<u>Original Paper</u>

Spelling Errors and Shouting Capitalization Lead to Additive Penalties to Trustworthiness of Online Health Information: Randomized Experiment With Laypersons

Harry J Witchel[1], PhD; Georgina A Thompson[1], BMBS; Christopher I Jones[2], CSci, CStat, PhD; Carina E I Westling[3], PhD; Juan Romero[4]; Alessia Nicotra[4], MBBS, PhD, FRCP; Bruno Maag[4], TGmeF, HFG; Hugo D Critchley[1], MBChB, DPhil, FRCPsych

[1]Department of Neuroscience, Brighton and Sussex Medical School, Brighton, United Kingdom

[2]Department of Primary Care and Public Health, Brighton and Sussex Medical School, Brighton, United Kingdom

[3]Faculty of Media and Communication, Bournemouth University, Bournemouth, United Kingdom

[4]Dalton Maag Ltd, London, United Kingdom

Corresponding Author:
Harry J Witchel, PhD
Department of Neuroscience
Brighton and Sussex Medical School
Trafford Centre for Medical Research
Brighton
United Kingdom
Phone: 44 1273 873 549
Email: h.witchel@bsms.ac.uk

Abstract

Background: The written format and literacy competence of screen-based texts can interfere with the perceived trustworthiness of health information in online forums, independent of the semantic content. Unlike in professional content, the format in unmoderated forums can regularly hint at incivility, perceived as deliberate rudeness or casual disregard toward the reader, for example, through spelling errors and unnecessary emphatic capitalization of whole words (online *shouting*).

Objective: This study aimed to quantify the comparative effects of spelling errors and inappropriate capitalization on ratings of trustworthiness independently of lay insight and to determine whether these changes act synergistically or additively on the ratings.

Methods: In web-based experiments, 301 UK-recruited participants rated 36 randomized short stimulus excerpts (in the format of information from an unmoderated health forum about multiple sclerosis) for trustworthiness using a semantic differential slider. A total of 9 control excerpts were compared with matching error-containing excerpts. Each matching error-containing excerpt included 5 instances of misspelling, or 5 instances of inappropriate capitalization (*shouting*), or a combination of 5 misspelling plus 5 inappropriate capitalization errors. Data were analyzed in a linear mixed effects model.

Results: The mean trustworthiness ratings of the control excerpts ranged from 32.59 to 62.31 (rating scale 0-100). Compared with the control excerpts, excerpts containing only misspellings were rated as being 8.86 points less trustworthy, those containing inappropriate capitalization were rated as 6.41 points less trustworthy, and those containing the combination of misspelling and capitalization were rated as 14.33 points less trustworthy (*P*<.001 for all). Misspelling and inappropriate capitalization show an additive effect.

Conclusions: Distinct indicators of incivility independently and additively penalize the perceived trustworthiness of online text independently of lay insight, eliciting a medium effect size.

XSL·FO
RenderX

Introduction

Trustworthiness of Online Health Information: Background, Context, and Importance

As of 2019, 90% of all British adults use the internet at least weekly [1], and as patients, they often search online for health information to solve their medical problems [2]; furthermore, they are likely to be influenced by the online information, including changing their health care decisions and their frequency of ambulatory care visits [3]. In a survey of 2200 adults with chronic health conditions in the United States who were active social media users, 57% used a health condition–specific website (eg, specializing in multiple sclerosis or rheumatoid arthritis) on a monthly basis, and 5% used such sites daily; half of the patients surveyed had asked a health-related question to others online within the previous 6 months, and 87% of those were seeking responses from other patients with the health condition [4].

In response to this explosion of unvetted potential sources providing online health care information that is acted upon, researchers, experts, and medical professionals have repeatedly expressed concern about the inaccuracy of online information and the limited ability of lay consumers to adequately assess its validity [2,5,6]. In particular, it has long been known that when lay users determine whether to use and trust online health care information, they are strongly influenced by nonmedical criteria that experts do not use [7-11]. Broadly, although academics favor a checklist approach of transparency criteria [12], the approach of nonexperts appears to be more variable and situation dependent, and it prioritizes factors such as ease of understanding and the attractiveness of graphic design [9]. More generally, in research on the factors that influence judgments of trust, all aspects of trustworthiness can be relational and depend on the type of person (or group being studied); major relational factors include accessibility, both cognitive as well as physical [13], and correctly accommodating language to the intended audience [14].

Understanding lay assessment of the trustworthiness of online health information is important because false online information presented to the general public, if believed, has the potential to undermine correct medical advice [15], to elicit unhealthy behavior [16], and to influence sociopolitical discourses on health care and other topics [17-19]. Sbaffi and Rowley's [20] recent review of how laypeople assess the trustworthiness and credibility of online health information concluded that so far, much less research has been performed to understand what interferes with trust (we describe these as penalties to trustworthiness [21]), compared with what causes trust.

Trust and credibility are closely related to one another and information quality, although there remains disagreement among researchers as to the exact definitions of these terms. In general, the definitions emphasize the likelihood of information use, believability, reliability, and dependability [20,22]. Correspondingly, there is little evidence to suggest that the general population reliably make fine conceptual distinctions between trustworthiness and credibility of information.

Online health support often is divided into seeking information or emotional reassurance and can be gender specific (eg, prostate cancer vs breast cancer) and person specific [23,24]. Trustworthiness remains pertinent to online emotional reassurance and sharing, as shown by the many occurrences of large-scale hoaxes designed to manipulate emotions [8]. For example, between 2010 and 2011, a Macmillan cancer forum was inundated with posts in response to an elaborate hoax by a purported mother about her 6-year-old daughter struggling with cancer. On the exposure of the hoax (perpetrated by a lonely 16-year-old girl), many users of the forum—who had formed close online relationships with the supposed mother—refused to believe it had all been a hoax [25]. In light of the range of uses that online health information has for the general public, we contend that it is important to build knowledge of the factors that influence how trust and its absence are formed online beyond source analysis and fact checking. As this study shows, both linguistic and metalinguistic factors affect how trustworthiness is instinctively rated.

Theoretical Underpinnings for Factors Influencing Trustworthiness

The range of elements that influence online trustworthiness includes issues related to the sources (eg, author identification and the absence of advertising), issues related to the content (eg, a date stamp and inclusion of medical evidence), issues related to design and engagement (eg, inclusion of images), and issues affecting all the above (eg, the absence of typographical errors) [20,26]. For some time, credibility has been subdivided into aspects such as source credibility, message credibility, and media credibility, which strongly influence each other [22]. All such aspects of credibility include accessibility, which can be relational and cognitive as well as physical [13]. More recently, Sun et al [6] have divided the elements that influence consumer evaluation of online health information into 25 *criteria* and 165 *quality indicators*; criteria are rules that reflect notions of value and worth (eg, expertise and objectivity), whereas quality indicators are properties of information objects to which criteria are applied to form judgments (eg, the owner of the website and inclusion of statistics). In line with Diviani et al [5], Sun et al [6] have suggested that indicators can be positive or negative (in terms of trustworthiness), and that consumers' perceived online health information quality could conceivably be measured by a small set of core dimensions (ie, a few groups of criteria might incorporate many of the quality indicators that explain most of the trustworthiness judgments).

So far, Lederman et al [8] have proposed a five-category model that highlights verification processes via the comparison of different websites or other online statements. This model includes argument quality, source credibility, source literacy competence, and crowd consensus. An extended six-category model [26] proposes that the lay reader may assess some or all of the following: reputation, endorsement, consistency with other sources, self-confirmation (agreement with the reader's own opinions), persuasive intent, and expectancy violation. For this research, we have adopted Lederman et al's [8] model of credibility; however, the choice among these models for this research is moot because all the models [5,6,20,22,26] are concordant with the idea that spelling errors will detract from

judgments of trustworthiness (ie, they are negative quality indicators).

Incivility, Literacy Competence, and Errors of Writing Mechanics

Although institutionally produced websites and curated online health content (cf. [6]) will usually be both grammatically correct and civil, the responses by the general public may be uncivil [27], and patient-authored text is known to have many misspellings [28] (and so do discharge summaries written by doctors [29]). Inappropriate capitalization (including inappropriate capitalization of entire words, sometimes called online *shouting*) and misspelling are examples of errors in literacy competence [8] and writing mechanics [30]; writing mechanics is defined as elements of a language that only manifest when communication is in written form. Both inappropriate capitalization and misspelling have been highlighted by qualitative investigations as explicit criteria used by lay readers in judgments of online credibility [8,31]. The rationales given to explain why these two error types undermine trustworthiness are that either (1) the errors imply a lack of intelligence (expertise, ability, authority, and education) [8,32,33] or (2) that they suggest a lack of motivation (objectivity, attention to detail, and conscientiousness) to be trustworthy [34,35]. The term incivility is used to describe this latter lack of motivation or effort to make statements that are compliant with rules of communication. Hargittai [33] distinguishes between spelling errors (errors because of the levels of education or social inequalities) and typographical errors (errors because of accidentally hitting the wrong keys on the keyboard). The fact that these errors are not corrected during proofreading is another form of incivility. Incivility implies a lack of respect for the reader, the platform, and the rules of social exchange [36,37], and it is fundamentally relational. Quantitative research on incivility in the mainstream World Wide Web demonstrates that civil statements are rated as more trustworthy and influential than uncivil ones [38,39].

Integration Versus Heuristics: Lay Judgments Based on Multiple Cues

The reader who must judge unvetted online health information is faced with a wealth of cues that indicate the degree of trustworthiness, and those cues may have contradictory effects (eg, a cogent message that is misspelled). There are three broad theories for how individuals (both lay and expert) make judgments based on multiple cues. The rational approaches are represented by the information integration model [40], in which an individual accounts for all the different pieces of information by a complex (but often subconscious) mathematical process that is usually based on addition, multiplication, or averaging; extensive observations of integration in judgments occur across cultures and individuals. This is the process that Sun et al [6] allude to when they propose that the consumer integrates the relevant trustworthiness criteria and quality indicators in a "complex cost-benefit analysis." In 2003, Fogg's [41] Prominence-Interpretation Theory proposed explicit mathematical relationships for how to predict the effects of multiple factors on credibility, but the theory never detailed how to measure the relevant quantities independently.

Computational models typically assume that the elements of incivility (eg, inappropriate capitalization) act together either additively or nonlinearly [42-44], although hypothesis-led proof for this assumption is minimal. The elucidation of this integration is only just starting in the literature [45], and the relative importance of each indicator in this intuitive cost-benefit analysis remains unknown; the relative values for each cue may be elucidated empirically by statistical methods such as regression, but it is unlikely that these values would be explicit or transparent in the minds of most lay decision makers. This is an area that needs to be further researched.

An alternative theory to cost-benefit analysis (ie, for how judgments are made based on multiple cues) is the process-level cognitive perspective [46]; this has been made famous by the heuristics and biases research program from behavioral economics [47]. Heuristics are rapid cognitive strategies or shortcuts (either explicit or subconscious) formulated as practical, bounded rational decision systems for multiple cues that can be more transparent than complex cost-benefit analyses, for example, hierarchical lexicographic decision models [48]. In a take-the-best judgment [49], first, a single cue (the most important one) is searched for in the environment and considered independently of all others, and if a tie or no clear result occurs, then the second most important cue is sought and decided upon, and so on. For example, when deciding whether you have the right of way when driving your car through an intersection, first, you follow the signal of any policemen present, and only if there are no police present, do you seek and consider a traffic light (including a temporary traffic light for construction), and only if there are no traffic lights present, do you then consider static road markings and the positions of the other cars.

In biased heuristics, only a limited subset of information (often only one cue) is used to make a fast and ecological judgment outside of conscious awareness [46]; when biased, these heuristics are used to support preferred or preconceived outcomes. With biased heuristics, the prioritization of the cue, and even the cue's basic validity for the judgment being made, is dubious. Such biased heuristics have been used to explain seemingly irrational preferences that individuals make in situations involving slot machines and organ donation [50]. Examples of biased heuristic processes include representativeness, where some cues are weighted disproportionately compared with their real representativeness—and availability—where a cue that is easily recalled (such as occurs with recency and news) determines the judgment. None of these judgment models so far proposed have explicitly assessed specific issues within literacy competence.

Determining the Criteria and Weighting of Judgments Based on Multiple Cues

A key feature of heuristics is that they are typically made subconsciously, and that post hoc explanations for such choices are often self-serving justifications or rationalizations [51]. That is, in the case of heuristics, the decision maker does not have privileged information on how the judgment was made, and furthermore, the decision maker can be wrong about themselves [52]. For example, university students have been shown to greatly overestimate how much they actually learned from

excellent lecturers (conflating it with how much they feel they learned, which is discounted by effort and exertion) [53]. This creates potential issues for researchers, in which insight-based techniques (both qualitative interviews and quantitative questionnaires) can lead to judgments and explanations of causes that are inaccurate because of demand characteristics or social desirability [54,55]. It has long been known that when evaluating credibility (eg, website privacy policy), the importance of factors that lay individuals say they would use to make their judgment diverge from the factors they are observed to use [41]. Recent sophisticated measurements suggest that although multiple factors can interact when quantifying perceived credibility (of information for an online health forum), these interactions do not support Petty and Cacioppo's Elaboration Likelihood Model (ELM) [56]. Furthermore, there is a gap in the literature for any data showing how spelling errors interact with other writing mechanics violations that might also affect trustworthiness [57]. This suggests that lay insight into the influences on their perceptions of trustworthiness are imperfect, and that research on trustworthiness should be supplemented by approaches that do not rely on such insight.

To avoid insight bias from our lay participants, the approach of this study is to compare and contrast *marginal differences* in penalties to trustworthiness elicited by different combinations of literacy or writing mechanics violations. Our experiments on changes in marginal trustworthiness were specifically organized so that the participating healthy volunteers were unaware that the experiment tested the effects of capitalization or misspelling *per se* (with full institutional ethical approval). We did not explicitly ask lay individuals for their beliefs regarding how their process for judging message credibility incorporates quality indicators relating to source credibility and media credibility. That is, instead of asking directly, "How much less would you trust a web post that is misspelled?" we simply presented the participants with some posts that had misspellings and asked the same question as usual, "How trustworthy do you find this information?" while subtly varying misspelling and capitalization (ie, *shouting text*, not acronyms or beginnings of sentences).

Research Questions

The research questions (RQs) were as follows:

- RQ1: Do errors in writing mechanics and incivility lead to marginal changes (ie, without signposting) in judgments for message trustworthiness?
- RQ2: Are the marginal effects of incivility, such as with inappropriate capitalization, on trustworthiness judgments quantitatively comparable with the known effects of writing mechanics such as spelling errors?
- RQ3: How, if at all, do the effects of writing mechanics (eg, spelling errors) and incivility (eg, inappropriate capitalization) integrate? Is there a ceiling effect or a binary effect in which once a message source is judged as incompetent, no further trustworthiness penalty is added to the judgment (similar to a take-the-best heuristic) [49]? Or

is there some additive, multiplicative, synergistic, or otherwise integrative function that increases the penalty on the judgment to a new higher level when both cues are present [40]?

Methods

### Participants, Recruitment and Ethical Approval

The project was approved by our local ethics committee (Brighton and Sussex Medical School's Research Governance and Ethics Committee, University of Sussex, approval 16/044/WIT). All experiments were performed according to the Declaration of Helsinki. All individuals provided informed consent via a welcome page in each online study. Participants were recruited via the Prolific website. We specified that the recruitment should focus on UK-based members of the public. Participants consented to participate with the understanding that the research concerned *responses to online text*; none of the advertising, web URLs or experimental information to the participants mentioned that the experiment was related to graphics, formatting, incivility, and spelling. This feature of the advertising was approved by our ethics committee, not least because the paragraphs were not considered misleading or potentially emotionally adverse.

Study Design Process

The study design was a confirmatory, cross-sectional experiment with a balanced incomplete block design; it was a randomized experiment with lay participants, each of whom experienced only a limited number of the possible options. The total number of experimental excerpts that were tested in the entire cohort was 36; there were 9 excerpts, each in 4 possible versions: no errors, inappropriate capitalization only (Caps only), misspelling only, or the combination of both errors. The stimulus texts (and the versions) are shown in the Multimedia Appendix 1. There were two additional paragraphs that always appeared as the first two stimulus texts, which were training stimuli. The training stimuli were not labeled as being different from other stimuli in any way and were never included in the statistical analyses. The only purpose of the training stimuli was to allow participants to familiarize themselves with the rating task and with the range of the trustworthiness scale: training stimulus 1 was quite believable (mean trustworthiness rating 54.33, SD 23.39; n=301), and training stimulus 2 was less plausible (mean trustworthiness 40.20, SD 25.80; n=301; $P<.001$, paired t test).

Each participant experienced and rated only 9 (of the possible 36) experimental excerpts, and those 9 included exactly one version of each excerpt, and among those 9, that participant would experience a mix of different error types (Figure 1). For example, participant number 001 experienced the capitalization-only version of excerpt E07 and then the *both errors* (capitalization and misspelling) version of excerpt E02. This double randomization prevented any participant from seeing the same excerpt twice.

Figure 1. Experiment design: what each participant experiences.
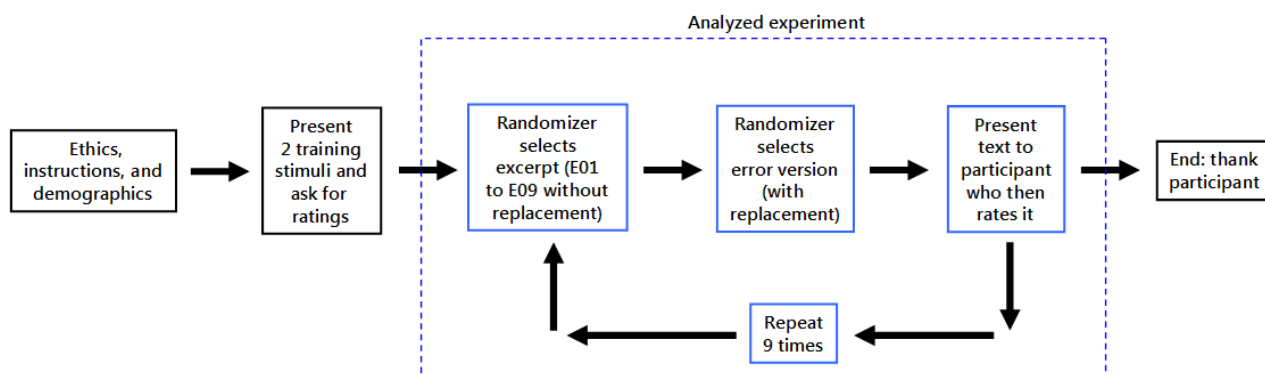
The goal was to present to participants a set of coherent text excerpts of similar lengths (70-100 words) on the topic of multiple sclerosis, each excerpt being a coherent answer to a question. The rationale for presenting excerpts in a question-then-answer format was that it was possible to ask whether readers trusted the advice enough to believe it or act upon it. To make these excerpts, a range of comments were found in the public domain (Table 1); these texts often had to be edited substantially to fit within the word count or to avoid explicitly endorsing commercial products (see Multimedia Appendix 2 for a comparison of the original and presented texts).

Initially, the excerpts and the presentation system were tested online by a small group of testers, who then provided verbal feedback on the test to the experimental team. After that, a cohort of 40 participants was recruited online to test the paragraphs and demonstrate that the paragraphs elicited similar standard deviations in ratings (19-28 units out of 100) and elicited a wide range of mean trustworthiness ratings (from 20 to 70). These excerpts were to be presented either as they were (*no errors* or the negative control) or in one of the three error versions stated above.

Table 1. Sources of excerpts on multiple sclerosis.

Code	Brief topic description	Website	Length
L01	Numerous artificial sweeteners	blogspot.com	78
L02	Hoax about artificial sweeteners	quora.com	81
E01	Triggers of the immune system	healingchronicles.com	81
E02	Programmer's intelligence	dailystrength.org	89
E03	Epstein Barr virus	medicaldaily.com	92
E04	Avonex patient	quora.com	74
E05	Up there in risk	quora.com	96
E06	Mental exercises	dailystrength.org	85
E07	Vitamin D	quora.com	100
E08	Small risk of Progressive Multifocal Leukoencephalopathy	my-ms.org	90
E09	Half of all people	ms.pitt.edu	71

Text Interventions

For the error versions of the texts (inappropriate capitalization or misspelling), we wanted to include 5 of the relevant errors per excerpt (a total of 10 errors for the combination of both errors version), with the errors spread throughout the excerpt (rather than bunched together). In excerpts where inappropriate capitalization was required, there would be 5 *sets* of words or phrases. Normally, a set was 1 or 2 words, although one of the 5 sets had to be a 4-word series. The priorities for selecting words to capitalize were (in this order) as follows:

- Adverbs (especially those suggesting extremity such as *very* or *never*)
- Judgments (*rubbish, hopeless,* and *horror*)

- Strong emotions (*worried* and *angry*)
- Words implying danger (*fatal* and *death*)
- Amounts (*all, every*, and *ten*)
- Adjectives (rather than nouns)
- Action verbs (especially gerunds)
- Conjunctions (*and*)

To verify that each word that we capitalized was naturally capitalized on the web, we analyzed words that were capitalized online on Twitter. We used the Claritin corpus, which is a crowdsourced data set of all the Twitter tweets that contained the word *claritin* in the month of October 2012 [58]. This corpus has some 4900 tweets, and we used MATLAB (MathWorks) to find all the words that were in all capitals (which did not have a hashtag or an at-sign in them); this resulted in a list of 343

capitalized words (Multimedia Appendix 3), many of which were short words, acronyms, and internet memes. From this list of words spontaneously capitalized on the web, we selected words in our excerpts to capitalize.

The rationale for how we selected words to misspell was that misspellings should be quite noticeable, and that the meaning of the words should remain clear to the reader even when misspelled. We avoided homonyms and words that looked plausibly English when misspelled. To make sure that misspelled words were noticeable, short words were preferred, or we placed the misspellings in the first syllable of a multisyllable word. In addition, one of the misspelled words had to be in the first 5 words of a paragraph. The misspelled word had to be completely understandable (in the absence of other words or context) even when misspelt. Thus, a misspelled word with missing or added letters should be pronounceable in English (eg, *yu* plainly means *you*). The types of misspellings were as follows:

- Swap one letter for another letter that is next to it on a qwerty keyboard (*pisitive*)
- Double a consonant (*esstimate*)
- Double a vowel or add an extra vowel (*theere*)
- Leave out a vowel (*expsure*)

To verify that each word that we misspelled was naturally misspelled on the web, we searched for the misspelled word along with the words *health* and *forum*; if we could not find at least two examples of a misspelling on online health forums, we did not use it. A complete listing of the misspellings and where we found them online is in Multimedia Appendix 4.

Study Delivery

The study was presented to participants using the Qualtrics portal, which allows for a wide range of question types and keeps track of answers and total response time. A full description of the survey in the Checklist for Reporting Results of Internet E-Surveys format [59] is included as part of the Multimedia Appendix 5 for this paper. The web-based study welcome page explained in brief what the study was about and what it entailed (estimated 8 min participation time, including reading the ethics and filling in demographics), the ethics of the study (include

the ability to withdraw at any time), and a brief complaints procedure. The ethics page explicitly excluded participants aged younger than 18 years or those from vulnerable populations. A pointer to a full-length participant information sheet (3 pages) was shown; the welcome/ethics page had an "I agree" button at the bottom. After the welcome page, participants filled in a brief multiple-choice demographics page, which included questions on gender, age, field of work (eg, health care, agriculture, and retired), and familiarity with the English language/Roman alphabet. All demographic questions included an option for "rather not say." After the demographics page, participants saw the instructions page and then were launched into the experimental ratings pages.

Each ratings page consisted of a short excerpt of text (which was randomized as to whether or not it had the spelling or capitalization errors), followed by a horizontal slider for rating how *trustworthy* the participant found the statement to be; the slider had anchors of *completely untrustworthy* (Figure 2, left) and *completely trustworthy* (Figure 2, right). Although the data collection was numerical (0-100, left to right), there were no numerical cues or tic marks seen by the participants. The instructions to the participants for the trustworthiness ratings were, "If you find something trustworthy, you would be prepared to act upon it; an untrustworthy statement you would ignore, and a rating in the middle represents information where you would want more proof or confirmation that it is correct." As explained in the instructions to participants, each stimulus excerpt was written as if it was an answer to a question written on an unmoderated health forum, with a specific focus on multiple sclerosis. Multiple sclerosis was chosen as a topic because the information was obviously important, but healthy participants would be unfamiliar with the veracity of each statement; thus, we predicted that the trustworthiness ratings would be more susceptible to nonverbal cues. The questions were as follows: (1) Is multiple sclerosis preventable? (2) How risky is Tecfidera as a treatment for multiple sclerosis? and (3) Does multiple sclerosis decrease intelligence/IQ? The nominal responses to these questions were the experimental stimulus excerpts being rated.

Figure 2. Unnumbered horizontal slider for trustworthiness ratings.

completely untrustworthy   completely trustworthy

Study Design, Analysis, and Statistics

Sample Size

To detect a difference in rating scores between two groups (ie, no errors vs misspellings, inappropriate capitalization, or both), with 80% power at the 5% significance level, assuming the standard deviation within each group is 25 and the difference between groups is 15 (equivalent to a medium effect size of 0.517), would require 60 participants per group (120 total). Each

participant was asked to rate 9 text excerpts, randomly divided between the 4 conditions, and assuming an intraclass (ie, within participant) correlation coefficient of 0.185 gives a design effect of 2.48. The product of the design effect and the sample size for a nonrepeating experiment is $2.48 \times 120 \approx 300$ total.

Modeling

Linear mixed effects (LME) models were fitted in Stata version 16.0 using the *mixed* command. Residuals were checked for

normality and homoscedasticity at the cluster and individual levels. Where the assumption of residual homoscedasticity was not appropriate, robust standard errors were used to allow for the calculation of appropriate 95% CIs and *P* values [60].

Results

Differences Between the Paragraphs

We ran an experiment in which we gathered data from 301 volunteers, each making 9 experimental observations (2709 ratings in total). We estimated that this task (rating 11 excerpts plus instructions and demographics) would take 8 min; in fact, it took a mean time of 6.25 min (mean 375.74 seconds, SE of mean 12.94 seconds). The median trustworthiness ratings of each of the excerpts (E01 to E09) in the negative control condition (ie, without any errors or incivility) are shown in the box and whisker plot in Figure 3. On each box, the red line in the center indicates the median, and the bottom and top edges

of the box indicate the first and third quartiles, respectively. When the notches for two boxes do not overlap, this implies that the true medians do differ with 95% CI. The whiskers extend to the most extreme data points not considered outliers, whereas outliers are shown using red plus signs; outliers are any points that fall more than 1.5 times the IQR away from the main box. Excerpt 01 generally elicits low ratings of trustworthiness (median 27.5), whereas excerpts 08 and 09 generally elicit high ratings of trustworthiness (medians 62 and 63, respectively). As illustrated by the nonoverlapping notches, these two sets of excerpts elicit significantly different ratings of trustworthiness, which has been true in all our previous cohorts rating these excerpts for trustworthiness (data not shown). Excerpts 03, 04, 05, and 06 all elicited median ratings in the middle range from 45 to 55, whereas excerpt 02 is a transitional excerpt between E01 and the middle range and E07 is a transitional excerpt between the middle range and the most trustworthy excerpts.

Figure 3. Trustworthiness ratings of the different excerpts in the negative control (no errors) condition. For each box, N=75.
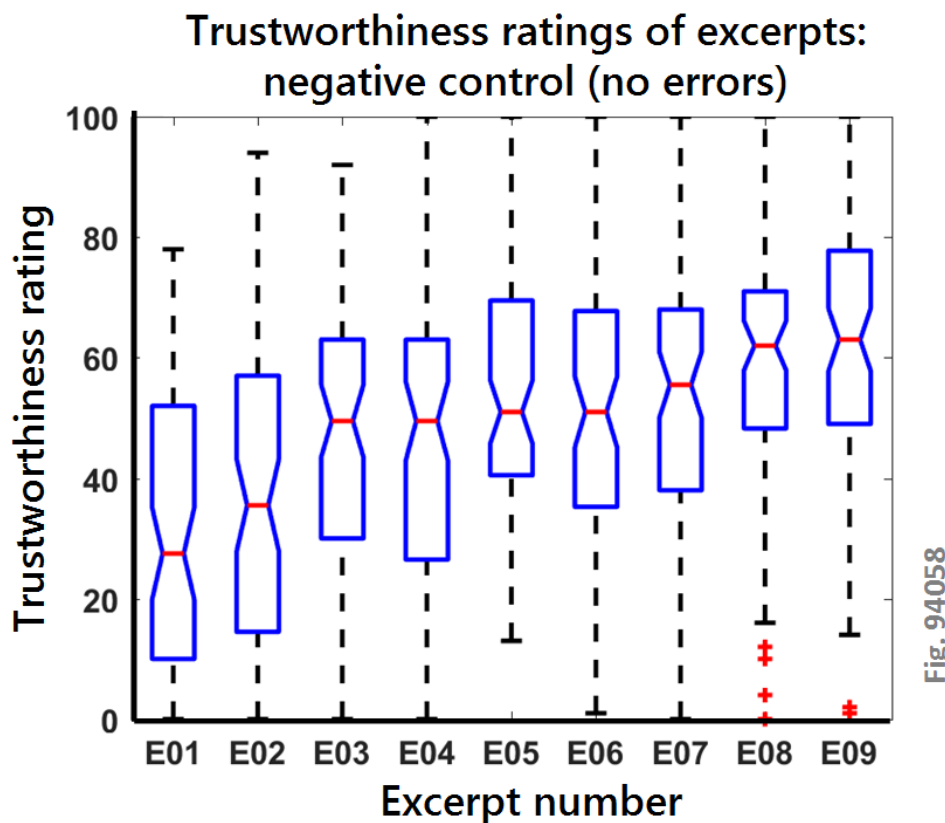
Cumulative Distributions Shifted Left by Errors

Figure 4 shows how the errors in writing mechanics and incivility led to changes in the cumulative probability distributions for each of the excerpts. As expected, for each excerpt, compared with the negative control with no errors (black, thick dashed line), all three alterations (inappropriate capitalization (blue, thin dot-dash line), misspelling (magenta, thin dotted line), and both errors together (red, thick solid line) led to a decrease in the ratings of trustworthiness (ie, a

leftward-upward shift in the curve). For most of the excerpts, at most points on the cumulative distribution curve, the combination of errors (both inappropriate capitalization and misspelling) led to decreases in trustworthiness ratings (ie, a shift left and up) compared with either of the single errors. That is, the lines for inappropriate capitalization only (blue, thin dot-dash) and misspelling only (magenta, thin dotted) fall between the thick black dashed line (no errors) and the red solid line (both errors); this result suggests that the two errors together have an additive or integrative effect.

Figure 4. Cumulative probability distribution plots for each excerpt (E01 to E09), comparing the alternative writing errors (lighter colored lines) with negative control (thick, black dashed line).
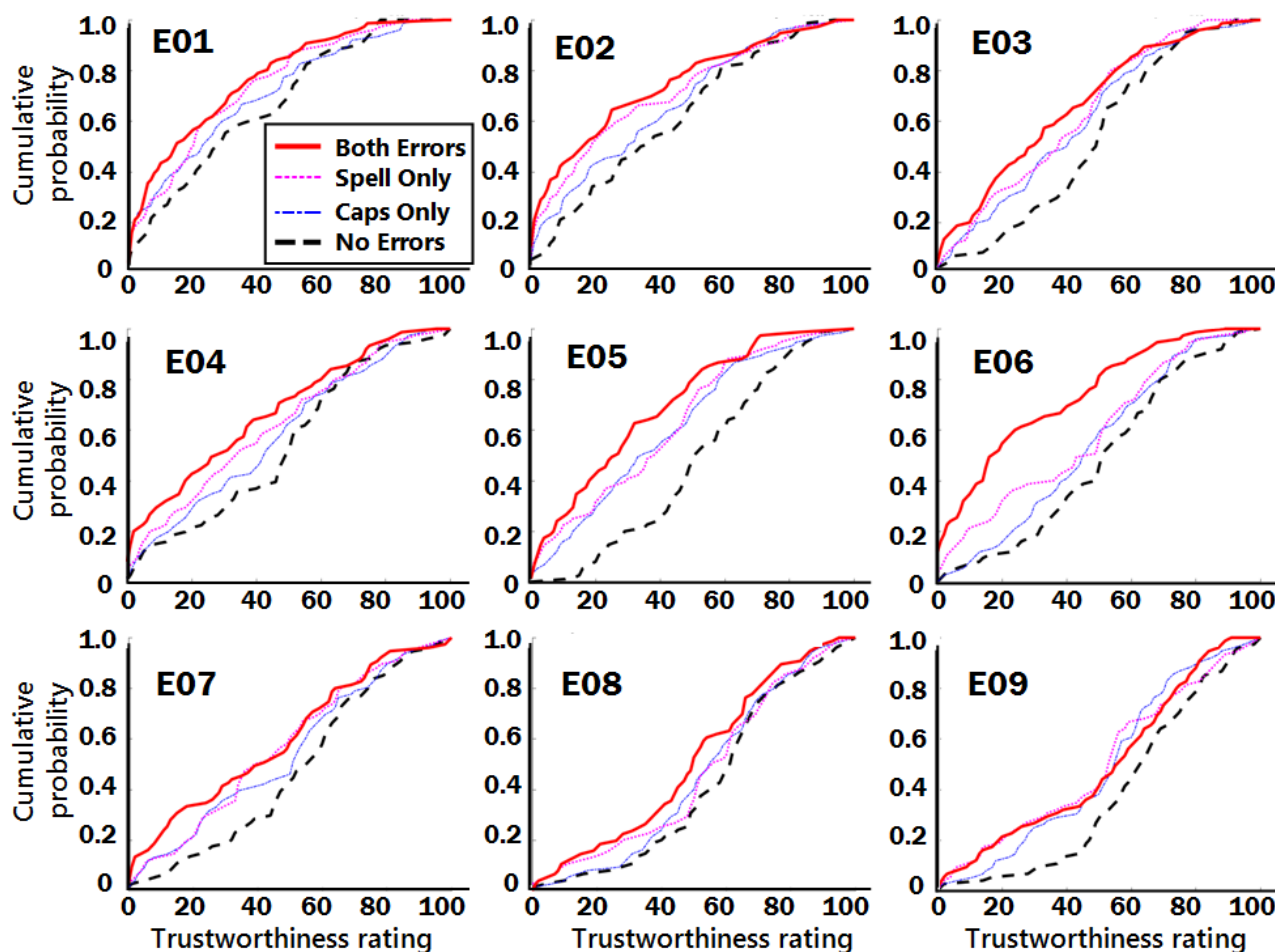
Mixed Linear Effects Model With Four Conditions of Alteration

We tested these data in a mixed linear effects model (model 1), where trustworthiness rating was the dependent variable, whereas alteration (no errors/misspellings/inappropriate capitalization/both misspellings and inappropriate capitalization) and excerpts (1-9) were included as fixed effects. A random effect for participant was included in the model to account for clustering of observations within volunteers. The reference group/condition for this model was E05, with no literacy errors. This model (and the following model) was calculated with robust standard errors [60] to allow for the heteroskedasticity of the residuals. The results of model 1 are shown in Table 2. The intracluster correlation (correlation within the individuals) coefficient estimate for model 1 is 0.334 (95% CI 0.287 to 0.384).

There is strong evidence against each of the null hypotheses that inappropriate capitalization only (result 1), misspelled only (result 2), and both errors (result 3) do not affect trustworthiness ratings compared with the negative control group; stated positively, our data suggest that there is a statistically significant penalty to trustworthiness for inappropriate capitalization (result 1), misspelling (result 2), and for both errors together (result 3). Inappropriate capitalization reduces trustworthiness ratings

by −6.41 (95% CI −8.96 to −3.86), and misspelling reduces trustworthiness ratings by −8.86 (95% CI −11.61 to −6.11). The effect on trustworthiness ratings of combining both inappropriate capitalization and misspellings together is −14.33 (95% CI −17.11 to −11.55), which appears to be an additive effect.

Our further analysis aimed to test whether there was likely to be either an additive or integrative effect [40] of combining inappropriate capitalization and misspelling. Such an effect should lead to a significantly larger trustworthiness penalty when both error types are combined compared with either error individually. To test for this, an alternative specification of the LME model of the same data was formulated (model 2). In model 2, the condition *both errors* was specified as an interaction between 2 binary variables, (1) misspelling and (2) capitalization errors, instead of considering all the errors as 4 conditions in a categorical independent variable. The specification of the rest of the model was identical to model 1. In this model (Table 3), the main effects and an interaction between them provide no evidence for an interaction effect between the 2 variables; that is, the main effects for inappropriate capitalization only and for misspelled only were as in the original model 1 (Table 2), whereas the coefficient for the interaction was not significantly different from zero. This supports the interpretation that the effects of the two error types are additive, rather than partially summative or synergistic.

Table 2. Mixed effect model 1 for trustworthiness rating, with errors and excerpts as fixed effects and a random effect for the clustering of data by participant.

Categorical variables	Coefficient	Robust SE	*P* value	95% CI
Alteration				
Caps only	−6.41	1.30	<.001	−8.96 to −3.86
Misspelled	−8.86	1.40	<.001	−11.61 to −6.11
Both errors	−14.33	1.42	<.001	−17.11 to −11.55
Excerpt				
E01	−11.94	1.71	<.001	−15.28 to −8.59
E02	−8.29	1.58	<.001	−11.38 to −5.19
E03	−0.70	1.56	.65	−3.77 to 2.36
E04	0.03	1.54	.98	−2.99 to 3.05
E06	1.80	1.62	.27	−1.39 to 4.98
E07	6.19	1.71	<.001	2.83 to 9.55
E08	14.75	1.53	<.001	11.75 to 17.75
E09	14.22	1.46	<.001	11.36 to 17.09
Constant	47.06	1.53	<.001	44.06 to 50.06

Table 3. Model 2: Alternatively specified linear mixed effect model with 2 binary variables for capitalization and spelling errors and an interaction term to account for combining both types of error (all unlisted values are identical to above).

Categorical variables	Coefficient	Robust SE	*P* value	95% CI
Alteration				
Caps only	−6.41	1.30	<.001	−8.96 to −3.86
Misspelled	−8.86	1.40	<.001	−11.61 to −6.11
Interaction	0.94	1.68	.58	−2.34 to 4.24

The output for model 1 shows the following comparisons: (1) capitalization only versus no errors, (2) misspelled only versus no errors, and (3) both errors versus no errors. Table 4 shows the additional comparisons between (4) capitalization only versus misspelled only, (5) both errors versus misspelled only, and (6) capitalization only versus both errors.

Comparison 4 indicates there is weak evidence (*P*=.06) against the null hypothesis of no difference between the effects of capitalization only and misspelled only. That is, misspelled only leads to a larger trustworthiness penalty by −2.45 (95% CI −5.02 to 0.12) compared with capitalization only. There is strong evidence against the null hypothesis of there being no difference between the effects of capitalization only versus both errors combined (comparison 6). Compared with capitalization only, the combination of both errors significantly reduces trustworthiness ratings by a further −7.92 (95% CI −10.28 to −5.56). Similarly, there is strong evidence against the null hypothesis of there being no difference between the effects of misspelled only and the combination of both errors (comparison 5). Compared with misspelling only, the combination of both errors leads to a further penalty to trustworthiness ratings of −5.47 (95% CI: −7.83 to −3.11).

Table 4. All possible alterations tested in between-group comparisons (model 1).

Comparison	Coefficient	Robust SE	*P* value	95% CI
(1) Caps only versus no errors	−6.41	1.30	<.001	−8.96 to −3.86
(2) Misspelled versus no errors	−8.86	1.40	<.001	−11.61 to −6.11
(3) Both errors versus no errors	−14.33	1.42	<.001	−17.11 to −11.55
(4) Caps only versus misspelled	2.45	1.31	.06	−0.12 to 5.02
(5) Both errors versus misspelled	−5.47	1.20	<.001	−7.83 to −3.11
(6) Caps only versus both errors	7.92	1.20	<.001	5.56 to 10.28

Statistically Significant Differences Between the Excerpts

In model 1, when compared with E05, the effect of the various excerpts' contents on trustworthiness ratings varies from –11.93 to 14.75 (a range of 26.68). This range is roughly twice as large as the effect of both errors (–14.33), suggesting that the errors in incivility and writing mechanics that we tested can together have an overall effect of nearly half of the effects of the content of the excerpts we tested.

Discussion

Original Contributions

This study sought to quantitatively determine how two different errors of writing mechanics (contributing to incivility) combine to penalize subjective ratings of trustworthiness in the medically relevant context of materials typical of an unmoderated online health forum. Using an LME model of a suitably powered study, we found that all three interventions (inappropriate capitalization, misspelling, and the combination of the two) were significantly different from the negative control (no added errors or incivility), which clearly answers RQ1. The data also show that (for these 70- to 100-word long excerpts about multiple sclerosis), the trustworthiness penalty for 5 instances of inappropriate capitalization was of a similar magnitude to the penalty for 5 instances of misspelling. Note that there was a trend for the penalty of misspellings to be larger, but as a generalized rule, the 2 are similar in magnitude, and the precise difference will depend on exactly how many words and which words are capitalized or misspelled. This finding answers RQ2. Finally, with a combination of different LME models, the data show that the combination of two different types of errors had a significantly greater trustworthiness penalty than either of the error types alone, and that the effect in this study was almost perfectly additive (RQ3); thus, the effects of the combination of errors were integrative [40] rather than a simplified heuristic such as take-the-best [48], multiplicative, or affected by ceiling effects in this study. This begins to answer the question recently posed of how spelling interacts with other quality indicators on credibility [57]. To the best of our knowledge, this is the first study that was specifically designed to test and quantify these kinds of specific additive effects on trustworthiness independently of the lay participants' insight.

We also showed that the stimulus excerpts that we designed are appropriate for studies that test the trustworthiness of online health information independently of lay insight. Although our study did not preclude lay insight (ie, participants might notice that some words were misspelled), the study was not dependent on such insight, which is useful for interrogating intuitive evaluations of information (ie, cost-benefit analyses [6]). In this study, the semantic content of the excerpts engendered consistent effects on trustworthiness ratings (at least among this type of online psychology experiment cohort, see Figure 3). In addition, this is the first time that numerical values have been gathered for the isolated effects of inappropriate capitalization.

The Results in Context

This study reaffirms an earlier observation that incivility decreases message credibility [38,39]. As suggested previously, inappropriate capitalization (shouting) is histrionic and induces a strong effect of incivility on readers' subjective ratings [37]. In our study, the effect of shouting text showed a trend for eliciting a smaller trustworthiness penalty than the effect of similar quantity of misspelling. One could easily speculate about new experiments where we might change the quantities of literacy errors; our experiment used either 5 misspellings or 5 shouting phrases, but one could run experiments to titrate errors, for example, to determine the relative effects of 3 misspellings or 10 inappropriate capitalizations. Furthermore, the effects of text shouting may be moderated by whether the statements are controversial [38]. We deliberately chose statements about multiple sclerosis that would be unfamiliar to the general public. This lack of familiarity makes the content and context not particularly emotional or controversial, likely weakening the effect of inappropriate capitalization.

The debate between how accessibility (ie, cost and speed) versus information quality (ie, accuracy and presumed benefit) quantitatively affect the use of (and search for) information continues [13]. Categorical frameworks have long been proposed to provide a theoretical underpinning for the factors, grouping specific elements that influence perceptions of trustworthiness in a variety of ways. The most well-known 2-category grouping of factors affecting how people respond to communication is Petty and Cacioppo's ELM for persuasion [56], in which elements contributing to a central pathway (eg, argument quality) are complemented by seemingly less rational elements that contribute to a peripheral pathway (eg, website design) [20,41]. Another two-category persuasion model that has been used to explain online trustworthiness is Chaiken's dichotomy of heuristics versus systematic information [61,62]. In both the Heuristic-Systematic Model and the ELM, diminishing motivation/involvement (or diminishing user dependency [45]) is associated with a switch from focusing on the effortful systematic evaluation of information (quality) to low-effort heuristics (accessibility).

In purchasing decisions, the type of product affects how strongly grammar and mechanics errors affect credibility [30]. In particular, *experience goods* (nontechnical items such as body lotion that are used personally) are more affected by grammar and writing mechanics than *search goods* (technical items such as printers). The implication is that when objective signals about trustworthiness are absent, heuristics play a stronger role [56,61]. In this study, where laypersons made judgments about unfamiliar medical issues, we might expect to find a stronger response to misspelling and capitalization. A necessary future approach is to repeat this experiment with multiple sclerosis patients who would exhibit user dependence when evaluating the statements [45].

Limitations

Our study had several important constraints. We deliberately tested laypersons' judgments on unfamiliar ideas about multiple sclerosis and showed that literacy errors can have a strong effect. Nevertheless, this effect may be smaller in a cohort that is more

dependent on knowing the information. In particular, if readers are dependent, then preexisting points of view and feeling of *homophily* [63] will influence perceived credibility, in a way that would not influence the general public with less interest in statements for and about multiple sclerosis.

It is also notable that in this study, the highest mean trustworthiness rating was 62.3 (0-100 scale), despite the statement being medically correct. Multiple factors may account for why this mean rating is not higher for trustworthiness. For example, participants in this psychology experiment saw the statements *in vacuo*, so that they could not verify the statement, check source credibility, or look for crowd consensus [8]. In real-life situations, other aspects of trustworthiness may dominate, and among some individuals, there may be ceiling and floor effects within this dataset, given the very wide standard deviations.

Conclusions

Incivility and literacy competence are proposed factors in how lay web users assess the trustworthiness of online health information. These results support Lederman et al's [8] theory of credibility assessment in online forums; the results also fit with Anderson's [40] description of integrative assessment of multiple cues. Here, we have shown that literacy competence errors have additive effects. The additive effects are strikingly precise. This implies that when people make seemingly rapid judgments about the trustworthiness of online text—no more than rough estimates—that their intuitive estimates seemingly add up quite accurately, at least at the population level. This completely vitiates binary models of judgment where a writer is judged as either competent or incompetent, with no further penalty for additional errors in writing mechanics [57]. The implication for writers is that for this level of errors (5 in 1 paragraph), multiple errors can add up.

Many other factors also contribute to trustworthiness, notably the argument quality of the content (logic), verification with other sources, reference credibility, and crowd consensus [8]. How these additional factors interplay with literacy competence will require further extensive research. A start would be to determine how persons with multiple sclerosis (with similar education levels to our current cohort) respond to these stimulus excerpts, as those readers would understand the information in a more self-relevant context.

Authors' Contributions

HC, HW, GT, BM, AN, and CW conceived the experiments. HW, GT, CW, and JR conducted the experiments. HW, CJ, and GT analyzed the results. CJ performed the statistical analyses. HW, GT, and HC drafted the manuscript. All authors reviewed the manuscript.

Conflicts of Interest

BM owns and JR worked for a commercial enterprise (Dalton Maag) that creates typefaces and graphical designs for commercial clients. AN acts as an ad hoc Scientific Consultant for Dalton Maag. This enterprise and those authors will not benefit directly or indirectly from data supporting incivility, misspelling, or shouting text. HW, GT, CW, CJ, and HC have no competing interests to declare.

Multimedia Appendix 1

Stimulus Paragraphs and errors list.
[DOC File , 51 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Excerpts with original sources.
[DOC File , 88 KB-Multimedia Appendix 2]

Multimedia Appendix 3

claritinCapitalizedWords.
[XLS File (Microsoft Excel File), 83 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Misspellings and Verifications.
[XLS File (Microsoft Excel File), 63 KB-Multimedia Appendix 4]

XSL•FO

RenderX

Multimedia Appendix 5

Supplementary Methods: CHERRIES.

[DOCX File , 28 KB-Multimedia Appendix 5]

References

1. Office for National Statistics UK. 2019. Internet Access – Households and Individuals, Great Britain: 2019 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2019 [accessed 2019-11-20]

2. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]

3. Hsieh RW, Chen L, Chen T, Liang J, Lin T, Chen Y, et al. The association between internet use and ambulatory care-seeking behaviors in Taiwan: a cross-sectional study. J Med Internet Res 2016 Dec 7;18(12):e319 [FREE Full text] [doi: 10.2196/jmir.5498] [Medline: 27927606]

4. The Health Union. 2016. Social Media for Health: What Socially-Active Patients Really Want URL: https://health-union.com/wp-content/uploads/2016/06/Social-Media-for-Health.pdf [accessed 2019-05-23] [WebCite Cache ID 78aKBnoMO]

5. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 7;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]

6. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 2;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]

7. Fogg B, Soohoo C, Danielson D, Marable L, Stanford J, Tauber E. How Do Users Evaluate the Credibility of Web Sites?: A Study With Over 2,500 Participants. In: Proceedings of the 2003 conference on Designing for user experiences. 2003 Presented at: DUX'03; June 6-7, 2003; San Francisco, CA. [doi: 10.1145/997078.997097]

8. Lederman R, Fan H, Smith S, Chang S. Who can you trust? Credibility assessment in online health forums. Health Policy Technol 2014 Mar;3(1):13-25. [doi: 10.1016/j.hlpt.2013.11.003]

9. Stanford J, Tauber E, Fogg B, Marable L. Advocacy - Consumer Reports. 2002. Experts vs Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites URL: https://advocacy.consumerreports.org/wp-content/uploads/2013/05/expert-vs-online-consumers.pdf [accessed 2019-05-23] [WebCite Cache ID 78aNK0qXf]

10. Brand-Gruwel S, Kammerer Y, van Meeuwen L, van Gog T. Source evaluation of domain experts and novices during web search. J Comput Assist Learn 2017 Jan 5;33(3):234-251. [doi: 10.1111/jcal.12162]

11. Sillence E, Briggs P, Harris P, Fishwick L. Going online for health advice: changes in usage and trust practices over the last five years. Interact Comput 2007 May;19(3):397-406 [FREE Full text] [doi: 10.1016/j.intcom.2006.10.002]

12. England CY, Nicholls AM. Advice available on the internet for people with coeliac disease: an evaluation of the quality of websites. J Hum Nutr Diet 2004 Dec;17(6):547-559. [doi: 10.1111/j.1365-277X.2004.00561.x] [Medline: 15546433]

13. Woudstra L, van den Hooff B, Schouten A. The quality versus accessibility debate revisited: a contingency perspective on human information source selection. J Assn Inf Sci Tec 2015 May 13;67(9):2060-2071. [doi: 10.1002/asi.23536]

14. Zimmermann M, Jucks R. How experts' use of medical technical jargon in different types of online health forums affects perceived information credibility: randomized experiment with laypersons. J Med Internet Res 2018 Jan 23;20(1):e30 [FREE Full text] [doi: 10.2196/jmir.8346] [Medline: 29362212]

15. Kata A. Anti-vaccine activists, web 2.0, and the postmodern paradigm-an overview of tactics and tropes used online by the anti-vaccination movement. Vaccine 2012 May 28;30(25):3778-3789. [doi: 10.1016/j.vaccine.2011.11.112] [Medline: 22172504]

16. Borzekowski DL, Schenk S, Wilson JL, Peebles R. e-Ana and e-Mia: a content analysis of pro-eating disorder web sites. Am J Public Health 2010 Aug;100(8):1526-1534. [doi: 10.2105/AJPH.2009.172700] [Medline: 20558807]

17. Flynn D, Nyhan B, Reifler J. The nature and origins of misperceptions: understanding false and unsupported beliefs about politics. Adv Political Psychol 2017 Jan 26;38:127-150. [doi: 10.1111/pops.12394]

18. Pickard V. The Political Economy of Communication. 2017. Media Failures in the Age of Trump URL: http://www.polecom.org/index.php/polecom/article/view/74/264 [accessed 2018-08-26]

19. Hobbs J, Kittler A, Fox S, Middleton B, Bates DW. Communicating health information to an alarmed public facing a threat such as a bioterrorist attack. J Health Commun 2004;9(1):67-75. [doi: 10.1080/10810730490271638] [Medline: 14761834]

20. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]

21. Albuja AF, Sanchez DT, Gaither SE. Fluid racial presentation: perceptions of contextual 'passing' among biracial people. J Exp Soc Psychol 2018 Jul;77:132-142. [doi: 10.1016/j.jesp.2018.04.010]

22. Rieh SY, Danielson DR. Credibility: a multidisciplinary framework. Ann Rev Info Sci Tech 2008 Oct 24;41(1):307-364. [doi: 10.1002/aris.2007.1440410114]

23. Gray R, Fitch M, Davis C, Phillips C. Breast cancer and prostate cancer self‑help groups: reflections on differences. Psycho‑Oncology 1996;5(2):137-142. [doi: 10.1002/(sici)1099-1611(199606)5:2<137::aid-pon222>3.3.co;2-5]

24. Seale C, Ziebland S, Charteris-Black J. Gender, cancer experience and internet use: a comparative keyword analysis of interviews and online cancer support groups. Soc Sci Med 2006 May;62(10):2577-2590. [doi: 10.1016/j.socscimed.2005.11.016] [Medline: 16361029]

25. BBC News Magazine. 2012. The Cruellest of Internet Hoaxes URL: https://www.bbc.co.uk/news/magazine-18282277 [accessed 2019-05-23] [WebCite Cache ID 78aSUcf7v]

26. Metzger MJ. Making sense of credibility on the web: models for evaluating online information and recommendations for future research. J Am Soc Inf Sci 2007 Nov;58(13):2078-2091. [doi: 10.1002/asi.20672]

27. Kornfield R, Smith KC, Szczypka G, Vera L, Emery S. Earned media and public engagement with CDC's 'tips from former smokers' campaign: an analysis of online news and blog coverage. J Med Internet Res 2015 Jan 20;17(1):e12 [FREE Full text] [doi: 10.2196/jmir.3645] [Medline: 25604520]

28. Gupta S, MacLean DL, Heer J, Manning CD. Induced lexico-syntactic patterns improve information extraction from online medical forums. J Am Med Inform Assoc 2014;21(5):902-909 [FREE Full text] [doi: 10.1136/amiajnl-2014-002669] [Medline: 24970840]

29. Foufi V, Gaudet-Blavignac C, Chevrier R, Lovis C. De-identification of medical narrative data. Stud Health Technol Inform 2017;244:23-27. [Medline: 29039370]

30. Ketron S. Investigating the effect of quality of grammar and mechanics (QGAM) in online reviews: the mediating role of reviewer credibility. J Bus Res 2017 Dec;81:51-59. [doi: 10.1016/j.jbusres.2017.08.008]

31. Schindler RM, Bickart B. Perceived helpfulness of online consumer reviews: the role of message content and style. J Consumer Behav 2012 Mar 21;11(3):234-243. [doi: 10.1002/cb.1372]

32. Figueredo L, Varnhagen CK. Didn't you run the spell checker? Effects of type of spelling error and use of a spell checker on perceptions of the author. Read Psychol 2005 Sep;26(4-5):441-458. [doi: 10.1080/02702710500400495]

33. Hargittai E. Hurdles to information seeking: spelling and typographical mistakes during users' online behavior. J Assoc Inf Syst 2006 Jan;7(1):52-67. [doi: 10.17705/1jais.00076]

34. Morin-Lessard E, McKelvie SJ. Does writeing rite matter? Effects of textual errors on personality trait attributions. Curr Psychol 2017 Mar 24;38(1):21-32. [doi: 10.1007/s12144-017-9582-z]

35. Vignovic JA, Thompson LF. Computer-mediated cross-cultural collaboration: attributing communication errors to the person versus the situation. J Appl Psychol 2010 Mar;95(2):265-276. [doi: 10.1037/a0018628] [Medline: 20230068]

36. Sobieraj S, Berry JM. From incivility to outrage: political discourse in blogs, talk radio, and cable news. Political Commun 2011 Feb 9;28(1):19-41. [doi: 10.1080/10584609.2010.542360]

37. Gervais BT. Incivility online: affective and behavioral reactions to uncivil political posts in a web-based experiment. J Inf Technol Politics 2015 Jan 14;12(2):167-185. [doi: 10.1080/19331681.2014.997416]

38. Graf J, Erba J, Harn R. The role of civility and anonymity on perceptions of online comments. Mass Commun Soc 2017 Feb 21;20(4):526-549. [doi: 10.1080/15205436.2016.1274763]

39. Thorson K, Vraga E, Ekdale B. Credibility in context: how uncivil online commentary affects news credibility. Mass Commun Soc 2010 Jun 3;13(3):289-313. [doi: 10.1080/15205430903225571]

40. Anderson N. Unified psychology based on three laws of information integration. Rev Gen Psychol 2013 Jun;17(2):125-132 [FREE Full text] [doi: 10.1037/a0032921]

41. Fogg B. Prominence-Interpretation Theory: Explaining How People Assess Credibility Online. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems. 2003 Presented at: CHI EA'03; April 5-10, 2003; Florida, USA p. 722-723. [doi: 10.1145/765891.765951]

42. Wanas N, El-Saban M, Ashour H, Ammar W. Automatic Scoring of Online Discussion Posts. In: Proceedings of the 2nd ACM Workshop on Information Credibility on the Web. 2008 Presented at: WICOW'08; October 30, 2008; Napa Valley, California, USA p. 19-26. [doi: 10.1145/1458527.1458534]

43. Castillo C, Mendoza M, Poblete B. Predicting information credibility in time-sensitive social media. Internet Res 2013 Oct 14;23(5):560-588. [doi: 10.1108/IntR-05-2012-0095]

44. Weerkamp W, de Rijke M. Credibility-inspired ranking for blog post retrieval. Inf Retrieval 2012 Feb 11;15(3-4):243-277. [doi: 10.1007/s10791-011-9182-8]

45. Johnson F, Rowley J, Sbaffi L. Modelling trust formation in health information contexts. J Inf Sci 2015 Mar 25;41(4):415-429. [doi: 10.1177/0165551515577914]

46. Pachur T, Bröder A. Judgment: a cognitive processing perspective. Wiley Interdiscip Rev Cogn Sci 2013 Nov;4(6):665-681. [doi: 10.1002/wcs.1259] [Medline: 26304271]

47. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974 Sep 27;185(4157):1124-1131. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]

48. Gigerenzer G, Goldstein DG. Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 1996 Oct;103(4):650-669. [doi: 10.1037/0033-295x.103.4.650] [Medline: 8888650]

49. Gigerenzer G, Goldstein DG. The recognition heuristic: a decade of research. Judgm Decis Mak 2011;6:100-121 [FREE Full text] [doi: 10.1093/acprof:oso/9780199390076.003.0008]

XSL•FO

RenderX

50. Bennis W, Katsikopoulos K, Goldstein D, Dieckmann AN, Berg N. Designed to fit minds. In: Todd PM, Gigerenzer G, editors. Ecological Rationality: Intelligence in the World. New York, USA: Oxford University Press; 2012:409-427.

51. Oreg S, Bayazit M. Prone to bias: development of a bias taxonomy from an individual differences perspective. Rev Gen Psychol 2009 Sep;13(3):175-193. [doi: 10.1037/a0015656]

52. Koriat A. Metacognition: decision-making processes in self-monitoring and self-regulation. In: Keren G, Wu G, editors. The Wiley Blackwell Handbook of Judgment and Decision Making. Chichester, UK: John Wiley & Sons; 2015:356-379.

53. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci U S A 2019 Sep 24;116(39):19251-19257 [FREE Full text] [doi: 10.1073/pnas.1821936116] [Medline: 31484770]

54. Orne MT. Demand characteristics and the concept of quasi-controls. In: Rosenthal R, Rosnow RL, editors. Artifacts in Behavioral Research. Oxford, UK: Oxford University Press; 2009:110-137.

55. Krefting L. Rigor in qualitative research: the assessment of trustworthiness. Am J Occup Ther 1991 Mar;45(3):214-222. [doi: 10.5014/ajot.45.3.214] [Medline: 2031523]

56. Petty RE, Cacioppo JT. The effects of involvement on responses to argument quantity and quality: central and peripheral routes to persuasion. J Pers Soc Psychol 1984;46(1):69-81. [doi: 10.1037/0022-3514.46.1.69]

57. Sauls M. TigerPrints: Clemson University Research. 2018. Perceived Credibility of Information on Internet Health Forums URL: https://tigerprints.clemson.edu/all_dissertations/2110 [accessed 2019-05-24]

58. Oleson D. World Bank Open Data. 2013. Discovering Drug Side Effects With Crowdsourcing URL: https://data.world/crowdflower/claritin-twitter/workspace/file?filename=1384367161_claritin_october_twitter_side_effects-1.csv [accessed 2019-11-23]

59. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]

60. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics 2000 Jun;56(2):645-646. [doi: 10.1111/j.0006-341x.2000.00645.x] [Medline: 10877330]

61. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]

62. Sundar S. The MAIN model: a heuristic approach to understanding technology effects on credibility. In: Metzger MJ, Flanagin AJ, editors. Digital Media, Youth, and Credibility. Cambridge, MA: The MIT Press; 2008:73-100.

63. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the internet: web sites versus discussion groups. Health Commun 2008 Jul;23(4):358-368. [doi: 10.1080/10410230802229738] [Medline: 18702000]

Abbreviations

ELM: Elaboration Likelihood Model
LME: linear mixed effects
RQ: research question