Treatment choices for depression: Young people's response to a traditional e-health versus a Health 2.0 website

Faye Scanlan[1], Anthony Jorm[2], Nicola Reavley[3], Denny Meyer[4] and Sunil Bhar[5]

Abstract

Objective: This exploratory experimental study compared young people's credibility appraisals and behavioural intentions following exposure to depression treatment information on a Health 2.0 website versus a traditional website. The traditional website listed evidence-based treatment recommendations for depression as judged by field experts. The Health 2.0 website contained information about how helpful each treatment was, as aggregated from feedback from young people with lived experience of depression.

Method: Participants ($n = 279$) were provided with a vignette asking them to imagine that they had just received a diagnosis of depression and they had gone online to find information to guide their treatment choices. They were randomly allocated to view either the traditional or the Health 2.0 website, and were asked to rate the credibility of the depression treatment information provided. They were also asked to indicate the extent to which they would be likely to act on the advice of the website.

Results: Participants in the traditional website condition rated their website as significantly more influential than did participants presented with the Health 2.0 website. This difference in treatment influence was fully accounted for the participants' perception of credibility of the information provided by the websites.

Conclusion: The traditional website was rated as significantly more credible and influential than the Health 2.0 website. Treatment decisions appeared to be based on the extent to which online information appears credible. In conclusion, health-related content was perceived by users as more credible when endorsed by experts than by other users, and perceived message credibility appears to be a powerful determinant of behavioural intentions within the e-health setting.

Keywords

depression, young adult, crowdsourcing, Internet, mental health

Although young people value and use the Internet to seek information on mental health,[1-3] surprisingly little is known about the impact of such information on young people's intentions to seek help,[4-6] and factors that influence such impact. The advent of Health 2.0 has provided users with an alternative source of information about health compared with traditional expert-sponsored websites. Given the variety of online information sources now available, it is important to understand the impact of different types of information on young people's decisions to seek treatment, and the

[1]Swinburne University of Technology, Melbourne, Australia; Orygen, The National Centre of Excellence in Youth Mental Health, Parkville, Melbourne, Australia; Centre for Youth Mental Health, The University of Melbourne, Carlton, Australia
[2]Centre for Mental Health, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Australia
[3]Centre for Mental Health, Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Australia
[4]Swinburne University of Technology, Melbourne, Australia
[5]Swinburne University of Technology, Melbourne, Australia

Corresponding author:
Faye Scanlan, Orygen, The National Centre of Excellence in Youth Mental Health, 35 Poplar Road, VIC 3052, Australia.
Email: fayeascanlan@gmail.com

factors that account for such impact. As the e-health landscape continues to evolve, several questions emerge regarding the impact of website information that is crowdsourced (that is, based on user-generated reviews and opinions) compared with the more traditional websites that are based on expert consensus and research.

The terms Medicine 2.0 and Health 2.0 are sometimes used synonymously and at other times they are defined differently. Health 2.0 is the more prevalent and widely accepted term.[7] In the absence of a consensus on an appropriate definition,[7] we use the following definition of Medicine 2.0 provided by Hughes and colleagues:

> '*The use of a specific set of Web tools (blogs, Podcasts, tagging, search, wikis, etc.) by actors in health care including doctors, patients, and scientists, using principles of open source and generation of content by users, and the power of networks in order to personalize health care, collaborate, and promote health education.*' (Hughes et al.,[8] Results section, para. 4)

The communications literature indicates that credibility is one of the strongest predictors of persuasion,[9–11] hence it is vital to better understand how young people appraise the credibility of mental health advice that they access online.

Credibility appraisals are appraisals about the extent to which information or an information source is believable.[12] The emergence of Health 2.0 has highlighted the importance of evaluating the credibility of e-health content,[13–15] as there is often no quality control procedure in place to assess the content prior to it being uploaded. Furthermore, individuals publishing content are not required to have any particular expertise in healthcare. As the e-health landscape continues to evolve, researchers have shifted from examining the role of content quality to the role of the user's perceived credibility of the website and content in treatment decisions.

Several research questions need to be examined in order to better understand how e-health users appraise the credibility of different types of e-health content.

First, there is a need to explore the extent to which e-health consumers believe that e-health content derived from lay expertise (i.e. 'user-generated' data) is as credible as that derived from professional expertise (i.e. traditional e-health data).[13,14] Emerging research indicates that content provided by patients on online health forums is valued as a credible source of health advice by e-health users.[16,17] Investigating how e-health consumers appraise the credibility of user-generated health data, including crowdsourced treatment reviews, has been identified as a research priority.[14,15,17] Crowdsourced medical data is frequently accessed.[18,19]

Moreover, evidence indicates that medical treatment advice derived from crowdsourced data is likely to differ from evidence-based guidelines.[20–23] For example, one study found that medications that were subject to United States Food and Drug Administration black box warnings were rated more poorly in crowdsourced reviews than in the empirical literature, whereas alternative medications or those with addictive properties were rated more positively.[21] Studies comparing crowdsourced treatment reviews of dermatological treatments to randomised controlled trial results found that, for some treatments, crowdsourced reviews were more positive than would be expected on the basis of their empirical support, while for others, treatments were rated more negatively.[20,22,24] This raises the question as to how e-health users rate the credibility of crowdsourced treatment data compared with evidence-based guidelines.

A second question that needs to be examined is the extent to which such crowdsourced e-health content influences young people's help-seeking intentions, and the extent to which credibility appraisal of such content plays a role in such influence. Internationally, there has been an increased emphasis on attempting to better harness the potential of online technologies to meet the needs of young people experiencing mental health difficulties.[25] While some websites provide e-therapy, many others provide mental health information with the intent of improving young people's awareness of their mental health needs, and their ability to source appropriate interventions offline. Most young people search for mental health information by entering a general search term on a search engine (e.g. 'depression' on Google) rather than going straight to a trusted website.[2,26,27] The retrieved results are likely to vary in information quality.[6] Therefore, it is important to understand which sources of online advice young people are likely to adopt.

Third, there is a need to explore how e-health users appraise the credibility of the content that they find online. The credibility literature has largely focused on the perceived credibility of information sources (i.e. perceived source credibility), with far fewer studies examining how individuals appraise the credibility of messages themselves (i.e. perceived message credibility).[10] However, an individual's credibility appraisal methods are likely to vary dependent on the information domain (e.g. entertainment versus health) and whether they are acquiring information online or offline.[14,28–30] Researchers have cautioned against the risk of overlooking the importance of message features in contributing to credibility perceptions of e-health content.[31,32] Indeed, there is evidence to suggest that e-health users focus on message rather than source credibility cues when appraising information derived from

both experiential expertise and scientific expertise in online health forms.[17]

Within the communications literature, research has consistently demonstrated that sources that are perceived to be credible have a greater impact on information receivers' attitude and behaviour than those that are not.[10,11] However, research suggests that in conditions of high user involvement (i.e. when information receivers are heavily invested in the topic), it is not the most important predictor of persuasion.[11] Rather, the perceived credibility of message content (i.e. message credibility) may be more influential. Wilson and Sherrell[11] conducted a meta-analysis examining the impact of source credibility on persuasion when user involvement was considered. They found that in 67% of the studies, source credibility only had an impact on persuasion in low user involvement conditions (i.e. when information receivers had low levels of investment in the topic being discussed). This finding is consistent with the Elaboration Likelihood Model of information processing (ELM).[33,34] ELM suggests that in low involvement conditions, individuals may use source cues (e.g. the website domain name) as a simple heuristic (or 'rule of thumb') to make a quick decision on whether or not their message is credible (e.g. the source looks professional). In contrast, when individuals are highly involved in a topic, if they have the ability to do so, they will invest cognitive effort in critically appraising message content (i.e. using the central route of cognitive processing) rather than relying on simple heuristics. A similar dual processing model has been applied to the context of assessing credibility in the online environment, emphasising the key role that user motivation and ability play.[35]

The online context and the nature of the task of searching for medical treatment information promote an emphasis on critically evaluating the characteristics of e-health content itself (i.e. perceived message credibility) in informing behavioural intentions.[17,33,36] There are two reasons for this. First, adopting inaccurate or misguided treatment advice could be very costly to the wellbeing of e-health users. Therefore, they are by definition a highly involved user group who are motivated to make an accurate credibility appraisal.[14,16,17,29] It can be argued that young people seeking mental health advice online are a particularly involved user group compared with other age groups, as they are particularly reluctant to seek mental health advice offline.[37] This suggests that they are less likely than older adults to verify the accuracy of the information they find online with offline sources (e.g. medical professionals, family members or peers). This creates extra vulnerability, as verification has been found to be the most important strategy that e-health users use to appraise the credibility of both scientific and experiential e-health content.[17] Second, e-health users are gathering treatment information in an environment in which source cues are often ambiguous or absent and, even when provided, they may not be trustworthy.[14,29,30,38,39] This promotes an emphasis on the influence of message over source cues in influencing credibility appraisals.[30]

Even when relying on heuristics to make credibility appraisals, in the context of searching for medical treatment information, e-health users are more likely to rely on message rather than source cues. This is because they are 'accuracy-motivated' users.[40] One heuristic that is likely to be particularly salient when searching for medical treatment information is that of 'self-confirmation', which reflects 'a tendency to notice and place greater value on information that supports one's beliefs, while discounting information that refutes those beliefs.' (Metzger and Flanagin,[29]p.453). The 'consistency heuristic' is also likely to be important; this reflects a tendency to believe information is credible if it is consistent with information provided by other sources.[40] In contrast, heuristic cues related to the surface characteristics of websites (e.g. the design and layout, interactivity and domain name) that have been found to influence Internet users' credibility appraisals[29,41] are likely to be less influential when an individual is searching for information in a situation where a poor decision could come at a high personal cost.[29,40]

Health 2.0 websites that utilise crowdsourcing allow e-health users to share and access experiential expertise in unprecedented and potentially more powerful ways than ever before. The 'shared experiential expertise'[42] that is available through these websites may be particularly valued by e-health users as a credible source of health advice. First, it may be perceived to be more credible than other forms of user-generated content, as it is more difficult to manipulate. As the volume of consumer feedback increases, the overall output (i.e. the 'rating') cannot be easily manipulated by any one reviewer's input.[42] Second, as the volume of feedback increases, it may be perceived to be less subjective and therefore less biased than any one individual's feedback.[17] Third, it may be more current than content on traditional websites, as it can be updated very quickly to consider new evidence or controversy.[43]

One study has compared the relative perceived credibility of crowdsourced data versus 'expert-generated' data. The perceived credibility of content based on 'expert-generated' versus 'user-generated' movie ratings was investigated.[42] User-generated ratings were only perceived to be more credible than expert-generated ratings if the volume of ratings was high. This was in contrast to a previous study which found that users neglected to attend to the number of ratings when considering reviews of e-commerce products.[44] Participants

were more likely to perceive the crowdsourced reviews to be credible if they had contributed crowdsourced data themselves (i.e. posted online reviews and recommendations) and reported familiarity with crowdsourced content.[42] As credibility appraisal methods are likely to vary between information domains, it may not be appropriate to generalise research from an e-commerce to an e-health setting.

The current exploratory study aimed to explore the extent to which traditional and Health 2.0 websites were perceived by young consumers as influential in treatment decisions and if so why. First, it explored whether treatment intentions differed after viewing the two types of website (traditional versus Health 2.0). Second, it examined whether the association between such intentions and website type was mediated by perceived message credibility. It was hypothesised that if behavioural intentions varied as a function of website allocation, this effect would be fully mediated by perceived message credibility.

The study attempted to promote external validity as the majority of research on the credibility of e-health content has lacked ecological validity.[10,45,46] The study used live and pre-existing websites in order to maximize the external validity of the information sources. A priori power analyses indicated that in order to determine whether there was a difference between experimental groups in behavioural intentions, a sample size of 210 would be required (anticipating a medium effect size and desired statistical power level of 0.95). While it would have been ideal to recruit a sample of actual users of the websites, or of individuals with current diagnoses of depression, it was not possible to do so within the resource constraints of this study. Therefore, a convenience sample was recruited, primarily comprising female university students. Given the exploratory nature of this study, this approach was deemed to be appropriate.

Method

Participants

Participants were required to be between 18 and 25 years of age and able to read English. They were recruited from social, family and professional networks of the researchers, and from a pool of undergraduate psychology students in the Research Experience Program (REP) at Swinburne University. Student participants were provided with course credit, while others were entered into a draw to win AU$50 gift vouchers. Participation was voluntary and participants could withdraw from the study at any time.

Only participants who were recruited through networks of the authors were offered the option of entering the prize draw and completing the experiment in their own time, as it was assumed that they would comply with the experiment instructions. In order to ensure that all other participants complied with instructions, they completed the experiment in a laboratory session with one of the authors present in the background.

The only check for data validity was for participants to enter the name of the website that they had just viewed. Any participant who failed to do so was excluded from all further analyses.

Design and procedure

The study used a between-subjects experimental design. Using an online algorithm (www.randomizer.org), participants were randomly assigned to one of two e-health websites: *WhatWorks4u* (www.whatworks4u.org), a prototype of a Health 2.0 website, or a more traditional e-health website, *BluePages* (http://BluePages.anu.edu.au/). In this study, *BluePages* will be described as a prototype of a 'traditional' e-health website, while *WhatWorks4u* will be described as a Health 2.0 website. These descriptions have been determined on the basis of how the depression treatment information on each website was derived and what objectives the website aims to fulfil in disseminating it.

BluePages has two objectives.[47] First, it provides information on treatments for depression based on the available scientific evidence and resources. Second, it offers 'screening tests for depression and anxiety', and links to other helpful resources. The functions of the provision of screening tests and links to other helpful content, including a forum for people with lived experiences of depression to interact, fall under the category of Health 2.0. However, the depression treatment ratings that were the focus in this study were derived solely from scientific expertise on the basis of 'systematic reviews of available scientific evidence and available resources'.[48] The objective of *BluePages* in relation to the dissemination of treatment information regarding depression can be described as improving knowledge of the scientific merit of different treatments. Therefore, while it is acknowledged that other sections of the *BluePages* website as a whole provide some Health 2.0 functions, for the purpose of this study it will be described as a 'traditional e-health website'.

In contrast, the primary purpose of the *WhatWorks4u* website is to collect, aggregate and share direct feedback from young people who have experienced mental health problems about their experiences of treatment.[49] In contrast to *BluePages*, the depression treatment ratings are solely derived from experiential expertise reflecting the Health 2.0 values of empowerment of healthcare consumers, open-source generation of medical content and the dissemination of user-generated content.

Participants were not informed that the research was comparing responses to a Health 2.0 versus a traditional e-health website. As can be seen in Table 1, while the websites differed in several ways, they were similar in some of the variables of interest in this study. Namely they were both sponsored by reputable Australian universities, provided summaries of 'what works' for treating depression, reviewed medical, psychological and self-help/lifestyle interventions, and were free of commercial sponsorship and advertising.

The *WhatWorks4u* website provided 'experiential expertise' from young people about what interventions 'worked for them'. This data was aggregated using crowd-sourcing technology and presented in the form of a 'star rating' scale. The overall star rating was presented alongside the number of reviews it was derived from and whether any adverse side effects to the treatment were reported. The *BluePages* website provided information about 'what works' for treating depression based on scientific expertise. Each intervention was rated on a 'smiley face' rating scale according to the scientific evidence base for the 'usefulness' of each intervention.

Participants were asked to: 'Imagine that you have recently been diagnosed with depression. You are unsure of what treatments you should try so you decide to go online to find some information about "what works" for treating depression. *You are looking for information that will help you decide what treatments to try*.' They were then invited to look for information to guide their treatment choices on their allocated website. Participants completed the experiment online or at a laboratory.

A manipulation check was used to determine whether participants had correctly perceived whether the information on the website was provided by an expert ('researchers with expertise in depression' for *BluePages*) or by consumers ('young people who have experienced depression' for *WhatWorks4u*). Participants also completed a range of questionnaires prior to, and immediately following their interaction with their assigned website.

Table 1. Summary of some key similarities and differences between the two websites.

Website	BluePages	WhatWorks4u
Domain name	.edu.au	.org
Developed by	Australian National University	University of Melbourne
Treatment rating displayed as	'Smiley face' rating scale	'Star rating' scale
Treatment ratings derived from	Scientific evidence	Aggregated user-generated treatment reviews
Treatment types included	Medical, psychological and self-help	Medical, psychological, complementary/alternative therapies, and alcohol and drug use
Commercial interest	No commercial interest or advertising	No commercial interest or advertising
Mental health content	Specific to depression	Covers all mental illnesses
Target age-group	All ages	Young people (12–30 years of age)
Tailored content provision	No ability to refine treatment reviews to the user's age, gender	Treatment reviews can be narrowed down by age, gender, mental illness, and location of the reviewer(s) using a search function
Content provided on the treatment section of the website	Provides (i) an explanation of the rating system, (ii) an explanation of different levels of scientific evidence, and (iii) individual treatment reviews. In addition to the smiley face rating, users can choose to read a narrative of the treatment review under the headings of: 'What is it?' 'How does it work?' 'Is it effective?' 'What are the advantages and disadvantages?' 'Where do you get it?' 'Recommendation' (i.e. the overall recommendation on the basis of the evidence). 'References'.	Provides (i) a star rating alongside the number of reviews it was derived from, and (ii) details of any adverse side effects noted and how many reviewers experienced each one. Users can also click on the treatment name to find out more about the treatment.

Measures

Participants completed a demographics form (sex, age, education level and employment status) and the following questionnaires prior to viewing their allocated website.

Internet literacy was assessed by three items, with all responses rated on a seven-point scale. First, participants were asked, 'How often do you use the Internet?' (*1 = 'never', 7 = 'very often'*); 'How much experience do you have in using the Internet?' (*1 = 'no experience', 7 = 'a great deal of experience'*); and 'How much expertise do you have in using the Internet?' (*1 = 'I am not at all expert', 7 = 'I am completely expert'*). An additional questionnaire[50] measured whether participants had any previous experience engaging in a number of Health 2.0 activities (see Table 2). Response options were 'yes', 'no' and 'I don't know'.

Perceived e-health literacy was measured by the e-Health Literacy Scale (eHEALS).[51] The total score of the questionnaire ranged from 8 to 40 with higher scores reflecting higher levels of perceived e-health literacy skills). Following the suggestion of Norman[52] two items were added to the measure to assess perceived skills in using ratings provided by others to judge the quality of online content, and in integrating non-professional online health advice with professional recommendations. All items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The eight-item eHEALS scale demonstrated good internal reliability (Cronbach's $\alpha = 0.84$).

Table 2. Percentage of participants who had previously engaged in Health 2.0 activities.

Health 2.0 activity	BluePages ($n = 140$)	WhatWorks4u ($n = 139$)
Signed up to receive an email alert/update about a health or medical issue	24.3%	19.4%
Read someone else's commentary or experiences about health or medical issues on an online newsgroup, website or blog	76.4%	75.5%
Gone online to find someone who may have similar health concerns to your own	75%	72.7%
Tracked your weight, diet or exercise routine online	45%	52.5%
Tracked any other health indicator online	59.3%	66.9%
Watched an online video about health or medical issues	82.1%	77.7%

Topic salience[53] was measured by four items that were used to determine how relevant the topic of depression was to participants' lives. Participants were asked, 'have you ever been diagnosed with depression?' and 'has someone close to you (e.g. a family member or friend) ever been diagnosed with depression?' Response options were 'yes', 'no' and 'I don't know'. Participants were asked to rate their perceived knowledge about depression and its treatment on a scale of 1 (*almost nothing*) to 4 (*a great deal*), and their interest in learning more about depression and its treatment on a scale of 1 (*not at all interested*) to 4 (*very interested*). Each of these four items was considered separately in the descriptive statistics and analyses.

Disposition to trust[54] was measured by the four-item Disposition to Trust Scale which assessed participants' general tendency toward trusting others. This scale was initially developed as a five-item scale on the basis of the literature pertaining to trait-based trust. Following initial psychometric analyses of the subscale within a broader e-commerce questionnaire using confirmatory factor analysis, one item was deleted on the basis of shared variance.[54] Unidimensionality of the four-item scale was confirmed using goodness of fit indices for the full structural model (GFI = 0.92, AGFI = 0.88, NFI = 0.93, and RMR = 0.04;[54] with factor loadings ranging from 0.61 to 0.89). The four-item scale demonstrated acceptable internal reliability in the current study (Cronbach's $\alpha = 0.75$).

Confirmation of prior belief[55] was measured by a modified two-item scale asking participants if the content confirmed their prior belief about the effectiveness of different treatments for depression. Responses were rated on a five-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). It was not possible to test the unidimensionality of the original three-item scale using confirmatory factor analysis as the single-factor congeneric model could not be computed. Therefore, an exploratory factor analysis was conducted resulting in a single factor as expected. The factor loading for the only reverse coded item 'Information from the ratings contradicted what I had known before viewing them' was very low (0.26). This item also had low correlations ($r < 0.22$) with both of the other items. Therefore, this item was deleted from the scale. The remaining two-items ('The ratings supported my impression of the effectiveness of different interventions for depression' and 'The ratings reinforced information I had previously known about interventions for depression') had strong factor loadings (0.79 and 0.82, respectively). The internal consistency of the revised scale was tested using Spearman Brown coefficient as this is considered to be a more appropriate statistic than Cronbach's alpha for two-item scales.[56] The two-item scale demonstrated acceptable internal consistency in this study ($r_{SB2} = 0.78$).

Perceived credibility[57] was measured at the level of the website, the message and the sponsors with three scales: message, sponsor and website credibility. Following criticism about poor construct validity of the scales,[10,58] each scale was subjected to confirmatory factor analyses. Perceived message and sponsor credibility were established as unidimensional constructs. In contrast, the website credibility scale had three factors. Consistent with the theoretical literature,[10,58] one reflected the core construct of credibility, while the other two related to cues that are better understood as characteristics that can enhance the credibility of a website or source, in this case, positive ratings of website design and dynamism (i.e. interactivity).

Perceived sponsor and message credibility[57] were measured on two separate scales. Perceived sponsor credibility was measured using a revised 3-item version of the original eight-item questionnaire.[57] The modified scale assessed the extent to which the website sponsors were perceived to 'be credible', 'have a positive reputation' and 'be trustworthy'. Each item was rated on a Likert scale of 1 (*not at all*) to 10 (*extremely*). This scale demonstrated good internal reliability (Cronbach's $\alpha = 0.92$), unidimensionality and metric invariance across the websites. Perceived message credibility (i.e. the perceived credibility of the depression treatment ratings) was measured on the original five-item scale.[57] Responses were rated on a seven-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The item measuring bias was reverse-coded, so that higher scores reflected higher levels of perceived credibility. The scale demonstrated an acceptable level of internal consistency (Cronbach's $\alpha = 0.78$). Both of these scales were found to be unifactorial on the basis of confirmatory factor analyses.

Website Dynamism and Design measures were derived from the Website Credibility Scale.[57] On the basis of an exploratory factor analysis, this scale was found to have three factors (website credibility, design and dynamism), but only two (website design and website dynamism) were invariant across conditions.

A three-item measure of website design assessed the degree to which the website was perceived to be: 'attractive', 'colourful' and 'likeable'. The scale demonstrated acceptable internal reliability (Cronbach's $\alpha = 0.79$). A four-item measure of website dynamism assessed the degree to which the website was perceived to be: 'interactive', 'interesting', 'involving' and 'bold' (Cronbach's $\alpha = 0.78$). Responses for both scales were rated on a six-point Likert scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*).

The 12-item website credibility subscale was refined using confirmatory factor analyses. Several items had to be deleted due to metric invariance between the two experimental groups. This indicated that the measure was being interpreted differently across the two experimental groups. As a result, it was not valid to compare the two groups on the subscale. In order to achieve metric invariance, it was necessary to remove the problematic item 'trustworthiness' from the scale, as it loaded significantly more heavily on the construct of website credibility for the *BluePages* group than it did for the *WhatWorks4u* group. Trustworthiness is a core component of source credibility, therefore, deleting this item would have resulted in a subscale that did not sufficiently represent the theoretical construct of Website Credibility as it has been defined in the literature to-date.[36,58,59] As a result, the decision was made to exclude this scale from further analyses.

Behavioural intention was measured using a three-item questionnaire.[60] Participants were asked: 'If you were experiencing depression, how likely would you be to: (i) act on the advice that was offered in the ratings, (ii) recommend the website to a friend who was experiencing depression, and (iii) forward the website to a friend or family member.' Responses were rated on a 10-point Likert scale, ranging from 1 (*extremely unlikely*) to 10 (*extremely likely*). This scale demonstrated good internal reliability (Cronbach's $\alpha = 0.89$), unidimensionality and metric invariance across the websites.

Analytical approach

Preliminary analyses comprised three steps. First, psychometric analyses were conducted to determine the reliability and validity of the included scales. Second, in order to establish whether randomisation was successful, a series of analyses were conducted to examine whether there were any differences between the experimental groups on user characteristics. Third, frequencies were used to establish the manipulation check results (i.e. did the group allocated to *BluePages* identify the source of the depression treatment ratings to be 'researchers with expertise in depression'; and did the group allocated to *WhatWorks4u* identify the source to be 'young people with experience of depression'?).

In order to examine the first research question, an independent samples *t*-test was conducted to determine whether there was a significant difference between-groups on behavioural intentions. As a significant difference in behavioural intentions was demonstrated between groups, Hayes' mediation analysis (PROCESS) was used to explore whether perceived message credibility mediated the impact of website allocation on behavioural intentions.[61]

Finally, a series of independent samples t-tests were used to ascertain whether there were any between-group differences on other website characteristics or in confirmation of prior belief, that may have

contributed to differences in behavioural intentions between the experimental groups.

Results

Participants

A desired sample size of 210 was based on a priori power analysis for a two-tailed student t-test, anticipating a medium effect size and desired statistical power level of 0.95;[62] Sample size requirements for regression analyses were also taken into account.[62]

From an initial pool of 305, 279 individuals were included in data analyses (mean age = 20.06 years, SD = 2.04; 219/279, 78.5% female) and 26 were excluded (14 were ineligible due to their age, 7 failed to complete the experiment, 3 failed to correctly identify their allocated website, and 2 had missing data). All participants had completed high school, and almost half (131/279, 46.9%) had pursued further qualifications. Most (256/279, 91.8%) were either working, or engaged in both study and employment at the time of the study. Participants were highly experienced and quite confident Internet users who were familiar with Health 2.0 technologies (Tables 2 and 3). Moreover, their perceived level of e-health literacy ($M = 3.6$,

SD = 0.6) was relatively high, consistent with previous research.[63,64] Participants reported high levels of exposure to Health 2.0 websites, and particularly to seeking lay expertise about illnesses and treatments (Table 2).

Depression was a salient topic; most participants (189/279, 67.7%) reported that a close friend or family member had been diagnosed with depression, and one in five (57/279, 20.4%) reported a lifetime history of depression. Almost three quarters of the sample (207/279, 74.5%) reported knowing 'a reasonable amount' or a 'great deal' about depression and its treatment. Most (244/279, 87.5%) reported being 'moderately' or 'very' interested in learning more about depression and its treatment.

Randomisation

A series of non-parametric Mann–Whitney U, chi-squared and Fisher's exact tests were used to test for between-group differences on user characteristics. Where multiple comparisons were being made, the significance criterion was divided by the number of variables that were being tested to protect against the risk of making a type I error.[62] There were no significant differences between groups on any user characteristics. An independent samples t-test indicated that there was

Table 3. Participants' disposition to trust, internet and e-health literacy skills (n = 279).

Variable	BluePages		WhatWorks4u		
	M (SD)	Median	M (SD)	Median	Item Range
Disposition to trust	3.54 (0.71)	3.75	3.52 (0.75)	3.5	1—7
Internet use	6.65 (0.70)	7	6.60 (0.95)	7	1—7
Internet experience	6.20 (0.87)	6	6.35 (0.98)	7	1—7
Internet expertise (SD)	5.24 (1.08)	5	5.44 (1.03)	6	1—7
eHEALS (8-item scale)	3.60 (0.56)	3.63	3.60 (0.62)	3.63	1—5*
Confidence in Web 2.0 use					
'I feel confident in my ability to combine professional and non-professional health advice that I find online'	3.47 (0.93)	4	3.27 (0.98)	4	1—5
'I feel confident in my ability to use other user's recommendations to filter relevant and trustworthy information online (e.g. using consumer review ratings when deciding whether or not to purchase a good or service)'	3.56 (0.88)	4	3.43 (0.99)	4	1—5

Note: *eHEALS scores were presented as means scores ranging from 1 to 5 rather than total scores ranging from 8 to 40.

no significant difference in eHEALS scores between the *BluePages* ($M = 3.6$, $SD = 0.57$) and the *WhatWorks4u* ($M = 3.6$, $SD = 0.62$) groups; $t(277) = 0.078$, $p = 0.938$ (two-tailed). Therefore, it can be concluded that any differences between groups on the outcome measures, or covariates, were not due to user characteristics.

Performance on the manipulation check

Failure rates for identifying the message source for *BluePages* and *WhatWorks4u* were 30% and 21%, respectively. A chi-squared test for independence (with Yates continuity correction) indicated that this difference was not statistically significant (χ^2 (1, 279) = 3.07, $p = 0.106$, $\phi = -0.11$). It was hypothesised that e-health users would similarly vary in their attendance to, and interpretation of information source in real-life situations. As this study prioritised external validity, participants were included in the main analyses regardless of their performance on the manipulation check.

Research questions

The mean behavioural intention rating was significantly higher for the group allocated to the *BluePages* website ($M = 6.22$, $SD = 2.30$) than the group allocated to the *WhatWorks4u* website ($M = 5.68$, $SD = 2.25$; $t = 1.987$, $df = 277$, $p = 0.048$, two-tailed). The magnitude of the differences in the means (mean difference = 0.54, 95% confidence interval (CI) 0.005–1.076) was small ($d = 0.24$, $\eta^2 = 0.01$).

A simple mediation analysis was conducted using ordinary least-squares path analysis and PROCESS statistical analysis to determine whether perceived message credibility mediated the impact of website allocation on behavioural intentions. The results are summarised in Table 4, and the regression coefficients are superimposed on the statistical diagram of the model (Figure 1). Participants assigned to the *BluePages* website ($X = 1$) reported significantly higher message credibility ratings than those assigned to the *WhatWorks4u* website ($a = 0.45$; $p < 0.001$). Participants who rated the perceived message credibility of the depression treatment ratings positively also reported significantly higher behavioural intentions toward acting on them ($b = 1.27$; $p < 0.001$). A bias corrected bootstrap confidence interval for the indirect effect of the manipulation of website on behavioural intentions through perceived message credibility ($ab = 0.57$) based on 10,000 samples was entirely above zero (0.31–0.89) indicating a significant effect. As can be seen in Table 4, there was no evidence that website allocation had a significant effect on behavioural intentions independent of its effect on perceived message credibility ($c^i = -0.03$, $SE = 0.24$, $p = 0.892$). Therefore, perceived message credibility fully mediated the effect of website on behavioural intention. Full mediation was also observed for the subset of participants who passed the manipulation check ($n = 208$).

As it was possible that other website covariates mediated the relationship between website and behavioural intentions, exploratory independent samples *t*-tests were conducted to compare the groups for differences on website covariates (see Table 5). No significant between-group differences emerged on perceived sponsor credibility or confirmation of prior belief. However, the mean ratings for website design and website dynamism were significantly higher in the *WhatWorks4u* group ($p < 0.001$). Bivariate correlations indicated that website design ($r = 0.22$, $p = 0.009$) and dynamism ($r = 0.42$, $p < 0.001$) were significantly positively correlated with behavioural intentions ($r = 0.33$, $p < 0.001$) for the *BluePages* group. Both variables were

Table 4. Results of mediation analysis for the total sample ($n = 279$).

Consequent										
		M (PMC)					*Y* (BI)			
Antecedent		Coeff.	SE	*p*	95% CI		Coeff.	SE	*p*	95% CI
X (WEB)	*a*	0.45	0.11	<0.001	(4.513, −0.237)	*c'*	−0.03	0.24	0.892	(−0.508, 0.443)
M (PMC)						*b*	1.27	0.13	0.000	(1.014, 1.521)
Constant	i_1	4.67	0.08	<0.001	(0.237, 0.668)	i_2	−0.23	0.62	0.7034	(−1.465, 0.989)
		$R^2 = 0.06$					$R^2 = 0.27$			
		$F(1, 277) = 17.11$, $p < 0.001$					$F(2, 276) = 51.08$, $p < 0.001$			

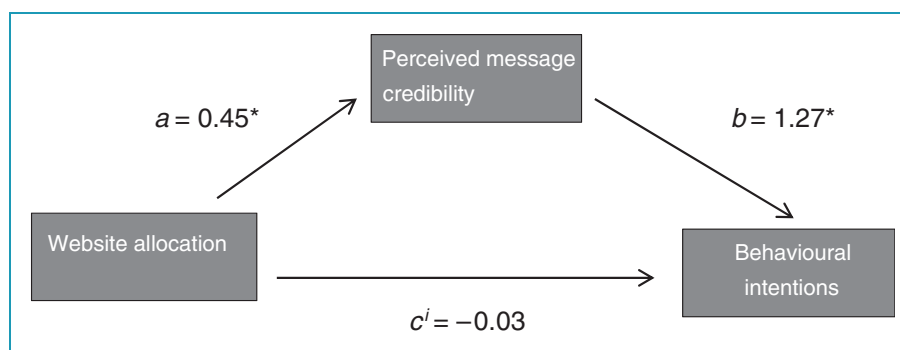PMC, perceived message credibility; BI, behavioural intentions.

Figure 1. Results of a simple mediation analysis (n = 279).
Note. * denotes significance at $p < 0.001$.

Table 5. Exploratory independent-samples t-tests (two-tailed) of between-group differences on website characteristics.

Website Variable	BluePages (n = 140) M (SD)	WhatWorks4u (n = 139) M (SD)	t (df)	p
Sponsor credibility	7.10 (1.96)	7.12 (1.51)	−0.118 (260)	0.906
Website dynamism	4.15 (0.98)	4.81 (0.77)	−6.284 (263)	<0.001
Website design	4.16 (1.17)	5.28 (0.67)	−9.829 (221)	<0.001
Confirmation of prior belief	5.01 (1.19)	4.82 (1.00)	1.446 (277)	0.132

also positively correlated with perceived message credibility for the *BluePages* group ($r = 0.13$, $p > 0.05$; and $r = 0.30$, $p < 0.001$; for the website design and dynamism, respectively). These results supported the inference that more positive website design and dynamism ratings did not account for the main effect of website on behavioural intentions.

Discussion

This exploratory experimental study compared young people's behavioural intention toward acting on depression treatment information provided on a traditional e-health website, compared with a Health 2.0 website. It was hypothesised that if behavioural intentions varied as a function of website allocation, this effect would be fully mediated by perceived message credibility. In response to calls for studies examining online credibility to be externally valid,[10] live and pre-existing websites were used in this study. As recommended by previous researchers,[65] efforts were made to choose websites that were similar on several important characteristics (e.g. similar website sponsors, lack of advertising, similar treatment rating scales used). The sample used in this study was comprised primarily of female university students. This sample limits the generalizability of the findings.

Behavioural Intention ratings were significantly higher among participants in the *BluePages* group (i.e. the traditional e-health website) than the *WhatWorks4u* group (i.e. the Health 2.0 website). As predicted, perceived message credibility fully mediated the impact of website allocation on behavioural intention. Several researchers have suggested that when evaluating medical content, individuals are likely to invest cognitive resources in appraising message credibility rather than simply relying on heuristics or 'cognitive short-cuts' to make quick judgements.[14,17,29,30] This is because the stakes are high when searching for content related to health in general, and medical treatments in particular, simply due to the potential repercussions of choosing to adopt inaccurate information. Dual processing models predict that in such conditions, if an e-health user has the necessary ability to critically appraise e-health content, they will do so rather than focusing on simple heuristic cues such as the surface features of the website (e.g. the design and interactivity).[34,35]

While the sample in this study was not clinical, participants were well suited to addressing the research

questions, as they had sufficient ability to complete the task as demonstrated by high levels of reported Internet use, experience, perceived expertise and adequate e-health literacy. In addition, the majority (87.5%) reported being 'moderately' or 'very' interested in learning more about depression and its treatment. Most (67.7%) had a friend or family member who had been diagnosed with depression. The prevalence of a lifetime history of depression (20.5%) was slightly higher than that found in the general population.[66] Previous research has found elevated levels of psychological distress among student populations.[67] Therefore, the sample succeeded in mirroring the high user-involvement condition that would be typical of most e-health users seeking treatment advice.

The finding that message credibility fully mediated the relationship between website allocation and behavioural intention is consistent with ELM[33,34] and Metzger's dual processing model of credibility assessment,[35] which posit that under conditions of high topic involvement, users will scrutinise message characteristics. Moreover, the results of this study were similar to those of McKnight and Kacmar[9] who found that message credibility fully mediated the relationship between initial impressions of a website (e.g. perceived website quality, credibility, and reputation) and the adoption of online legal advice. The results of this study also support other research that demonstrated that message characteristics were important predictors of the adoption of e-health advice.[31] They are also consistent with the finding that message credibility can be an important determinant of e-health users' appraisals of the credibility of e-health data regardless of whether it is derived from experiential and scientific expertise.[16,17]

Further analyses supported the inference that the difference in behavioural intentions ratings between groups was not due to differences in other website features. There was no difference in perceived sponsor credibility between the groups, supporting the inference that differences in the perceived credibility of the website sponsors did not account for the main effect of website on behavioural intention. While there were significant differences between groups in ratings of website design and dynamism features, both of these differences were in favour of the Health 2.0 website, and hence cannot explain the higher ratings of behavioural intention for the traditional website.

A significant minority of participants failed the manipulation check. These findings are similar to those of previous research that found that significant percentages of participants fail correctly identify the source of information.[60] As this study was prioritising external validity, participants' data was included in the main analyses regardless of their performance on the manipulation check. Full mediation was also evident in the subset of participants who passed the manipulation check.

Researchers are beginning to investigate the impact of age on how e-health users appraise the credibility of e-health data. There is some evidence to suggest that young people may be more critical consumers of traditional e-health content than older adults. However, when Health 2.0 content is considered, the impact of age on credibility appraisals seems to be more complex.[68]

Limitations and future research

This study had several limitations that should be considered in interpreting the results.

First, in promoting external validity of the information sources used, the internal validity of the study was attenuated. The websites differed in several ways in terms of the content offered, their domain names (.org versus .edu.au), website design and interactivity features. Some of these variables were measured and controlled for within the study design. However, it was not possible to control for all potential confounding variables. As such, the results of this study cannot be solely attributed to differences in credibility appraisals of treatment data derived from experiential versus professional expertise. Instead, they need to be conceptualised as overall differences in credibility appraisals and behavioural intentions related to the websites as a whole.

Second, this exploratory study only included one prototype of a Health 2.0 and a traditional e-health website. Therefore, it is not possible to generalise the results to other websites. As an example of a Health 2.0 website, *WhatWorks4u* had significant limitations that may have biased results toward more favourable ratings of the *BluePages* website. It provided aggregated treatment ratings in the absence of user reviews or access to individual reviewer's ratings. As a result, several important credibility cues, such as rating consistency, reviewer reputation, argument strength and reviewers' literary competence[17,69] were not available. Rating consistency may be a particularly important credibility cue in Health 2.0 websites, as it allows e-health users to assess 'crowd consensus' which has been found to be a strong predictor of credibility in the context of experiential e-health information.[17] Moreover, *WhatWorks4u* was a newly established website with a low volume of ratings of depression treatments. For example, only 12 reviews had been posted for cognitive behavioural therapy for major depressive disorder. At the time the study was designed, there was some evidence to suggest that end-users did not attend to rating volume when appraising the credibility of crowdsourced data.[44] However, more recent research

indicates that crowdsourced ratings may need to reach a critical threshold in order to be perceived as equally credible to ratings attributed to 'experts'.[42] The low volume of ratings on *WhatWorks4u* makes it difficult to comment on whether the treatment ratings were consistent across the two websites. Furthermore, *WhatWorks4u* included many more types of treatments than *BluePages*. However, taking the example of CBT, on the *WhatWorks4u* website, it had a four-star rating at the time of the study which is comparable to its three-star rating on *BluePages*. Future research should investigate the perceived credibility of a range of different Health 2.0 websites. Moreover, researchers comparing traditional e-health to Health 2.0 websites should carefully consider which websites to investigate and choose websites that are comparable in as many core features as possible to reduce the number of confounding variables in the study design. In this study, the two websites provided different levels of detail in their treatment ratings, and this information was presented in very different ways. For example, *BluePages* provided more detailed treatment reviews, while *WhatWorks4u* was somewhat difficult to navigate. These differences introduce a number of confounding variables that were not controlled for within the design of this study.

Third, participants were predominantly female and highly educated, therefore is not possible to generalise the findings to other populations. The results need to be interpreted bearing this in mind. There is evidence to suggest that references are a particularly important indicator of credibility to university students.[70,71] As the sample in this study was predominantly made up of students, they may have been particularly influenced by evidence-based information provided on *BluePages*.

Fourth, while the study attempted to enhance external validity, it was still conducted in an artificial situation. Participants recruited outside of the researchers' personal and professional networks were required to complete the experiment in a laboratory with one of the authors supervising in the background. This is likely to have promoted the validity of the data, however it also compromised the external validity of the study, as the experimental setting was more artificial than it would have been if all participants could complete the experiment privately. Moreover, in an artificial situation, certain variables cannot be controlled for. For example, young people usually use search engines to find e-health content[2,26,27,72] and the process by which young people find a website has been found to influence their credibility appraisals of online content.[73] Furthermore, while the use of a vignette can encourage goal-directed searching and encourage individuals to act as they would in a real-life situation, hypothetical situations cannot replicate real-world circumstances.

Depressive symptoms (e.g. amotivation, hopelessness and helplessness) may reduce the likelihood that individuals who are currently depressed would act on any e-health content accessed. Furthermore, cognitive symptoms of depression, such as poor concentration, and fatigue may also interfere with credibility appraisal processes that are observed in non-depressed populations. Future research should recruit more diverse samples of young people and include those with current diagnoses of mental illness. Researchers should consider recruiting directly from the websites in question to investigate the opinions of active users. Similarly, several important predictors of persuasion were not included in the analyses. Two particularly important variables that were not accounted for were perceived source expertise (lay expertise and professional expertise) and homophily.[60,74] Future research should also consider incorporating more complex data validity checks to improve the generalizability of results, particularly if self-report data is being used.

Finally, future research should go beyond measuring behavioural intention toward acting on treatment advice to measure actual behaviour change that results from accessing e-health data. While behavioural intentions are assumed to lead to behaviour, this is not always the case.[75]

Contributions of this research to the field

This is the first study investigating the relative impact of mental health treatment advice provided on a Health 2.0 versus a traditional e-health website. Furthermore, it is the first study to compare the perceived credibility of these two types of mental health treatment advice, from the perspective of end-users. Behavioural intentions ratings for the traditional e-health website exceeded those of the group allocated to the traditional e-health. However, the effect size was small and mean behavioural intentions ratings for both websites were above the midpoint on the rating scale. The Health 2.0 website used in this study had several limitations that may have detracted from the perceived credibility of the treatment ratings provided, therefore, the results suggest that e-health consumers are likely to perceive more established Health 2.0 websites as quite credible sources of medical treatment advice.

The results of this study also highlight the persuasive power of message credibility within e-health. They add weight to previous findings that e-health users rely heavily on message credibility cues when appraising the credibility of both user-generated and scientific health information online.[16,17] Consistent with previous research,[31] message characteristics were very important in influencing participants' behavioural intention toward adopting online advice. The ultimate aim of

e-health websites is to positively influence the wellbeing of their users. Therefore, it is imperative to understand how to promote message credibility to fully harness their potential.

References

1. Berger M, Wagner TH and Baker LC. Internet use and stigmatized illness. *Soc Sci Med* 2005; 61: 1821–1827.
2. Burns J, Davenport T, Christensen H, et al. *Game On: Exploring the Impact of Technologies on Young Men's Mental Health and Wellbeing. Findings from the First Young and Well National Survey.* Melbourne: Young and Well Cooperative Research Centre, 2013.
3. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R and Malla A. Preferences of young adults with first-episode psychosis for receiving specialized mental health services using technology: A survey study. *JMIR Mental Health* 2015; 2: e18.
4. Edwards Hart T and Chester A. Online mental health resources for adolescents: Overview of research and theory. *Austral Psychol* 2010; 45: 223–230.
5. Kauer SD, Mangan C and Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. *J Med Internet Res* 2014; 16: e66.
6. Reavley NJ and Jorm AF. The quality of mental disorder information websites: A review. *Patient Educ Counsel* 2011; 85: e16–e25.
7. Van De Belt TH, Engelen LJ, Berben SAA and Schoonhoven L. Definition of health 2.0 and medicine 2.0: A systematic review. *J Med Internet Res* 2010; 12.
8. Hughes B, Joshi I and Wareham J. Health 2.0 and Medicine 2.0: Tensions and controversies in the field. *J Med Internet Res* 2008; 10.
9. McKnight H and Kacmar C. Factors of information credibility for an Internet advice site. In *2006 HICSS'06 Proceedings of the 39th Annual Hawaii International Conference on System Sciences*. IEEE, 2006, pp. 1–10.
10. Pornpitakpan C. The persuasiveness of source credibility: A critical review of five decades' evidence. *J Appl Soc Psychol* 2004; 34: 243–81.
11. Wilson EJ and Sherrell DL. Source effects in communication and persuasion research: A meta-analysis of effect size. *J Acad Marketing Sci* 1993: 101–112.
12. Tseng S and Fogg B. Credibility and computing technology. *Commun ACM* 1999; 42: 39–44.
13. Adams SA. Revisiting the online health information reliability debate in the wake of "web 2.0": An interdisciplinary literature and website review. *Int J Med Informatics* 2010; 79: 391–400.
14. Eysenbach G. Credibility of health information and digital media: New perspectives and implications for youth. *John D Catherine T MacArthur Found Ser Digital Media Learning* 2007: 123–54.
15. Metzger MJ and Flanagin AJ. Using web 2.0 technologies to enhance evidence-based medical information. *J Health Commun* 2011; 16: 45–58.
16. Fan H, Leederman R, Smith S and Chang S. How online health forum users assess user-generated content: Mixed method research. In *Proceedings of ECIS*, 2013.
17. Lederman R, Fan H, Smith S and Chang S. Who can you trust? Credibility assessment in online health forums. *Health Policy Technol* 2014; 3: 13–25.
18. Van De Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M and Schoonhoven L. Internet and social media for health-related information and communication in health care: Preferences of the Dutch general population. *J Med Internet Res* 2013; 15: e220.
19. Thackeray R, Crookston BT and West JH. Correlates of health-related social media use among adults. *J Med Internet Res* 2013; 15: e21.
20. Armstrong AW, Harskamp CT, Cheeney S, Wu J and Schupp CW. Power of crowdsourcing: Novel methods of data collection in psoriasis and psoriatic arthritis. *J Amer Acad Dermatol* 2012; 67: 1273–1281. e9.
21. Adusumalli S, Lee H, Hoi Q, Koo S-L, Tan IB and Ng PC. Assessment of web-based consumer reviews as a resource for drug performance. *J Med Internet Res* 2015; 17: e211.
22. Armstrong AW, Wu J, Harskamp CT, Cheeney S and Schupp CW. Crowdsourcing for data collection: A pilot study comparing patient-reported experiences and clinical trial data for the treatment of seborrheic dermatitis. *Skin Res Technol* 2013; 19: 55–57.

23. Chou WY, Prestin A, Lyons C and Wen KY. Web 2.0 for health promotion: Reviewing the current evidence. *Amer J Public Health* 2013; 103: e9–e18.

24. Armstrong AW, Cheeney S, Wu J, Harskamp CT and Schupp CW. Harnessing the power of crowds: Crowdsourcing as a novel research method for evaluation of acne treatments. *Amer J Clin Dermatol* 2012; 13: 405–416.

25. Hickie IB. Youth mental health: We know where we are and we can now say where we need to go next. *Early Intervent Psych* 2011; 5: 63–69.

26. Burns JM, Davenport TA, Durkin LA, Luscombe GM and Hickie IB. The Internet as a setting for mental health service utilisation by young people. *Med J Australia* 2010; 192: S22–S26.

27. Neal DM, Campbell AJ, Williams LY, Liu Y and Nussbaumer D. "I did not realize so many options are available": Cognitive authority, emerging adults, and e-mental health. *Library Inform Sci Res* 2011; 33: 25–33.

28. Metzger MJ and Flanagin AJ. *Digital Media, Youth, and Credibility*. Cambridge, MA: The MIT Press, 2008.

29. Metzger MJ and Flanagin AJ. Psychological approaches to credibility assessment online. *Handbook Psychol Commun Technol* 2015; 32: 445.

30. Pure RA, Markov AR, Mangus JM, Metzger MJ, Flanagin AJ and Hartsell EH. Understanding and evaluating source expertise in an evolving media environment. In: *Social Software and the Evolution of User Expertise: Future Trends in Knowledge Creation and Dissemination*. Hershey, PA: IGI Publishing, 2013, pp. 37–51.

31. Dutta-Bergman MJ. The impact of completeness and web use motivation on the credibility of e-health information. *J Commun* 2004; 54: 253–269.

32. Eastin MS. Credibility assessments of online health information: the effects of source expertise and knowledge of content. *J Comput Mediated Commun* 2001; 6(4): tb00126.

33. Petty RE and Cacioppo JT. The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *J Personal Soc Psychol* 1984; 46: 69–81.

34. Petty RE and Cacioppo JT. *The Elaboration Likelihood Model of Persuasion*. New York: Springer, 1986.

35. Metzger MJ. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *J Amer Soc Informat Sci Technol* 2007; 58: 2078–2091.

36. Metzger MJ, Flanagin AJ, Eyal K, Lemus DR and McCann RM. Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Commun Yearbook* 2003; 27: 293–336.

37. Rickwood D, Deane FP, Wilson CJ and Ciarrochi J. Young people's help-seeking for mental health problems. *Adv Mental Health* 2005; 4: 218–251.

38. Willemsen LM, Neijens PC and Bronner F. The ironic effect of source identification on the perceived credibility of online product reviewers. *J Comput Mediated Commun* 2012; 18: 16–31.

39. Flanagin AJ and Metzger MJ. Digital media and youth: Unparalleled opportunity and unprecedented responsibility. *John D Catherine T MacArthur Found Ser Digital Media Learning*. 2007: 5–27.

40. Metzger MJ and Flanagin AJ. Credibility and trust of information in online environments: The use of cognitive heuristics. *J Pragmatics B* 2013; 59: 210–220.

41. Fogg B, Soohoo C, Danielson DR, Marable L, Stanford J and Tauber ER. 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*. New York: ACM Press, 2003, pp. 1–15.

42. Flanagin AJ and Metzger MJ. Trusting expert- versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Comput Human Behav* 2013; 29: 1626–1634.

43. Reavley NJ, Mackinnon AJ, Morgan AJ, et al. Quality of information sources about mental disorders: A comparison of Wikipedia with centrally controlled web and printed sources. *Psychol Med* 2012; 42: 1753–1762.

44. Flanagin AJ, Metzger MJ, Pure R and Markov A. User-generated ratings and the evaluation of credibility and product quality in ecommerce transactions. In *2011 44th Hawaii International Conference on System Sciences (HICSS)*. New York: IEEE, 2011, pp. 1–10.

45. Anker AE, Reinhart AM and Feeley TH. Health information seeking: A review of measures and methods. *Patient Educ Counsel* 2011; 82: 346–54.

46. Freeman KS and Spyridakis JH. An examination of factors that affect the credibility of online health information. *Tech Commun* 2004; 51: 239–263.

47. *BluePages*. About Us, http://*BluePages*.anu.edu.au/index.php?id=the-content-of-*BluePages*2016 (accessed 10 October 2016).

48. *BluePages*. The Content of *BluePages*, http://*BluePages*.anu.edu.au/index.php?id=the-content-of-*BluePages*2016 (accessed 10 October 2016).

49. *WhatWorks4u*. About *WhatWorks4u*, http://WhatWorks4u.org/about.html2016 (accessed 10 October 2016).

50. Fox S and Jones S. *The social life of health information*. Washington, DC: Pew Internet & American Life Project, 2009.

51. Norman CD and Skinner HA. eHEALS: The eHealth literacy scale. *J Med Internet Res* 2006; 8(4): e27.

52. Norman CD. eHealth Literacy 2.0: Problems and opportunities with an evolving concept. *J Med Internet Res* 2011; 13(4): e125.

53. Jorm A, Griffiths K, Christensen H, Korten A, Parslow R and Rodgers B. Providing information about the effectiveness of treatment options to depressed people in the community: A randomized controlled trial of effects on mental health literacy, help-seeking and symptoms. *Psychol Med* 2003; 33: 1071–1079.

54. Gefen D. E-commerce: The role of familiarity and trust. *Omega* 2000; 28: 725–737.

55. Cheung MY, Luo C, Sia CL and Chen H. Credibility of electronic word-of-mouth: Informational and normative determinants of on-line consumer recommendations. *Int J Electron Commerce* 2009; 13: 9–38.

56. Eisinga R, Grotenhuis Mt and Pelzer B. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *Int J Public Health* 2013; 58(4): 637–642.

57. Flanagin AJ and Metzger MJ. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Soc* 2007; 9: 319.

58. McCroskey JC and Young TJ. Ethos and credibility: The construct and its measurement after three decades. *Commun Stud* 1981; 32: 24–34.

59. Hellmueller L and Trilling D. The credibility of credibility measures: A meta-analysis in leading communication journals, 1951 to 2011. Paper presented at the *WAPOR 65th Annual Conference*, Hong Kong, 2012.

60. Hu Y and Shyam Sundar S. Effects of online health sources on credibility and behavioral intentions. *Commun Res* 2010; 37: 105.

61. Hayes AF. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press, 2013.

62. Tabachnick BG and Fidell LS. *Using multivariate statistics*, 5th ed. Boston, MA: Allyn & Bacon/Pearson Education, 2007.

63. Paek HJ and Hove T. Social cognitive factors and perceived social influences that improve adolescent e-health literacy. *Health Commun* 2012; 27: 727–737.

64. Ghaddar SF, Valerio MA, Garcia CM and Hansen L. Adolescent health literacy: The importance of credible sources for online health information. *J School Health* 2012; 82: 28–36.

65. Sillence E, Hardy C and Briggs P. Why don't we trust health websites that help us help each other? An analysis of online peer-to-peer healthcare. In *Proceedings of the 5th Annual ACM Web Science Conference,* 2013.

66. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR and Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psych* 2005; 62: 593.

67. Stallman HM. Psychological distress in university students: A comparison with general population data. *Austral Psychol* 2010; 45: 249–257.

68. Liao QV and Fu W-T. Age differences in credibility judgments of online health information. *ACM Trans Computer–Human Interact* 2014; 21: 2.

69. Cheung CMK and Thadani DR. The effectiveness of electronic word-of-mouth communication: A literature analysis. In *BLED 2010 Proceedings*, 2010, pp. 329–345.

70. Lucassen T, Noordzij ML and Schraagen JM. Reference blindness: The influence of references on trust in Wikipedia. In *Proceedings of the ACM WebSci'11*, 2011.

71. Lucassen T, Muilwijk R, Noordzij ML and Schraagen JM. Topic familiarity and information skills in online credibility evaluation. *J Amer Soc Inform Sci Technol* 2013; 64: 254–264.

72. O'Dea B and Campbell A. Healthy connections: Online social networks and their potential for peer support. *Stud Health Technol Informatics* 2011; 168: 133–140.

73. Hargittai E, Fullerton L, Menchen-Trevino E and Thomas KY. Trust online: Young adults' evaluation of web content. *Int J Commun* 2010; 4: 468–494.

74. Wang Z, Walther JB, Pingree S and Hawkins RP. Health information, credibility, homophily, and influence via the Internet: Web sites versus discussion groups. *Health Commun* 2008; 23: 358–368.

75. Webb TL and Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychol Bull* 2006; 132: 249.