American Psychological Association
Journal of Experimental Psychology: Applied

1076-898X/19/$12.00
2019, Vol. 25, No. 1, 41–61
http://dx.doi.org/10.1037/xap0000203

Contextualizing Numeric Clinical Test Results for Gist Comprehension: Implications for EHR Patient Portals

Daniel Morrow and Renato Ferreira Leitão Azevedo
University of Illinois at Urbana-Champaign

Rocio Garcia-Retamero
University of Granada

Mark Hasegawa-Johnson and Thomas Huang
University of Illinois at Urbana-Champaign

William Schuh
Carle Foundation Hospital, Urbana, Illinois

Kuangxiao Gu and Yang Zhang
University of Illinois at Urbana-Champaign

Patient portals to Electronic Health Record (EHR) systems are underused by older adults because of limited system usability and usefulness, including difficulty understanding numeric information. We investigated whether enhanced context for portal messages about test results improved responses to these messages, comparing verbally, graphically, and video-enhanced formats. Older adults viewed scenarios with fictitious patient profiles and messages describing results for these patients from cholesterol or diabetes screening tests indicating lower, borderline, or higher risk levels. These messages were conveyed by standard format (table of numerical test scores) or one of the enhanced formats. Verbatim and gist memory for test results, risk perception, affective response, attitude toward and intention to perform self-care behaviors, and satisfaction were measured. Verbally and video enhanced context improved older adults' gist but not verbatim memory compared to the standard format, suggesting we were successful in designing messages that highlight gist-based information. Little evidence was found for benefits related to the graphically enhanced format. Although verbally and video enhanced formats improved gist memory and message satisfaction, they had less impact on the other responses to the messages. However, these responses reflected level of risk: As risk associated with test results increased, positive affect decreased whereas negative affect, perceived risk, behavioral attitudes, and intentions increased, as predicted by behavioral change theories.

Public Significance Statement

The goal of this project is to make patient portals to Electronic Health Records more useful for older adults. Our findings show that older adults better remember numeric health information such as clinical test results when this information is presented in a format that provides context for interpreting the gist of the information for risk (lower, borderline, higher), compared to typical formats used in patient portals.

Keywords: risk communication, fuzzy trace theory, patient portal, electronic health record, aging

Self-managing illness requires understanding and using a wide range of health-relevant numeric information. For example, to understand their cholesterol test results, people need to interpret the implications of these numbers (total, triglycerides, HDL, LDL)

Editor's Note. Sean M. Lane served as Action Editor for this article.

This article was published Online First January 28, 2019.

Daniel Morrow and Renato Ferreira Leitão Azevedo, Department of Educational Psychology, University of Illinois at Urbana-Champaign; Rocio Garcia-Retamero, Department of Experimental Psychology, University of Granada; Mark Hasegawa-Johnson and Thomas Huang, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign; William Schuh, Carle Foundation Hospital, Urbana, Illinois; Kuangxiao Gu and Yang Zhang, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign.

This work was supported by the Agency for Health care Research and Quality under grant R21HS022948. The current research was also funded by the Ministerio de Economía, Industria y Competitividad (Spain; PSI2014-51842-R). We thank Bidisha Roy, Caleigh Silver, Cheryl Cheong, and Xue Yang for help with data collection. We are grateful to Charles Lansford for serving as the physician in the video recordings, and for his expert advice about the construction of the scenarios.

Correspondence concerning this article should be addressed to Daniel Morrow, Department of Educational Psychology, University of Illinois at Urbana-Champaign, Education Building, 1310 South 6th Street, Champaign, IL 61820. E-mail: dgm@illinois.edu

in terms of risk of cardiovascular disease, and decide how to respond to perceived health threats by engaging in self-care behaviors (e.g., taking prescribed medications, changing diet, exercising).

Understanding health-related numeric information can be challenging for patients. For cholesterol results, people need to understand the level of risk indicated by each score and integrate this information to understand overall level of risk. Moreover, higher LDL, total cholesterol, or triglycerides numbers indicate more risk, whereas higher HDL numbers indicate less risk, so people need to understand trade-offs between these numbers.

Older adults can be especially challenged by health-relevant numeric information (Peters et al., 2009). They often need to understand such information because they are more likely than younger adults to have chronic illness. Yet they tend to experience declines in the cognitive and literacy resources needed to understand and make decisions about this information (Reyna, 2011). Declines in numeracy, or the ability to make sense of numbers (Peters, 2012), may hamper mapping test scores to risk categories and integrating risk information across scores to derive a global understanding of risk and what to do about it.

Traditionally, providers help patients understand test results. During face-to-face interactions, they discuss what the numbers mean and help patients plan how to address the threats indicated by the results. However, patient/provider collaboration is often eroded by a health care system that reduces time for in-depth discussion and increases the number of specialists providing care to the patient, which complicates care coordination as well as provider/patient communication.

Technology such as patient portals to Electronic Health Record (EHR) systems has the potential to support patient/provider collaboration despite limited contact time. Portals provide patients ready access to their health information that may support understanding and planning (Institute of Medicine [IOM], 2012; Gibbons & Casale, 2010; Stead & Lin, 2009). Indeed, meaningful use EHR systems are supposed to promote patient-centered care and support patient/provider collaboration essential to this care (IOM, 2012). Unfortunately, portals currently function more as patient information repositories than collaborative tools. It is not surprising then that patients, especially older adults, underutilize portals. For example, in one study, 82% of adults 65–69 years old registered to use a portal, 83% of these logged on at least once, and 69% viewed at least one lab result, whereas 72% of adults 75–79 years old registered, 76% logged on at least once, and 56% viewed at least one lab result (Gordon & Hornbrook, 2016; also see Neuner, Fedders, Caravella, Bradford, & Schapira, 2015). This underutilization reflects limited system usability and usefulness more than access to the technology itself (Czaja, Zarcadoolas, Vaughon, Lee, Rockoff, & Levy, 2015; Haga et al., 2014; Lyles, Schillinger, & Sarkar, 2015; Taha, Czaja, Sharit, & Morrow, 2013; Taha, Sharit, & Czaja, 2014). Most relevant to the current study, difficulty understanding numeric information presented on portals contributes to underutilization (Haga et al., 2014).

In the present article, we investigate how to support older adult comprehension of and response to patient portal-based numeric information. This goal requires explaining how patients understand numeric health information in order to perceive and plan to mitigate risk, the impact of aging on the cognitive resources needed for these processes, and how to help older adults understand and make decisions about numeric health information in order to manage their illness.

Understanding and Acting on Numeric Health Information

The framework in Figure 1 integrates health literacy, fuzzy trace memory, text comprehension, and health behavior theories to

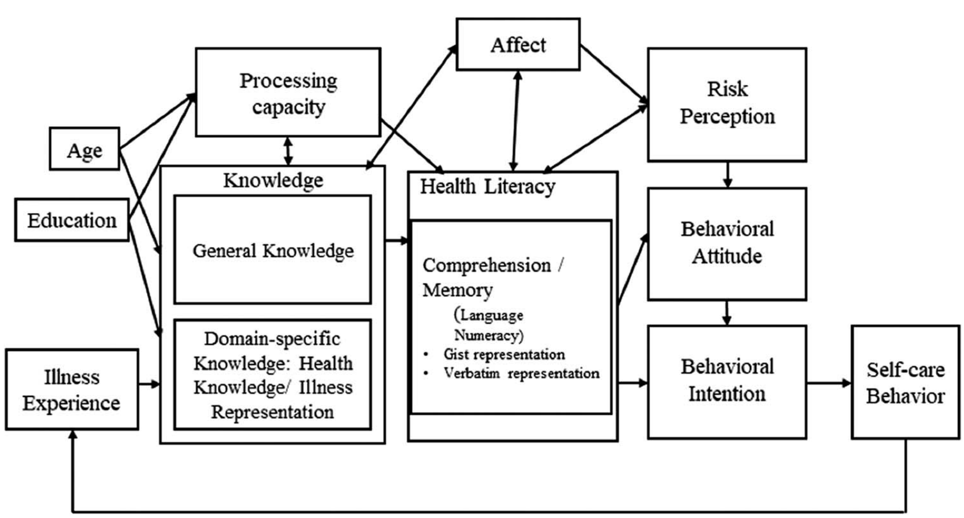

Figure 1. Framework guiding the design and evaluation of portal messages From "A multidisciplinary approach to designing and evaluating Electronic Medical Record portal messages that support patient self-care," by D. Morrow, M. Hasegawa-Johnson, T. Huang, W. Schuh, R. F. L. Azevedo, K. Gu, . . . R. Garcia-Retamero, 2017, *Journal of Biomedical Informatics, 69,* pp. 63–74. Copyright 2017 by the Elsevier Ltd. Adapted with permission. Arrows between boxes indicate possible associative as well as causal relationships between concepts.

identify processes involved in understanding and responding to health information, and the cognitive abilities/resources that influence these processes (for more detail, see Morrow et al., 2017). The framework is similar to the Integrated Behavior Model proposed by Montaño and Kasprzyk (2008) because it links patient abilities to factors that directly influence their health behaviors. In addition, our framework focuses on processes involved in understanding and remembering health information. According to fuzzy trace theory (Reyna, 2008; Reyna & Brainerd, 1995), people process and remember information at multiple levels. The verbatim level captures the literal facts or "surface form," preserving information about precise numeric values (or exact linguistic representations in the case of language comprehension). In addition, making sense of numbers to support action requires understanding the information in terms of goals and knowledge to create gist representations that capture bottom-line implications of the numbers for health and that are organized around qualitative, often affective and evaluative, dimensions. Therefore, gist often reflects integral affect, which is directly relevant to the evaluation of the health information, rather than incidental affect (Peters et al., 2009). For instance, gist representations of cholesterol test results may capture ordinal risk values (e.g., lower/borderline/higher) for heart disease that are associated with evaluative responses (e.g., better/worse). Because gist representations are relatively simple and rooted in experience, they may be easier to create than verbatim representations.

People generally prefer to operate on the least precise memory representation to accomplish a task. Therefore, gist that captures categorical or ordinal relationships, or gist combined with verbatim representations, may be most effective for understanding implications of test results for risk, depending on task requirements (Peters et al., 2009; Reyna, 2008). Gist representations may be especially important for numeric comprehension by patients with limited numeracy.

The process-knowledge model of health literacy (Chin et al., 2011; Morrow & Chin, 2015) draws on text comprehension theories to identify processes involved in understanding linguistic health information. These include recognizing words and integrating the concepts associated with these words into propositions, or idea units. Understanding numeric information requires similar processes (Peters, 2012; Reyna, Nelson, Han, & Dieckmann, 2009). For example, understanding cholesterol scores involves encoding exact numeric values (verbatim representation) and mapping them to categorical or ordinal risk categories (gist representation). The gist representations for individual scores may be integrated into an overall or global gist representation of the risk-related implications of the message. Information from the scores must also be integrated with health knowledge (Morrow & Chin, 2015).

As seen in the left part of Figure 1, comprehension processes depend on cognitive and other resources, which in turn depend on broader characteristics such as age, education, and illness experience. Comprehension requires processing capacity (e.g., working memory) and knowledge (of language and health-related concepts), as well as noncognitive resources such as affect, which are important for responding to the evaluative dimensions of gist representations (Morrow & Chin, 2015; Reyna, 2008).

Aging influences comprehension processes because of age-related changes in the cognitive and affective resources needed for comprehension (Finucane et al., 2002). Processing capacity tends to decline with age, and these age-related changes may be exacerbated by illness experience. For example, uncontrolled hypertension can impair processing capacity (Van den Berg, Kloppenborg, Kessels, Kappelle, & Biessels, 2009). On the other hand, general knowledge (e.g., about language or verbal ability) and domain-specific knowledge (e.g., about health) tend to grow with age-related experience (e.g., Beier & Ackerman, 2005; Hartshorne & Germine, 2015). Age-related resource limitations and gains interact to influence health literacy (Chin et al., 2011) and memory for health information (Chin et al., 2017), such that high levels of knowledge can offset processing capacity limits. Aging is also accompanied by increasing focus on affect and emotion that influences comprehension of and decisions about health information (Charles & Carstensen, 2010; Morrow & Chin, 2015). For example, older adults are more likely than younger adults to remember emotionally charged information, especially when emotions are positive.

Although the left side of the framework in Figure 1 identifies abilities that influence comprehension and memory at verbatim and gist levels, the right side shows that perception of the risk associated with test results or other presented health information is influenced by comprehension and memory for this information, suggesting that risk perception also depends on the cognitive and affective resources that influence gist. In addition, perception of the risk associated with test results is shaped by factors such as beliefs about illness (e.g., How susceptible and vulnerable to the illness am I?; Brewer et al., 2007; Meyer, Leventhal, & Gutmann, 1985).

The right side of Figure 1 also identifies other possible responses to health information that influence self-care behaviors. According to theories of behavior change (e.g., the theory of reasoned action/planned behavior, Ajzen & Fishbein, 1977; Ajzen & Madden, 1986; Montaño & Kasprzyk, 2008), risk perception shapes attitudes toward actions that may mitigate perceived risk (see also risk-as-feelings hypothesis; Loewenstein, Weber, Hsee, & Welch, 2001). Risk perception has been linked to getting vaccinations in part because it influences behavioral attitudes (Brewer et al., 2007; Crano & Prislin, 2008). It has also been linked to behavioral intentions. For example, perception of risk associated with out-of-range test results is linked with intention to call a physician (Zikmund-Fisher, Exe, & Witteman, 2014).

Attitudes are also influenced by factors such as beliefs about whether the actions are likely to influence illness (Brewer et al., 2007; Webb & Sheeran, 2006). Patients may understand that their cholesterol scores indicate high risk but do not think exercise or diet can reduce this risk. Finally, behavioral attitudes predict intentions to act, which predict performance of the behaviors (Crano & Prislin, 2008; Webb & Sheeran, 2006).

In short, the framework suggests comprehension of and memory for health information influence health decisions and behaviors, which underlines the importance of presenting this information in patient portals so as to support comprehension, memory, and action.

Improving Gist Memory and Supporting Action

We argue that comprehension of and memory for risk-related information often influences risk perception and self-care behavior. Moreover, the literature on risk perception suggests that people better extract gist that supports action when numbers are presented in a context that supports comprehension (e.g., mapping numbers to risk categories that guide appropriate affective/evaluative responses (Peters et al., 2009; Reyna, 2008; Slovic & Peters, 2006). Such contexts may also benefit older adults by reducing comprehension demands on processing capacity and by leveraging prior knowledge. Older adults may rely on gist because their verbatim memories are less robust (Castel, 2005; Reyna, 2008; Tanius, Wood, Hanoch, & Rice, 2009), and because they focus on affect in decision making (Mikels et al., 2010). Accurate gist may support older adults' risk perception, their attitude toward behavior that addresses this risk, and intention to act.

Traditionally, providers help patients understand test results. They use verbal (e.g., words that emphasize key concepts) and nonverbal cues (tone of voice, facial expressions) to indicate what information is most important, and to support patients' affective and evaluative reactions that scaffold gist comprehension (Henneman, Marteau, & Timmermans, 2008; Reyna et al., 2009) and increase their satisfaction (Ambady, Koo, Rosenthal, & Winograd, 2002; Roter, Frankel, Hall, & Sluyter, 2006). Unfortunately, providers do not use these strategies consistently because of limited time and training, so that patients leave office visits not receiving and/or not remembering critical information (Kessels, 2003). Patient portals may exacerbate this problem because test results are usually delivered as a set of numbers with minimal context, often in a table format. Table formats can impair ability to detect out-of-range values for a test result embedded in multiple test results (Zikmund-Fisher et al., 2014, 2017); for example, see Figure 2.

We investigated whether enhanced context for standard portal messages about cholesterol and diabetes screening test results improve older adult gist comprehension and boost attitudes toward and intention to perform self-care behaviors. We compared verbally, graphically, and video-enhanced formats that provide increasing levels of support for understanding and remembering test results at an ordinal level of gist (lower, borderline, higher risk). These formats may also help calibrate risk perception and increase motivation to make decisions about behaviors that mitigate risk, although those format features that improve comprehension do not always increase motivation (Ancker, Senathirajah, Kukafka, & Starren, 2006).

The verbally enhanced format should support ordinal gist comprehension of risk associated with test scores (see Figure 3). Labels for evaluative categories (lower, borderline, or higher risk) were added to facilitate the process of mapping scores to appropriate regions of risk on the scale. Such evaluative labels have been shown to facilitate interpretation of health data, in part by promoting emotional processing of quantitative information (Peters et al., 2009). Therefore, the labels may benefit older adults, who tend to have similar or higher levels of knowledge about language/verbal ability (Hartshorne & Germine, 2015) and to focus more on emotional meaning (Charles & Carstensen, 2010) relative to younger adults.

Component	Your Value	Standard Range	Units
Total Cholesterol	184	< 200 -	mg/dl
Triglycerides	42	< 150 -	mg/dl
HDL Cholesterol	47	40 - 60	mg/dl
LDL Cholesterol	130	< 100 -	mg/dl

"Your test results require discussion to assess your future plan of care.
A follow up appointment is recommended to discuss your results."

Figure 2. Standard message format (Morrow et al., 2017). HDL = high density lipoproteins; LDL = low density lipoproteins.

Component	Your Value (mg/dl)	Range of Scores (mg/dl)
Total Cholesterol	184 **Desirable**	< 200 (Desirable) 200-240 (Borderline) >240 (High)
Triglycerides	42 **Optimal**	< 100 (Optimal) 101-149 (Normal) 150-199 (Borderline) 200-499 (High) >500 (Very High)
HDL Cholesterol	47 **Borderline**	< 40 (Low/Bad) 41-59 (Borderline) >60 (High/Good)
LDL Cholesterol	130 **Borderline**	< 100 (Optimal) 101-129 (Near Optimal) 130-159 (Borderline) 160-189 (High) >190 (Very High)

"Your risk for heart disease is borderline. I recommend a follow up appointment to discuss your future plan of care."

Figure 3. Verbally enhanced message format (Morrow et al., 2017). HDL = high density lipoproteins; LDL = low density lipoproteins.

The graphically enhanced format provided graphical as well as verbal cues (see Figure 4). The test scores were embedded in graphic representations of the scale for each score (number lines), with color coding and facial icons reinforcing the verbal labels. Because the order of the regions from lower to higher risk (or higher to lower risk for HDL) is emphasized, this format should further support ordinal gist understanding (Ancker et al., 2006; Peters et al., 2009). Similar graphics support gist understanding organized around evaluative and affective dimensions (Fagerlin et al., 2007; Garcia-Retamero & Galesic, 2009; Galesic & Garcia-Retamero, 2011; Reyna et al., 2009) and improve patients' understanding of cholesterol scores and other health parameters (Douglas & Caldwell, 2011; Tao, Yuan, & Qu, 2018), including improving their ability to calibrate scores to risk level compared to tabular formats typical of portals (Zikmund-Fisher et al., 2017).

In the most enhanced condition, the same graphic display (with the labeled test scores) is accompanied by a video of a physician discussing the results (see Figure 5), using nonverbal cues (prosody, facial expressions) and verbal cues (the same risk category labels as in the other enhanced formats) to signal information relevance and to guide affective interpretation, as in ideal face-to-face communication.

Based on our framework, we expected enhanced messages to be better understood and remembered, and more acceptable to patients (meeting their informational and affective needs) by providing context that guides understanding of test values, particularly at the gist level. Adding verbal labels for risk categories to the standard messages should help older adults integrate numeric information into gist, reducing the need for high levels of numeracy and processing capacity (e.g., working memory) to develop gist representations (Peters et al., 2009). Evaluative labels also help older adults interpret and integrate numbers in terms of affective significance, so they can develop affectively as well as cognitively organized gist representations that support decisions (Peters et al., 2009). In the graphic condition,

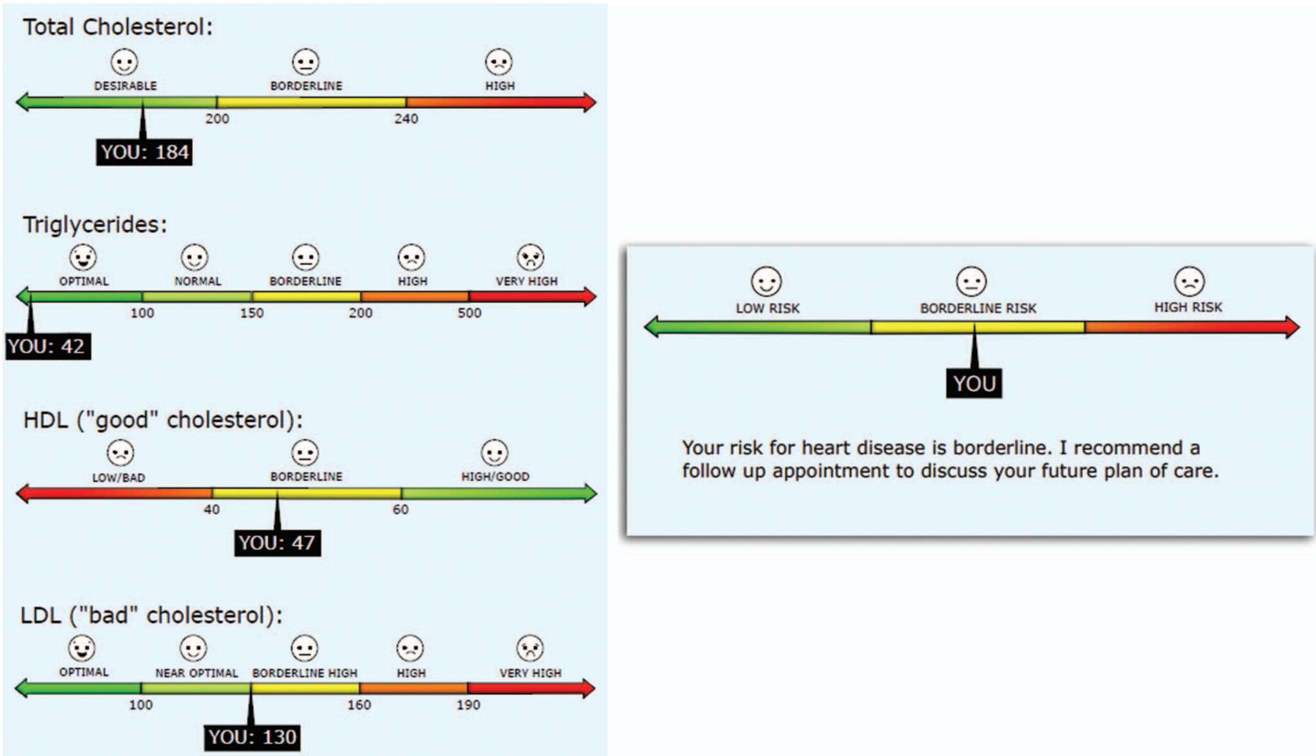

Figure 4. Graphically enhanced message format (Morrow et al., 2017). HDL = high density lipoproteins; LDL = low density lipoproteins. See the online article for the color version of this figure.

embedding the labeled test scores in the graphic number lines help convey quantitative information more efficiently than verbal and numeric formats do by making key gist features more cognitively accessible (Peters et al., 2009; Reyna et al., 2009), helping patients integrate risk information across test scores. Finally, the multimedia video format may be most effective because the verbal and nonverbal cues reinforce each other (Van Gerven, Paas, Van Merriënboer, Hendriks, & Schmidt, 2003).

We also varied the risk level indicated by the test scores in the scenarios (lower, borderline, higher risk). Borderline results were expected to be most difficult to understand because uncertainty (and variability) of risk is higher for each score, and integration across scores is more demanding, although this difficulty may be mitigated by enhanced formats if they guide integration of gist interpretation of component scores. We also investigated whether more enhanced messages would be more trusted by participants (satisfied with information provided) because they are easier to understand and prompt affective response to message information.

Enhanced messages should also improve risk perception and increase attitude toward and intent to perform self-care behaviors that may mitigate risk (increased risk perception and behavioral attitude and intent as risk increases). Because gist memory (Reyna et al., 2009) and risk perception (Brewer et al., 2007; Peters et al., 2009) involve affective response to risk, we also explored whether negative affect would increase and positive affect decrease with increasing risk associated with test results. The enhanced messages may also increase sensitivity of affect to risk level, reflecting more accurate gist representation of risk. Moreover, participants viewing the video message may be most sensitive to risk because the physician's commentary about the test results emphasizes risk and reinforces affective response.

Method

Participants

Participants were 144 community-dwelling adults (average age of 71.9 years, range = 60–94 years; 71.5% females). All were native English speakers, with no physical, cognitive or visual/auditory impairments that could limit participation. 18.8% had high school level of education or lower, 13.2% had some college and 68.0% had at least a college degree. The study was approved by the University of Illinois at Urbana-Champaign and Carle Hospital Foundation institutional review boards. Participants provided consent before participation.

Scenarios

Twenty-four scenarios that contained fictitious patient profiles and messages that described results for these patients from cholesterol or HbA1c diabetes screening tests were developed in collaboration with two physicians from our partner health care system to ensure that the patterns of test results were typical of older adults. Messages for diabetes screening as well as cholesterol test results were developed because Type 2 diabetes is a common age-related chronic illness that is often comorbid with cholesterol-

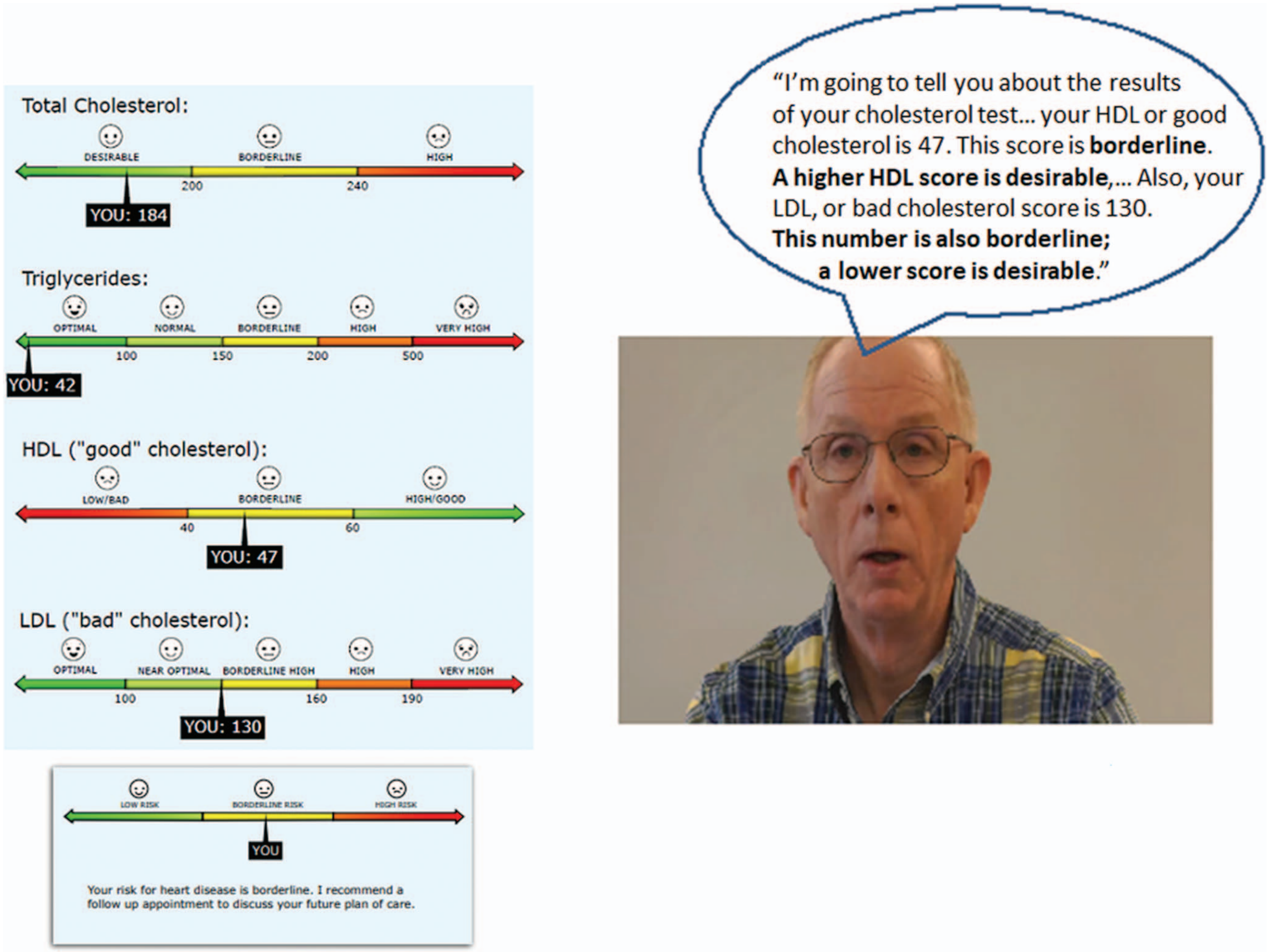

Figure 5. Video enhanced message format (Morrow et al., 2017). The speech balloon is for illustration purposes only. Participants in the study hear the speech. HDL = high density lipoproteins; LDL = low density lipoproteins. Photos are used with permissions. See the online article for the color version of this figure.

related illness. The cholesterol messages described complex patterns of scores on multiple tests (total cholesterol, triglycerides, high density lipoproteins [HDL], and low density lipoproteins [LDL]) that suggested low, borderline, or high risk for cardiovascular illness, according to standard cut-off values for risk categories (National Institutes of Health, National Heart, Lung, and Blood Institute, 2001). To help participants understand the overall risk associated with each message, the message ended with a summary of the overall risk for heart disease associated with the test scores. Therefore, each message was divided in two segments: (1) description of all test result components and the associated level of risk for each component, and (2) an overall summary of risk.

There were equal numbers of cholesterol and diabetes messages reporting test results from each level of risk. Because these risk levels depend on patient-specific risk factors (e.g., coronary artery disease, hypertension, family history of heart disease) as well as the scores, patient profiles accompanied each message. A practice scenario preceded both the cholesterol and diabetes messages to familiarize participants with the structure of the trials and the measures.

Message Formats

We compared comprehension of and responses to clinical test result messages presented in standard, verbally enhanced, graphically enhanced, and video-enhanced formats.

Standard. The standard format presented test results as a table of numbers with some text information, typical of the actual portal in our partner health organization (see Figure 2). Such table formats are commonly used in many actual patient portals and are difficult to interpret by patients (Zikmund-Fisher et al., 2017).

Verbally enhanced. In the verbally enhanced format (see Figure 3), labels for evaluative categories (more or less risk) as well as more information about the regions of risk associated with the scale for each score were added to the table to provide context for interpreting the specific numbers. The labels and cut-off values for these categories were based on recommendations from the

National Institutes of Health, National Heart, Lung, and Blood Institute (2001). These labels should promote emotional processing of the quantitative information (Peters et al., 2009), and may help older adults understand the results because they tend to have high levels of verbal ability (Schwartz, Woloshin, Black, & Welch, 1997) and focus on emotional meaning (Charles & Carstensen, 2010).

Graphically enhanced. In the graphically enhanced condition, graphics that convey key relational features (larger/smaller than) were included. Similar to Leckart (2010) and Zikmund-Fisher et al. (2017), the test scores were embedded within graphic number line representations of each scale. Zikmund-Fisher et al. (2017) investigated several cues for enhancing the impact of number lines on patient understanding of test scores, such as block versus gradient lines to distinguish risk categories as well as different color schemes (green for in-range and gray for out-of-range scores, vs. a stoplight condition with green for in-range scores and yellow-orange-red to indicate increasing risk levels in the out-of-range region). Our graphically enhanced format was similar to the block line graph with the stoplight color scheme in Zikmund-Fisher et al. (2017). In addition, like Leckart (2010), we used facial icons as well as color to reinforce the verbal labels for each risk region, which should support gist representation organized around evaluative and affective dimensions (Garcia-Retamero & Cokely, 2017; Garcia-Retamero & Galesic, 2009; Reyna, 2008). Unlike Zikmund-Fisher et al. (2017), we also used these graphs to convey multicomponent test results (cholesterol) as well as single-component results (diabetes).

Video enhanced. The same graphics were accompanied by the video of a physician who provided commentary about the test scores, with nonverbal cues (prosody, facial expressions) signaling information relevance and guiding affective interpretation, as in ideal face-to-face communication. The graphics and numbers were also included in this condition. The video messages were recorded by a physician, who presented the information as if he were discussing the results and their implications with his patient. In the script for the messages, important information was italicized and the most important/relevant information was bolded to indicate what the physician should emphasize when talking to the patient (see Table 1 for an example). As each test score was discussed, the corresponding part of the graphic loomed to help patients link the verbal commentary with the relevant graphic information. This multimedia format should be most effective because the verbal and nonverbal cues reinforce each other (Van Gerven et al., 2003).

Before the primary study, a pilot study was conducted to ensure that the video-recorded messages would be appropriate for older adults (Azevedo et al., 2015). Older adults (same inclusion criteria as primary study) viewed the 24 video messages used in the primary study. As in the primary study, each message was divided into two parts: (1) the physician first described the test results, followed by questions about the risk information in the message; (2) then, the summary statement was presented, followed by additional questions about the complete message's content and presentation. Older adults were generally able to understand the gist of the video test messages for both cholesterol and diabetes tests, especially after hearing the summary risk statement. Participants' affective responses to the messages were appropriate to the message's level of risk: As the level of risk associated with the test results increased, positive affect decreased and negative affect increased. The same pattern occurred for the cholesterol and for the diabetes messages. Moreover, participants thought the video messages were informative and that the physician's delivery (e.g., tone of voice) was appropriate for the level of risk conveyed.

Study Design

The 12 cholesterol and 12 diabetes messages were presented in blocks. Within each block there were messages at each of the three risk levels (low, borderline, and high). Message format was a between-groups factor.

For each scenario, after seeing the test results message in one of the four randomly assigned message formats, participants responded to questions about memory for the test results, risk perception, affective response, behavioral attitudes, behavioral intentions, and message satisfaction, in that order. To ensure participants knew which scenario was tested, the questions included the name of the patient in the preceding scenario (e.g., "If you were the patient Peter, how would you feel as you viewed this message?"). Thus, participants responded to the questions as if they were the patient in the scenario.

Measures

Message memory. Verbatim memory for the individual component scores in the message was probed by asking participants to recall the exact numeric value of the component score. e.g., "Sam's LDL score was = ______ (numerical score)". Verbatim memory for the individual test scores (e.g., HDL) was measured, scored both strictly (correct = exact value of the test score) and more liberally (correct = within absolute value of 5% of the correct value). In addition, gist memory for risk associated with both individual test scores (e.g., HDL) and with the overall message (global risk) was evaluated (ordinal level gist: low/borderline/high; Reyna et al., 2009). For example, "Considering Sam's cho-

Table 1
Example of Script for Cholesterol Messages in the Video-Enhanced Condition

Example
"I'm going to tell you about the results of your cholesterol test. Your *total cholesterol score* is *145*. This is a ***good score*** and shows that you have **low risk for heart disease**. *Your triglycerides* score is *148*. This score is ***normal, and is another sign that you have low risk for heart disease***. Your *HDL or good cholesterol is 43*. This score is ***borderline. A higher HDL score is desirable*** because these lipoproteins help remove bad cholesterol from your bloodstream and artery walls, which lowers your risk. **In addition**, *your LDL, or bad cholesterol*, *score is 72*. This score is also good in terms of your *risk for heart disease*. ***Overall***, ***I feel your risk for heart disease is low. Your test results are good so there is no need to worry about your results at this time.***"

lesterol test results, his HDL score indicated by the set of results in the message was: (ordinal level gist: high/good; borderline; low/bad). Global risk was probed before as well as after the second part of the message (summary statement) in order to evaluate how well participants could extract overall gist only from the component scores. For example, "Considering Sam's cholesterol test results, his overall level of risk indicated by the set of results in the message was = _________ (ordinal level gist: low/borderline/high)".

Affective reactions. Participants indicated their affective response to the messages. They indicated to what extent seven negative and seven positive emotions were experienced by responding, for each emotion, to a 9-point scale ranging from 1 (*not at all*) to 9 (*very much*), as follows: "If you were the patient Sam, how would you feel as you watched this message? Indicate the extent that you felt: (assured, calm, cheerful, happy, hopeful, relaxed, and relieved; or anxious, afraid, discouraged, disturbed, sad, troubled, and worried" (Garcia-Retamero & Cokely, 2011). As in Garcia-Retamero and Cokely (2011), a composite score was created by reverse scoring negative emotion ratings and combining with the positive emotion ratings. The composite score ranged from 1 (*most negative*) to 5 (*neutral*) to 9 (*most positive*).

Risk perception. Participants indicated the perceived risk associated with the reported test results by ranking on a 9-point scale, ranging from 1 (*very unlikely*) to 9 (*very likely*) the likelihood of developing heart disease and heart-related complications if nothing was done to reduce the reported cholesterol levels, if they were the patient in the scenario (Garcia-Retamero & Cokely, 2011).

Attitude toward taking medication. Participants indicated attitude toward taking medications by ranking on a 9-point scale, ranging from 1 (*not at all*) to 9 (*very much*) how favorable they would feel about taking medications prescribed for lowering cholesterol, if they were the patient in the scenarios (Garcia-Retamero & Cokely, 2011).

Intention to perform self-care behaviors. Similarly, behavioral intent was measured by asking participants to rank the following on a 9-point scale, ranging from 1 (*I have no intention of doing this*) to 9 *(I am certain that I would do this*): If they were the patient in the scenario, (a) how likely were they to take medication prescribed to reduce cholesterol; (b) how likely were they to change their diet; and (c) how likely were they to increase their level of exercise? (adapted from Garcia-Retamero & Cokely, 2011).

Satisfaction with the message. Participants indicated the extent to which they considered the information conveyed in the messages was useful for the patient in the scenario on a 9-point scales ranging from 1 (*not at all useful*) to 9 (*very useful*), as follows: "How useful do you think was the information conveyed in this message?" (Garcia-Retamero & Cokely, 2011).

Results

Message Format Effects on Memory for Test Results

Memory for the component scores in the cholesterol messages (total, triglycerides, HDL, LDL) was analyzed by a 4 (format: standard, verbally enhanced, graphically enhanced, video-enhanced) × 3 (risk level: lower, borderline, high) × 2 (score type: verbatim, gist) mixed design analysis of variance (ANOVA; risk level and score type were repeated measures).[1] We collapsed across type of component scores because they were presented in a fixed order in the messages and because they varied in number of risk categories (three or five), making it difficult to interpret any differences in memory as due to the type of score (e.g., total vs. HDL) or the order of presentation of the score in the message (e.g., first vs. third). For purposes of scoring participant memory, component scores with five possible risk categories (triglycerides and LDL) were simplified to three categories (optimal or near optimal/normal was scored as low level of risk; high or very high represent was scored a high level of risk).

Memory was influenced by format, $F(3, 140) = 4.2$, $p < .01$, $\eta_p^2 = .02$, video ($M = .50$ correct) > graph enhanced ($M = .42$), video > standard ($M = .40$), video = verbally enhanced ($M = 47$); risk level, $F(2, 700) = 4.5$, $p < .05$, $\eta_p^2 < .01$, low ($M = .48$) > borderline (.42) = high ($M = .44$); and score type, $F(1, 700) = 244.2$, $p < .001$, $\eta_p^2 = .20$, gist ($M = .57$) > verbatim ($M = .33$). Most important, these effects were qualified by a Format × Risk Level × Score interaction, $F(6, 700) = 4.4$, $p < .001$, $\eta_p^2 = .03$, suggesting the impact of message format and risk level differed for verbatim and gist scores. This interaction was analyzed by conducting separate Format × Risk Level ANOVAs for verbatim and gist scores.

For verbatim scores, risk level, $F(2, 280) = 7.0$, $p < .01$, $\eta_p^2 = .02$, high < borderline, but not message format, $F(3, 140) = 1.8$, $p > .10$, $\eta_p^2 = .02$, influenced memory, although the effect of risk level was small. The Format × Risk Level interaction was also not significant, $F(6, 280) = 1.3$, $p > .10$, $\eta_p^2 = .01$. Unlike verbatim memory, gist memory was influenced by message format, $F(3, 140) = 12.2$, $p < .001$, $\eta_p^2 = .09$, enhanced formats > standard, no difference among enhanced formats, as well as risk level, $F(2, 280) = 16.3$, $p < .001$, $\eta_p^2 = .07$, low = high > borderline. These effects on gist memory were qualified by a Format × Risk Level interaction, $F(6, 280) = 4.9$, $p < .001$, $\eta_p^2 = .06$. This interaction was analyzed by separate format ANOVAs for each risk level. Format influenced gist memory for the lower risk level scenarios, $F(3, 140) = 2.8$, $p < .05$, $\eta_p^2 = .06$ (video > standard) and borderline risk level scenarios, $F(3, 140) = 22.4$, $p < .001$, $\eta_p^2 = .32$ (all enhanced formats > standard), but not for high risk level, $F(3, 140) = 0.17$, $p > .10$, $\eta_p^2 = .004$ (see Figure 6).

Gist memory for global risk associated with the test results was also probed, both before the summary statement about overall risk (which required participants to combine risk information across the four component scores) and after the statement (making it easier to identify global risk for the message). Note that extracting global risk without support from the summary is often required in patient portals when only component test scores are provided. Global gist memory was analyzed by a Format × Risk Level × Probe Position (before/after summary statement) mixed design ANOVA (risk and position were repeated measures). Global gist was influenced by message format, $F(3, 140) = 9.1$, $p < .001$, $\eta_p^2 = .06$ (video enhanced > all other formats; verbally enhanced > standard) and risk level, $F(2, 700) = 44.2$, $p < .001$, $\eta_p^2 = .08$ (high > low >borderline). Probe position was also significant, $F(1, 700) = 43.1$, $p < .001$, $\eta_p^2 = .04$, with more

[1] This analysis was performed on the cholesterol dataset only, because diabetes messages had just one component score (A1C). For all other parallel analyses performed on diabetes message data, see the Appendix.

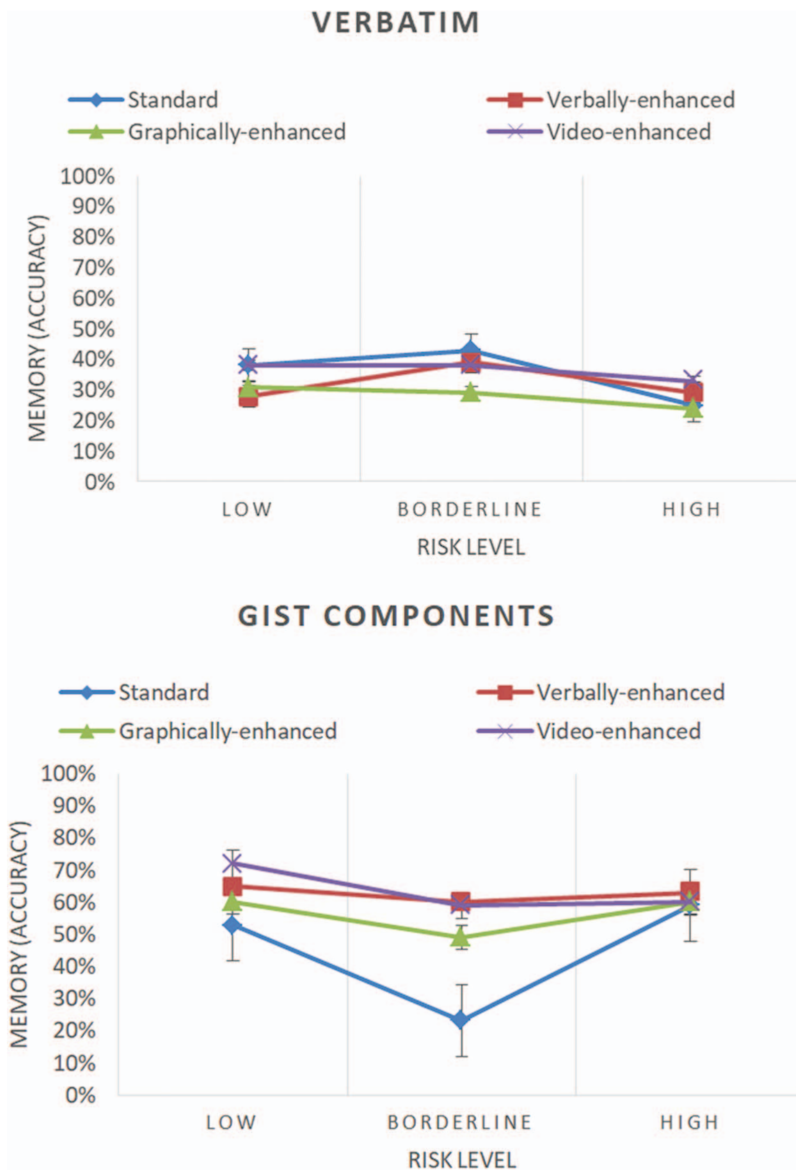

Figure 6. Verbatim and gist memory for the component scores. See the online article for the color version of this figure.

accurate memory after rather than before the summary statement. A Format × Risk interaction, $F(6, 700) = 4.6$, $p < .001$, $\eta_p^2 = .03$, was analyzed by separate format ANOVAs for each level of risk (see Figure 7). Although message format did not influence gist memory for the higher risk scenarios, $F(3, 140) = 2.0$, $p > .10$, $\eta_p^2 = .04$, it influenced gist memory for the lower risk messages, $F(3, 140) = 5.1$, $p < .01$, $\eta_p^2 = .10$ (video > graphic enhanced) and for the borderline risk messages, $F(3, 140) = 6.8$, $p < .001$, $\eta_p^2 = .13$ (video > standard; verbally enhanced > standard). Thus, analysis of the global as well as the component gist scores revealed an advantage for the video and verbally enhanced formats compared to the standard format messages, especially for the more challenging borderline test results condition. Somewhat surprisingly, the graphically enhanced format did not improve gist memory, especially in the lower risk condition. Analysis of gist memory errors suggested that participants tended to overestimate risk for the lower risk (normal) test result messages in this condition.

Message Format Effects on Responses to Test Results

To examine whether format influenced participant responses to, as well as memory for the risk information in the portal messages, parallel ANOVAs were conducted for affective response, risk perception, attitude toward taking medication, and intent to perform self-care behaviors that may mitigate risk (medication, exercise, and diet). We note that for all of these variables, the effect of risk level was reliable ($p < .001$) and consistent, showing a monotonic increase in ratings with level of risk (Bonferroni comparisons: higher > borderline > lower

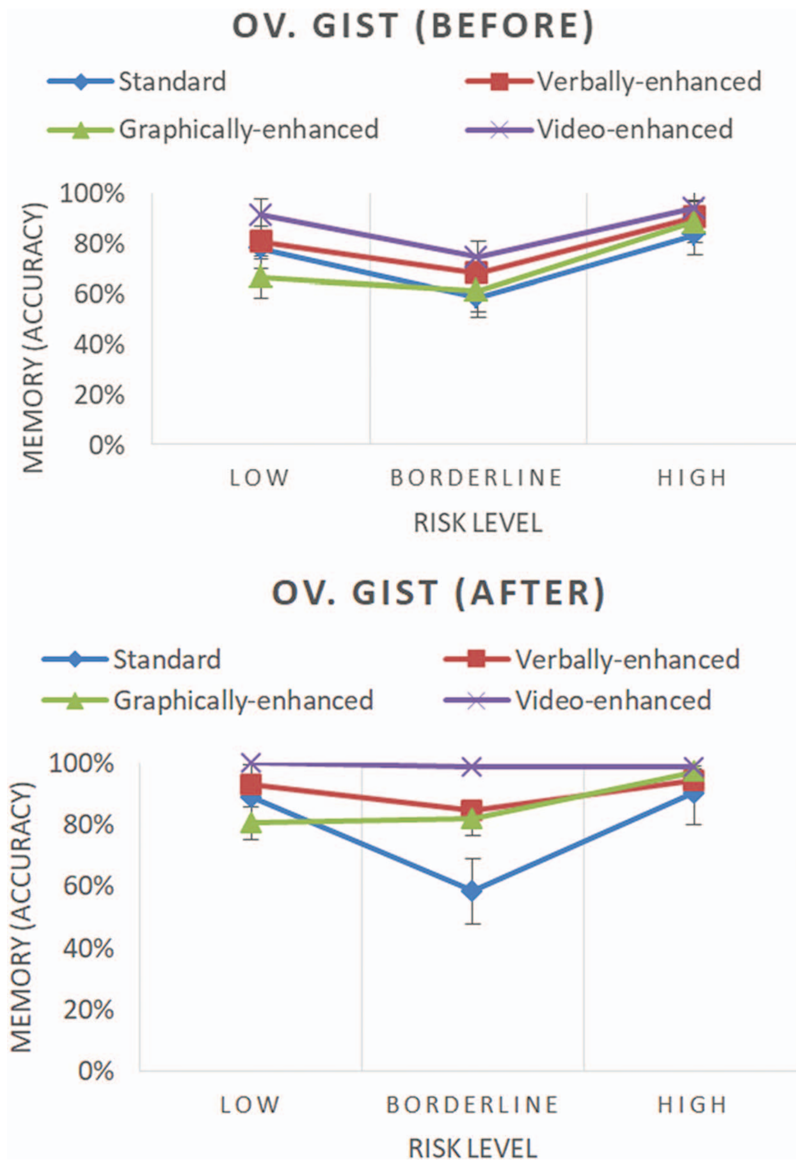

Figure 7. Overall gist memory. See the online article for the color version of this figure.

risk scenarios). Therefore, overall risk level effects are not included for the individual Format × Risk Level ANOVAs reported below, except for the affective response measure.

Affective responses. Affective responses were influenced by risk level ($F(2, 420) = 455.4$, $p < .001$, $\eta_p^2 = .68$; low > borderline > high) with very high responses (positive affect) for low risk scenarios that decreased as risk levels increased (see Figure 8). Affect was not influenced by message format, $F(3, 420) = 0.9$, $p > .10$, $\eta_p^2 = .01$. The Format × Risk interaction approached significance, $F(6, 420) = 2.0$, $p > .05$, $\eta_p^2 = .03$.

Risk perception. Message format influenced risk perception, $F(3, 420) = 7.5$, $p < .001$, $\eta_p^2 = .05$ (graphic > verbally enhanced), although the format effect depended on risk level, $F(6, 420) = 3.2$, $p < .001$, $\eta_p^2 = .04$. This Format × Risk interaction was analyzed by conducting separate format one-way ANOVAs for each level of risk. For the lower risk scenarios, participants' perceived risk was greater for graphic than for verbally and video enhanced formats, $F(3, 140) = 6.0$, $p < .001$, $\eta_p^2 = .11$ (graphic > verbally enhanced and video, graphic = standard). A somewhat similar pattern occurred for borderline risk scenarios, $F(3, 140) = 3.0$, $p < .05$, $\eta_p^2 = .06$ (graphic > verbally enhanced, graphic = video = standard). There were no format-related differences in perceived risk for the higher risk scenarios, $F(3, 140) = 2.7$, $p > .10$, $\eta_p^2 = .06$. These results are consistent with the pattern of gist memory errors, suggesting participants tended to estimate higher levels of risk for the low risk (normal) test result messages in the graphic compared to the other format conditions (see Figure 9).

Medication attitude. Format influenced attitude toward taking medication, $F(3, 420) = 4.9$, $p < .001$, $\eta_p^2 = .03$, although the Bonferroni comparisons were not significant. There was a

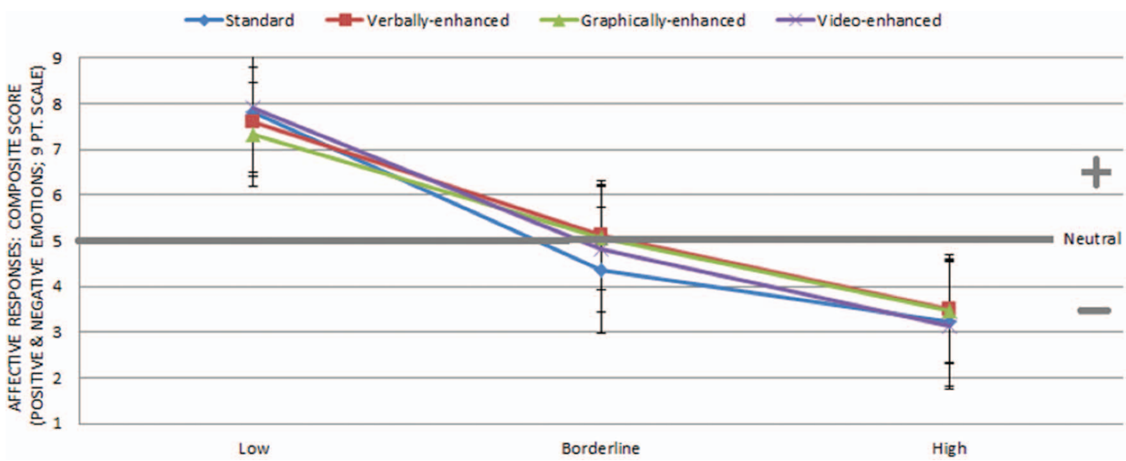

Figure 8. Affective responses (1 = most negative, 5 = neutral, 9 = most positive). See the online article for the color version of this figure.

numeric trend for more positive attitudes in the graphic than in the other format conditions. However, a Format × Risk interaction, $F(6, 420) = 2.7$, $p < .05$, $\eta_p^2 = .04$, showed that participants in the graphic condition had a more positive attitude toward taking medication compared to the other format conditions only for the lower risk messages (graph > all message formats), which is consistent with their exaggerated perception of risk in this condition. Format differences were not significant for the borderline and higher risk scenarios (see Figure 10).

Behavioral intention. Like attitude toward taking medication, format influenced participants' intention to take medication, $F(3, 420) = 3.9$, $p < .001$, $\eta_p^2 = .03$. Although the Bonferroni comparisons were not significant, again there was a numeric trend for more positive attitudes in the graphic than in the other format conditions. However, neither intention to exercise, $F(3, 420) = 2.0$, $p > .10$, $\eta_p^2 = .01$, nor intention to change diet was influenced by message format, $F(3, 420) = 2.1$, $p > .10$, $\eta_p^2 = .02$ (also see Figure 10).

Message satisfaction. Finally, format influenced participants' satisfaction with the messages, $F(3, 420) = 27.0$, $p < .001$, $\eta_p^2 = .16$. Participants were most satisfied with the graphic compared to the other message formats (graphic > all formats; video > verbally enhanced and standard), even though this format tended to decrease gist memory accuracy and increase risk perception (primarily for lower risk messages) compared to the other formats (see Figure 11). Similar graphic formats have been shown to be preferred in comparison with the standard tables (Zikmund-Fisher et al., 2017). Our results extend this finding, showing that participants are more satisfied with the graphic format even in comparison with other enhanced message formats, perhaps because satisfaction is associated with perceived ease of understanding and usefulness of the graphic format (Tao et al., 2018). In addition, scenario risk level had an

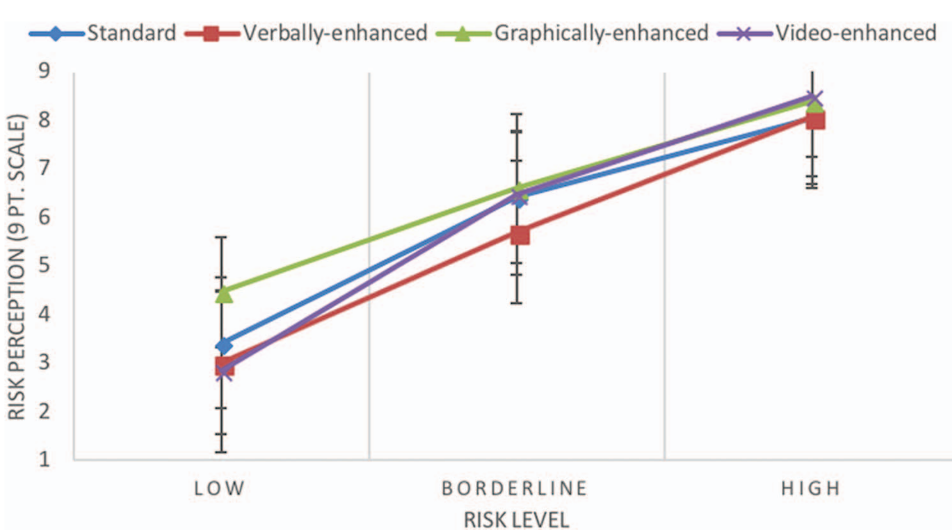

Figure 9. Risk perception (1 = lowest, 9 = highest). See the online article for the color version of this figure.

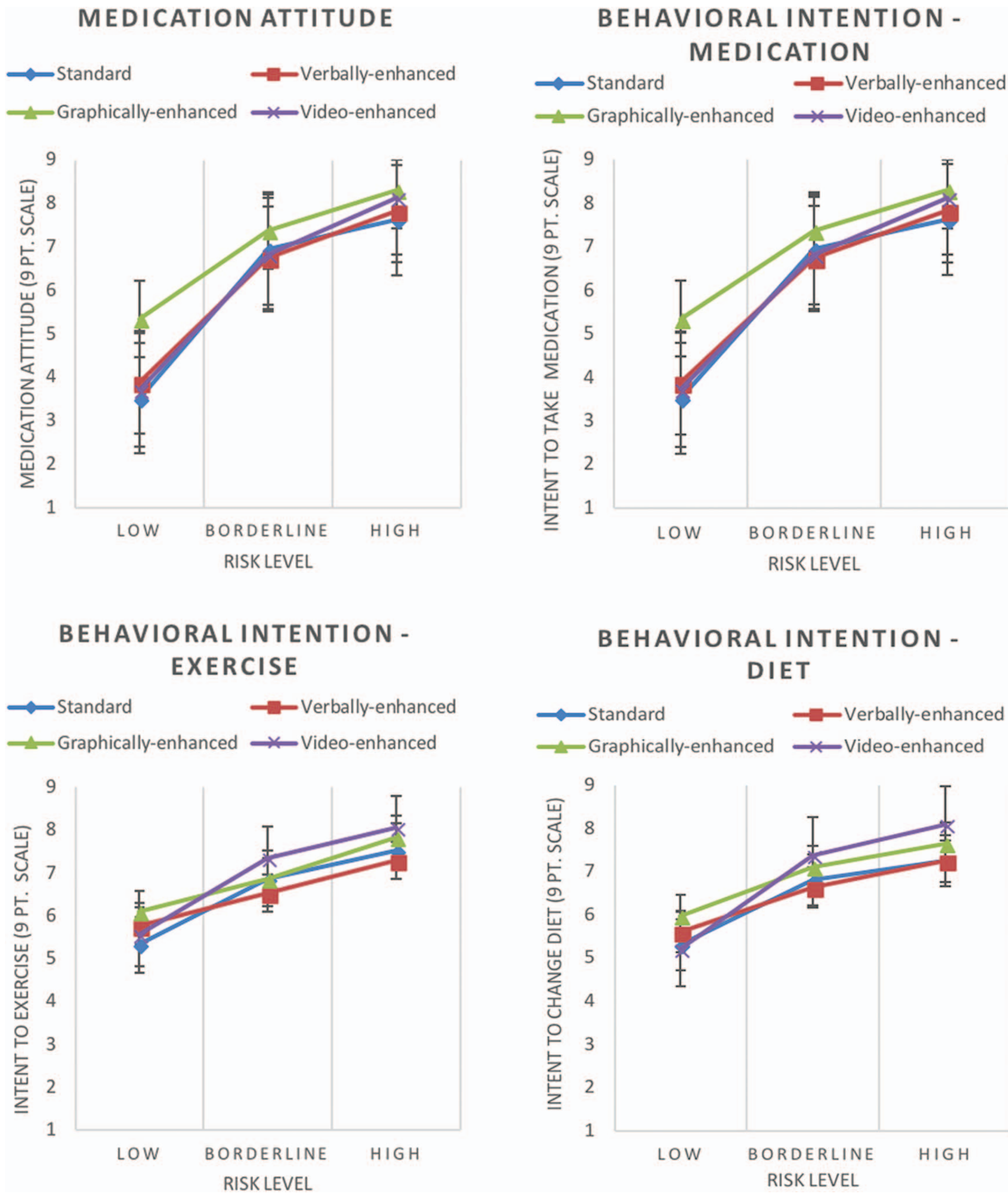

Figure 10. Attitude and behavioral intentions (1 = lowest, 9 = highest). See the online article for the color version of this figure.

attenuated impact on satisfaction compared to the other format conditions, $F(2, 420) = 3.7$, $p < .05$, $\eta_p^2 = .02$ (follow-up Bonferroni comparisons not significant).

Comparing Cholesterol and Diabetes Message Results

Parallel analyses were conducted for the diabetes message outcome variables. For the sake of brevity, results of these analyses are presented in the Appendix, with a summary of the cholesterol data results for comparison. The pattern of results for the diabetes messages was very similar to the cholesterol message results, with the following minor exceptions. The effect of risk level for verbatim memory scores and the effect of format for overall gist scores did not reach significance. Furthermore, the Format × Risk interaction for medication attitude was not significant. These differences may reflect in part a ceiling effect on memory for the diabetes scores because these messages were much simpler than the cholesterol messages.

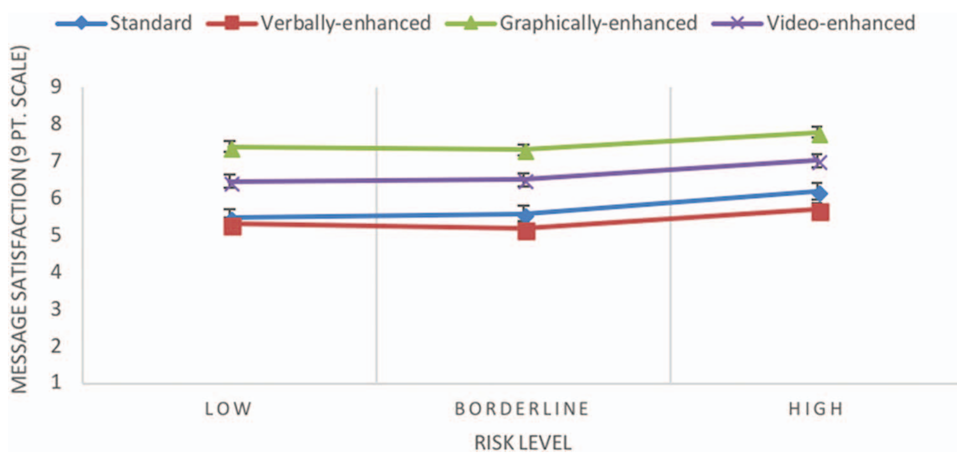

Figure 11. Message satisfaction (1 = lowest, 9 = highest). See the online article for the color version of this figure.

Notably, participants were equally satisfied with the graphic enhanced and video-enhanced diabetes messages in comparison to the standard and verbally enhanced formats, $F(3, 420) = 27.1$, $p < .001$, $\eta_p^2 = .16$, which can be seen as additional evidence that the video format was most effective (comprehension and message satisfaction measures). These findings increase our confidence that the message format results generalize to other clinical test result information.

Relationships Among Participant Ability and Message Outcome Measures

Simple correlations among the participant ability measures replicate patterns found in the cognitive aging literature (see Table 2). Measures of vocabulary ability (Ekstrom, French, Harman, & Dermen, 1976) and literacy (Author Recognition Test; Stanovich, West, & Harrison, 1995) were correlated, suggesting a verbal ability/knowledge construct, whereas measures of processing speed (letter comparison and pattern comparison tests, Salthouse & Babcock, 1991) were correlated, suggesting a fluid mental ability/processing capacity construct. Subjective numeracy (Fagerlin et al., 2007) and objective numeracy (Berlin Numeracy Test, Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012) were moderately correlated. The health literacy measure (Short Test of Functional Health Literacy in Adults; Baker et al., 1999) correlated with all of the cognitive ability measures, consistent with our framework that views health literacy as a complex function of broader cognitive abilities (also see Chin et al., 2011; Morrow & Chin, 2015). In addition, the correlational analyses increase our confidence that our participant sample were representative of the population of older adults.

We also examined associations of these ability variables with individual differences in memory for and responses to the portal messages (see Table 2). Gist memory was associated with most of the cognitive, numeracy, and health literacy measures, consistent with theories of numeric comprehension (Reyna, 2011). It was also correlated with risk perception, which in turn was correlated with intention to perform self-care behaviors. Path analyses of these correlations were conducted to further evaluate predictions based on our framework and the results will be described in a future paper.

Discussion

The present study provides evidence that enhancing the context of numeric information in patient portal messages improves older adults' memory for and response to clinical test results, compared to a table-based format often used in patient portals to EHR systems. More specifically, verbally and video enhanced messages improved older adults' gist but not verbatim memory for test results, suggesting we were successful in designing portal messages that highlight gist-based risk information. Ordinal gist features of risk for illness associated with test scores were emphasized by verbal labels for risk categories, graphical features that emphasized risk categories (color coding, lines that highlight ordinal relationships among risk levels) and a video physician who used verbal and nonverbal cues to gist categories in order to elaborate the graphical representation of risk.

On the whole, the video-enhanced and verbally enhanced formats were more effective than the graphic format, with some evidence that the video format was most effective (overall memory for component scores in the messages; global gist memory). An unexpected finding was that the graphic format prompted older adults to estimate higher levels of risk than indicated by the scores in the lower risk messages compared to the other formats. This result differs from Zikmund-Fisher et al. (2017), who found similar graphics improved the ability to distinguish levels of urgency in clinical test scores relative to a similar table format. This difference may reflect an age differences in samples (our study focused on older adults while Zikmund-Fisher et al. (2017) included a broader age range with a younger mean age), as well as some differences between the risk levels and the types of graphs used to display test information in the two studies. Risk associated with test scores in our study also ranged from low/normal to very high, while the scores in Zikmund-Fisher et al. (2017) indicated borderline and higher risk. Finally, Zikmund-Fisher et al. (2017) used a comprehension measure, whereas our study measured verbatim and gist memory. It is possible that demands on processing

Table 2
Correlation Matrix for Variables, Cholesterol Dataset (N = 144)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Age	—	.08	−.21**	−.14†	−.10	−.31***	−.44***	−.19*	−.11	−.20**	−.21**	−.29***	−.20*	.31***	−.20*	−.19*	−.35***	.05
2. Gender		—	.09	−.09	.01	−.05	−.12	.12	.23**	−.08	−.02	−.22**	−.07	−.12	−.01	−.01	.14†	.07
3. Education			—	.45***	.33***	.33***	.22**	.19*	.36***	.17*	.26***	.18*	.16†	.29***	.07	.05	.28***	−.05
4. Verbal ability construct				—	.21**	.41***	.25***	.18*	.23**	.14†	.26***	.29***	.30***	.32***	.18*	.17*	.28***	−.24***
5. Self-reported health					—	.35***	.24***	.07	.45***	.14†	.22**	.19*	.18*	.23*	.16†	.22**	.17*	−.09
6. Health literacy (STOFHLA)						—	.38***	.10	.34***	.14†	.25***	.15†	.28***	.33***	.09	.12	.26***	−.20*
7. Fluid ability construct							—	.20*	.24***	.25***	.16†	.23**	.20*	.25***	.19*	.24***	.29***	−.02
8. Objective numeracy								—	.33***	.20**	.25***	.17*	.23**	.25***	.11	.11	.28***	.10
9. Subjective numeracy									—	.33***	.34***	.11	.25***	.23**	.09	.07	.21*	−.05
10. Verbatim memory										—	.42***	.24***	.21**	.23**	.17*	.22**	.21*	−.01
11. Gist comprehension (before)											—	.38***	.46***	.38***	.20*	.27***	.33***	−.05
12. Affective responses												—	.58***	.47***	.45***	.46***	.41***	−.06
13. Risk perception													—	.39***	.21**	.33***	.37***	.05
14. Medication attitude														—	.46***	.50***	.86***	.18*
15. Behavior intention (medication)															—	.85***	.45***	.17*
16. Behavior intention (exercise)																—	.46***	−.08
17. Behavior intention (diet)																	—	−.13
18. Message satisfaction																		—
M	71.9	—	16.1	11.8	5.4	34.7	12.6	.2	4.5	.5	.8	2.2	2.4	2.3	1.0	1.0	1.9	5.8
SD	7.7		2.5	5.6	1.3	2.6	2.7	.2	1.0	.1	.2	.9	1.0	1.3	1.1	1.1	1.4	2.2

Note. STOFHLA = Short Test of Functional Health Literacy in Adults.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

capacity (working memory) in our study reduced potential benefits of the graphic format for older adults.

Although the enhanced message formats had robust effects on gist memory in our study, they had less impact on older adults' responses to the messages. However, the graphically enhanced format did increase risk perception and attitude toward taking medication in the lower risk (and sometimes borderline risk) messages. The link between risk perception and attitude toward behavior that mitigates the risk is consistent with behavioral change theories, which hypothesize that patients view health behaviors more positively when they believe these behaviors address risk of illness.

The level of risk associated with the test results had a robust effect on all measures except for message satisfaction. However, risk level had different effects on message memory compared to participants' responses to this information: Although participants recalled borderline results less accurately than either lower or higher risk, their affect, risk perception, attitude toward, and intent to perform health behaviors increased monotonically with risk level. They may have remembered borderline results less accurately because the component scores varied in level of risk indicated, which complicated the process of integrating the information into an overall estimate of risk for these messages. In addition, participants tended to remember the gist of borderline risk messages as higher than indicated in the messages, which may have had the effect of increasing perceived risk, affective response, behavioral attitude, and intent in this condition.

Theoretical Implications

Several of the findings are consistent with cognitive and behavioral theories relevant to self-care. First, message format improved gist but not verbatim memory for the test results, which aligns with predictions from fuzzy trace theory. Second, the contribution of affective response to risk perception, as well as the finding that affect was associated with gist memory (see Table 2) is consistent with fuzzy trace theory's claim that gist is partly organized in terms of affective dimensions (Reyna, 2011). Third, the covariation of risk perception, behavioral attitude, and behavioral intention across level of risk associated with the test results is consistent with behavior change theories. Moreover, the relationships between these variables and gist memory error reflecting risk overestimation is consistent with Fuzzy Trace theory and helps link behavioral change theories to theories of cognition in self-care, as articulated by our framework (see Figure 1). Finally, we found that older adults' cognitive abilities are related to memory for and responses to risk information in the portal messages, as predicted our framework. Gist memory was predicted by subjective numeracy and health literacy. Similarly, Zikmund-Fisher et al. (2017) found that lower health literacy (Chew, Bradley, & Boyko, 2004) and graph literacy (Galesic & Garcia-Retamero, 2011) was associated with reduced ability to distinguish levels of urgency in scores when interpreting tabular and graphic displays of test results.

Collectively, these findings support our theory-guided approach to designing enhanced formats for patient portal messages. Verbal (categorical risk labels) and graphical (color coded number lines) features of messages were important for conveying gist-related features. However, the finding that enhanced formats had little direct impact on affective, risk perception and behavioral responses to the messages suggests the need to consider additional features of message content and format in relation to patient motivation and engagement. For example, the content of the test result messages used in the study (cholesterol and diabetes test scores indicating risk level) appeared to be more important than message format for directly influencing affective response, risk perception, and behavior. Other content, such as information about medication effects or messages designed to influence patient beliefs, might be more effectively conveyed by some formats than others. In addition, other types of message formats such as narrative (De Graaf, Sanders, & Hoeken, 2016) may have more impact on patient responses related to motivation.

An unexpected finding was the tendency for older adults to estimate higher levels of risk for the good news/low risk messages in the graphically enhanced condition compared to the other formats. This was not the case in the video-enhanced condition even though it contained the same graphic. This result is consistent with findings in the risk perception literature suggesting that some types of graphic displays can encourage risk avoidance (e.g., perhaps by emphasizing the full range of risk values for a score, including the high-risk regions; Schirillo & Stone, 2005). This may be especially the case for older adults because of an age-related decline in controlled attention processes that inhibit irrelevant information, so that they misremember low-risk scores as higher risk. Moreover, the challenge may have been exacerbated by the fact that older adults tend to have lower graphic literacy (Garcia-Retamero & Muñoz, 2013; Rodríguez et al., 2013), which would hamper their ability to interpret the graph in the first place. These difficulties in interpreting the graphic display may have been countered by the physician commentary in the video condition, which emphasized how the test scores were associated with specific regions of the scale, reducing the impact of the salient higher risk regions on the number line and supporting the ability to inhibit attention to this information. In addition, physician commentary may have helped older adults to make sense of the graphic format. This explanation is also consistent with evidence that people with lower graph literacy are better able to use graphically displayed risk information when provided verbal explanation or other forms of support (Okan, Garcia-Retamero, Cokely, & Maldonado, 2015). Such findings suggest the value of multimedia for conveying risk information to older adults in patient portals. However, we note that because we did not have a measure of 'ground truth' that validates absolute risk for the three risk level conditions, we do not know whether the graphic format led to miscalibration of risk compared to the other conditions among our participants.[2] Moreover, given that some types of graphic aids have been found to reduce rather than increase perceived risk (e.g., Galesic, Garcia-Retamero, & Gigerenzer, 2009), it is important to further investigate the impact of different types of graphic formats on risk perception among older adults with diverse abilities.

Practical Implications

Our findings suggest the value of a theory-guided approach to designing messages that improve patients' understanding and use of numeric health information. Older adults were better able to remember clinical test information at a gist level. The finding that portal message format effects occurred for diabetes as well as cholesterol test results, diagnostic tests related to two common chronic illnesses among older adults, increases confidence in the generalizability of the findings.

Although the link between remembering and responding to clinical test results on the one hand, and patient portal utilization on the other hand, was not tested in the present paper, it is possible that making portal information more cognitively accessible will increase utilization by boosting perceived usefulness of the technology. Theories of planned behavior have been adapted to explain patient acceptance and use of technology that supports self-care. For example, Or et al. (2011) found that older adult beliefs about ease of and usefulness of web-based services predicted intent to use these services. These models may be useful for exploring whether enhanced patient portal messages not only help patients understand and remember risk information, but also motivate them to use portals to support self-care. However, an important caveat is that patients sometimes prefer to receive numeric health information in formats that are not the most effective for improving their memory for this information (e.g., Greene, Peters, Mertz, & Hibbard, 2008; Garcia-Retamero & Cokely, 2017). Indeed, older adults in our study were most satisfied with the graphic format, even though there was little evidence that this format was more effective than the standard portal format (also see Tao et al., 2018).

Addressing barriers to portal use, not only through making portal information easier to understand but improving the usability of the portal system itself, has the potential to reduce disparities in using and benefiting from this form of health technology. More usable and useful patient portal systems can engage and empower patients by supporting their ability to make effective decision about and to act on portal-based information (Lyles et al., 2015; IOM, 2012). However, this is only like to be the case if redesigned portal systems build on existing patient/provider relationships that engender trust rather than replacing this relationship (Lyles et al., 2013). To the extent patients can more fully understand health information presented in their portal, they will be better prepared to engage with providers during clinic visits and able to manage their self-care. They will also have less need to call provider offices for clarification, which will reduce provider workload.

From this perspective, the video-based format may be especially valuable because patient interaction with health information in portal environments retains some aspects of the patient/provider relationship. However, there are clear limitations in how video messages could be implemented in actual portal systems because new videos would need to be developed for different patients and test results, which is very time consuming. To address this issue, we are developing a conversational agent (CA), or virtual provider, to deliver health information in portal environments. Ideally, the CA will emulate best practices for face-to-face communication but will also be capable of delivering a wide range of health information to diverse patients (see Morrow et al., 2017, for more detail, and Azevedo et al., 2017, for an initial evaluation of a CA prototype for patient portals).

[2] From a clinical perspective, such calibration, with risk perception and intention to perform self-care behaviors rising as risk of illness associated with test results increases, would be appropriate. Patients receiving test results in the normal range should have a lower intention to take medication because they are at low risk and there may be side effects associated with taking the medication.

We next plan to scale up evaluation of our approach to improving portal information for patients. The present study had clear limitations on generalizing the results directly to portal environments. The simulation was abstract, without a web-based environment for patients to interact with. This approach may underestimate the cognitive demands of accessing as well as understanding portal messages (e.g., navigation demands may swamp benefits of message format for comprehension). Of course, the use of fictitious patient scenarios also constrains ability to generalize the findings. For example, because participants did not respond to their actual test results, we may underestimate the role of health knowledge and beliefs in how patients understand and respond to health information. Finally, only older adults participated, limiting ability to generalize to other age groups. On the other hand, this age group is now the least likely to use portals, so addressing this group may have greatest impact on patient portal use in health care.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888–918. http://dx.doi.org/10.1037/0033-2909.84.5.888

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions and perceived behavioral control. *Journal of Experimental Social Psychology, 22,* 453–474.

Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. H. (2002). Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. *Psychology and Aging, 17,* 443–452. http://dx.doi.org/10.1037/0882-7974.17.3.443

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: A systematic review. *Journal of the American Medical Informatics Association, 13,* 608–618. http://dx.doi.org/10.1197/jamia.M2115

Azevedo, R. F. L., Gu, K., Zhang, Y., Sadauskas, V., Sakakini, T., Morrow, D., . . . Schuh, W. (2017, November). *Using computer agents to explain clinical test results*. Paper presented at the American Medical Informatics Association Annual Symposium Proceedings, Washington, DC.

Azevedo, R. F. L., Morrow, D., Hasegawa-Johnson, M., Gu, K., Soberal, D., Huang, T., . . . Garcia-Retamero, R. (2015). Improving Patient Comprehension of Numeric Health Information. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59,* 488–492. http://dx.doi.org/10.1177/1541931215591105

Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., & Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Education and Counseling, 38,* 33–42. http://dx.doi.org/10.1016/S0738-3991(98)00116-5

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging, 20,* 341–355. http://dx.doi.org/10.1037/0882-7974.20.2.341

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

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20,* 718–721. http://dx.doi.org/10.1037/0882-7974.20.4.718

Charles, S. T., & Carstensen, L. L. (2010). Social and emotional aging. *Annual Review of Psychology, 61,* 383–409. http://dx.doi.org/10.1146/annurev.psych.093008.100448

Chew, L. D., Bradley, K. A., & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine, 36,* 588–594.

Chin, J., Madison, A., Gao, X., Graumlich, J. F., Conner-Garcia, T., Murray, M. D., . . . Morrow, D. G. (2017). Cognition and health literacy in older adults' recall of self-care information. *The Gerontologist, 57,* 261–268.

Chin, J., Morrow, D. G., Stine-Morrow, E. A. L., Conner-Garcia, T., Graumlich, J. F., & Murray, M. D. (2011). The process-knowledge model of health literacy: Evidence from a componential analysis of two commonly used measures. *Journal of Health Communication, 16*(Suppl. 3), 222–241. http://dx.doi.org/10.1080/10810730.2011.604702

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7,* 25–47.

Crano, W. D., & Prislin, R. (Eds.). (2008). *Attitudes and attitude change*. New York, NY: Psychology Press.

Czaja, S. J., Zarcadoolas, C., Vaughon, W. L., Lee, C. C., Rockoff, M. L., & Levy, J. (2015). The usability of electronic personal health record systems for an underserved adult population. *Human Factors, 57,* 491–506. http://dx.doi.org/10.1177/0018720814549238

De Graaf, A., Sanders, J., & Hoeken, H. (2016). Characteristics of narrative interventions and health effects: A review of the content, form, and context of narratives in health-related narrative persuasion research. *Review of Communication Research, 4,* 88–131.

Douglas, S. E., & Caldwell, B. S. (2011). Design and validation of an individual health report (IHR). *International Journal of Industrial Ergonomics, 41,* 352–359. http://dx.doi.org/10.1016/j.ergon.2011.02.004

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests* (pp. 109–113). Princeton, NJ: Educational testing service.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the Subjective Numeracy Scale. *Medical Decision Making, 27,* 672–680. http://dx.doi.org/10.1177/0272989X07304449

Finucane, M. L., Slovic, P., Hibbard, J. H., Peters, E., Mertz, C., & MacGregor, D. G. (2002). Aging and decision-making competence: An analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. *Journal of Behavioral Decision Making, 15,* 141–164. http://dx.doi.org/10.1002/bdm.407

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: A cross-cultural comparison. *Medical Decision Making, 31,* 444–457. http://dx.doi.org/10.1177/0272989X10373805

Galesic, M., Garcia-Retamero, R., & Gigerenzer, G. (2009). Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology, 28,* 210–216. http://dx.doi.org/10.1037/a0014474

Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied, 17,* 270–287. http://dx.doi.org/10.1037/a0023677

Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: A systematic review of health research and evidence-based design heuristics. *Human Factors, 59,* 582–627. http://dx.doi.org/10.1177/0018720817690634

Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: A cross-cultural comparison. *American Journal of Public Health, 99,* 2196–2202. http://dx.doi.org/10.2105/AJPH.2009.160234

Garcia-Retamero, R., & Muñoz, R. (2013). Cómo Mejorar la Comprensión de los Riesgos Médicos en Personas Mayores. *Revista Latinoamericana de Psicología, 45,* 253–264. http://dx.doi.org/10.14349/rlp.v45i2.1071

Gibbons, M. C., & Casale, C. R. (2010). Reducing disparities in health care quality: The role of health IT in underresourced settings. *Medical Care*

Research and Review, 67(5, Suppl.), 155S–162S. http://dx.doi.org/10.1177/1077558710376202

Gordon, N. P., & Hornbrook, M. C. (2016). Differences in access to and preferences for using patient portals and other eHealth technologies based on race, ethnicity, and age: A database and survey study of seniors in a large health plan. *Journal of Medical Internet Research, 18*(3), e50. http://dx.doi.org/10.2196/jmir.5105

Greene, J., Peters, E., Mertz, C. K., & Hibbard, J. H. (2008). Comprehension and choice of a consumer-directed health plan: An experimental study. *The American Journal of Managed Care, 14,* 369–376.

Haga, S. B., Mills, R., Pollak, K. I., Rehder, C., Buchanan, A. H., Lipkus, I. M., . . . Datto, M. (2014). Developing patient-friendly genetic and genomic test reports: Formats to promote patient engagement and understanding. *Genome Medicine, 6,* 58. http://dx.doi.org/10.1186/s13073-014-0058-6

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science, 26,* 433–443. http://dx.doi.org/10.1177/0956797614567339

Henneman, L., Marteau, T. M., & Timmermans, D. R. (2008). Clinical geneticists' and genetic counselors' views on the communication of genetic risks: A qualitative study. *Patient Education and Counseling, 73,* 42–49. http://dx.doi.org/10.1016/j.pec.2008.05.009

Institute of Medicine (IOM). (2012). *Health IT and patient safety: Building safer systems for better care*. Washington, DC: The National Academies Press.

Kessels, R. P. (2003). Patients' memory for medical information. *Journal of the Royal Society of Medicine, 96,* 219–222.

Leckart, S. (2010). The blood test gets a makeover. *Wired*. Retrieved from http://www.wired.com/magazine/2010/11/ff_bloodwork/

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127,* 267–286. http://dx.doi.org/10.1037/0033-2909.127.2.267

Lyles, C. R., Sarkar, U., Ralston, J. D., Adler, N., Schillinger, D., Moffet, H. H., . . . Karter, A. J. (2013). Patient–provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study. *Journal of the American Medical Informatics Association, 20,* 1128–1131. http://dx.doi.org/10.1136/amiajnl-2012-001567

Lyles, C., Schillinger, D., & Sarkar, U. (2015). Connecting the dots: Health information technology expansion and health disparities. *PLoS Medicine, 12*(7), e1001852. http://dx.doi.org/10.1371/journal.pmed.1001852

Meyer, D., Leventhal, H., & Gutmann, M. (1985). Common-sense models of illness: The example of hypertension. *Health Psychology, 4,* 115–135. http://dx.doi.org/10.1037/0278-6133.4.2.115

Mikels, J. A., Löckenhoff, C. E., Maglio, S. J., Carstensen, L. L. Goldstein, M. K., & Garber, A. (2010). Following your heart or your head: Focusing on emotions versus information differentially influences the decisions of younger and older adults. *Journal of Experimental Psychology: Applied, 16,* 87–95. http://dx.doi.org/10.1037/a0018500

Montaño, D. E., & Kasprzyk, D. (2008). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), *Health behavior and health education: Theory, research, and practice* (4th ed., pp. 67–96). San Francisco, CA: Jossey-Bass.

Morrow, D. G., & Chin, J. (2015). Health literacy and health decision making among older adults. In T. Hess, J. Strough, & C. Lockenhoff (Eds.), *Aging and decision-making: Empirical and applied perspectives* (pp. 261–282). London, UK: Elsevier. http://dx.doi.org/10.1016/B978-0-12-417148-0.00013-3

Morrow, D., Hasegawa-Johnson, M., Huang, T., Schuh, W., Azevedo, R. F. L., Gu, K., . . . Garcia-Retamero, R. (2017). A multidisciplinary approach to designing and evaluating Electronic Medical Record portal messages that support patient self-care. *Journal of Biomedical Informatics, 69,* 63–74. http://dx.doi.org/10.1016/j.jbi.2017.03.015

National Institutes of Health, National Heart, Lung, and Blood Institute. (2001). *ATP III At-A-Glance: Quick Desk Reference (01–3305)*. Retrieved from https://catalog.nhlbi.nih.gov/sites/default/files/publicationfiles/01-3305.pdf

Neuner, J., Fedders, M., Caravella, M., Bradford, L., & Schapira, M. (2015). Meaningful use and the patient portal: Patient enrollment, use, and satisfaction with patient portals at a later-adopting center. *American Journal of Medical Quality, 30,* 105–113. http://dx.doi.org/10.1177/1062860614523488

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied, 21,* 178–194. http://dx.doi.org/10.1037/xap0000045

Or, C. K., Karsh, B. T., Severtson, D. J., Burke, L. J., Brown, R. L., & Brennan, P. F. (2011). Factors affecting home care patients' acceptance of a web-based interactive self-management technology. *Journal of the American Medical Informatics Association, 18,* 51–59. http://dx.doi.org/10.1136/jamia.2010.007336

Peters, E. (2012). Beyond comprehension: The role of numeracy in judgments and decisions. *Current Directions in Psychological Science, 21,* 31–35. http://dx.doi.org/10.1177/0963721411429960

Peters, E., Dieckmann, N. F., Västfjäll, D., Mertz, C. K., Slovic, P., & Hibbard, J. H. (2009). Bringing meaning to numbers: The impact of evaluative categories on decisions. *Journal of Experimental Psychology: Applied, 15,* 213–227. http://dx.doi.org/10.1037/a0016978

Reyna, V. F. (2008). A theory of medical decision making and health: Fuzzy trace theory. *Medical Decision Making, 28,* 850–865. http://dx.doi.org/10.1177/0272989X08327066

Reyna, V. (2011). Across the life span. In B. Fischoff, N. Brewer, & J. Downs (Eds.), *Communicating risks and benefits: A users guide* (pp. 111–120). Washington, DC: FDA.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75. http://dx.doi.org/10.1016/1041-6080(95)90031-4

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135,* 943–973. http://dx.doi.org/10.1037/a0017327

Rodríguez, V., Andrade, A. D., García-Retamero, R., Anam, R., Rodríguez, R., Lisigurski, M., . . . Ruiz, J. G. (2013). Health literacy, numeracy, and graphical literacy among veterans in primary care and their effect on shared decision making and trust in physicians. *Journal of Health Communication, 18*(Suppl. 1), 273–289. http://dx.doi.org/10.1080/10810730.2013.829137

Roter, D. L., Frankel, R. M., Hall, J. A., & Sluyter, D. (2006). The expression of emotion through nonverbal behavior in medical visits. Mechanisms and outcomes. *Journal of General Internal Medicine, 21*(Suppl. 1), S28–S34. http://dx.doi.org/10.1111/j.1525-1497.2006.00306.x

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776. http://dx.doi.org/10.1037/0012-1649.27.5.763

Schirillo, J. A., & Stone, E. R. (2005). The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism. *Risk Analysis, 25,* 555–566. http://dx.doi.org/10.1111/j.1539-6924.2005.00624.x

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127,* 966–972. http://dx.doi.org/10.7326/0003-4819-127-11-199712010-00003

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science, 15,* 322–325. http://dx.doi.org/10.1111/j.1467-8721.2006.00461.x

Stanovich, K. E., West, R. F., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology, 31,* 811–826. http://dx.doi.org/10.1037/0012-1649.31.5.811

Stead, W. W., & Lin, H. S. (Eds.). (2009). *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: National Academies Press.

Taha, J., Czaja, S. J., Sharit, J., & Morrow, D. G. (2013). Factors affecting usage of a personal health record (PHR) to manage health. *Psychology and Aging, 28,* 1124–1139. http://dx.doi.org/10.1037/a0033911

Taha, J., Sharit, J., & Czaja, S. J. (2014). The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. *Journal of Applied Gerontology, 33,* 416–436. http://dx.doi.org/10.1177/0733464812447283

Tanius, B. E., Wood, S., Hanoch, Y., & Rice, T. (2009). Aging and choice: Applications to Medicare Part D. *Judgment and Decision Making, 4,* 92–101.

Tao, D., Yuan, J., & Qu, X. (2018). Presenting self-monitoring test results for consumers: The effects of graphical formats and age. *Journal of the American Medical Informatics Association, 25,* 1036–1046. http://dx.doi.org/10.1093/jamia/ocy046

van den Berg, E., Kloppenborg, R. P., Kessels, R. P., Kappelle, L. J., & Biessels, G. J. (2009). Type 2 diabetes mellitus, hypertension, dyslipidemia and obesity: A systematic comparison of their impact on cognition. *Molecular Basis of Disease, 1792,* 470–481. http://dx.doi.org/10.1016/j.bbadis.2008.09.004

Van Gerven, P. W. M., Paas, F., Van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology, 73*(Part 4), 489–505. http://dx.doi.org/10.1348/000709903322591208

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin, 132,* 249–268. http://dx.doi.org/10.1037/0033-2909.132.2.249

Zikmund-Fisher, B. J., Exe, N. L., & Witteman, H. O. (2014). Numeracy and literacy independently predict patients' ability to identify out-of-range test results. *Journal of Medical Internet Research, 16*(8), e187. http://dx.doi.org/10.2196/jmir.3241

Zikmund-Fisher, B. J., Scherer, A. M., Witteman, H. O., Solomon, J. B., Exe, N. L., Tarini, B. A., & Fagerlin, A. (2017). Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. *Journal of the American Medical Informatics Association, 24,* 520–528.

Appendix

Diabetes Dataset

Table A1
Parallel Results Across Message Formats (Cholesterol and Diabetes Datasets)

Cholesterol	Diabetes	Comparisons
Verbatim scores		
Main effect: Format		
$F(3, 140) = 1.7$, ns, $p > .10$, $\eta_p^2 = .02$.	$F(3, 140) = .6$, ns, $p > .10$, $\eta_p^2 = .005$	Same pattern.
Main effect: Risk level		Differences – Risk level not significant for diabetes.
$F(2, 280) = 7.0$, $p < .01$, $\eta_p^2 = .02$, high < borderline	$F(2, 280) = .3$, ns, $p > .10$, $\eta_p^2 = .001$	*Obs: Probably due to ceiling effects on memory.*
Interaction: Format × Risk Level		
$F(6, 280) = 1.3$, ns, $p > .10$, $\eta_p^2 = .01$.	$F(6, 280) = 1.6$, ns, $p > .10$, $\eta_p^2 = .02$	Same pattern.
Overall gist comprehension/global gist memory		
Main effect: Format		Some differences on direct comparisons, overall similar trends.
$F(3, 140) = 9.1$, $p < .001$, $\eta_p^2 = .06$; video enhanced > all other formats; verbally enhanced > standard	$F(3, 140)=9.8$, $p < .001$, $\eta_p^2 = .06$; enhanced formats > standard	*Obs: Probably due to ceiling effects on memory (diabetes)/reduced variance.*
Main effect: Risk level		
$F(2, 700) = 44.2$, $p < .001$, $\eta_p^2 = .08$; high > low > borderline.	$F(2, 700) = 17.1$, $p < .001$, $\eta_p^2 = .03$; high > low and high > borderline.	Low risk not different than borderline. But similar trend.
Main effect: Probe position		
$F(1, 700) = 43.1$, $p < .001$, $\eta_p^2 = .04$, after > before.	$F(1, 700) = 18.9$, $p < .001$, $\eta_p^2 = .02$; after > before	Same pattern.
Format × Risk interaction		Similar patterns.
$F(6, 700) = 4.6$, $p < .001$, $\eta_p^2 = .03$	$F(6, 700) = 2.7$, $p < .01$, $\eta_p^2 = .01$	
Low risk	Low risk	With some minor differences on direct comparisons.
$F(3, 140) = 5.1$, $p < .01$, $\eta_p^2 = .10$; video > graphic enhanced	$F(3, 140) = 4.2$, $p < .01$, $\eta_p^2 = .08$; video > standard and verbally enhanced > standard	
Borderline risk	Borderline risk	
$F(3, 140) = 6.8$, $p < .001$, $\eta_p^2 = .13$, video > standard; verbally enhanced > standard.	$F(3, 140) = 6.5$, $p < .001$, $\eta_p^2 = .12$, enhanced formats > standard	
High risk	High risk	
$F(3, 140) = 2.0$, ns, $p > .10$, $\eta_p^2 = .04$	$F(3, 140) = 1.7$, ns, $p > .10$, $\eta_p^2 = .03$	
Affective responses		
Main effect: Risk level		Same patterns.
Positive responses	Positive responses	
$F(2, 420) = 399.7$, $p < .001$, $\eta_p^2 = .66$	$F(2, 420) = 495.9$, $p < .001$, $\eta_p^2 = .70$	
Negative responses	Negative responses	
$F(2, 420) = 303.2$, $p < .001$, $\eta_p^2 = .59$	$F(2, 420) = 459.6$, $p < .001$, $\eta_p^2 = .69$	
Main effect: Format		
Positive	Positive	
$F(3, 420)=1.1$, ns, $p > .10$, $\eta_p^2 = .01$;	$F(3, 420) = 1.6$, ns, $p > .10$, $\eta_p^2 = .01$	
Negative	Negative	
$F(3, 420) = 1.6$, ns, $p > .10$, $\eta_p^2 = .01$	$F(3, 420) = 1.5$, ns, $p > .10$, $\eta_p^2 = .01$	
No interactions		
Positive	Positive	
$F(6, 420)=1.8$, ns, $p > .10$, $\eta_p^2 = .02$;	$F(6, 420) = 1.7$, ns, $p > .10$, $\eta_p^2 = .02$	
Negative	Negative	
$F(6, 420) = 1.5$, ns, $p > .10$, $\eta_p^2 = .02$.	$F(6, 420) = 1.2$, ns, $p > .10$, $\eta_p^2 = .02$.	

(*Appendix continues*)

Appendix (*continued*)

Cholesterol	Diabetes	Comparisons
Risk perception		
Main effect: Format	Main effect: Format	Similar patterns.
$F(3, 420) = 7.5$, $p < .001$, $\eta_p^2 = .05$; graphic > verbally enhanced	$F(3, 420) = 7.8$, $p < .001$, $\eta_p^2 = .03$; graphic > verbally enhanced	
Format × Risk interaction	Format × Risk interaction	
$F(6, 420) = 3.2$, $p < .001$, $\eta_p^2 = .04$.	$F(6, 420) = 3.3$, $p < .01$, $\eta_p^2 = .03$.	
Low risk	Low risk	
$F(3, 140) = 6.0$, $p < .001$, $\eta_p^2 = .11$, graphic > verbally enhanced and video, graphic = standard.	$F(3, 140) = 3.9$, $p < .001$, $\eta_p^2 = .08$, graphic > verbally enhanced and video, graphic = standard.	Difference:
Borderline risk	Borderline risk	Borderline risk: Marginally significant. Same trend.
$F(3, 140) = 3.0$, $p < .05$, $\eta_p^2 = .06$; graphic > verbally enhanced, graphic = video = standard.	$F(3, 140) = 2.4$, $p < .10$, $\eta_p^2 = .05$; graphic > verbally enhanced.	*Obs: Probably due to ceiling effects on memory (diabetes)/reduced variance.*
High risk	High risk	
$F(3, 140) = 2.7$, ns, $p > .10$, $\eta_p^2 = .06$.	$F(3, 140) = .9$, ns, $p > .10$, $\eta_p^2 = .02$.	
Medication attitude		
Main effect: Format	Main effect: Format	Same patterns.
$F(3, 420) = 4.9$, $p < .001$, $\eta_p^2 = .03$, Bonferroni comparisons were not significant. Numeric trend (graphic > other formats).	$F(3, 420) = 4.4$, $p < .001$, $\eta_p^2 = .03$, Bonferroni comparisons were not significant. Numeric trend (graphic > other formats).	Difference:
Format × Risk interaction	Format × Risk interaction was not significant	Interaction was not significant.
$F(6, 420) = 2.7$, $p < .05$, $\eta_p^2 = .04$	$F(6, 420) = .5$, ns, $p > .10$, $\eta_p^2 = .01$.	
Low risk (graph > all message formats)		
Behavioral intention		
Main effect: Format		
Intention to take medication, $F(3, 420) = 3.9$, $p < .001$, $\eta_p^2 = .03$, Bonferroni comparisons were not significant. Numeric trend (graphic > other formats).	$F(3, 420) = 2.6$, $p < .05$, $\eta_p^2 = .02$, Bonferroni comparisons were not significant. Numeric trend (graphic > other formats).	Same patterns.
Intention to exercise	Intention to exercise	
$F(3, 420) = 2.0$, ns, $p > .10$, $\eta_p^2 = .01$	$F(3, 420) = .8$, ns, $p > .10$, $\eta_p^2 = .01$ Intention to change diet	
Intention to change diet		
$F(3, 420) = 2.1$, ns, $p > .10$, $\eta_p^2 = .02$.	$F(3, 420) = .7$, ns, $p > .10$, $\eta_p^2 = .01$.	
Message satisfaction		
Main effect: Format		
$F(3, 420) = 27.0$, $p < .001$, $\eta_p^2 = .16$. (graphic > all formats; video > verbally enhanced and standard)	$F(3, 420) = 27.1$, $p < .001$, $\eta_p^2 = .16$. (graphic and video > verbally enhanced and standard).	Similar patterns. However, for diabetes video and graphic are equally preferred in comparison with standard and verbally enhanced formats.
Main effect: Risk level	Main effect: Risk level	
$F(2, 420) = 3.7$, $p < .05$, $\eta_p^2 = .02$; follow-up Bonferroni comparisons not significant; trend: higher satisfaction on high risk levels.	$F(2, 420) = 3.5$, $p < .05$, $\eta_p^2 = .01$; follow-up Bonferroni comparisons not significant; trend: higher satisfaction on high risk levels	

(*Appendix continues*)

Table A2

Correlation Matrix for Variables, Diabetes Dataset (N = 144)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Age	—	.08	−.21**	−.14†	−.10	−.31***	−.44***	−.19*	−.11	−.09	−.19*	−.30***	−.30***	−.43***	−.43***	−.28***	−.23**	.11
2. Gender		—	.09	−.09	.01	−.05	−.12	.12	.23**	−.01	.10	−.22**	−.07	−.07	−.12	−.03	−.01	.02
3. Education			—	.45***	.33***	.33***	.22**	.19*	.36***	−.05	.16†	.22**	.22**	.23**	.25***	−.06	−.05	−.14†
4. Verbal ability construct				—	.21**	.41***	.25***	.18*	.23**	−.11	.15†	.37***	.36***	.21*	.23**	.07	.10	−.35***
5. Self-reported health					—	.35***	.24***	.07	.45***	.14†	.20*	.29***	.30***	.22*	.17*	.02	.04	−.06
6. Health literacy (STOFHLA)						—	.38***	.10	.34***	.04	.28***	.26***	.38***	.26***	.26***	−.06	.08	−.26***
7. Fluid ability construct							—	.20*	.24***	.15†	.14†	.28***	.22**	.24***	.27***	.11	.11	−.11
8. Objective numeracy								—	.33***	−.01	.13	.12	.16†	.18*	.20*	.04	.05	.00
9. Subjective numeracy									—	.02	.20*	.15†	.25***	.10	.11	−.07	−.05	−.05
10. Verbal memory (A1C)									—	.09	−.01	.01	.10	.03	.04	.06	.04	
11. Gist Comp. (Before)											—	.25***	.38***	.34***	.29***	.24***	.25***	.01
12. Affective responses												—	.60***	.40***	.41***	.32***	.36***	−.05
13. Risk perception													—	.50***	.48***	.30***	.36***	.00
14. Medication attitude														—	.91***	.56***	.58***	−.20*
15. Behavior intention (Meds)															—	.60***	.60***	−.22*
16. Behavior intention (Exercise)																—	.95***	−.08
17. Behavior intention (Diet)																	—	−.10
18. Message satisfaction																		—
M	71.9	—	16.1	11.8	5.4	34.7	12.6	.2	4.5	1.0	.9	2.5	3.2	2.7	2.6	1.5	1.6	6.0
SD	7.7		2.5	5.6	1.3	2.6	2.7	.2	1.0	.1	.2	1.0	.9	1.3	1.4	1.2	1.2	2.3

Note. STOFHLA = Short Test of Functional Health Literacy in Adults.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Received October 2, 2017
Revision received September 6, 2018
Accepted October 28, 2018