for how to improve the
measurement of PME with respect to each criterion. I use brief pack messages about the toxic chemicals in cigarette smoke (chemical messages) as a case study throughout this review because such messages are a recent example of a health communication intervention that relies on high-impact messages that were selected using PME. Chemical messages are also an ideal case study because a sizable literature has emerged on developing such messages, their behavioral impact, and underlying psychological mechanisms. Ultimately, this review seeks to provide an agenda for future research on PME that will improve its use in increasingly common online message evaluation studies in which participants are briefly exposed to multiple messages.

**Conceptual criteria for perceived message effectiveness measures**

Researchers have conceptualized PME in several different ways, but all conceptualizations share the common goal of identifying promising messages. I begin this section with a brief review of the corresponding definitions to concisely introduce the challenges in using PME in formative research. Many definitions of PME lack conceptual clarity and specificity. Two examples are, “Evaluative assessments of each message’s persuasive qualities,” and, “The extent to which a message has been favorably received and evaluated.” A third definition, "A summative perception of message quality based on various message content and format elements, which can include the soundness, strength, and/or novelty of the presented evidence and arguments as well as edits, cuts, visual images, etc," additionally treats design elements that may influence PME as its indicators and suggests that the judgment is multidimensional. The previous three definitions are largely silent on the role of the message recipient and the target behavior. Two more definitions, “An estimate of the degree to which a persuasive message will be favorably evaluated—in terms of its persuasive potential—by
recipients of that message,”49 and, “The extent to which a message recipient believes that a health message will affect him or her personally in terms of the particular message objectives,”152 focus on the message recipient, but do not refer to the target behavior explicitly. Thus, some of the challenges in using PME are theoretically unclear definitions, inconsistent use of referents, and unknown dimensionality. I address each of these challenges and others in the next few sections by focusing on the qualities that an adequate measure of PME should have.

**Sensitive to differences in messages**

A measure of PME should practically differentiate among candidate health messages. Formative research on chemical messages often relies on brief exposure to such messages, for example, in an online message evaluation study. Brief exposure does not usually impact constructs that effective messages can be expected to change (e.g., risk perceptions or quit intentions) in part because of the cognitive constraints (e.g., limited viewing time or distraction) associated with such message testing scenarios.134,145 PME is an exception, as prior experimental studies have shown that brief exposure to health messages impacts it. For example, a study in which U.S. young adults evaluated 12 oral health promotion messages found that PME ratings sufficiently differentiated between the individual messages.2 As another example, a study in which U.S. adolescents evaluated 30 anti-drug public service announcements (PSAs) found that 16 of the anti-drug PSAs had higher PME ratings than the control, nine had lower PME ratings, and wide variation in mean PME ratings of all PSAs.60 While some existing measures of PME practically differentiate among messages leading researchers to generally support its use in decision making about health messages,11,49 at least one researcher has recommended against using PME ratings in formative research on health messages due to PME measures that do not practically differentiate among messages.114
It is not enough for a PME measure to just be sensitive to health messages because the early stages of development often focus on identifying meaningful content, potent communication frames, and other useful message features. A PME measure should also practically differentiate among message elements. For example, the recent development of chemical messages proceeded through a number of stages including the selection of discouraging chemicals, comparing communication frames, as well as testing different combinations of elements together in a single message. Prior studies have shown that familiar toxic chemicals in cigarette smoke discourage people from wanting to smoke more than unfamiliar chemicals. Communication frames about the health harms associated with exposure to chemicals or common unappealing products that chemicals are found in discourage smoking more than only mentioning a chemical name by itself. A PME measure should be sensitive to toxic chemicals and corresponding communication frames and produce ratings that are sufficient to rank individual chemical messages that vary by these features. As another example, prior studies have found that pictorial cigarette warning increase quit intentions more than text-only warnings. Thus, a PME measure should be sensitive to the presence of graphics in cigarette warning labels. As a theoretical example, in substantive contexts where tailoring or targeting increase the effectiveness of messages, a PME measure should produce ratings that are higher for tailored or targeted messages than general messages. These examples demonstrate that, for a given context, a PME measure should be sensitive to a variety of message elements in order for that measure to be applicable in the early and late stages of message development.

Predicts behavior

A general consensus among researchers is that PME should predict behavior for it to be
meaningful in message selection. A small number of recent longitudinal studies have examined the relationships between PME and intentions and or behavior mostly in the context of cigarette smoking. In a four-month study that exposed U.S. adult smokers to at least one of 14 *Tips From Former Smokers* campaign advertisements, PME at baseline predicted quit attempts at follow-up. A two-week study that exposed Australian adult smokers to anti-smoking advertisements found that PME predicted changes in quit intentions from baseline. Another two-week study in which U.S. adult smokers saw smoking cessation advertisements found that baseline PME predicted quit intentions at follow-up. In a three-week study that exposed Australian adult smokers to anti-smoking advertisements, researchers found that one measure of PME predicted quit intentions and changes in smoking behavior at follow-up after controlling for baseline intentions while a second measure did not predict either outcome. Outside of tobacco control, 428 African-American women evaluated social marketing and educational advertisements from the *Take the Charge. Take the Test.* campaign about HIV testing. The four-week study found that PME predicted 6-month and 12-month HIV testing intentions at follow-up after controlling for baseline intentions and the most recent HIV test. However, PME did not predict HIV testing at follow-up. These studies generally support the predictive validity of PME, and further evidence should come from longitudinal studies with diverse substantive contexts. These studies generally support the predictive validity of PME, although this may vary depending on the measure being used.

A larger number of cross-sectional studies in a number of contexts, not just cigarette smoking, provide additional evidence of the predictive validity of PME. They generally do so by examining the relationship of PME with intentions, which is the strongest predictor of behavior, because behavior is difficult to measure in a cross-sectional study. In a study
of U.S. adult smokers who watched four PSAs about smoking, researchers found that PME aggregated over the four PSAs was associated with overall intentions to quit smoking. A study in which U.S. adults viewed video messages promoting colonoscopy and had the option to visit a colonoscopy information page afterwards found that PME was associated with intention to seek information about colonoscopy, which in turn was associated with information-seeking behavior. A meta-analysis of 40 studies found a modest overall correlation between PME and behavioral attitudes, the main predictor of behavior change from the conceptual model of PME that informed their study. This meta-analysis included studies of messages about a range of topics, including nutrition, self-control, and sexually transmitted diseases, presented on various forms of media. Another meta-analysis of 23 studies on the impact of verbal person-centered messages on social support outcomes found that PME was strongly correlated with social support outcomes although this effect was based on only two of the 23 studies included in the meta-analysis.

In addition to the growing evidence for the predictive validity of PME in a variety of behavioral contexts, a few studies have shown that PME is a mediator of the impact of health messages on behavior. For example, four cross-sectional studies found positive to strong support that PME mediated the effect of message exposure on attitude towards the behavior in the context of 15 messages about separate behaviors including flossing, binge drinking, television addiction, and exercising. An experimental study from the same authors found strong support for PME mediating the impact of message exposure on intentions in the context of two messages about risky sexual behaviors and social networking. In addition, a seven-week evaluation study of 128 intervention strategies for encouraging healthy food choices found that PME mediated the effect of exposure to an intervention on its acceptance.
selected interventions provided nutritional information to participants directly via product labeling or campaigns or indirectly via announcing incentives or disincentives to guide food choices. Finally, a previously mentioned meta-analysis of 23 studies found that PME mediated the effect of verbal person-centered messages on social support outcomes.

Despite the cross-sectional and longitudinal evidence for the predictive validity of PME, understanding the psychological mechanisms underpinning PME is essential to appropriately using it in message selection. To this end, researchers continue to debate what it means for PME to be a mediator of the impact of messages on behavior or even just a predictor of behavior. Much of the earlier conceptual work treats PME as a necessary and a sufficient condition for a health message to affect behavior. This conceptualization of PME is based on communication theories of persuasion, such as the Elaboration Likelihood Model (ELM). A number of these theories identify attitudes as a major proximal determinant of behavior and define persuasion as changing attitudes that are relevant to behavior change. According to the ELM, people process messages centrally via elaboration or peripherally via heuristics, with central processing having a greater likelihood of causing sustained persuasion. Central and peripheral processing exist on a continuum with central processing dominating when motivation and ability to process messages are high and peripheral processing dominating when both are low. Motivation and ability vary as functions of message features such as argument quality and understandability. The earlier theoretical work essentially identifies PME as an indicator of the extent to which a message has these and other desirable features that motivate and facilitate higher levels of elaboration. Thus, PME is a necessary condition for message impact in that it functions as a gatekeeper to higher levels of elaboration. It is also a sufficient condition for message impact because, when considering population-level effects, some percentage of the
people who view a message with high PME and will centrally process and be persuaded by it. Accordingly, interventionists have often selected promising messages based on PME evaluations without further post-testing of their behavioral impact.

In more recent conceptual work, researchers have generally walked back the claim that PME is a necessary and sufficient condition for a message to achieve its desired outcome. This is partly because the claim implies that a person must think that a health message is effective in order for that message to impact behavior and implies a high level of elaboration on the part of the viewer. Print advertising and marketing often contain health messages, providing naturalistic exposure to such messages. Consumer psychological research on how people view and process print marketing and advertising shows that people pay little to no attention to messages embedded in ads\textsuperscript{118} and process them heuristically with minimal elaboration.\textsuperscript{12,55} In addition, research at the intersection of economics, neuroscience, and psychology has demonstrated the importance of affect\textsuperscript{135,136} and emotion\textsuperscript{115,137} as well as feeling\textsuperscript{29,30} as sources of information in processing a variety of stimuli and decision making. These streams of research suggest that people view a message for a short period of time and use their initial affective and emotional responses to and feelings towards that message to summarily understand it. Although the level of elaboration is relatively low in this type of processing, people likely achieve a level of understanding that is necessary for a message to impact behavior at all. Therefore, it can be argued that PME that captures this overall impression of a message mediates the impact of a message as a necessary, but not a sufficient condition. This represents the ideal scenario in which PME is a marker for the likelihood that a given message will impact behavior. In this scenario, researchers could use PME to select a few promising health messages from a much larger pool of candidate messages that would then be further tested in a longitudinal experiment. These
observations correspond to a recent meta-analysis suggesting that PME evaluations may not substitute actual studies of behavioral impact.\textsuperscript{112}

Another possibility is that messages affect PME through the outcome that they are designed to change. In other words, people process a message, which causes changes in the corresponding behavior to varying degrees. As a result, when asked about PME, people infer the extent to which a message impacted them by first thinking about their behavior. This account of PME corresponds to Self-Perception Theory,\textsuperscript{9,10,128} and, if true, would imply that PME is a retrospective judgment. To my knowledge, only one study, also mentioned previously, has examined whether behavior predicts PME.\textsuperscript{48} This study found strong evidence against PME being a retrospective judgment in the context of messages about drinking and driving, television addiction, and other behaviors.\textsuperscript{48} More studies are needed to confirm that PME is not a retrospective judgment. The other possibility is that messages affect behavior and PME in parallel as separate outcomes that have a bidirectional relationship. This account of PME receives theoretical support from dual-process theories that specify that outcomes can come about through implicit (often automatic) or explicit (often controlled) processes.\textsuperscript{55} The latter has only been mentioned in one study as a logical possibility.\textsuperscript{152} In both alternate accounts, because PME would not demonstrate predictive validity with behavior, it would be harder to pinpoint what PME stands for and confidently use it to select important messages in formative research. It may still be possible to use PME in the case that it has a bidirectional relationship with behavior because it would at least capture some information about behavior while not explicitly violating predictive validity with behavior.

**Focuses on behavior**

A measure of PME should focus on behavior, and the most straightforward way to do this
is to use items that assess the perceived effects of a message on established determinants of behavior. These perceptions are often called *effects perceptions*, and researchers have traditionally assessed PME using them or *message perceptions*. An example of an effects perception is the extent to which viewers of an anti-smoking message think it makes smoking less appealing to them. Another example is how much the recipients of a marijuana risk message think it encourages them to avoid marijuana use. In comparison, the extent to which a message is perceived as compelling or informative are two examples of message perceptions. PME assessed using effects perceptions, also called *impact* or *personalized* PME, quantify the recipient’s overall impression of a message and is a judgment about how much they think that the message would change behavior or its important antecedents. On the other hand, PME assessed using message perceptions, also called *attribute* or *ad-directed* PME, is a judgment about whether a message has characteristics that should enable further processing of that message. PME assessed using effects perceptions is often more narrowly concerned with predicting changes in behavior among the target audience and follows from previously cited research on PME as a necessary but not a sufficient condition for message impact. In contrast, PME assessed using message perceptions is more general and may be a marker for the broad persuasive potential of a message and follows from previously cited research on PME as a necessary and a sufficient condition for message impact.

Although the distinction between the two types of PME is not new, a fair amount of applied research has not adhered to the distinction, often treating PME assessed using effects or message perceptions interchangeably. Recent meta-analyses have shown that approximately half of the existing PME measures focus on effects perceptions while the other half focus on message perceptions. There is also a notable tendency to combine items measuring both types of
PME in a single scale without considering that they are conceptually distinct constructs.\textsuperscript{105,152} Future psychometric studies should compare the validity of PME measures focusing on impact with that of PME measures focusing on message attributes. I propose that only effects perceptions be labeled as PME because they focus on behavior as the target of PME judgments and often draw on research on behavioral intentions,\textsuperscript{129,131,147} outcome expectancies of performing a the behavior,\textsuperscript{61} and risk perceptions.\textsuperscript{57,130} Thus, effects perceptions are more proximal to behavior than message perceptions, which are appraisals of specific message features (e.g., argument strength). Message perceptions may even be a determinant of effects perceptions, a suggestion that other researchers have also made.\textsuperscript{50,152}

Another way to focus on behavior is by specifying the behavior of interest in the actual items being used to assess PME. Effects perceptions items commonly reference the target behavior while message perceptions items generally do not do so.\textsuperscript{105} The suggestion to use behavioral referents corresponds to the principle of compatibility originally articulated as a fundamental requirement for the predictive validity of a measure of intention with the corresponding behavior in the context of the Theory of Reasoned Action.\textsuperscript{59} According to this principle, a construct must be measured at the same level of generality or specificity as the behavior of interest for it to demonstrate predictive validity with that behavior.\textsuperscript{59} Generality or specificity of behavior are characterized by four criteria: action (e.g., refrain from), target (e.g., cigarette smoking), context (e.g., on or near school premises), and time (e.g., the next time that there is a social opportunity to do so).\textsuperscript{59} Context and time are most likely inapplicable to the measurement of PME because health messages often address risky behaviors at a population-level so that their occurrence varies by setting and people have different patterns of engagement. Depending on how PME is operationalized, action may be ambiguous and difficult to specify.
However, specifying the behavior (e.g., cigarette smoking) as the target of the judgment would ensure compatibility of a measure of PME with behavior and improve predictive validity. To my knowledge, no experimental studies have examined how specification of a behavioral referent impacts the reliability and predictive validity of a measure of PME.

About the respondent

Another way to improve the measurement of PME is to include personal referents in the items on a measure. Effects perceptions items commonly use personal referents to direct the message recipient to consider its effects on their own attitudes, beliefs, and perceptions while message perceptions items generally do not do so.\textsuperscript{105} The clearest rationale for including personal referents comes from the \textit{perceptual component of the third-person effect}, which states that people tend to think that media messages have a stronger impact on others than themselves.\textsuperscript{44} More than three decades of research documented in individual studies, systematic reviews, and meta-analyses demonstrate that this self-other asymmetry in media perception is generally robust to variations in substantive topics and research methodology.\textsuperscript{36,37,54,132,139} Usually, researchers are interested in the extent to which the person who views a message thinks it will impact them. Therefore, measuring PME without using personal referents in the items may not suit researchers’ purpose in using it in message development because the resulting PME judgments may be about variable focal groups (e.g., respondents or the general public). Furthermore, leaving out personal referents may accentuate the third-person effect and yield an overestimate of PME that could lead to less than ideal decisions about message selection. Thus, a PME measure should include personal referents in all of its items.
Psychometric criteria for perceived message effectiveness measures

Construct validity

Many measures of PME have not been rigorously validated even though they continue to be used in formative research on health messages.\textsuperscript{105,152} The distinction between effects and message perceptions should be central to any validation efforts. While effects and message perceptions are conceptually distinct, the only study to my knowledge that has formally tested the dimensionality of both types of perceptions found mixed support for their measuring a single construct or separate, but strongly correlated constructs.\textsuperscript{50} The study did not formally examine sources of measurement error or nuisance multidimensionality that, when incorporated into a measurement model, could have provided additional insights about the dimensionality of both types of perceptions. Future psychometric studies should demonstrate the distinction between effects and message perceptions in a variety of behavioral contexts.

A measure of PME should also demonstrate predictive validity with behavior. Related validation efforts would ideally include PME assessed using effects perceptions or message perceptions and extend to early, middle, and late cognitive determinants of behavioral impact such as those from the UNC Tobacco Warnings Model (TWM) in the context of chemical messages.\textsuperscript{65} Focusing on the predictive validity of PME assessed using effects perceptions or message perceptions with early, middle, and late cognitive determinants of behavioral impact would provide meaning to the distinction between effects and message perceptions even if measurement models are not fully sensitive to it. For example, paying attention to and noticing chemical messages is an early cognitive determinant of behavioral impact from the TWM, and thinking about the chemical messages is a late cognitive determinant. Research findings that effects and message perceptions similarly predict attention to and noticing of chemical messages,
but effects perceptions more strongly predict thinking about the chemical messages would suggest that effects perceptions are more proximal to behavior than message perceptions. This suggestion would stand even if neither effects nor message perceptions predicted behavior in the context of chemical messages.

Future validation efforts would also benefit from a focus on identifying plausible psychological mechanisms underlying PME, as suggested in a recent meta-analysis. More specifically, these efforts would focus on determining the role, if any, that effects and message perceptions have in message processing and provide further insights about the distinction between them. For example, research findings in a number of substantive contexts that message perceptions have no role in message processing while effects perceptions mediate the behavioral impact of messages would give precedence to the use of effects perceptions in assessing PME over message perceptions. These findings would indicate that PME assessed with message perceptions is a criterion variable while PME assessed with effects perceptions is not, provided that a plausible theoretical explanation for the mediation exists. Thus, validation efforts that focus on potential psychological mechanisms underlying PME could provide a conceptual foundation that guides its use in message development.

Minimal measurement error

Two generally understudied aspects of most psychological scales are differential item functioning (DIF) and differential test functioning (DTF). DIF denotes the phenomenon that a given item functions differently in one context as compared to another. DTF refers to the phenomenon that DIF in one or more items on a scale aggregates and thereby causes unreliable scale scores in one context as compared to another. All self-report measures exhibit some level of DIF and DTF, but the thrust of DIF and DTF analyses is to identify levels that cannot be
ignored. Problematic DIF or DTF can make it difficult to accept a statistically significant difference on a scale because that difference may actually be caused by measurement error.\textsuperscript{140} DIF and DTF are equally valid concerns for PME assessed using effects or message perceptions. While DIF and DTF analyses are becoming more common in other areas of psychological assessment, I am aware of no studies that have looked at the presence of DIF and DTF in existing PME measures.

DIF and DTF in a PME measure may originate from subgroups interpreting one or more items differently. Because researchers are often interested in developing messages among diverse populations, including those with health disparities, a PME measure should exhibit similar psychometric properties among subgroups of interest. DIF and DTF analyses would verify that the psychometric properties of a PME measure do not vary in selected subgroups or provide corrected scaled scores in the case that they do for use in between-group comparisons of PME ratings. They may also identify items that are particularly susceptible to DIF so that they can be dropped from future iterations of a PME measure.

DIF and DTF also applies to evaluation studies in which each participant rates multiple messages. The psychometric properties of a PME measure should not change in the context of specific health messages. In other words, the characteristics of specific health messages should not influence the psychometric properties of a PME measure. If a PME measure yields statistically significantly different rankings for a set of messages and the measure exhibits DIF or DTF in the context of one or more messages, then at least some of the statistically significant differences contributing to the rankings are attributable to measurement error. Thus, the presence of DIF and DTF in a PME scale originating from candidate messages could lead to less than ideal decisions about those messages.
Particular item features may also contribute to measurement error in a PME scale. For example, not specifying the personal referent may contribute to a lack of clarity in PME items and thereby increase measurement error. This applies to scales that uniformly exclude personal referents or include them only in some items. The latter case is arguably more problematic because any measurement error from excluding the personal referent would be confined to some items and not be distributed over the entire scale. Similar concerns apply to the use of behavioral referents in PME scales. It is also possible that effects perceptions are more prone to measurement error than message perceptions or vice versa. Thus, the method of assessing PME may also contribute to measurement error on a scale. Future studies should explore these possibilities in a variety of substantive contexts with the goal of minimizing measurement error in PME scales.

**Brief in length**

Existing measures of PME tend to be lengthy with shorter scales having six items and longer scales having ten.\(^{49,105,152}\) For PME to be an efficient tool for evaluating a large pool of messages, a measure must be brief. Shorter scales should enable researchers to repeatedly expose participants to multiple messages and have them rate each of the messages on PME. This is in contrast to longer measures, which are likely to pose a higher cognitive burden on the participant. Longer scales are also likely to contribute to survey fatigue. Both factors could cause participants’ item response behavior to change as they progress through a message evaluation study and introduce unknown sources of measurement error that affect the quality of PME ratings. Both factors could also impose a lower ceiling on how many messages participants can meaningfully evaluate in a single study.
Discussion

This review points to three key observations that may guide future research on PME as an efficient tool for message development. First, PME has been traditionally assessed using two conceptually distinct types of perceptions, effects perceptions and message perceptions. This distinction has often been overlooked in the measurement of PME so that many existing measures include items that assess effects and message perceptions. The distinction has also often been overlooked in studies that use PME, which tend to treat both types of PME as equally valid tools for message development. I suggest that only effects perceptions are PME given the broad goal of interventions that rely on brief messages to change viewer behavior. In contrast, message perceptions may be an antecedent of PME, which is a hypothesis that future validation studies should address.

The second observation is that efforts to validate PME scales would benefit from focusing on identifying plausible mechanisms underlying any role in message processing instead of providing correlational evidence for the PME-behavior correspondence. A recent meta-analysis of PME suggested that the wealth of correlational data for the PME-behavior correspondence does not adequately demonstrate its construct validity, nor its utility in message development. It is premature to discard all correlational construct validity data because PME scales suffer from a lack of standardization. Focusing validation efforts on underlying psychological mechanisms may identify a theoretical importance for PME in message processing, thereby transforming it from the criterion variable that it is currently treated as.

The last observation is that psychometric studies of PME scales would benefit from focusing on lesser studied aspects of self-report measures such as whether individual items on a scale or the scale as a whole function differently among various sub-populations. These
phenomena would arise from measurement error in a PME scale and may obscure comparisons among types of brief messages, individual messages, and subgroups. Because the number of potential applications of PME are quite large, a PME scale should exhibit minimal measurement error so that researchers can use it to make optimal decisions about messages in a variety of applications.
CHAPTER 3: VALIDATION OF THE UNC PERCEIVED MESSAGE EFFECTIVENESS SCALE

Overview

Background. Interventionists commonly identify promising messages for health communication efforts based on audience members’ ratings of perceived message effectiveness (PME).

Purpose. We sought to validate a new PME measure that improved on existing scales by focusing on the behavior and respondent, being brief, and having strong psychometric properties.

Methods. Participants were a national convenience sample of 999 adults and national probability samples of 1,692 adults and 869 adolescents recruited in 2015. Smokers and non-smokers rated up to six brief messages about the chemicals in cigarette smoke on two PME scales. The first was the new 3-item UNC PME Scale that assessed effects perception. The second was an established 6-item PME scale that assessed message perceptions. We examined the UNC PME Scale’s psychometric properties and compared both scales using item factor analysis.

Results. The UNC PME Scale measured the same construct across multiple chemical messages (all factor loadings ≥ 0.86). It exhibited high reliability (> 0.85) over very low to moderate levels of PME (z = -2.5–0.2), a range that is useful for identifying more promising messages. Samples of adults and adolescents showed a similar pattern of results. As expected,

the UNC PME Scale was strongly positively correlated with the message perceptions scale ($r = 0.84$). It also exhibited strong psychometric properties among participants regardless of education, reactance, sex, and smoking status.

**Discussion.** The UNC PME Scale reliably and validly measured PME among adults and adolescents from diverse groups. This brief scale may be used to efficiently evaluate candidate anti-smoking messages and may be suitable for adaptation to other risky behaviors.

**Keywords:** health communication, message development, formative research, item response theory

**Validation of the UNC Perceived Message Effectiveness Scale**

Perceived message effectiveness (PME) is concerned with the perception that candidate messages will or will not achieve their objectives, and the use of PME as a tool for message selection has become increasingly common since year 2000. Many researchers specifically use PME as an early indicator of a health message’s potential to change behavior. There is growing evidence for the predictive validity of PME, with a small number of longitudinal studies demonstrating that PME predicts changes in smoking behavior in the context of anti-smoking messages. A comparatively larger number of cross-sectional studies show that PME is associated with attitudes and behavioral intentions in a variety of health messaging contexts, including those seeking to promote colonoscopy, improving social support outcomes, and preventing sexually transmitted diseases.

PME measures have traditionally used two different types of perceptions to inform their measurement. The first type are *message perceptions*, two examples being the extent to which a message is thought of as compelling or informative. Message perceptions are rooted in communication theories of persuasion, such as the Elaboration Likelihood Model, which posit
that important characteristics of a message such as argument quality function as a gateway to
greater elaboration about the message.\textsuperscript{58, 79, 116} Greater elaboration facilitates, in turn, attitude
change that may affect performance of a behavior. PME assessments that use message
perceptions (i.e., attribute or ad-directed PME) are judgments about whether a message has
characteristics that should enable further processing of that message.\textsuperscript{125} In addition, this type of
PME is more general and may be a marker for the broad persuasive potential of a message.\textsuperscript{153}

The second type of perceptions used in PME measures are \textit{effects perceptions}. The extent
to which the viewers of an anti-smoking message think that it makes smoking less appealing to
them is an example of an effects perception. Another example is how much the recipients of a
marijuana risk message think that it encourages them to avoid marijuana use. Effects perceptions
are rooted in research at the intersection of economics, neuroscience, and psychology suggesting
that people view a message for a brief period of time and use their initial affective responses to
summarily process it.\textsuperscript{30, 135, 137} The resulting overall impression may motivate behavior change.
PME assessments that use effect perceptions (i.e., impact or personalized PME) quantify the
overall impression of a message and are judgments about a message’s potential to change
important antecedents of behavior. Furthermore, this type of PME is often more narrowly
concerned with predicting changes in behavior among the target audience. As a result, effects
perception items use behavioral and personal referents that direct respondents to consider the
effects of messages on their own attitudes, beliefs, thoughts, or behaviors.\textsuperscript{105}

Although there is a clear conceptual distinction between message and effects perceptions,
there is a notable tendency among researchers to use them interchangeably or combine them in a
single measure of PME.\textsuperscript{50, 105, 152} Due to this tendency, some researchers have criticized the
current measurement of PME as inadequate\textsuperscript{152} and questioned the meaningfulness and utility of
PME judgments in message development altogether.\textsuperscript{112,114} Given that many health messages aim to change behavior and effect perceptions are conceptually proximal to behavior, PME scales with a clear effects orientation are a promising direction for the literature. Additional concerns about the measurement of PME are that existing scales are either lengthy or too generic and do not uniformly use behavioral or personal referents in their items. Both practices may increase measurement error in a scale by increasing the cognitive burden on respondents or reducing the clarity and precision of PME judgments. A high cognitive burden on respondents may also limit the number of messages that can be efficiently evaluated using PME judgments in a single study.

To address issues in the current measurement of PME, we developed the UNC Perceived Message Effectiveness Scale, conceptualizing PME as the extent to which a person believes that a health message will affect them in ways that are consistent with message objectives, particularly changing behavior.\textsuperscript{152} The UNC PME Scale has only three items that uniformly focus on behavior and the respondent. We sought to examine its psychometric properties in the context of brief messages about the chemicals in cigarette smoke that were designed to discourage smoking. Three main goals for the use of PME guided our study. First, the UNC PME Scale should measure the same construct across different messages so that researchers can meaningfully compare individual messages using PME ratings. Second, it should function similarly among diverse populations so that researchers can compare messages among subgroups of interest, such as those with health disparities. Finally, unlike many PME measures used in the literature,\textsuperscript{105,152} it should demonstrate construct validity.

**Methods**

**Participants**

Participants were two samples of adults and one sample of adolescents recruited in 2015.
We recruited a convenience sample of U.S. adults (ages ≥ 18; n = 1,034) using Amazon Mechanical Turk. In addition, the Carolina Survey Research Laboratory invited all 13–25-year-olds, all smokers, and a randomly selected subset of adult (ages ≥ 25) non-smokers, who had previously completed a tobacco-related phone survey to participate in an online follow-up survey. Data collection was multimodal (desktop computers, mobile devices, and by mail), with nonresponders being contacted up to three times through telephone reminder calls and priority mailings. The follow-up survey had an overall response rate of 73% (2637/3612). We treated the data from adults (18+; n = 1,758) and adolescents (13–17; n = 877) as separate probability samples. After eliminating participants with missing data on the UNC PME Scale, construct validators, or demographic characteristics, the three analytic samples had 999 (adult convenience), 1,692 (adult probability), and 869 (adolescent probability) participants.

**Procedures**

In a repeated-measures design, participants in the adult convenience sample rated two chemical messages. Those in the probability samples rated six chemical messages in one of five randomly assigned orders. The messages varied by chemical and associated contextual information (Table 3.1). Adult convenience participants received $3 while adult and adolescent probability participants received $45 for completing their respective surveys, which were much longer. The Institutional Review Board at the University of North Carolina approved the procedures for all three samples.

**Measures**

**UNC PME Scale.** In the adult convenience study, we developed twelve candidate items to assess various perceptions of chemical messages designed to discourage smoking. We worded the items such that they could be answered by both smokers and non-smokers. The five-point
Exploratory factor analysis using maximum likelihood estimation and promax rotation revealed a four-factor solution of effects perceptions, message perceptions, message reactance, and message credibility. 

The factor for effects perceptions had four items that assessed respondents’ perceptions of discouragement, concern, unpleasantness, and appeal as related to the contents of the chemical messages. The discouragement item, "This message discourages me from wanting to smoke," came from our previous work on cigarette warnings and is theoretically derived from work on behavioral intentions. The concern item, "This message makes me concerned about the health effects of smoking," focused on the health consequences of smoking, and is derived from work on affect and risk perception. The unpleasantness item, "This message makes smoking seem unpleasant to me," and the appeal item, "This message makes smoking seem less appealing to me," were focused on reduced pleasure from smoking and are derived from work on smoking expectancies. Due to the overlap between these last two items (r = .84) and cognitive testing revealing greater clarity in the unpleasantness item, we dropped the appeal item yielding the 3-item UNC PME Scale (α = .93). For clarity, we generally refer to the UNC PME Scale as our effects perceptions scale in the remainder of this paper to emphasize the conceptual difference between it and the message perceptions scale that we utilized.

Other measures. To support analyses of construct validity, we assessed message perceptions, message reactance, and message credibility for each chemical message. The 6-item message perceptions scale (assessed in adult convenience sample only) references the respondent in one item only and does not use behavioral referents at all. We measured message reactance, or resistance to the message, using the Brief Reactance to Health Warnings Scale. Finally, we
measured message credibility (adult convenience sample only) using two items, "This message is believable to me," and "This message seems credible to me." The five-point response scale for all items ranged from "strongly disagree" to "strongly agree" (coded as 1–5). We predicted that our effect perceptions scale would correlate positively with the message perceptions scale and message credibility but would be negatively correlated with message reactance.

The survey also assessed smoking status and standard demographic variables. Adult smokers were those individuals who had smoked at least 100 cigarettes in their lifetime and currently smoke every day or some days, and adolescent ever smokers were those who had ever tried smoking cigarettes, even one or two puffs.

**Data analysis**

Analyses used R (ver. 3.4.3) with three selected add-on packages, lavaan (ver. 0.5-23.1097) and mirt (ver. 1.27.1) for estimating psychometric models and ggplot2 (ver. 2.2.1) for plotting related mathematical functions.

**Psychometric properties.** To parse variability in the items on our effects perceptions scale that is inherent to PME from variability that is specific to chemical messages, we used a two-tier item bifactor analytic (IFA) model with a general factor for PME spanning all chemical messages and orthogonal message-specific factors. We compared the loadings on and variance accounted for by the general factor with those for the message-specific factors to determine the extent to which the scale may function differently in the context of specific chemical messages. We also examined information curves from the IFA model to characterize scale and item reliability. The information score is a quantification of the variability that a measure captures about the construct of interest and varies across the possible range (standardized) of the construct. Higher information points to lower standard error of measurement and, thereby,
greater reliability.

To arrive at the preferred IFA model with acceptably low levels of measurement non-invariance across messages in the message-specific and general PME factors, we estimated a series of increasingly constrained IFA models and compared them using the likelihood ratio (LR) test for nested models. We confirmed model selection by examining global fit of the preferred model using the appropriate IFA $\chi^2$ analog and the root mean square error of approximation (RMSEA), item fit with graded response parameterization using the $S-\chi^2$ index, and person fit using the $Z_h$ index. The preferred IFA model incorporated strong invariance across chemical messages in the general and specific dimensions and had adequate global fit in the adult convenience ($G^2 = 1839, df = 15606, p > .05$), adult probability ($M_2 = 973, df = 153, p < .001$), and adolescent probability ($M_2 = 277, df = 153, p < .001$) samples. The RMSEA (range = 0–0.056) was small in the three samples. The IFA model did not exhibit systematic deviations in item fit in the adult convenience (range $S-\chi^2 = 35.7–46.5$, range $df = 28–33$), adult probability (range $S-\chi^2 = 134–213$, range $df = 101–112$), and adolescent probability (range $S-\chi^2 = 22.4–56.0$, range $df = 21–29$) samples. The model also fit better than expected for a large majority of participants (range $Z_h > 0 = 77.8–82.9\%$) in each of the three samples.

**Differential item and test functioning.** To determine whether individual items on our effects perceptions scale had similar psychometric properties among subgroups that differed by education (adults: $\leq$ some college, or $>$ some college; adolescents: middle school, or high school), reactance ($\leq$ "neither disagree or agree," or higher), sex, and smoking status (adults: smoker or not; adolescents: ever-smoker or not), we conducted differential item functioning (DIF) analyses. We treated each message as a potential instance of DIF and conducted separate analyses for each instance using multiple-group unidimensional graded response models. We
used the LR/f ratio to select anchor items. Next, we estimated a series of more constrained models and used the LR test for nested models to identify items with any DIF. Additional LR tests revealed whether an instance of DIF was related primarily to the reliability or dimensionality of the involved item or both.

To assess whether any observed DIF caused our effects perceptions scale as a whole to function differently for subgroups in terms of reliability and dimensionality, we conducted differential test functioning (DTF) analyses using effect sizes. Specifically, we calculated Cohen’s $d$ based on a final model with between-group constraints for DIF to characterize the magnitude of any instance of DIF as well as DIF. Effect sizes with absolute values of 0.2 amounted to negligible DIF or DTF while those greater than 0.2 warranted further investigation. DIF testing involves many comparisons, inflating the false-discovery rate. DIF on individual items may cancel out at the scale level if the direction of DIF varies across items within the scale or the magnitude is small. Conducting DTF analyses using effect sizes allowed us to avoid unnecessarily flagging items and pursue all possibilities of DIF without having to correct for inflated false-discovery rates.

**Construct validity.** We evaluated the construct validity of our effects perceptions scale by examining the average correlations across messages between our scale and the message perceptions scale, message credibility, and message reactance. To take advantage of our multitrait-multioccasion data, we used the correlated trait-correlated uniqueness (CTCU) model to estimate all factor correlations and variance components. We also compared the unexplained variance (uniqueness) in the items on our effects perceptions scale and the message perceptions scale to assess the measures’ relative susceptibility to measurement error. The CTCU model in the adult convenience sample had four correlated factors for both measures of PME, message
reactance, and message credibility that spanned the two messages. The CTCU model in the probability samples had two correlated factors for our message perceptions scale and message reactance that spanned the six messages. The models retained relevant constraints for measurement invariance from the IFA model for our effects perceptions scale and applied similar constraints for all construct validators. In the three samples, the CTCU model had adequate global fit (RMSEA = 0.074–0.10; CFI = 0.95–0.98).

Results

The mean ages of adult convenience and probability participants were 33.8 (SD = 11.0) and 43.1 (SD = 17.7), respectively; adolescents had a mean age of 15.0 (SD = 1.37). Fewer than half of adult convenience (46.7%) and probability (30.0%) participants had a bachelor’s or advanced degree (Table 3.2). In both adult samples, around one-third of participants (convenience = 31.1%, probability = 37.4%) were current smokers, and 10.2% of adolescents had ever tried smoking cigarettes.

Psychometric properties

Our effects perceptions scale measured the same construct in the context of six unique chemical messages. In the adult convenience sample, the three items strongly loaded on the general factor for PME (0.89–0.92) and weakly loaded on the message-specific factors (0.18–0.27; Table 3.3). The general factor for PME accounted for the vast majority of the variance in the items (82.6%). In comparison, the two message-specific factors together explained an additional 5.2% of the variance in the items. These patterns indicated that participants understood the scale similarly in the context of two chemical messages. The adult and adolescent probability samples also replicated these findings across the larger set of six chemical messages.

Among adults, our effects perceptions scale measured very low (convenience: z = -2.5;
probability $z = -2.3$) to mean levels (convenience: $z = 0.2$; probability: $z = -0.1$) of PME with large amounts of information that corresponded to high reliability ($\geq 0.85$; Figure 3.1). In contrast, the scale reliably measured extremely low ($z = -2.9$) to somewhat low ($z = -0.7$) levels of PME among adolescents. This is because the majority of participants ($\geq 55\%$) in the three samples responded to each of the three items with the highest option, "strongly agree" (coded as 5) irrespective of message resulting in left skewed response distributions. Thus, our effects perceptions scale did not provide information about individuals who were likely to elicit higher PME than the five-point response scale allowed. This ceiling effect was more pronounced among adolescents and present in all items in the three samples. In the three samples, concern contributed the least amount of information (max = 5.1–7.3) to the scale. Discouragement contributed the most information (max = 7.1) in the adult convenience sample while unpleasant did so in the adult (max = 10.3) and adolescent (max = 17.5) probability samples.

**Differential item and test functioning**

Among adults and adolescents who varied by education, reactance, sex, or smoking status, our effects perceptions scale exhibited similarly strong psychometric properties (Figure 3.2). Across all three samples, the items on our effects perceptions scale exhibited negligible to small DIF (absolute value Cohen’s $d = 0.003–0.18$) in 34 out of 168 potential instances of DIF (all $p < .01$) and larger DIF (absolute value $d = 0.22–0.36$) in eight instances (all $p < .01$). The 34 instances of negligible to small DIF were distributed over all three samples, all three items, all four grouping variables, and all six chemical messages. Similarly, the eight instances of larger DIF were distributed over all three samples and all chemical messages even though they involved the concern and discouragement items and smoking status only. These patterns suggested that the items on our effects perceptions scale generally maintained strong
psychometric properties among adult and adolescent participants who varied by the aforementioned characteristics.

Most of the statistically significant instances of DIF may be attributable to high power to detect even negligible DIF. DTF analyses provided evidence in support of this possibility. Any observed DIF canceled out due to the presence of an item with DIF in the opposite direction or was minimized by the presence of items with no DIF. As a result, across all 56 potential instances of DTF, our effects perceptions scale did not exhibit DTF in 25 instances and exhibited only negligible to small DTF that did not warrant further investigation in 31 (absolute value $d = 0.00004–0.18$). Thus, our effects perceptions scale as a whole maintained strong psychometric properties among adults and adolescents who varied by selected demographic characteristics.

**Construct validity**

Our effects perceptions scale demonstrated convergent validity through a high positive correlation with the message perceptions scale ($r = 0.84$) and a moderate positive correlation with message credibility ($r = 0.62$) in the adult convenience sample. It also demonstrated convergent validity through weak to moderate negative correlations with message reactance in the adult convenience ($r = -0.49$), adult probability ($r = -0.40$), and adolescent probability samples ($r = -0.30$). The items on our effects perceptions scale also had lower amounts of unexplained variance (uniqueness = 0.19–0.21) than those on the message perceptions scale (uniqueness = 0.29–0.39) in the adult convenience sample. Estimates of unexplained variance in the items on our effects perceptions scale in the adult (uniqueness = 0.18–0.24) and adolescent (uniqueness = 0.19–0.21) probability samples were comparable to those in the adult convenience sample.
Discussion

In three large national samples, the UNC PME Scale reliably and validly measured PME among adult and adolescent smokers and non-smokers and diverse populations. The scale also retained desirable psychometric properties in each sample across all messages describing the harms of chemicals in cigarettes. Thus, researchers should be able to use the scale to evaluate many messages in a single study while being confident that differences in PME ratings between messages are not due to measurement error. Our brief 3-item effects perceptions scale should also enable researchers to minimize the cognitive burden on participants from repeated assessments. In situations that require a single-item measure of PME, researchers should use the discouragement or unpleasantness item because both items performed similarly well and better than the concern item.

The finding that the items on the message perceptions scale had greater amounts of unexplained variance than our effects perceptions scale after accounting for PME across the chemical messages suggests that the latter scale had greater conceptual clarity. Moreover, it was less prone than the message perceptions scale to being influenced by methods effects originating from specific messages and or unknown sources of multidimensionality. It is possible that specifying the behavior and the rater enabled the effects perceptions scale to elicit clearer PME judgments than the message perceptions scale, which does not use behavioral or personal referents. While researchers have theorized that specifying behavioral and or personal referents should improve the measurement of PME, prior studies have not formally examined whether doing so reduces measurement error. Our study provides preliminary evidence that the use of referents can improve the psychometric properties of PME measures. Future studies should investigate the effects of specifying different types of referents on the psychometric properties of
PME scales using explanatory item response modeling\textsuperscript{31} or generalizability theory.\textsuperscript{144}

While our effects perceptions scale was strongly positively correlated with the message perceptions scale, there are clear conceptual differences between general perceptions of a message and perceptions that a message would affect the respondent. It is conceivable that message perceptions are further from behavior, while effects perceptions are more proximal to behavior. Moreover, message perceptions may impact effects perceptions such that the extent to which a person believes a message would affect them may be based, at least in part, on their perceptions of whether the message is believable, compelling, and so forth. However, from a practical standpoint, the purpose of PME measures is to guide selection of a small set of messages that have the most potential to change behavior, from a larger pool of messages. For that purpose, there may be an advantage in using effects perceptions measures which directly ask participants whether messages are likely to have effects. While some recent work has questioned the validity of PME judgements in predicting actual effectiveness, a close examination of the corresponding\textsuperscript{112} and other\textsuperscript{105} measures reveals a striking lack of consistent measurement of PME and little evidence of strong psychometric properties. Improving measurement properties is likely to improve the precision of PME measures in guiding the selection of promising messages.

**Strengths and limitations**

The main strengths of our study were the three large, diverse national samples and repeated-measures designs, which provided us with opportunities to replicate our basic findings about the psychometric properties of the UNC PME Scale in the context of several chemical messages and populations. Another major strength of our study was the use of latent variable models that took advantage of all of the data in a sample (e.g., two-tier item bifactor analysis) to provide robust inferences about the performance of our brief scale.
Our study also had a number of important limitations that should be addressed in future research. The main limitation is that the construct validity data exclusively came from comparison to one other PME scale41 and cognitive measures that are theoretically relevant to message development. Because researchers are often interested in the potential impact of persuasive messages on behavior, future studies should examine the predictive validity of the UNC PME Scale with behavior. Another limitation is that our study examined messages about cigarette smoking but not other behaviors. Future studies should replicate our findings in the context of messages about other health behaviors to confirm that the UNC PME Scale is applicable to diverse health messaging contexts.

Conclusions

The UNC PME Scale is a brief measure with three strongly performing items that should enable researchers to efficiently and effectively evaluate messages for health behaviors among diverse populations, including those with health disparities. While our findings are specific to chemical messages that seek to prevent cigarette smoking, the wording of the PME items is general so that the scale should be readily adaptable to messages for other behaviors.
<table>
<thead>
<tr>
<th>Label</th>
<th>Chemical message</th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>Cigarette smoke contains ammonia. This is used as a pesticide and causes breathing problems.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AR1</td>
<td>Cigarette smoke contains arsenic. This is found in rat poison and causes heart damage.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>AR2</td>
<td>Cigarette smoke contains arsenic. This causes lung tumors.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FO</td>
<td>Cigarette smoke contains formaldehyde. This causes throat cancer.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LE</td>
<td>Cigarette smoke contains lead. This causes cancer and brain disorders.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UR</td>
<td>Cigarette smoke contains uranium. This causes lung tumors and kidney damage.</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Table 3.2

Participant characteristics

<table>
<thead>
<tr>
<th></th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 999$</td>
<td>$n = 1,692$</td>
<td>$n = 869$</td>
</tr>
<tr>
<td>Age</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>13–17</td>
<td>-</td>
<td>-</td>
<td>100</td>
</tr>
<tr>
<td>18–25</td>
<td>24.9</td>
<td>24.9</td>
<td>-</td>
</tr>
<tr>
<td>26–34</td>
<td>40.1</td>
<td>14.1</td>
<td>-</td>
</tr>
<tr>
<td>35–44</td>
<td>17.4</td>
<td>12.8</td>
<td>-</td>
</tr>
<tr>
<td>45–54</td>
<td>10.2</td>
<td>18.4</td>
<td>-</td>
</tr>
<tr>
<td>55–64</td>
<td>6.31</td>
<td>16.3</td>
<td>-</td>
</tr>
<tr>
<td>65+</td>
<td>1.00</td>
<td>13.5</td>
<td>-</td>
</tr>
<tr>
<td>Male</td>
<td>54.9</td>
<td>45.7</td>
<td>48.3</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>82.4</td>
<td>70.2</td>
<td>81.9</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>6.71</td>
<td>19.5</td>
<td>9.44</td>
</tr>
<tr>
<td>Native American</td>
<td>0.80</td>
<td>2.38</td>
<td>1.84</td>
</tr>
<tr>
<td>Asian</td>
<td>6.81</td>
<td>2.14</td>
<td>2.07</td>
</tr>
<tr>
<td>Other</td>
<td>3.30</td>
<td>5.70</td>
<td>4.72</td>
</tr>
<tr>
<td>Hispanic</td>
<td>8.52</td>
<td>7.82</td>
<td>6.44</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; high school</td>
<td>0.70</td>
<td>10.2</td>
<td>99.4</td>
</tr>
<tr>
<td>High school</td>
<td>12.5</td>
<td>25.8</td>
<td>0.60</td>
</tr>
<tr>
<td>Some college</td>
<td>28.7</td>
<td>23.2</td>
<td>-</td>
</tr>
<tr>
<td>Associate’s degree</td>
<td>11.4</td>
<td>10.3</td>
<td>-</td>
</tr>
<tr>
<td>College degree</td>
<td>36.4</td>
<td>20.0</td>
<td>-</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>8.61</td>
<td>7.74</td>
<td>-</td>
</tr>
<tr>
<td>Professional or doctoral degree</td>
<td>1.70</td>
<td>2.19</td>
<td>-</td>
</tr>
<tr>
<td>Income, annual household</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0–$24,999</td>
<td>26.1</td>
<td>33.0</td>
<td>-</td>
</tr>
<tr>
<td>$25,000–$49,999</td>
<td>35.5</td>
<td>26.5</td>
<td>-</td>
</tr>
<tr>
<td>$50,000–$74,999</td>
<td>21.2</td>
<td>17.4</td>
<td>-</td>
</tr>
<tr>
<td>$75,000–$99,999</td>
<td>10.1</td>
<td>9.99</td>
<td>-</td>
</tr>
<tr>
<td>$100,000 or more</td>
<td>7.10</td>
<td>13.0</td>
<td>-</td>
</tr>
<tr>
<td>Smoker</td>
<td>31.1</td>
<td>37.4</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Notes. Adult smokers were defined as having ever smoked at least 100 cigarettes and currently smoking every day or some days. Adolescent ever smokers were defined as having ever tried cigarettes, even one or two puffs. Missingness was negligible for all participant characteristics and highest for annual household income in Study 2 (4.79%).
## Table 3.3

Factor loadings for the UNC Perceived Message Effectiveness Scale

<table>
<thead>
<tr>
<th></th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PME</td>
<td>AR1</td>
<td>FO</td>
</tr>
<tr>
<td>Discouragement</td>
<td>0.92</td>
<td>0.23</td>
<td>0.93</td>
</tr>
<tr>
<td>Unpleasantness</td>
<td>0.91</td>
<td>0.18</td>
<td>0.91</td>
</tr>
<tr>
<td>Concern</td>
<td>0.89</td>
<td>0.27</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PME</td>
<td>AR1</td>
<td>FO</td>
</tr>
<tr>
<td>Discouragement</td>
<td>0.92</td>
<td>0.23</td>
<td>0.93</td>
</tr>
<tr>
<td>Unpleasantness</td>
<td>0.91</td>
<td>0.18</td>
<td>0.91</td>
</tr>
<tr>
<td>Concern</td>
<td>0.89</td>
<td>0.27</td>
<td>0.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PME</td>
<td>AR1</td>
<td>FO</td>
</tr>
<tr>
<td>Discouragement</td>
<td>0.93</td>
<td>0.19</td>
<td>0.92</td>
</tr>
<tr>
<td>Unpleasantness</td>
<td>0.91</td>
<td>0.31</td>
<td>0.91</td>
</tr>
<tr>
<td>Concern</td>
<td>0.86</td>
<td>0.37</td>
<td>0.89</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Adult convenience</th>
<th>Adult probability</th>
<th>Adolescent probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PME</td>
<td>AR1</td>
<td>FO</td>
</tr>
<tr>
<td>Discouragement</td>
<td>0.92</td>
<td>0.26</td>
<td>0.92</td>
</tr>
<tr>
<td>Unpleasantness</td>
<td>0.91</td>
<td>0.36</td>
<td>0.91</td>
</tr>
<tr>
<td>Concern</td>
<td>0.89</td>
<td>0.35</td>
<td>0.89</td>
</tr>
</tbody>
</table>

### Notes
Items loaded on a general factor for perceived message effectiveness (PME) and uncorrelated factors for each of the rated messages (AM, AR1, AR2, FO, LE, or UR) to control for message-specific variance. $\chi^2$ = likelihood ratio (adult convenience) and limited information (adult and adolescent probability) goodness-of-fit indices; RMSEA = root mean squared error of approximation; $Z_h$ = person-fit index summarized as percentage of cases with good fit.
Figure 3.1 Information curves for the UNC Perceived Message Effectiveness Scale. Information is a measure of the variability that a scale captures about the construct of interest and varies across the possible range (standardized) of the construct. Higher information points to lower standard error of measurement and greater reliability.
Figure 3.2 Effect sizes for differential item functioning (DIF) and differential test functioning (DTF) exhibited by the UNC Perceived Message Effectiveness Scale. Interpretation of Cohen’s $d$ (with respect to absolute value): negligible DIF or DTF, $d \leq 0.2$; and DIF or DTF requiring further investigation, $d > 0.2$. 


CHAPTER 4: RESEARCH DESIGN AND METHODS

As mentioned in Chapter 2, researchers have seldom empirically investigated the construct validity of effects and message perceptions while continuing to use them interchangeably to measure PME. The validation study of the UNC PME Scale, which focuses on effects perceptions, found that it demonstrated convergent validity with message perceptions as well as message reactance and message credibility (see Chapter 3). Building on this foundation, Aims 2 and 3 will examine the construct validity of the UNC PME Scale and an established message perceptions scale in two ways. Aim 2 will examine the predictive validity (i.e., the $b$-pathway in a simple mediation model) of both scales with a set of constructs that are relevant to chemical message processing. Aim 3 will subsequently identify which scale is a better predictor of the impact of chemical messages on a number of important tobacco control outcomes (through mediation analyses). For clarity, I refer to the UNC PME Scale the effects perceptions scale throughout in the remainder of this chapter to highlight the conceptual difference between it and the existing message perceptions scale.

Overview of parent study

The proposed research will use data from a four-week randomized clinical trial funded by the FDA and NCI (P50CA180907), which aimed to assess the impact of cigarette pack chemical messages on quit intentions. The trial recruited a convenience sample of 719 adult (ages $\geq 21$) smokers in California through Craigslist, Facebook, and newspaper postings as well as in-person methods with data collection ending in March 2017. The sample had a mean age of 42 ($SD = 13$)
and was diverse, including many African Americans (30%), low-income participants (57%), low-education participants (22%), and sexual minorities (25%).

Smokers were randomized to receive three chemical messages \((n = 360)\) or three control messages about properly disposing cigarette litter \((n = 359)\) in essentially equal numbers. Control messages were attention-matched to the chemical messages and had similar word lengths, literacy requirements, and visual features (Figure 4.1). Participants completed the baseline survey at visit 1 and a brief survey at each of visits 2 to 5 over the next four weeks. During the first four visits, participants brought eight days’ worth of cigarettes. While participants in the intervention condition completed the follow-up survey at visits 2 through 4, study personnel applied a chemical message on the side of their cigarette packs. Smokers in the intervention condition rotated through the three chemical messages, being given one chemical message at each of visits 2 through 4 in one of six randomly selected orders. Procedures in the control arm of the trial were identical to those in the intervention arm. Additional details on the trial design and protocol are available in a separate paper.22

**Data preparation**

Measures for Aims 2 and 3 can be found in the Appendix. All analyses will be conducted using data from visit 5 only. Data preparation will follow the last-observation-carried-forward procedures (as part of the intent-to-treat analysis) used in the main trial paper to maintain parity in terms of sample size and results on any primary and secondary outcomes also examined here. These procedures will maximize the sample size while minimizing missingness across all measures. Any missingness at this stage, which is likely to be minimal, will be handled via full information maximum likelihood (FIML) estimation. In contrast to other estimation methods that use only the complete dataset, FIML constructs a model likelihood function using the incomplete
dataset by summating across case-wise likelihood functions for all persons with responses on at least some variables in the model.\textsuperscript{35}

I will conduct preliminary analyses on data from the effects and message perceptions scales because of their fundamental importance to this research. I will examine the distribution of responses to each of the items on both scales to confirm that the items should be specified as ordinal in all analyses similar to the validation study. Consistent deviations from the standard normal distribution in skewness and kurtosis for several items would support the ordinal specification over the continuous specification. Assuming that the ordinal specification is optimal, I will calculate ordinal Cronbach’s alpha\textsuperscript{33} for both scales separately to confirm that the approximate reliability of the scales is similarly high as in the validation study.

I will use R (ver. 3.5.1) for data preparation, preliminary analyses, and replicating randomization checks from the main trial paper using simple or multinomial logistic regression. The main trial paper found that trial condition did not differ by a set of standard demographic variables.\textsuperscript{22} I will use Mplus (ver. 8) to estimate all other models. All analyses will use a critical alpha of .05. Finally, I will adjust the methods for Aims 2 and 3 as needed.

Analysis plan for Aim 2

Aim 2 will examine the construct validity of the effects and message perceptions scales using confirmatory factor analytic models. The psychometric literature on incorporating multioccasion into CFA models,\textsuperscript{69,81,92} special CFA models for assessing construct validity,\textsuperscript{28,53,120} and measurement invariance testing\textsuperscript{76,78,94} informed the proposed methodological approach for Aim 2. In pursuing Aim 2, I will examine two hypotheses the first of which is that the effects perceptions and message perceptions scales measure distinct, albeit somewhat overlapping, constructs (Hypothesis 1). I will estimate unidimensional and bidimensional CFA
models. The bidimensional model will have two correlated factors, one for the items from the effects perceptions scale and the other for the items from the message perceptions scales. Items from one scale will not cross-load onto the other scale’s factor because I am strictly interested in testing the overall dimensionality of the nine items from both scales. In the unidimensional model, the nine items will load onto a general factor for PME. The unidimensional model will be nested within the bidimensional model because the bidimensional model will estimate one additional parameter, the correlation between both factors. Thus, I will use the likelihood ratio (LR) test for nested models to confirm the following prediction:

**Prediction 1.1:** The bidimensional CFA model with correlated individual factors for effects perceptions and message perceptions will exhibit better fit than the unidimensional model with one general factor for PME.

A statistically significant LR test would indicate that the bidimensional model has better fit than the unidimensional model. I will confirm model selection by examining the Bayesian information criterion (BIC). A negative BIC for the bidimensional model and a positive BIC for the unidimensional model would amount to absolute support for the bidimensional model over the unidimensional one. In the event that both BICs are positive, a lower BIC for the bidimensional model would amount to relative support for the bidimensional model over the unidimensional one. If the LR test and the BIC support the bidimensional model, I will interpret the correlation between the factors to assess whether the nine items are practically bidimensional. A factor correlation of .90 or higher would suggest that the items are measuring a single construct and further exploration of the unidimensional model is warranted.

Regardless of whether the retained model is unidimensional or bidimensional, I will examine modification indices involving only items on the message perceptions scale. There are
two substantive reasons for limiting the examination of modification indices in this way. First, unlike the effects perceptions scale, the message perceptions scale generally does not use behavioral or personal referents, which may introduce unknown sources of multidimensionality in or increase measurement error in at least some of the items. Second, the validation study of the effects perceptions scale found that the scale had noticeably lower measurement error than the message perceptions scale (see Chapter 3). The statistical reason for limiting the examination of modification indices is that many can be data-driven anomalies owing to the extensive power afforded by the CFA framework to detect minute deviations in model fit that do not reflect true specification errors. I will start with the largest statistically significant modification index involving only items from the message perceptions scale and use the BIC to confirm that estimating the associated path improves model fit. A meaningful reduction in the BIC (> 10) would suggest that the corresponding path should be retained in the model. I will continue to sequentially examine modification indices until a suggested modification does not lead to a meaningful reduction in the BIC or none involving only items assessing message perceptions are statistically significant after respecification. At this point, I will examine the root mean square error of approximation (RMSEA; < .06) and comparative fit index (CFI; ≥ .95) to ensure that I have arrived at an appropriate stopping point.

The final model I will estimate is a multiple-group version of the respecified CFA model with trial condition as the grouping variable. Multiple-group procedures simultaneously fit a given CFA model in two or more groups while providing pooled absolute and relative fit statistics that follow from the rules of probability. They are commonly used to examine the extent of measurement non-invariance in a scale when administered among different groups. As the intervention group received chemical messages while the control group received littering
messages, it is possible that the factorial structure of the nine items differs between the two types of messages. Multiple-group procedures will allow me to formally examine this possibility. After estimating a multiple-group version of the respecified CFA model, I will compare it to the pooled CFA model using the LR test to confirm the following prediction:

**Prediction 1.2:** The multiple-group bidimensional CFA model with separate parameter estimates for the intervention and control groups will not exhibit better fit than the bidimensional CFA model with pooled estimates.

A statistically significant LR test would indicate that at least one estimated parameter varies between the intervention and control groups. I will also examine the BIC so that a lower BIC for the multiple-group model would confirm the results of a statistically significant LR test and require further measurement invariance testing. On the other hand, a statistically significant LR test but a higher BIC for the multiple-group model would suggest that the model is capturing minor discrepancies in estimated parameters between the intervention and control groups that do not reflect true differences in factorial structure. The validation study of the effects perceptions scale found minimal measurement non-invariance (i.e., differential item functioning) in the scale across a total of six unique chemical messages and similar results for the message perceptions scale in the context of special CFA models for construct validity testing (see Chapter 3). Based on these findings and the greater power to detect measurement non-invariance in the validation study, I strongly expect that I will not need to pursue further measurement invariance testing in Aim 2. Thus, for the sake of brevity, the corresponding quantitative procedures, which are similar to those used in the validation study, are not outlined here.

The second hypothesis is that the effects perceptions scale measures a construct that is closer to behavior than the message perceptions scale (Hypothesis 2). To evaluate this
hypothesis, I will examine the association of both scales simultaneously with 14 construct validators. The selection of validators is addressed later in this section, but one example of a validator is number of times butting out or forgoing a cigarette in the past week (micro-quitting), which I use to describe the analytic approach for examining the construct validity of both scales. The validation study of the effects perceptions scale found that the relevant measurement models with separate dimensions for the effects and message perceptions scales had adequate global fit strongly suggesting that Prediction 1.1 will be formally confirmed (see Chapter 3). Thus, the starting point for the analysis will be the bidimensional CFA model used in the examination of Hypothesis 1. I will extend the bidimensional model by adding micro-quitting as a count (i.e., Poisson) manifest variable. Specifically, I will regress micro-quitting on the effects and message perceptions scales and use a $z$-test of the difference between the associations to formally confirm the relevant construct validity prediction (provided below). I will also confirm that the extended model has admissible fit using the BIC, RMSEA, and CFI criteria mentioned previously. I will use the analytic approach for micro-quitting for construct validity analyses involving the other 13 validators with minor modifications as necessary. In the case that a validator was measured using a multi-item scale, I will use a latent variable for that validator. The type of latent or manifest variable will correspond to the response scale(s) of the corresponding item(s).

The selection of construct validators was guided primarily by the Message Impact Framework (MIF) and the UNC Tobacco Warnings Model (TWM). The MIF is an organizational framework for reviewing evidence on how brief messages impact behavior. The TWM, which was partly derived from the MIF and developed in the context of pictorial cigarette warnings, suggests that particular characteristics of chemical messages increase attention to and noticing the messages, which increase negative affective reactions to and social interactions...
about the messages. The previous changes increase thinking about the warnings and thereby motivation to quit in the context of repeated exposure to the messages, often during product use. The MIF differs from the TWM in that it does not make any formal predictions about how brief messages operate and permits a larger range of constructs to potentially determine their impact on behavior.

Based on the MIF and TWM, I have grouped the 14 validators into four categories (Table 4.1). Early mediators will be perceived understandability, attention or noticing, and recognition of the chemical messages. Intermediate mediators will be negative affective reactions to the messages, negative pack attitudes, avoidance of cigarette pack labels, and number of conversations about the labels. Negative affective and related message reactions are important because they enable people to create meaning around abstract information. Similarly, avoidance of chemical messages is a behavior that is largely driven by negative affect and is a marker for deeper message processing. Finally, late mediators will be thinking about the harms of smoking, thinking about the chemicals in cigarettes, and quit intentions.

Behavioral outcomes will be chemical information seeking in addition to micro-quitting as mentioned previously. Prior theoretical and empirical research has demonstrated the importance of these behavioral outcomes in tobacco control. First, a limited number of prior studies have shown that micro-quitting behaviors, including butting out and forgoing a cigarette, are predictive of subsequent quit attempts. Second, information-seeking behaviors can reinforce the effects of health communication interventions for cigarette smoking and other contexts. I will also examine two informative outcomes, awareness of chemicals in intervention messages and awareness of health effects in intervention messages because
chemical messages are legally intended to accurately communicate the risks of exposure to the toxic chemicals in cigarette smoke as well as motivate behavior.\textsuperscript{1,22}

In the remainder of this section, I develop predictive validity expectations for the effects and message perceptions scale with respect to the five categories of construct validators beginning with the early mediators. The TWM and MIF posit that message characteristics function as gatekeepers to paying attention to, noticing,\textsuperscript{24} and later recalling a chemical message.\textsuperscript{108} The message perceptions scale focuses on the extent to which a recipient thinks that a message has characteristics that enable further processing of it. Thus, the message perceptions scale conceptually overlaps with message characteristics themselves and is likely associated with other early mediators that message characteristics are associated with as well. In contrast, there is minimal conceptual support for the effects perceptions scale being associated with the early mediators mainly because it primarily focuses on perceived changes in constructs that could be conceptualized as late mediators such as motivation to quit smoking. These observations lead to the following joint predictions:

Prediction 2.1: The message perceptions scale will be associated with the early mediators.

Prediction 2.2: The effects perceptions scale will not be associated with the early mediators.

If these predictions hold, then the message perceptions scale will demonstrate predictive validity with the early mediators through moderate to moderately high positive associations, whereas the effects perceptions scale will not demonstrate predictive validity with the early mediators.

Next, I address intermediate mediators of chemical message impact. The TWM and MIF posit that paying attention to and noticing the chemical messages leads to negative affect and other message reactions such as negative pack attitudes and avoidance.\textsuperscript{24,108} The effects
perceptions scale has items that appear to sample aspects of negative affect such as the unpleasant item and, to a lesser extent, the concern item. Thus, the effects perceptions scale conceptually overlaps with negative affective reactions to the chemical messages and negative pack attitudes. The effects perceptions scale is also conceptually close to avoidance largely due to their mutual overlap with negative affect. In contrast, the message perceptions scale does not have any items related to negative affective or other message reactions. Given that negative affect is an intermediate mediator, these differences also suggest that the effects perceptions scale captures information about later stages of the causal pathway through which chemical messages operate and the message perceptions scale captures information mostly about the earlier stages. The TWM and MIF also emphasize that chemical messages may influence smoking behavior through message-directed social interactions similar to pictorial cigarette warnings. The effects and message perceptions scales do not conceptually overlap with these social interactions and likely do not provide any relevant information about them. These observations lead to the following joint predictions:

**Prediction 2.3**: The effects perceptions scale will be more positively associated with the intermediate mediators except for social interactions than the message perceptions scale.

**Prediction 2.4**: Neither the effects perceptions scale nor the message perceptions will be associated with social interactions.

If these predictions hold, then the effects perceptions scale will demonstrate predictive validity with intermediate mediators except social interactions through moderate to moderately high positive associations. The message perceptions scale will do the same except through weak to moderate positive associations.

The TWM also posits that negative affective reactions to and social interactions about
chemical messages lead to increases in thinking about the messages and, in turn, greater quit intentions.\textsuperscript{24} These effects on late mediators most likely come about through deliberative processes that the effects perceptions scale also captures information about. A comparison of the effects perceptions scale with the Tripartite Model of Risk Perception (TRIRISK) suggests some conceptual overlap between the effects perceptions scale and the deliberative sub-scale of the TRIRISK model.\textsuperscript{57} In addition, the \textit{discouragement} item on the effects perceptions scale most likely captures motivation to quit smoking.\textsuperscript{131} Therefore, the effects perceptions scale is conceptually close to the late mediators and captures some information about the impact of chemical messages on them. In contrast, the message perceptions scale does not appear to sample any aspects of the relevant deliberative processes or constructs that are theoretically similar to quit intentions. These observations lead to the following joint predictions:

\textbf{Prediction 2.5}: The effects perceptions scale will be associated with the late mediators.

\textbf{Prediction 2.6}: The message perceptions scale will not be associated with the late mediators.

If these predictions hold, then the effects perceptions scale will demonstrate predictive validity with the late mediators through moderate to moderately high positive associations, whereas the message perceptions scale will not demonstrate predictive validity with the late mediators.

Finally, the TWM and MIF posit that greater quit intentions from exposure to chemical messages ultimately lead to changes in smoking behavior. The effects perceptions scale is conceptually closer to behavior than the message perceptions scale for two reasons. First, it references smoking uniformly in each of its items, and specifying behavior is a formal requirement for a measure to demonstrate predictive validity with behavior.\textsuperscript{59} Second, the effects perceptions scale focuses on perceived changes in proximal determinants of behavior from
exposure to chemical messages. In contrast, the message perceptions scale does not satisfy either criterion. Both criteria are not relevant to predicting changes in informative outcomes from exposure to chemical messages. Although, the MIF and TWM are not explicitly concerned with informative outcomes, some of the early and intermediate mediators may also play roles in health information seeking and processing.\textsuperscript{83,151} Furthermore, informative outcomes are likely easier for chemical messages to change than behavioral outcomes and may only require prior increases in some of the early and intermediate mediators like attention or noticing. Both effects and message perceptions scales theoretically overlap with most intermediate mediators and the message perceptions scale also does so with the early mediators, suggesting that both scales are close to informative outcomes. These observations lead to the following joint predictions:

- **Prediction 2.7**: The effects perceptions scale will be associated with the behavioral outcomes.

- **Prediction 2.8**: The message perceptions scale will not be associated with the behavioral outcomes.

- **Prediction 2.9**: The effects perceptions and message perceptions scales will be similarly positively associated with the informative outcomes.

If these predictions hold, then the effects perceptions scale will demonstrate predictive validity with both behavioral outcomes and both informative outcomes through weak to moderate positive associations. In addition, I expect the message perceptions scale not to demonstrate predictive validity with both behavioral outcomes through no associations and to demonstrate predictive validity with both informative outcomes through weak to moderate positive associations.
Analysis plan for Aim 3

In Aim 3, I will identify whether the effects perceptions scale or the message perceptions scale is a better proxy for the behavioral impact of brief messages. The specific hypothesis that I will test using structural equation modeling (SEM) is that only the effects perceptions scale is a proxy for the impact of chemical messages on smoking behavior (Hypothesis 3). The proposed analytic approach for Aim 3 is essentially an extension of that for Aim 2 and was informed by the methodological literature on mediation analyses, causal effects, and general SEM. The criteria for fit statistics mentioned in the proposed analytic approach for Aim 2 will also apply here.

I will estimate a complex mediation model with number of times butting out or forgoing a cigarette in the past week as the behavioral outcome (Figure 4.2). The corresponding measurement model will have already been tested in Aim 2. Text-only cigarette pack labels do not motivate quitting to the same extent that pictorial cigarette warnings do, and chemical messages specifically increase micro-quitting. Therefore, micro-quitting appears to be a more appropriate behavioral outcome for these analyses than sustained quitting. The complex mediation model will formally test a number of key predictions:

Prediction 3.1: Chemical messages will elicit greater message perceptions than control messages.

Prediction 3.2: Chemical messages will elicit greater effects perceptions than control messages.

Prediction 3.3: Message perceptions will be associated with effects perceptions.

Prediction 3.4: Message perceptions will not be associated with number of times butting out or forgoing a cigarette in the past week.
Prediction 3.5: Effects perceptions will be associated with number of times butting out or forgoing a cigarette in the past week.

Prediction 3.6: Message perceptions will not mediate the effect of chemical messages on number of times butting out or forgoing a cigarette in the past week, neither individually nor serially through effects perceptions.

Prediction 3.7: Effects perceptions will mediate the effect of chemical messages on number of times butting out or forgoing a cigarette in the past week.

These predictions follow from a number of important observations. First, as mentioned previously, the RCT used attention-matched control messages with similar word lengths, literacy requirements, and visual features as the intervention messages. Second, chemical messages generally elicit greater message perceptions than other control messages like the existing Surgeon General’s warning about carbon monoxide.6 Third, some researchers have suggested that message perceptions may be an antecedent to effects perceptions. Finally, as mentioned previously, effects perceptions are conceptually closer to behavior than message perceptions.

As observed, this model will specify mediational paths involving effects and message perceptions simultaneously, allowing both to compete against each other for which is the better behavioral marker. Following standard procedures for mediation and path analysis, mediated effects will be calculated as the product of two or more relevant path coefficients, and their statistical significance will be examined using z-tests.117,126 Given that both scales were strongly positively correlated in the validation study, this approach is advantageous because, for a given scale, it will provide corresponding path coefficients while controlling for the effects of the other scale particularly on micro-quitting behavior. By extension, the model should enable pure statistical tests of mediation for both effects and message perceptions.
Finally, I address potential concerns regarding the temporal ordering of both scales and micro-quitting. Effects and message perceptions were assessed at visit 5 while micro-quitting was assessed at each visit. To support mediation analyses, effects and message perceptions would have ideally been assessed during at least one of visits 2 to 4. This was not done to isolate the effect of chemical messages from any testing effects on trial outcomes. Although there is a strong conceptual basis for at least effects perceptions to determine behavior and not the other way around (see Chapter 2), I will also conduct similar analyses for chemical information seeking, social interactions about the messages, and quit intentions and base general inferences on patterns across the four models.

**Power calculations**

Power analyses for SEM (and CFA) models used a semi-iterative approach that operationalizes the null and alternative models using the RMSEA and uses this information to determine whether a sample size is sufficient to differentiate between misspecified and adequately specified SEM models.\(^{88,89}\) Preliminary analyses conducted using a subset \((n = 652)\) of the RCT data found that an SEM model with some important minor specification errors had an RMSEA of \(.085\) (95% CI: \(.074–.096\)) and an adequately specified SEM model had an RMSEA of \(.052\) (95% CI: \(.040–.065\)). I used this information to conduct two sets of power analyses for Aims 2 and 3. The first set of analyses operationalized null and alternative models using point estimates for the RMSEA from preliminary analyses. The second set of analyses used the boundaries of the confidence interval for the RMSEA of the adequately specified SEM model. According to both versions, I will have at least 98% power for an adequately specified SEM model with 41 or more degrees of freedom, given the final RCT sample size of 719 smokers.
Table 4.1

Construct validators categorized according to relevance in message processing based on the Message Impact Framework and UNC Tobacco Warnings Model.

<table>
<thead>
<tr>
<th>Role in message impact</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early mediator</td>
<td><em>Perceived understandability</em></td>
</tr>
<tr>
<td>Early mediator</td>
<td><em>Attention and noticing</em></td>
</tr>
<tr>
<td>Early mediator</td>
<td><em>Recognition</em></td>
</tr>
<tr>
<td>Intermediate mediator</td>
<td><em>Negative affect</em></td>
</tr>
<tr>
<td>Intermediate mediator</td>
<td><em>Pack attitudes</em></td>
</tr>
<tr>
<td>Intermediate mediator</td>
<td><em>Avoidance of cigarette pack labels</em></td>
</tr>
<tr>
<td>Intermediate mediator</td>
<td><em>Number of conversations about the label in the past week</em></td>
</tr>
<tr>
<td>Late mediator</td>
<td><em>Quit intentions</em></td>
</tr>
<tr>
<td>Late mediator</td>
<td><em>Thinking about the chemicals in cigarettes</em></td>
</tr>
<tr>
<td>Late mediator</td>
<td><em>Thinking about the harms of smoking</em></td>
</tr>
<tr>
<td>Behavioral outcome</td>
<td><em>Number of times butting out or forgoing a cigarette in the past week</em></td>
</tr>
<tr>
<td>Behavioral outcome</td>
<td><em>Chemical information seeking</em></td>
</tr>
<tr>
<td>Informative outcome</td>
<td><em>Awareness of health effects in intervention messages</em></td>
</tr>
<tr>
<td>Informative outcome</td>
<td><em>Awareness of chemicals in intervention messages</em></td>
</tr>
</tbody>
</table>
Figure 4.1. Cigarette pack labels placed on smokers’ cigarette packs in intervention (A) and control (B) arms.
Figure 4.2. Complex mediation model of the impact of chemical messages on micro-quitting through effects perceptions and message perceptions.
CHAPTER 5: PREDICTIVE VALIDITY OF MESSAGE PERCEPTIONS AND EFFECTS PERCEPTIONS IN THE CONTEXT OF ANTI-SMOKING MESSAGES

Overview

Background. To select the most promising messages from a larger pool of candidate messages, formative research has often relied on perceived message effectiveness (PME) scales assessing either of two related constructs, message perceptions (persuasive potential) and effects perceptions (potential for behavioral impact). We sought to examine the predictive validity of these two constructs.

Methods. Participants were 703 U.S. adult smokers (ages ≥ 21). We randomly assigned them to receive brief messages about toxic chemicals in cigarette smoke or properly disposing cigarette litter on their cigarette packs for three weeks. The survey assessed message perceptions and effects perceptions with two established scales. The final visit survey assessed message perceptions, effects perceptions, and important outcomes. Structural equation models examined message perceptions and effects perceptions as simultaneous predictors.

Results. More positive message perceptions were associated with greater attention to the message ($\beta = 0.82, p < .001$), but effects perceptions were not. More positive effects perceptions only were associated with more negative affective message reactions, more thinking about the chemicals in cigarette smoke, more thinking about the harms of smoking, and stronger quit intentions (range $\beta = 0.74–0.87$, all $p < .01$). More positive effects perceptions were also associated with more frequent butting out a cigarette, more frequent forgoing a cigarette, and being more likely to engage in a quit attempt ($\beta = 0.36–0.66$, all $p < .001$). Message perceptions
were not associated with these outcomes.

**Discussion.** Message perceptions demonstrated predictive validity with attention, an early behavioral antecedent from the UNC Tobacco Warnings Model, while effects perceptions demonstrated predictive validity with later antecedents as well as quitting behaviors. Formative research on tobacco risk messages may benefit from focusing on message perceptions early on and shifting to effects perceptions to characterize messages’ potential for behavior change.

**Keywords:** Effects perceptions, formative research, measurement, perceived message effectiveness, structural equation modeling

Predictive Validity of Message Perceptions and Effects Perceptions in the Context of Anti-Smoking Messages

Health communication efforts commonly employ brief messages designed to discourage health risk behaviors. Researchers often use audience ratings of health messages or perceived message effectiveness (PME) to select the most promising ones for further clinical testing or dissemination. For example, previous studies have used PME to identify potent messages about the chemicals in cigarette smoke (chemical messages) and evaluate HIV testing messages already out in the field. In the absence of behavioral data on message impact, PME ratings are a cost-effective means to efficiently evaluating many candidate messages.

While use of PME has steadily increased over the last two decades, little is definitively known about the extent to which PME is indicative of message impact on behavior. A major hurdle to evaluating the predictive validity of PME ratings is the heterogeneity of PME measures, which have traditionally assessed either of two constructs, message perceptions and effects perceptions. Message perceptions are judgments about whether a message will promote further processing that leads to persuasion, and effects perceptions are judgments about
a message’s potential to change important antecedents of behavior or behavior itself. To our knowledge, only one study has formally examined the structure of both types of perceptions and found that they constituted distinct constructs.\textsuperscript{50} However, due to their high correlation, the researchers ultimately concluded that message perceptions and effects perceptions could be used interchangeably in applied settings.\textsuperscript{50}

Message perceptions and effects perceptions have important differences, which may preclude their interchangeable use in formative research. One difference is that message perceptions focus on characteristics that facilitate initial processing of the message itself. These characteristics may include message credibility (e.g., "How believable was the message in this?")\textsuperscript{52} and perceived argument strength (e.g., "This ad was convincing.").\textsuperscript{85} In contrast, effects perceptions focus on the behavior that a message is designed to influence. Dimensions of effects perceptions may include perceived impact on behavioral motivation (e.g., "This ad makes me want to quit smoking.")\textsuperscript{101} or an antecedent such as concern or worry (e.g., "This ad made me feel concerned about my smoking.").\textsuperscript{19} Based on this difference in focus (the message or behavior), we hypothesize that message perceptions and effects perceptions are distinct constructs (\textit{bidimensionality hypothesis}).

Another difference is that effects perceptions items ask recipients of a message to think about how that message affects their own behavior using behavioral and personal referents. The use of behavioral referents shifts the meaning of effects perceptions away from general persuasive potential, as in the case of message perceptions, towards motivation to change behavior. The use of personal referents further increases the specificity of effects perceptions to a message recipient’s motivation to change their own behavior. The principle of compatibility, originally articulated in the context of the Theory of Reasoned Action, indicates that a construct
must be measured at the same level of specificity as the behavior of interest for it to demonstrate predictive validity with that behavior. Therefore, we hypothesize that effects perceptions are a better predictor of behavior than message perceptions (behavioral compatibility hypothesis).

By pursuing the bidimensionality and behavioral compatibility hypotheses, our study ultimately sought to shed light on the distinction between message perceptions and effects perceptions and which of these constructs is more relevant to developing messages for behavior change. We examined the predictive validity of message perceptions and effects perceptions in the context of anti-smoking messages about the chemicals in cigarette smoke (chemical messages) as a case study. Our data came from the final visit of a randomized clinical trial (RCT) of the impact of chemical messages on intentions to quit smoking.

Methods

Participants and procedures

A convenience sample of 719 U.S. adult smokers from the San Francisco Bay Area participated in an RCT that compared the impact of chemical messages versus anti-littering messages. Specifically, participants were ages 21 or older following the legal age of buying tobacco products in California, proficient in English, and current smokers (i.e., had smoked at least 100 cigarettes in their lifetime and now smoke every or some days). Exclusion criteria were pregnancy, enrollment in an ongoing smoking cessation trial, living in the same household as another trial participant, and smoking fewer than seven cigarettes per week or only roll-your-own cigarettes.

Smokers attended visits at the trial office and brought an eight-day supply of cigarettes to all but the last visit. Participants were randomized to receive chemical messages (intervention) or messages about properly disposing cigarette litter (littering messages; control) on the sides of
their cigarette packs each week for three weeks. An example of a chemical message is, "Cigarette smoke contains formaldehyde. This causes throat cancer"; and an example of a littering message is, "Cigarette butts don’t biodegrade. Please do not litter." While participants in the intervention condition completed weekly surveys, study personnel applied a new chemical message on the side of their cigarette packs. Thus, the smokers in the intervention arm rotated through three chemical messages during the course of the RCT in a random order. The control arm followed identical procedures. Smokers received up to $300 for their participation in the trial. Data collection lasted from September 2016 to March 2017. Additional details on design and protocol are available in the main RCT paper\textsuperscript{22}.

**Measures**

The baseline survey assessed participant demographics and smoking behavior. The final visit survey assessed message perceptions using an established scale with six items: "This message is worth remembering"; "This message grabbed my attention"; "This message is powerful"; "This message is informative"; "This message is meaningful"; and "This message is convincing."\textsuperscript{41} We assessed effects perceptions also at Visit 5 using the UNC Perceived Message Effectiveness Scale with three items: "This message discourages me from wanting to smoke"; "This message makes smoking seem unpleasant to me"; and "This message makes me concerned about the health effects of smoking."\textsuperscript{7} The response options for both scales ranged from "strongly disagree" to "strongly agree" (coded as 1–5).

The UNC Tobacco Warnings Model (TWM) informed the selection of major construct validators.\textsuperscript{22} These were attention to the labels, an early antecedent to behavior from the TWM; the two intermediate antecedents number of conversations about the labels in the past week and negative affect; and the three late antecedents thinking about the chemicals in cigarette smoke,
thinking about the harms of smoking, and quit intentions (Table 5.1). Given our behavioral compatibility hypothesis, we expected message perceptions to be positively associated with the early and intermediate antecedents and not be associated with the late antecedents. In contrast, we expected effects perceptions to not be associated with the early antecedents and be positively associated with the intermediate and later antecedents. The main behavioral outcomes were number of times butting out a cigarette in the past week, number of times forgoing a cigarette in the past week, and weekly recall of quit attempts. Given that quitting initiation exists on a spectrum,113 examining multiple quitting behaviors that differ in intensity is likely to add depth to the predictive validity findings. We expected message perceptions to not be associated with and effects perceptions to be positively associated with the behavioral outcomes.

Because chemical messages can have a range of effects,22 we also included recognition of the labels, avoidance of the labels, and seeking information about the chemicals in cigarette smoke as construct validators (Table 5.1).108 Recognition conceptually overlaps with attention; avoidance is a largely affect-driven behavior that is productive for tobacco warning impact;63 and chemical information seeking may reinforce some of the educational effects of tobacco warnings.83 The final visit survey assessed all TWM constructs, behavioral outcomes, and other construct validators.

Data analysis

Analyses used R (ver. 3.5.1)121 with three add-on packages, psych (ver. 1.8.4)122 for calculating ordinal reliability coefficients, lavaan (ver. 0.6-2)124 for estimating measurement and structural models, and ggplot2 (ver. 3.0.0)148 for creating figures.

Participants with complete missingness on both predictors, message perceptions and effects perceptions, were dropped from the analyses resulting in an analytic sample of 703
smokers. Single-item outcomes did not have any missingness, and partial missingness on multi-item outcomes was negligible (< 3%). Therefore, we used pairwise deletion to handle missingness on the outcomes. Doing so enabled us to use ordinal estimation for all models without having to drop any additional participants. Analyses used data exclusively from the final visit, and statistical tests used a critical alpha of .01 (or a 99% confidence interval) to account for the possibility of artificially inflated validity associations due to assessing predictors and validators at a single time point.

**Dimensionality.** We examined the dimensionality of the message perceptions and effects perceptions scales by estimating unidimensional and bidimensional confirmatory factor analytic (CFA) models. The unidimensional model permitted the nine items from both scales to load onto a general factor for PME. The bidimensional model allowed the six message perceptions items and the three effects perceptions items to load onto two separate but correlated factors without any cross-loadings. We used the likelihood ratio (LR) test for nested models to compare both CFA models.

Respecification of the preferred CFA model followed an iterative process in which we tested correlated errors based on substantial modification indices (MIs), beginning with the largest and sequentially moving to the smallest. For each MI, we confirmed that estimating the associated path statistically significantly improved model fit via the LR test. After exhausting all substantial MIs, we examined the Bayesian Information Criterion (BIC), Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI) for adequacy (BIC < 0; RMSEA < .06; CFI > .95). We also substantively confirmed that suggested respecifications should be retained in the final model by examining relevant item wordings. Finally, we confirmed that the message perceptions and effects perceptions scales functioned similarly in the
context of chemical and littering messages using standard procedures for testing measurement invariance involving comparisons of increasingly constrained multiple-group versions of the respecified CFA model.  

**Predictive validity.** We examined the predictive validity of message perceptions and effects perceptions using structural equation modeling (SEM). For a given outcome, we estimated a structural model that simultaneously regressed the outcome on the latent variables for message perceptions and effects perceptions using the best fitting bidimensional CFA model as the measurement model. The resulting associations represented the extent to which one type of perception predicted the outcome after controlling for the other type of perception. This was a departure from the small number of longitudinal and meta-analytic PME validation studies, 19, 40, 49, 112 which have focused on independent associations between predictors and outcomes, making our validation efforts inherently comparative. The analytic approach was identical for all outcomes except that we specified an outcome measured with a single item as a manifest variable and one measured with a multi-item scale as a latent variable. All multi-item outcomes demonstrated high reliability ($\alpha \geq 0.84$). We confirmed model fit using the BIC, RMSEA, and CFI based on previously mentioned criteria.

**Results**

More than a third of the participants were white (37.4%) or African-American (35.6%; Table 5.2). Nearly a quarter were gay, lesbian, or bisexual (25.3%). A sizable minority of the participants smoked only some days (22.9%).

**Dimensionality**

Message and effects perceptions represented conceptually distinct but overlapping constructs (Figure 5.1). In a bidimensional measurement model, the message perceptions and
effects perceptions items loaded strongly (0.78–0.92) onto separate, but highly correlated factors ($r = 0.82$). The message and effects perceptions items also loaded strongly (0.80–0.90) onto a general factor for perceived message effectiveness in a unidimensional measurement model. However, the bidimensional model had better fit to the data than the unidimensional model ($\chi^2 = 136, p < .001$). Model fit indices improved in the bidimensional model (BIC = 106; RMSEA = .12; CFI = .99) as compared to the unidimensional model (BIC = 517; RMSEA = .19; CFI = .98) thereby confirming the result of the LR test.

The bidimensional model needed to account for local dependence mostly in the message perceptions items to achieve adequate global fit. Thus, we correlated the errors on the attention-grabbing and powerful, the informative and meaningful, and the meaningful and convincing message perceptions items. LR tests confirmed that these three modifications improved model fit (range $\chi^2 = 19.9–56.4$, all $p < .001$). We also correlated errors on the discouragement and unpleasant effects perceptions items, which improved model fit ($\chi^2 = 21.6, p < .001$). The final bidimensional model had adequate fit (BIC = -64.2; RMSEA = .061; CFI = 1.00).

The message perceptions and effects perceptions scales functioned similarly among participants who received chemical or littering messages. More specifically, LR tests for measurement invariance found that the final bidimensional model held exactly among participants who received either type of message except for the latent means for both constructs differing by message type ($\chi^2(2) = 16.6, p < .001$). Common fit indices for variously constrained multiple-group models confirmed that the model with freely estimated latent means while constraining all other parameters to be equal across both groups of participants had the best fit (BIC = -420; RMSEA = 0.055; CFI = 1.00).
Predictive validity

The SEM model for attention to the labels had marginally acceptable fit (BIC = 52.5; RMSEA = .10; CFI = .99). In comparison, the SEM models for the other 11 outcomes had more than adequate fit (BIC median [range] = -90.7 [-198, -27.8]; RMSEA = .06 [.05, .08]; CFI = 1.00 [.99, 1.00]).

Message perceptions demonstrated predictive validity with the TWM construct that is the farthest away from behavior, while effects perceptions demonstrated predictive validity with constructs that are conceptually closer to behavior and with behavior itself. Specifically, more positive message perceptions were associated with greater attention ($\beta = 0.82$, $p < .001$; Figure 5.2). Neither message perceptions nor effects perceptions were associated with conversations about the labels (both $p \geq .27$). More positive effects perceptions were associated with more negative affect, more thinking about the chemicals in cigarette smoke, more thinking about the harms of smoking, and stronger quit intentions (range $\beta = 0.74$–0.87, all $p < .01$). Finally, more positive effects perceptions were associated with more frequent butting out a cigarette ($\beta = 0.36$, $p < .001$), more frequent forgoing a cigarette ($\beta = 0.53$, $p < .001$), and being more likely to engage in a quit attempt ($\beta = 0.66$, $p < .001$). Unexpectedly, more positive message perceptions were also associated with weaker quit intentions ($\beta = -0.29$, $p = .006$) and being less likely to engage in a quit attempt ($\beta = -0.35$, $p = .003$).

A similar pattern of distinct correlates for message perceptions and effects perceptions held for the other outcomes we examined. More positive message perceptions were associated with greater recognition of the messages ($\beta = 0.35$, $p = .003$). More positive effects perceptions were associated with greater message avoidance ($\beta = 0.78$, $p < .001$) and more information seeking about chemicals ($\beta = 0.44$, $p < .001$). Unexpectedly, more positive message perceptions
were also associated with weaker avoidance ($\beta = -0.27, p = .004$). Analyses stratified by trial arm yielded similar findings for all outcomes, although some associations lost statistical significance due to smaller sample size.

**Discussion**

In a diverse sample of adult smokers, message perceptions demonstrated predictive validity with attention to and recognition of the messages. In contrast, effects perceptions represented a distinct construct that demonstrated predictive validity with negative affective reactions to the messages, thinking about the chemicals in cigarette smoke, thinking about the harms of smoking, avoidance of the messages, seeking chemical information, quit intentions, butting out a cigarette, forgoing a cigarette, and quit attempts. Thus, message perceptions were associated with constructs most distal to behavior, and effects perceptions were associated with constructs more proximal to behavior as well as behavior itself (quitting behaviors and other behaviors that are productive for successful quitting). Formative research may use message perceptions in the earliest stages of message development and shift to effects perceptions when the focus is explicitly on behavior change.

Message perceptions and effects perceptions represented separate but highly correlated constructs in the context of both chemical and littering messages, supporting the bidimensionality hypothesis. The effects perceptions scale used behavioral and personal referents while the message perceptions scale did not. This differential use of referents may have contributed to the bidimensionality between message perceptions and effects perceptions. In addition, message perceptions were relevant in the earliest stage of the TWM while effects perceptions were relevant in the later stages, which is consistent with the behavioral compatibility hypothesis. The differential use of behavioral and personal referents may have also
shifted the meaning of effects perceptions closer to the respondent’s behavior as compared to message perceptions, thereby enhancing the correspondence of effects perceptions with behavioral antecedents and behavior. These observations raise questions about whether focusing on the message recipient and referencing the target behavior are defining features of effects perceptions. Furthermore, some existing message perceptions items may be readily adapted to use personal referents (e.g., "This message is meaningful to me.") or behavioral referents (e.g., "This ad was informative about the harms of smoking."). It is unknown how incorporating referents into a measure of message perceptions would change the meaning of the underlying construct. Future studies should pursue these research questions using explanatory item response modeling or generalizability theory, which allow researchers to identify the psychological processes underlying responses to self-report measures.

Our study also found that more negative message perceptions were unexpectedly associated with stronger message avoidance, stronger quit intentions, and being more likely to engage in a quit attempt (after controlling for effects perceptions). These associations may represent statistical anomalies from controlling for effects perceptions, always a possibility when controlling for highly correlated constructs. In addition, the confidence intervals for these unexpected associations included practically null values unlike for other statistically significant associations. Thus, these marginal findings should be interpreted with caution. Given these caveats, we offer a speculation on the unexpected findings. Perhaps audience members who think a message is problematic spend more time on it, leading to greater central processing and thereby greater behavioral motivation. This would be a dual process model in which promising messages motivate behavior, but their impact is somewhat undercut by additional attention to problematic messages.
While our study strongly supports evaluating candidate messages using effects perceptions measures, the actual validity associations with behavioral antecedents were strong and with quitting and other behaviors were weak to moderate. Thus, effects perceptions did not explain substantial amounts of variance in some constructs of interest. This suggests that a researcher could not conclude that messages with adequate effects perceptions will change behavior (i.e., adequate effects perceptions are not a sufficient condition for message impact as theorized by a number of researchers). However, a researcher could conclude that messages without adequate effects perceptions will not change behavior (i.e., adequate effects perceptions are a necessary condition for message impact). This points to the practical value of evaluating candidate messages on effects perceptions: doing so can help researchers identify potent messages to further test for behavioral impact in field trials or longitudinal experiments. Effects perceptions being close to behavior provides a conceptual basis for their use in formative research in this capacity. Future studies should explore what constitutes adequate effects perceptions for message exclusion. Future studies should also identify contexts in which it is useful to assess message perceptions, given that our findings do not allow us to rule out their use in formative research altogether.

Strengths of our study are the large number of theoretically informed validators, including early to late behavioral antecedents and behavioral outcomes, as well as the use of message perceptions and effects perceptions scales with adequate psychometric properties and latent variable models that suited the aim of evaluating the predictive validity of both constructs in a comparative framework. The cross-sectional assessment of both predictors and all construct validators is a limitation of our study. Future studies should replicate our findings using appropriate longitudinal data. Two other limitations are that our study only used one message
perceptions scale and one effects perceptions scale and included only chemical and littering messages. Therefore, our findings may not be generalizable to other PME measures or message testing scenarios. Future validation studies should focus on multiple message perceptions and effects perceptions scales and include diverse messages about many health behaviors and relevant antecedents.

Conclusions

Even though message perceptions and effects perceptions are highly correlated, our study shows that they are indeed distinct constructs. In the context of tobacco use, message perceptions are important in the earliest stages of message processing while effects perceptions are important in the intermediate and late stages. Therefore, we recommend focusing on effects perceptions when evaluating messages for their potential to change smoking behavior and possibly other behaviors.
Table 5.1

Measures for all construct validators with ordinal reliability coefficients for multi-item scales.

<table>
<thead>
<tr>
<th>Construct, measure(s)</th>
<th>Response</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attention to the messages (α = 0.93)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much did the labels grab your attention?</td>
<td>&quot;Not at all&quot;/&quot;Never&quot; (1)</td>
<td>111, 56</td>
</tr>
<tr>
<td>How often did you notice the labels?</td>
<td>&quot;Very much/All the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td>How often did you read or look closely at the labels?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recognition of the messages</strong></td>
<td>&quot;Correct recognition&quot; (1)</td>
<td></td>
</tr>
<tr>
<td>Select the text from the label we put on the side of your cigarette packs at your last visit.</td>
<td>&quot;Incorrect recognition&quot; (0)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of conversations about the label in the past week</strong></td>
<td>0–100 times</td>
<td></td>
</tr>
<tr>
<td>In the last week, how many times did you talk to other people about the label on your cigarette packs?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Negative affect (α = 0.93)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much did the labels on your cigarette packs make you feel?</td>
<td>&quot;Not at all&quot; (1)</td>
<td>77, 46, 11</td>
</tr>
<tr>
<td>How much did the labels on your cigarette packs make you feel?</td>
<td>&quot;Extremely&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Thinking about the chemicals in cigarettes</strong></td>
<td>&quot;Never&quot; (1)</td>
<td></td>
</tr>
<tr>
<td>In the last week, how much did you think about the chemicals in the smoke from your cigarettes?</td>
<td>&quot;All of the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Thinking about the harms of smoking (α = 0.84)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much did the labels make you think about the health problems caused by smoking?</td>
<td>&quot;Not at all&quot;/&quot;Never&quot; (1)</td>
<td>56</td>
</tr>
<tr>
<td>In the last week, how much did you think about the harm your smoking might be doing to you?</td>
<td>&quot;Very much&quot;/ &quot;All of the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td>In the last week, how much did you think about the harm your smoking be doing to other people?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Avoidance of cigarette pack labels (α = 0.89)</strong></td>
<td>&quot;Never&quot; (1)</td>
<td>72</td>
</tr>
<tr>
<td>How often did you try to avoid the labels on your cigarette packs?</td>
<td>&quot;All of the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td>How often did you try to avoid the labels on your cigarette packs?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often did you put your cigarettes away because you didn’t want others to see the labels on your packs?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Seeking information about the chemicals in cigarette smoke</strong></td>
<td>&quot;0 times&quot; (1)</td>
<td>100</td>
</tr>
<tr>
<td>In the last 3 weeks, how many times have you looked for information about the chemicals in cigarettes or cigarette smoke?</td>
<td>&quot;6 or more times&quot; (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Quit intentions (α = 0.96)</strong></td>
<td>&quot;Not at all [item stem]&quot; (1)</td>
<td>80</td>
</tr>
<tr>
<td>How interested are you in quitting smoking in the next month?</td>
<td>&quot;Very [item stem]&quot; (4)</td>
<td></td>
</tr>
<tr>
<td>How much do you plan to quit smoking in the next month?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Response</td>
<td>Count</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-------</td>
</tr>
<tr>
<td>How likely are you to quit smoking in the next month?</td>
<td>&quot;Never&quot; (1)</td>
<td>36</td>
</tr>
<tr>
<td>Number of times butting out a cigarette in the past week</td>
<td>&quot;10 or more times&quot; (5)</td>
<td>36</td>
</tr>
<tr>
<td>In the last week, how often have you butted out a cigarette before you finished it because you wanted to?</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Number of times forgoing a cigarette in the past week</td>
<td>&quot;Never&quot; (1)</td>
<td>36</td>
</tr>
<tr>
<td>In the last week, how often have you stopped yourself from having a cigarette because you wanted to?</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>Weekly recall of quit attempts</td>
<td>&quot;Yes&quot; (1)</td>
<td>51</td>
</tr>
<tr>
<td>In the last week, did you stop smoking for 1 day or longer because you were trying to quit smoking?</td>
<td>&quot;No&quot; (0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2

Participant characteristics

<table>
<thead>
<tr>
<th></th>
<th>n = 703</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>21–29</td>
<td>22.6</td>
</tr>
<tr>
<td>30–39</td>
<td>22.0</td>
</tr>
<tr>
<td>40–49</td>
<td>18.9</td>
</tr>
<tr>
<td>50–59</td>
<td>25.6</td>
</tr>
<tr>
<td>60+</td>
<td>10.9</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50.8</td>
</tr>
<tr>
<td>Female</td>
<td>44.8</td>
</tr>
<tr>
<td>Transgender (includes other gender identity)</td>
<td>4.4</td>
</tr>
<tr>
<td>Gay, lesbian or bisexual</td>
<td>25.3</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>37.4</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>35.6</td>
</tr>
<tr>
<td>Asian</td>
<td>8.4</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>4.8</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>3.3</td>
</tr>
<tr>
<td>Other</td>
<td>10.5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>13.8</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>High school graduate or less</td>
<td>21.8</td>
</tr>
<tr>
<td>Some college</td>
<td>37.8</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>33.1</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>7.3</td>
</tr>
<tr>
<td><strong>Household income, annual</strong></td>
<td></td>
</tr>
<tr>
<td>$0–$24,999</td>
<td>45.2</td>
</tr>
<tr>
<td>$25,000–$49,999</td>
<td>24.0</td>
</tr>
<tr>
<td>$50,000–$74,999</td>
<td>12.4</td>
</tr>
<tr>
<td>$75,000+</td>
<td>18.4</td>
</tr>
<tr>
<td>Low income, &lt; 200% of federal poverty level</td>
<td>57.6</td>
</tr>
<tr>
<td><strong>Smoking frequency</strong></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>77.1</td>
</tr>
<tr>
<td>Non-daily</td>
<td>22.9</td>
</tr>
</tbody>
</table>
Figure 5.1. Confirmatory factor analytic model of message perceptions and effects perceptions. Bayesian Information Criterion = -64.2; Root Mean Squared Error of Approximation = .061; and Comparative Fit Index = 1.00.
Figure 5.2. Predictive validity of message perceptions and effects perceptions with behavioral antecedents from the UNC Tobacco Warnings Model (*), quitting behavior, and related constructs. Error bars denote 99% confidence intervals.
CHAPTER 6: MESSAGE PERCEPTIONS AND EFFECTS PERCEPTIONS AS PROXIES FOR IMPACT ON BEHAVIOR

Overview

**Background.** Researchers commonly use *message perceptions* (persuasive potential) or *effects perceptions* (behavioral impact) in formative research to select tobacco risk messages. We sought to identify whether message perceptions or effects perceptions are more useful as proxies for the behavioral impact of tobacco risk messages.

**Methods.** In a three-week trial, 703 U.S. adult smokers (ages ≥ 21) were randomly assigned to receive brief messages about toxic chemicals in cigarette smoke (chemical messages) or control messages about properly disposing cigarette litter on their cigarette packs. The final follow-up survey assessed message perceptions, effects perceptions, quit intentions, and six behavioral outcomes. We conducted multiple mediation analysis in a structural equation modeling framework to test the indirect effects of messages by way of message perceptions and effects perceptions.

**Results.** Message perceptions did not independently mediate the impact of chemical messages on any of the outcomes (7 *p*s > .01). In contrast, effects perceptions mediated the impact of chemical messages on avoiding the messages, seeking chemical information, intentions to quit smoking, butting out a cigarette, forgoing a cigarette, and making a quit attempt (6 *p*s < .001). No mediation was present for social interactions about the message (*p* = .72). The effect sizes for these mediated effects were small to medium.

**Discussion.** Effects perceptions, but not message perceptions, were a proxy for risk messages’ impact on quit intentions and six quitting and related behaviors. These findings point
to the diagnostic value of effects perceptions in formative research on tobacco risk messages.

*Keywords*: Effects perceptions, formative research, perceived message effectiveness, multiple mediation, structural equation modeling

**Message Perceptions and Effects Perceptions as Proxies for Impact on Behavior**

Health messages that aim to change behavior are often iteratively developed through multiple rounds of testing, beginning with a large of pool of candidate messages.\(^5,^{41,110}\) It is usually unfeasible and inefficient to evaluate many candidate messages based on changes in behavioral outcomes, such as vaccination or quitting smoking, that may actually occur over a period of weeks or months. Thus, interventionists have used audience ratings of candidate messages’ persuasive potential (*message perceptions*) or of their potential to change behavioral antecedents and behavior itself (*effects perceptions*) to identify the most promising messages for further testing in a behavioral trial. Perceived message effectiveness (PME) measures have traditionally assessed either message perceptions or effects perceptions, occasionally combining items for both constructs within a single scale.\(^{50,105}\)

A burgeoning literature has begun to recommend the use of effects perceptions over message perceptions when developing messages that seek to change behavior.\(^7,^{105}\) Effects perceptions are conceptually closer to behavior, and the corresponding items generally use behavioral and personal referents to enhance predictive validity with behavior.\(^{105}\) Moreover, effects perceptions are associated with determinants of message impact that are closer to behavior (e.g., quit intentions) while message perceptions are associated with only some of the earliest determinants (e.g., attention to the message).\(^4\) Finally, effects perceptions scales may exhibit less measurement error than message perceptions scales due to greater conceptual
While the aforementioned observations support the use of effects perceptions in formative research, many researchers have expressed caution about the diagnostic value of PME ratings in identifying messages that are actually effective often due to under-conceptualization or unsatisfactory validity. The small number of PME validation studies in behavioral contexts have focused mostly on PME correlations or associations with outcomes of interest such as intentions to quit cigarette smoking and seeking information about colonoscopy, occasionally in longitudinal data. A more informative way to evaluate the diagnostic value of PME ratings is to examine whether they mediate message impact on behavioral or other meaningful outcomes. These validation efforts require a study design with control messages that are expected to be weaker than candidate messages and decompose the impact of candidate messages as a function of PME thereby determining the extent to which PME functions as proxy for message impact. These mediation analyses would also provide information on the substantive importance of PME when conducted across a variety of behaviors and their antecedents.

The current study sought to examine the diagnostic value of PME ratings using brief messages about the chemicals in cigarette smoke designed to discourage smoking (chemical messages) and control messages designed to discourage littering of cigarette butts (littering messages) as a case study. Because message perceptions and effects perceptions conceptually overlap, the current study also sought to compare the diagnostic value of these constructs. The main hypothesis guiding our study was that effects perceptions are a better proxy for behavioral impact than message perceptions. Briefly viewed messages often do not activate central pathways to persuasion and instead elicit affective reactions that are informative to the viewer. Some of the stated dimensions of effects perceptions (e.g.,
perceived message impact on worry) overlap with these affective reactions whereas the stated dimensions of message perceptions (e.g., message credibility) do not. Thus, we expected effects perceptions to mediate message impact and message perceptions to not mediate message impact.

In examining message perceptions and effects perceptions as potential proxies for the behavioral impact of chemical messages, our study sheds light on the relationship between PME and behavior. Doing so is essential to transforming PME from a criterion variable that is useful for message development into a substantive variable potentially with a defined role in message processing. Given limited guidelines on the use of PME in formative research, our study also provides suggestions for the optimal use of PME in message development.

Methods

Participants and procedures

A randomized clinical trial (RCT) comparing the effects of chemical messages and littering messages recruited a convenience sample of 719 U.S. adult smokers from the San Francisco Bay Area. Participants were ages 21 or older (given the legal age of buying tobacco products in California) and current smokers (i.e., had smoked at least 100 cigarettes in their lifetime and now smoke every or some days). Smokers who were enrolled in an ongoing smoking cessation trial, smoked less than seven cigarettes weekly or only roll-your-own cigarettes, or were pregnant were ineligible to participate in the RCT.

Participants attended visits at the trial office and brought an eight-day supply of cigarettes to all but the last visit. The trial randomized smokers to receive chemical messages (intervention) or littering messages (control) in nearly equal numbers. Littering messages were attention-matched to the chemical messages and had similar word lengths, literacy requirements, and visual features (Figure 6.1). Participants in both arms completed surveys at each trial visit and
rotated through three unique messages, one per week. After a run-in week, study staff applied a
different message at the next three weekly visits on the side of participants’ cigarette packs while
the participants completed a survey. Smokers received up to $300 for their trial participation.
Data collection began in September 2016 and finished in March 2017. Full details on design and
protocol are available in the main trial paper.22

Measures

The baseline survey assessed participant demographics and smoking behavior. At the last
visit, the survey assessed message perceptions using an established scale41 and effects
perceptions using the UNC Perceived Message Effectiveness Scale.7 The message perceptions
had six items: "This message is worth remembering"; "This message grabbed my attention";
"This message is powerful"; "This message is informative"; "This message is meaningful"; and
"This message is convincing."41 The effects perceptions scale had three items: "This message
discourages me from wanting to smoking"; "This message makes smoking seem unpleasant to
me"; and "This message makes me concerned about the health effects of smoking."7 Both
measures used a five-point response scale ranging from "strongly disagree" (coded as 1) to
"strongly agree" (5).

As tobacco warnings can have a range of effects,22,108 the final visit survey assessed quit
intentions and six behavioral outcomes: number of conversations about the messages in the past
week, avoidance of cigarette pack messages, seeking information about the chemicals in
cigarette smoke, number of times butting out a cigarette in the past week, number of times
forgoing a cigarette in the past week, and weekly recall of quit attempts. Social interactions
about the messages,96 avoidance of the messages,63 and seeking information about chemicals83
are behaviors that are potentially productive for quitting. Quitting initiation exists on a
spectrum, and forgoing or butting out a cigarette represent less intense micro-quitting behaviors than quit attempts. Each of the seven outcomes either appears on the UNC Tobacco Warnings Model (e.g., social interactions), which succinctly describes how tobacco warnings impact behavior, or has some correspondence with a behavioral antecedent in the model (e.g., avoidance with negative affect), further underscoring their importance. The survey assessed avoidance ($\alpha = 0.89$) and quit intentions ($\alpha = 0.96$) using separate three-item scales with desirable psychometric properties. The other five outcomes were assessed using single-item measures (Table 6.1).

**Data analysis**

Analyses used R (ver. 3.5.1) and the add-on packages lavaan (ver. 0.6-2) for estimating all path models, psych (ver. 1.8.4) for calculating ordinal reliability coefficients, nnet (ver. 7.3-12) for estimating multinomial logistic regression models for randomization checks, and ggplot2 (ver. 3.0.0) for creating figures.

We conducted multiple mediation analysis in a structural equation modeling (SEM) framework, estimating SEMs using mean- and variance-adjusted weighted least squares (WLSMV). Doing so allowed us to use pairwise deletion to handle missingness while treating all continuous variables as ordinal. Single-item outcomes did not have any missingness, and partial missingness on multi-item outcomes was negligible (< 3%). However, we dropped 16 cases with complete missingness on both mediators, message perceptions and effects perceptions, yielding an analytic sample of 703 smokers.

The measurement model specified orthogonal factors for message perceptions and effects perceptions (Figure 6.2). The measurement model also accounted for local dependence between four message perceptions items through pairwise correlated errors and likewise on one pair of
effects perceptions items. Baig et al.\textsuperscript{7} found the corresponding confirmatory factor analytic model to adequately capture the bidimensionality of message perceptions and effects perceptions. We specified a single-item outcome as a manifest variable and a multi-item outcome as a latent variable.

For a given outcome (e.g., quit intentions), the structural model simultaneously regressed the outcome onto message perceptions ($b$) and effects perceptions ($e$; Figure 6.2). The structural model also regressed message perceptions and effects perceptions onto trial condition ($a$ and $d$) as well as regressing effects perceptions onto message perceptions ($c$). Following standard procedures,\textsuperscript{68} we calculated three indirect effects of chemical messages on the outcome as the products of relevant path coefficients: through message perceptions alone ($a \times b$), through message perceptions and effect perceptions sequentially ($a \times c \times e$), and through effects perceptions alone ($d \times e$). We used model-based $z$-tests to examine the statistical significance of the indirect effects. We report path coefficients and indirect effects that were partially standardized based on the variances of endogenous manifest and latent variables and not the variance of trial condition, the exogenous binary covariate. In addition, we controlled for trial condition; the corresponding main effects are reported in the trial\textsuperscript{22} and other papers.\textsuperscript{63,65}

We confirmed SEM model fit using the Bayesian Information Criterion (BIC), Comparative Fit Index (CFI), and Root Mean Squared Error of Approximation (RMSEA). A negative BIC, large CFI ($\geq 0.95$), and small RMSEA ($\leq 0.06$) indicated adequate global fit. We also calculated the effect size $\nu$ for the three indirect effects.\textsuperscript{82} This is a novel effect size measure for mediation that represents the variance in the outcome jointly accounted for by the independent variable and mediator(s) after correcting "for spurious correlation induced by the ordering of the variables."\textsuperscript{82} Given its close relationship to $R^2_{\text{Med}}$, the usual verbal categories
facilitated interpretation: small (0.02), medium (0.15), and large (0.25). We used this approach to multiple mediation analysis for all seven outcomes. The lack of appropriate temporal ordering between the mediators and outcomes due to their assessment at a single time point meant that mediated effects could be artificially inflated. Thus, we used a critical alpha of .01 (or a 99% confidence interval) to provide for conservative tests of mediation.

**Results**

More than a third of participants in the control and intervention arms were white (≥ 37%) or African-American (≥ 34%). A minority of participants in both arms had at least a college degree (≥ 39%). Trial condition was not associated with standard demographics (Table 6.1) confirming that randomization succeeded in creating balanced conditions (26 tests p ≥ .06). All SEM models had adequate fit as indicated by the negative BICs (range -204 to -118), small RMSEAs (.05 to.06), and large CFI (.99 to 1.00).

**Mediation by message perceptions**

Chemical messages did not affect any of the outcomes through the potential mediator message perceptions (a \* b in Figure 6.2). With respect to the predictor-mediator pathway (a), chemical messages did not elicit message perceptions different from littering messages (p = .08). With respect to the mediator-outcome pathways (b), more positive message perceptions were associated with weaker avoidance (β = -0.28, p = .006) and being less likely to engage in a quit attempt (β = -0.38, p = .003). Message perceptions were not associated with conversations about the messages (p = .36), seeking information about chemicals (p = .55), quit intentions (p = .01), butting out a cigarette (p = .91), or forgoing a cigarette (p = .21). As a result, none of the indirect effects of chemical messages through message perceptions on these seven outcomes were statistically significant (7 ps ≥ .11; Figure 6.3). Furthermore, chemical messages explained
very small amounts of variance on average in the seven outcomes through message perceptions ($\nu$ median [range] = 0.002 [0.00, 0.003]).

**Sequential mediation by message perceptions through effects perceptions**

Chemical messages did not affect any of the outcomes sequentially through the potential mediators message perceptions and then effects perceptions ($a \times c \times e$ in Figure 6.1). As already reported for the (first) predictor-mediator pathway ($a$), chemical and littering messages did not differ on message perceptions. With respect to the mediator-mediator pathway ($c$), more positive message perceptions were associated with more positive effects perceptions ($\beta = 0.81, p < .001$). With respect to the final mediator-outcome pathways ($e$), more positive effects perceptions were associated with weaker avoidance ($\beta = 0.79, p < .001$), more frequent seeking information about chemicals ($\beta = 0.45, p < .001$), stronger quit intentions ($\beta = 0.91, p < .001$), more frequent butting out a cigarette ($\beta = 0.36, p < .001$), more frequent forgoing a cigarette ($\beta = 0.57, p < .001$), and being more likely to make a quit attempt ($\beta = 0.72, p < .001$). Effects perceptions were not associated with conversations about the messages ($p = .72$). Looking at the complete mediational pathway, none of the sequential indirect effects of chemical messages through message perceptions and effects perceptions on these seven outcomes were statistically significant ($7 ps \geq 0.09$; Figure 6.3). Additionally, chemical messages explained very small amounts of variance in the seven outcomes sequentially through message perceptions and effects perceptions ($\nu$ median [range] = 0.004 [0.00, 0.010]).

**Mediation by effects perceptions**

Chemical messages affected six of the seven outcomes through effects perceptions. With respect to the predictor-mediator pathway ($d$), chemical messages elicited more positive effects perceptions than littering messages ($\beta = 0.26, p < .001$). As reported previously for the mediator-
outcome pathways (e), effects perceptions were positively associated with all outcomes except conversations about the messages. As a result, effects perceptions mediated the impact of chemical messages on avoidance ($\beta = 0.20, p = .001$), seeking information about chemicals ($\beta = 0.11, p = .008$), quit intentions ($\beta = 0.23, p = .002$), butting out a cigarette ($\beta = 0.09, p = .004$), forgoing a cigarette ($\beta = 0.14, p = .002$), and quit attempts ($\beta = 0.18, p = .003$; Figure 6.3).

Effects perceptions did not mediate the impact of chemical messages on conversations about the messages ($p = .72$). Chemical messages explained small to medium amounts of variance in the seven outcomes through effects perceptions ($\nu$ median [range] = 0.02 [0.00, 0.053]).

**Discussion**

In a diverse sample of adult smokers, message perceptions did not independently mediate the impact of chemical messages on three behaviors that may play a role in quitting initiation, a motivator of quitting, and three quitting behaviors that varied in intensity. Message perceptions also did not sequentially mediate chemical message impact on these seven outcomes through effects perceptions. On the other hand, excluding conversations about the messages, effects perceptions mediated the impact of chemical messages on the six remaining outcomes.

Therefore, effects perceptions exclusively functioned as a proxy for chemical message impact on six outcomes that included various informative and quitting behaviors, supporting our behavioral proxy hypothesis. Formative research should prioritize effects perceptions for identifying promising messages for behavior change.

An open question is whether effects perceptions are a true mediator of message impact. In this regard, effects perceptions may actually represent an orientation among audience members to productively engaging with a message in ways that lead to behavior change. This orientation may come about via initial affective reactions to messages that inform the viewers’ receptivity to
the messages. Such an account corresponds to regulatory fit theory in which "feeling right" about a message or message receptivity may lead to sustained message engagement that is productive in a given context. The interpretation of effects perceptions as representing an orientation to productive message engagement corresponds to our mediational findings for effects perceptions. In line with these mediational findings, our cross-sectional validation study found that effects perceptions were associated with later behavioral antecedents (e.g., cognitive elaboration) from the TWM. Future studies should formally examine the relationship between regulatory fit and effects perceptions as a first step to understanding the hypothesized orientation.

Unlike effects perceptions, message perceptions did not mediate the impact of chemical messages on any of the outcomes with the corresponding effect sizes being very small. Thus, message perceptions were minimally diagnostic of a message’s potential for behavior change. A possible explanation for the lack of mediation is that chemical and littering messages did not sufficiently vary on message perceptions. Indeed, multiple rounds of message testing systematically excluded weaker messages yielding highly potent trial messages. However, chemical messages did elicit more positive effects perceptions than littering messages even though both types of messages were attention-matched. As such, an alternative explanation for the lack of mediation by message perceptions is that they are not sensitive to messages with similar design features (e.g., word length). Future studies should attempt to replicate our findings using a diverse pool of candidate messages. Doing so will shed further light on whether there is any value to assessing message perceptions in formative research.

Our study provides strong support for diagnostic value of effects perceptions in message testing. Chemical messages (i.e., the intervention) and effects perceptions (i.e., the mediator) jointly accounted for meaningful amounts of variance in avoidance of the messages, chemical
information seeking, quit intentions, butting out a cigarette, forgoing a cigarette, and quit attempts. Given that the corresponding effect sizes were small to medium, a more cautious view is that there are considerable amounts of variance in these outcomes that remain unexplained. As such, the extent to which effects perceptions capture a message’s potential to change behavior also remains uncertain. Such a view ignores the stated purpose of using effects perceptions (or PME in general), which is to identify potent messages for further testing in a behavioral trial or similar study. In other words, message testing with effects perceptions is not a substitute for studying behavioral impact, but a means to making the latter more efficient especially when dealing with large pools of candidate messages. Our findings generally support the use of effects perceptions in this capacity.

Strengths of our study include a diverse sample of smokers, randomization to trial condition, and structural equation models that uniquely estimated mediated effects. Focusing on a variety of outcomes provided ample opportunities to replicate our basic mediational findings. A limitation of our study is the assessment of mediators and outcomes at the last trial visit, which made it more difficult to confirm the temporal ordering of effects. However, similar mediational patterns across six outcomes generally rule out the possibility of reverse causation. The assessment of mediators and outcomes at a single time point may have also inflated their covariances contributing to potentially spurious indirect effects. Using a more conservative critical alpha of .01 (or 99% confidence intervals) and effect sizes to contextualize indirect effects may have mitigated this possibility. Future studies should replicate our findings using a longitudinal design with explicit temporal ordering of mediators and outcomes. Our study also did not sufficiently capture the heterogeneity in message testing scenarios, whether from candidate messages or PME instruments. Future studies should also replicate our findings in
other behavioral contexts using different message perceptions and effects perceptions scales.

**Conclusion**

Effects perceptions mediated the impact of chemical messages on behavior, pointing to their diagnostic value in formative research on health messages. Effects perceptions may represent an orientation among message recipients to further message engagement that is productive for behavior change. Thus, formative assessments of effects perceptions are practically useful for message development and may contribute to our understanding of message processing. The value of message perceptions above and beyond effects perceptions remains to be established.
Figure 6.1. Messages placed on smokers’ cigarette packs in the intervention (A) and control (B) arms.
Table 6.1

Measures for all outcomes in multiple mediation analyses with ordinal reliability coefficients for multi-item scales.

<table>
<thead>
<tr>
<th>Construct, measure(s)</th>
<th>Response</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of conversations about the message in the past week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last week, how many times did you talk to other people about the label on your cigarette packs?</td>
<td>0–100 times</td>
<td>–</td>
</tr>
<tr>
<td><strong>Avoidance of cigarette pack messages (α = 0.89)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often did you try to avoid the labels on your cigarette packs?</td>
<td>&quot;Never&quot; (1)</td>
<td>72</td>
</tr>
<tr>
<td>How often did you try to avoid the labels on your cigarette packs?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often did you put your cigarettes away because you didn’t want others to see the labels on your packs?</td>
<td>&quot;All of the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Seeking information about the chemicals in cigarette smoke</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last 3 weeks, how many times have you looked for information about the chemicals in cigarettes or cigarette smoke?</td>
<td>&quot;0 times&quot; (1)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Quit intentions (α = 0.96)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How interested are you in quitting smoking in the next month?</td>
<td>&quot;Not at all [item stem]&quot; (1)</td>
<td>80</td>
</tr>
<tr>
<td>How much do you plan to quit smoking in the next month?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How likely are you to quit smoking in the next month?</td>
<td>&quot;Very [item stem]&quot; (4)</td>
<td></td>
</tr>
<tr>
<td><strong>Number of times butting out a cigarette in the past week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last week, how often have you butted out a cigarette before you finished it because you wanted to smoke less?</td>
<td>&quot;Never&quot; (1)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number of times forgoing a cigarette in the past week</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last week, how often have you stopped yourself from having a cigarette because you wanted to smoke less?</td>
<td>&quot;Never&quot; (1)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Weekly recall of quit attempts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last week, did you stop smoking for 1 day or longer because you were trying to quit smoking?</td>
<td>&quot;Yes&quot; (1)</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>&quot;No&quot; (0)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.2

**Participant characteristics**

<table>
<thead>
<tr>
<th>Age, years</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>21–29</td>
<td>22.4</td>
<td>22.9</td>
</tr>
<tr>
<td>30–39</td>
<td>21.2</td>
<td>22.9</td>
</tr>
<tr>
<td>40–49</td>
<td>21.2</td>
<td>16.6</td>
</tr>
<tr>
<td>50–59</td>
<td>25.8</td>
<td>25.4</td>
</tr>
<tr>
<td>60+</td>
<td>9.4</td>
<td>12.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>54.1</td>
<td>47.4</td>
</tr>
<tr>
<td>Female</td>
<td>42.2</td>
<td>47.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>transgender (includes other gender identity)</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.7</td>
<td>5.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gay, lesbian or bisexual</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25.8</td>
<td>24.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>37.1</td>
<td>37.7</td>
</tr>
<tr>
<td>Black or African-American</td>
<td>36.8</td>
<td>34.3</td>
</tr>
<tr>
<td>Asian</td>
<td>8.5</td>
<td>8.3</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>4.8</td>
<td>4.9</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>3.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Other</td>
<td>9.4</td>
<td>11.7</td>
</tr>
<tr>
<td>Hispanic</td>
<td>11.9</td>
<td>15.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>High school graduate or less</td>
<td>24.6</td>
<td>18.9</td>
</tr>
<tr>
<td>Some college</td>
<td>34.0</td>
<td>41.7</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>34.3</td>
<td>32.0</td>
</tr>
<tr>
<td>Graduate degree</td>
<td>7.1</td>
<td>7.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Household income, annual</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0–$24,999</td>
<td>46.7</td>
<td>43.7</td>
</tr>
<tr>
<td>$25,000–$49,999</td>
<td>24.9</td>
<td>23.1</td>
</tr>
<tr>
<td>$50,000–$74,999</td>
<td>9.9</td>
<td>14.9</td>
</tr>
<tr>
<td>$75,000+</td>
<td>18.4</td>
<td>18.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low income, &lt; 200% of federal poverty level</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>59.8</td>
<td>55.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smoking frequency</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily</td>
<td>21.8</td>
<td>24.0</td>
</tr>
<tr>
<td>Non-daily</td>
<td>78.2</td>
<td>76.0</td>
</tr>
</tbody>
</table>
Figure 6.2. Path diagram for structural equation model with mediated effects of chemical messages on a given outcome through message perceptions alone ($a \times b$), message perceptions and effects perceptions sequentially ($a \times c \times e$), and effects perceptions alone ($d \times e$). Measurement model and additional paths included for completeness appear in gray.
Figure 6.3. Standardized indirect effects of chemical messages through message perceptions and effects perceptions with 99% confidence intervals.
CHAPTER 7: FUTURE DIRECTIONS FOR OPTIMIZING USE OF PERCEIVED MESSAGE EFFECTIVENESS IN FORMATIVE RESEARCH

The findings in this dissertation empirically establish a clear distinction between message perceptions and effects perceptions, two common perceived message effectiveness (PME) constructs, using brief messages about chemicals in cigarette smoke as a case study. The practical consequence of this distinction is that both constructs cannot be used interchangeably in formative research on message development. Formative research should prioritize effects perceptions when evaluating messages that are designed to change behavior. Aim 1 demonstrated that a novel effects perceptions scale reliably measured effects perceptions and had lower measurement error than a comparable message perceptions scale. Aim 2 found that effects perceptions were a better predictor of behavior than message perceptions. In Aim 3, effects perceptions behaved as a proxy for quitting and other related behaviors. These findings lead to additional insights concerning the role of message perceptions in formative research, the stability of both message perceptions and effects perceptions across diverse behavioral contexts, and the interpretability of PME scores. I discuss these insights in this chapter and end by touching on some strengths and limitations of this dissertation.

Message perceptions as a moderator

Message perceptions were not directly relevant to behavior change. In Aim 2, more positive message perceptions were associated with more attention to and greater recognition of the messages but were not associated with later behavioral antecedents or behavior. In Aim 3, message perceptions did not mediate the behavioral impact of messages. These findings indicate
that message perceptions are only predictive of initial message processing and are not a proxy for behavioral impact.

However, "adequate" message perceptions could be a pre-condition for messages to affect behavior. In other words, the findings of my dissertation do not rule out the possibility of messages having a greater impact on effects perceptions among participants with positive message perceptions. Empirical support for this moderation hypothesis may suggest that briefly viewed messages benefit from having persuasive characteristics even if these characteristics are not necessarily behavioral antecedents. This may be because such messages still need to draw enough attention that leads to further processing captured by effects perceptions, and persuasive characteristics likely facilitate attention.79

Empirical support for moderation would optimistically point to a two-stage message testing process in which messages are first evaluated on message perceptions. A smaller number of candidate messages that score adequately on message perceptions are then evaluated on effects perceptions. This two-stage process would facilitate the selection of messages that are optimized for both attention and behavior change. A more skeptical view of empirical support for moderation would be that assessing message perceptions in addition to effects perceptions provides little information that enhances message selection beyond assessing effects perceptions only, given their high correlation. In Aim 2, effects perceptions were strongly associated with most quitting behaviors and related outcomes to begin with. Therefore, moderation by message perceptions may also not translate into meaningfully larger indirect effects through effects perceptions. Future studies should examine a broader variety of messages (both low and high quality, whereas this dissertation examined only relatively high-quality messages). Such studies could productively examine message perceptions as a moderator to determine whether also
assessing message perceptions would augment the diagnostic value of effects perceptions in formative research.

**Effects perceptions and health promoting behaviors**

This dissertation examined effects perceptions in the context of a health risk behavior (i.e., smoking), but effects perceptions may shift in meaning in the context of health promoting behaviors (e.g., exercise). First, a reframed version of the effects perceptions scale would shift the action from stopping or reducing to starting or increasing the referenced behavior. Thus, "This message discourages me from wanting to smoke," would become "This message encourages me to want to exercise." Second, the scale may also need to assess different effects perceptions. The scale currently assesses effects perceptions that were informed by research on risk perceptions,

57 behavioral motivation,

131 and outcome expectancies as relevant to smoking. However, these effects perceptions may not map onto known determinants of a health promoting behavior such as exercise. Even if shifting the action does not lead to changes in the underlying construct measured by the effects perceptions scale, assessing different effects perceptions is very likely to do so. Any changes in the underlying construct would lead to changes in the psychometric properties of the original scale. A next step to examining the stability of effects perceptions in different behavioral contexts is to develop an effects perceptions scales for health promoting behaviors.

A similar concern about changes in the underlying construct most likely does not apply to the message perceptions scale because message perceptions are judgments about the messages themselves. Thus, message perceptions items do not reference behavior and or an action related to it (i.e., discouraging or encouraging a behavior). In addition, message perceptions items assess message characteristics that are universal across behavioral contexts (e.g., credibility). Therefore,
message perceptions may represent a more stable construct across different behavioral contexts. Although the generality of message perceptions is likely a disadvantage when evaluating messages that specifically seek to change behavior, it may mean that message perceptions are better suited to a variety of other message testing scenarios than effects perceptions. For example, message perceptions may be more suited to evaluating health communication materials that seek to change knowledge and awareness. Aim 2 did not examine the predictive validity of message perceptions with knowledge or awareness. Aim 3 also did not examine message perceptions as a proxy for knowledge or awareness. Subsequent research should examine the relationships between message perceptions and these outcomes.

**Effects perceptions item pool**

The interpretability of mean effects perceptions scores in formative research is a practical concern that the current studies did not address. One issue is the meaning of a score with respect to foretelling behavior change. For example, does a mean effects perceptions score of 3.2 presage behavior change? Another issue is the meaning of a score when compared against other scores and how large a difference is sufficient to support decisions about message selection. For example, if two messages have mean PME scores of 2.1 and 2.4, does the difference suggest that the higher scoring message is more promising in terms of behavior change? In practice, researchers have commonly interpreted PME scores for individual messages directly (i.e., a middling score of 3.2 on a PME scale points to a mediocre message), \(^{112}\) examined differences between mean PME scores, \(^{110}\) or compared mean PME scores to an ad hoc threshold.\(^ {45}\) However, these strategies may not facilitate message selection, especially when mean effects perceptions scores for candidate messages are right skewed or do not vary much (as in Aim 1).

A novel approach to this interpretability problem is to evaluate messages using an effects
perceptions item pool instead of a multi-item effects perceptions scale. The motivation for this strategy is twofold. First, in comparison to individual item scores or person response patterns, mean scale scores commonly lead to a loss of information about respondent standing on the construct of interest.\textsuperscript{141} Second, according to item response theory (IRT), items that assess a common construct can capture different levels of it. Thus, the effects perceptions item pool would include items that vary on the level of effects perceptions that they capture, with some items capturing very low levels of effects perceptions and others capturing progressively higher levels of effects perceptions. Messages with greater potential to change behavior would likely elicit higher scores on items that capture higher levels of effects perceptions. Thus, earlier stages of message development would focus on items that capture low levels of effects perceptions to exclude the poorest messages. Later stages would focus on items that capture progressively higher levels of effects perceptions to identify the most promising messages for field testing, an aspect of the current scale that is underdeveloped as seen in Aim 1. This strategy would likely provide more information about the candidate messages’ potential for changing behavior than focusing only on mean effects perceptions scores. This strategy would also likely overcome the need to interpret effects perceptions scores directly. Instead, researchers could retain a meaningful number of high-scoring messages from the previous stage, taking into account that the current stage will evaluate those messages on items that measure even higher levels of effects perceptions. Future studies should develop a large effects perceptions item pool as a first step to evaluating this strategy.

**Strengths and limitations**

A major strength of this dissertation is the use of a behavioral theory, the UNC Tobacco Warnings Model (TWM), to inform the selection of construct validators. Other validation studies
of PME constructs have mostly focused on criterion variables that are pertinent to specific applications. This dissertation also used four varied data sources, six chemical and three littering messages, and multiple measurement occasions that provided many opportunities for replicating basic findings within each aim. Some specific findings like the bidimensionality of message perceptions and effects perceptions or greater measurement error in the message perceptions scale were replicated across the aims in different ways. Finally, this dissertation relied on latent variable models that suited the research questions. For example, item factor analytic models in Aim 1 provided insight about the psychometric properties of the effects perceptions scale across multiple chemical messages. As another example, Aim 3 used structural equation models to estimate unique mediated effects for message perceptions and effects perceptions.

The major conceptual limitation of this dissertation is that it assumes that using self-report scales, whether assessing effects perceptions or some other PME construct, is sufficient to characterize messages’ potential for behavior change. This assumption may be too strong, as the parent randomized clinical trial (RCT) found that chemical messages that were promising on effects perceptions ultimately failed to change quit intentions, the main trial outcome, or quit attempts. However, chemical messages did increase forgoing and butting out a cigarette. This is despite the fact that effects perceptions demonstrated predictive validity with and functioned as a proxy for all four quitting-related outcomes in Aims 2 and 3, respectively. A simple explanation for this mismatch is that it confirms that evaluating messages on effects perceptions is not a substitute for studying behavioral impact, but a means to making the latter more efficient, as discussed in Aims 2 and 3. An alternative explanation for the mismatch is that the use of self-report scales to evaluate messages assumes that people have insight into their own behavior in the first place. This is problematic because a large body of research shows that automaticity and
unconscious processes may underlie a substantial portion of human behavior. Furthermore, people may perform a behavior first and infer their reasons for acting in that way when responding to items on a self-report scale. Thus, the use of self-report scales to evaluate messages still remains largely atheoretical even though there is reason to believe that effects perceptions capture an orientation towards sustained productive message engagement, as discussed in Aim 3. One way to establish a firmer theoretical basis for using effects perceptions beyond speculation is to triangulate effects perceptions with objective measures of behavioral antecedents from the TWM. Future studies should therefore examine the predictive validity of effects perceptions with attention to messages assessed via eye-tracking and biobehavioral measures of affective message reactions.

Another conceptual limitation is that the current studies included only highly developed messages suggesting that the findings may be specific to later message development stages. The three chemical messages in Aims 2 and 3 were selected from a larger set of nine, which were the most promising messages from a larger set of 76 chemical messages. Development of the littering messages also followed a similar process, but involved a much smaller pool of candidate messages. Highly developed messages are likely to have a greater impact on behavior and its antecedents than less developed messages and potentially rank higher on PME constructs in general. The immediate consequence of examining the predictive validity of message perceptions and effects perceptions in the context of highly developed messages is that the validity associations reported in Aim 2 may be higher than they would be in the context of more diverse messages. Furthermore, if the validity associations are specific to a highly developed subset of chemical messages that elicit high effects perceptions, the expectation that such messages change behavior should be lower than the validity associations actually imply. This makes the mismatch
between the null results of the RCT and the promising diagnostic value of effects perceptions reported in Aims 2 and 3 less surprising. Subsequent research should replicate the major findings of this dissertation in the context of more varied chemical and other messages to more accurately characterize the diagnostic value of PME constructs in general.

Conclusions

Effects perceptions are conceptually distinct from message perceptions and well suited to formatively evaluating tobacco risk or other messages that seek to change behavior. It may be possible to further improve the diagnostic value of effects perceptions through a more elaborate message testing process either involving item pools or message perceptions. Future research should explore these possibilities in addition to examining the predictive validity of effects perceptions in other behavioral contexts with more diverse messages.
APPENDIX: SURVEY MEASURES FOR PROPOSED ANALYSES IN AIMS 2 AND 3

<table>
<thead>
<tr>
<th>Construct, item(s), and response scale(s)</th>
<th>Visit 1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Message perceptions</strong></td>
<td>✓</td>
</tr>
<tr>
<td>The labels are worth remembering.</td>
<td></td>
</tr>
<tr>
<td>The labels grabbed my attention.</td>
<td></td>
</tr>
<tr>
<td>The labels are powerful.</td>
<td></td>
</tr>
<tr>
<td>The labels are informative.</td>
<td></td>
</tr>
<tr>
<td>The labels are meaningful.</td>
<td></td>
</tr>
<tr>
<td>The labels are convincing.</td>
<td></td>
</tr>
<tr>
<td>&quot;Strongly disagree&quot; (1) … &quot;Strongly agree&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Effects perceptions</strong></td>
<td>✓</td>
</tr>
<tr>
<td>The labels discourage me from wanting to smoke.</td>
<td></td>
</tr>
<tr>
<td>The labels make smoking seem unpleasant to me.</td>
<td></td>
</tr>
<tr>
<td>The labels make me concerned about the health effects of smoking.</td>
<td></td>
</tr>
<tr>
<td>&quot;Strongly disagree&quot; (1) … &quot;Strongly agree&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Perceived understandability</strong></td>
<td>✓</td>
</tr>
<tr>
<td>How easy were the labels to understand?</td>
<td></td>
</tr>
<tr>
<td>&quot;Not at all&quot; (1) … &quot;Extremely&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Attention and noticing</strong></td>
<td>✓</td>
</tr>
<tr>
<td>How much did the labels grab your attention?</td>
<td></td>
</tr>
<tr>
<td>&quot;Not at all&quot; (1) … &quot;Very much&quot; (5)</td>
<td></td>
</tr>
<tr>
<td>How often did you notice the labels?</td>
<td></td>
</tr>
<tr>
<td>How often did you read or look closely at the labels?</td>
<td></td>
</tr>
<tr>
<td>&quot;Never&quot; (1) … &quot;All the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Avoidance of cigarette pack labels</strong></td>
<td>✓</td>
</tr>
<tr>
<td>How often did you try to avoid the labels on your cigarette packs?</td>
<td></td>
</tr>
<tr>
<td>How often did you put your cigarettes away because you didn’t want others to see the labels on your packs?</td>
<td></td>
</tr>
<tr>
<td>How often did you remove the labels from your cigarette packs because you didn’t want to look at them?</td>
<td></td>
</tr>
<tr>
<td>How often did you transfer cigarettes to another container to avoid looking at the labels?</td>
<td></td>
</tr>
<tr>
<td>How often did you place a cover or case over your cigarette pack to avoid looking at the labels?</td>
<td></td>
</tr>
<tr>
<td>How often did you draw on the labels to deface them?</td>
<td></td>
</tr>
<tr>
<td>&quot;Never&quot; (1), &quot;Rarely&quot; (2), &quot;Sometimes&quot; (3), &quot;Often&quot; (4), or &quot;All of the time&quot; (5)</td>
<td></td>
</tr>
<tr>
<td><strong>Negative affect</strong></td>
<td>✓</td>
</tr>
<tr>
<td>How much did the labels on your cigarette packs make you feel anxious?</td>
<td></td>
</tr>
</tbody>
</table>
How much did the labels on your cigarette packs make you feel sad?
How much did the labels on your cigarette packs make you feel scared?
How much did the labels on your cigarette packs make you feel disgusted?
How much did the labels on your cigarette packs make you feel guilty?

"Not at all" (1), "A little" (2), "Somewhat" (3), "Very" (4), or "Extremely" (5)

Quit intentions

How interested are you in quitting smoking in the next month?
"Not at all interested" (1), "A little interested" (2), "Somewhat interested" (3), or "Very interested" (4)
How much do you plan to quit smoking in the next month?
"Not at all" (1), "A little" (2), "Somewhat" (3), or "Very much" (4)
How likely are you to quit smoking in the next month?
"Not at all likely" (1), "A little likely" (2), "Somewhat likely" (3), or "Very likely" (4)

Thinking about the chemicals in cigarettes

In the last week, how much did you think about the chemicals in the smoke from your cigarettes?
"Never" (1) … "All of the time" (5)

Thinking about the harms of smoking

In the last week, how much did you think about the harm your smoking might be doing to you?
"Never" (1) … "All of the time" (5)

Number of times butting out or forgoing a cigarette in the past week

In the last week, how often have you stopped yourself from having a cigarette because you wanted to smoke less?
In the last week, how often have you butted out a cigarette before you finished it because you wanted to smoke less?
"Never" (1), "1–2 times" (2), "3–4 times" (3), "5–9 times" (4), or "10 or more times" (5)

Number of conversations about the label in the past week

In the last week, how many times did you talk to other people about the label on your cigarette packs?
Open-ended number of times (restricted to 0–100)

Chemical information seeking

In the last 3 weeks, how many times have you looked for information about the chemicals in cigarettes or cigarette smoke?
"0 times" (1), "1–2 times" (2), "3–5 times" (3), or "6 or more times" (4)

Awareness of health effects in intervention messages

Are you aware that smoking cigarettes causes lung tumors?
Are you aware that smoking cigarettes causes lip cancer?
Are you aware that smoking cigarettes causes strokes?
Are you aware that smoking cigarettes causes throat cancer?
Are you aware that smoking cigarettes causes kidney damage?
Are you aware that smoking cigarettes causes heart damage?
Awareness of chemicals in intervention messages

Are you aware that cigarette smoke contains arsenic? ✓
Are you aware that cigarette smoke contains ammonia?
Are you aware that cigarette smoke contains formaldehyde?
Are you aware that cigarette smoke contains lead?
Are you aware that cigarette smoke contains uranium?
Are you aware that cigarette smoke contains benzene?

Recognition

Select the text from the label we put on the side of your cigarette packs at your last visit?

"Cigarette smoke contains uranium. This causes lung tumors and kidney damage."
"Cigarette smoke contains arsenic. This causes heart damage."
"Cigarette smoke contains formaldehyde. This causes throat cancer."
"Cigarette smoke contains lead. This causes cancer and brain disorders."
"Please refrain from littering. Cigarette butts are the most littered item."
"Cigarette butts don’t biodegrade. Please do not litter."
"Cigarette litter requires cleanup. Discard cigarette butts properly."
"Cigarette butts harm the environment."

Negative pack attitudes

How much do your cigarette packs with the labels make you feel embarrassed?
How much do your cigarette packs with the labels make you feel unaccepted?
How much do your cigarette packs with the labels make you feel ashamed?

"Not at all" (1) … "Extremely" (5)
REFERENCES


