![emerald **insight**]

Journal of Documentation

Investigation of Goodreads' reviews: Kakutanied, deceived or simply honest?
Lala Hajibayova,

Article information:

Users who downloaded this article also downloaded:

,"Patterns of citations for the growth of knowledge: a Foucauldian perspective", Journal of Documentation, Vol. 0 Iss 0 pp. - https://doi.org/10.1108/JD-08-2018-0125

,"Personal information organization in everyday life: modeling the process", Journal of Documentation, Vol. 0 Iss 0 pp. - https://doi.org/10.1108/JD-05-2018-0080

![ECU LIBRARIES]

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

Investigation of Goodreads' reviews: Kakutanied, deceived or simply honest?

Lala Hajibayova
Kent State University, Kent, Ohio, USA

Abstract

Purpose – The purpose of this paper is to present an analysis of Goodreads' user-generated book reviews from a linguistic perspective for insights into the psychological aspects of reviewers' perceptions and behaviors. This examination of users' language and perspectives may shed light on the role and value of user-generated reviews in complementing the traditional representation of resources and facilitating the discoverability of cultural objects.

Design/methodology/approach – This study involved a textual analysis of 474,803 unique reviews of Goodreads' 2015 top-rated books generated by 9,335 Goodreads' reviewers. In order to better understand the nuances of user-generated reviews, a content analysis was applied to 2,500 reviews of each of the five top-ranked titles in Goodreads' Fiction Literature genre category.

Findings – The analysis of user-generated reviews demonstrates that language is a quite stable and reliable dimension across Goodreads' users. The high rate of function words utilized, in particular I-words, coupled with positive emotion words, suggests that reviewers tended to convey their opinions in order to influence other individuals' reading choices, or in Bourdieu's (1985) terms, influence cultural production. In line with previous studies of user-generated reviews, the prevalence of positive reviews may also imply their unreliable nature. This study supports the importance of transparency regarding inclusion of user-generated reviews in traditional systems of knowledge representation, organization and discovery, such as WorldCat.

Originality/value – This study contributes to better understanding of linguistic characteristics of Goodreads' reviews, including the role and value of user-generated reviews in complementing traditional representation of resources and facilitating discoverability of cultural objects.

Keywords Goodreads, Goodreads' book reviews, Linguistic analysis of user-generated reviews, Linguistic characteristics of Goodreads' reviews, Theory of cultural production, User-generated book reviews

Paper type Research paper

Introduction

In her tribute to the *New York Times'* chief daily book Critic, whose name had long ago entered the lexicon as a transitive verb, Schwartz (2017) writes that Michiko Kakutani's critical assessments of literary works have not only guided generations of readers, but "been the hope and fear of more writers than could possibly be counted – a seriously big deal, or ordeal, as the case might be" (para. 1), though to be Kakutanied always bestowed cultural prestige. However, with the development of a networked information society, the reign of a few powerful voices has given way to participatory cultures that set relatively low barriers for expression and engagement and empowered ordinary users to connect with other members of their networked community and create and share their own critical contents (Jenkins *et al.*, 2009). In particular, the user-generated reviews, as "words of mouth," presumably affect the purchasing behavior of millions of consumers (Chevalier and Mayzlin, 2006), whereas businesses believe that a corporate website must provide community content to maintain customer loyalty in a competitive environment (McWilliam, 2000; Wu *et al.*, 2015).

Studies of user-generated reviews have considered a range of analyses of what makes a review helpful (e.g. Otterbacher, 2011; Quaschning *et al.*, 2015; Suleman and Vechtomova,

2015; Willemsen *et al.*, 2011), such as predicting utility (e.g. Zhang and Varadarajan, 2006), evaluating helpfulness (e.g. Lee and Koo, 2015), predicting future changes in sales (e.g. Archak *et al.*, 2011) and evaluating linguistic characteristics of reviews (e.g. Aerts *et al.*, 2017; Schellekens *et al.*, 2010; Zhu *et al.*, 2017). Mudambi and Schuff's (2010) analysis of Amazon. com[1] reviews, for instance, revealed that depth and product type affect the perceived helpfulness of a review, and reviews giving extreme positive and negative ratings are considered less helpful than those that give moderate ratings.

As with any commodities, the perception and evaluation of cultural products have become subject to crowd evaluation (van Dijck, 2009). In particular, Goodreads[2], an Amazon-owned book-based social website, has become a popular platform for readers and writers to connect with one another, read and post reviews, and carry on discussions of literary topics. With a record of 80m members, 2.3m books and 80m reviews, Goodreads justifiably claims to be one of the world's largest sites for readers' input and book recommendations (Goodreads, n.d.). Goodreads offers its users various levels of engagements, such as the role of author and/or librarian. Goodreads' users can apply to become a Goodreads' librarian after contributing 50 book reviews, a status that allows the user to edit data on authors, books and book covers, and various editions of books (Goodreads, n.d.). Goodreads also encourages authors to join their author program, in which they can create an author page, promote their publications and engage with readers (Goodreads, n.d.). As a book-based social network site (Thelwall and Kousha, 2017), Goodreads enables its users to engage with other Goodreads' users through friending and following their activities. In particular, the Goodreads' "meet people" function allows users to explore and connect with the most followed users; read the best reviews; check out the most recent statuses and reviews; and see who is currently online through the "online now" option.

As a corporate-sponsored enterprise, Goodreads also serves publishers' and booksellers' commercial interests. The affordances of its digital labor architecture (Postigo, 2016), which emerge through direct interactions with communicative technologies (Leonardi, 2011), can turn the generated content into monetary or social capital and serve various business interests. In particular, by defining such roles as librarian and author, Goodreads provides incentives for users to increase their participation by enabling following and sharing reviews through other social network sites (such as Facebook[3]), which allows it to extract the best value from the user-generated content. In this regard, some controversy surrounds the issue of commercial benefit from volunteer digital labor, which may be conceptualized as either empowerment or exploitation (Fish and Srinivasan, 2012). As empowerment, digital labor enables volunteers to gain autonomy, skills and engagement with the community (Fish and Srinivasan, 2012), and in Shirky's (2010) terms, volunteerism constitutes a form of leisure activity that makes productive use of "cognitive surplus." Csikszentmihalyi (1975, 1990) described the flow-like experiences of chosen labor, such as enjoyment and captivation that leads to a full absorption in the engaged task (Nakamura and Csikszentmihalyi, 2009), suggesting the personal rewards of digital volunteerism. Critics of digital labor, however, claim that digital labor constitutes a modern form of exploitation, which not only pays nothing for services rendered but downgrades human labor through fragmentation of the work, disassociation of workers from the final outcome and absence of protection. Traditional labor critique, however, may not be the most salient approach to the role Goodreads plays in recruiting the public to engage as intermediaries in the process of cultural production and consumption (Bourdieu, 1985).

As a prominent literary forum, Goodreads' reviews are considered a valuable resource for library acquisitions and reference services (Thelwall and Kousha, 2017). Reviews by Goodreads' users are also recognized by the major library cataloging services, such as WorldCat[4], which provide links to the available Goodreads' user-generated reviews of their cataloged resources. Given its scope and influence, this forum is, as conceptualized by Bourdieu (1985), a large-scale cultural production which, along with being a commodity with

commercial value, possesses the cultural value of a symbolic good. Depending on whether symbolic or economic considerations were foremost, Bourdieu (1985) divided the field of cultural production into two sectors: small-scale or restricted production and the large-scale production, according to which the sector of restricted production serves producers of cultural goods, following its "own criteria for the evaluation of its products," whereas the sector of large-scale cultural production serves non-producers of cultural goods, i.e., "the public at large" (p. 17), and "submits to the laws of completion for the conquest of the largest possible market" (p. 17). Thus, small-scale production has a relatively high degree of autonomy, whereas large-scale production or mass production is heteronomous, i.e., subject to external rules. Although neither small- nor large-scale productions is fully autonomous, Bourdieu (1985) characterizes small-scale production as more oriented toward "achieving the truly cultural recognition accorded by the peer group whose members are both privileged clients and competitors" (p. 17). In this regard, the question arises as to how Goodreads' reviewers attempt to influence the cultural production and consumption.

This study, therefore, is an investigation of linguistic characteristics of Goodreads' reviews through the lens of Bourdieu's (1985) theory of cultural production. The following sections review research on user-generated reviews, explicate Bourdieu's theory of cultural production and finally present and discuss an analysis of linguistic aspects of Goodreads' book reviews.

User-generated reviews

As part of the scholarly communication process that generates and distributes new knowledge, critical reviews have long been a primary source of information about new publications (Spink *et al.*, 1998). Steiner (2008) argues that, as private criticism in the public sphere, user-generated reviews are in many ways connected to the scholarly criticism of a work. However, the anonymity presumably granted by internet platforms is associated with a sense of liberation on the part of reviewers that encourages free and frequent expression of their feeling (Katz and Rice, 2002). Besides providing either positive or negative opinions about a book, user-generated reviews also provide insights into the book's content (Kousha *et al.*, 2017). Thus, in Bourdieu' (1985) terms, reviewers can be viewed as cultural intermediaries who, as literary critics, bridge cultural production and consumption.

Researchers have investigated various aspects of Goodreads' reviews, such as their characteristics (e.g. Dimitrov *et al.*, 2015), gender differences across reviewers and reviewed genres (e.g. Thelwall, 2017a, b), and Goodreads' ratings as alternative measures (i.e. altmetrics) of the wider cultural and educational impact of publications (e.g. Kousha *et al.*, 2017; Zuccala *et al.*, 2015). For example, Kousha *et al*'s (2017) analysis revealed that users' reviews of engineering and science publications provided more information about the professional and/or technical aspects of academic books than did reviews of arts and humanities and social sciences books. Kovács and Sharkey's (2014) analysis of thousands of Goodreads' reviews of 64 English-language books between 2007 and 2011 revealed that books that won prestigious book awards tended to attract more readers following the announcement of an award, and that readers' ratings of books that were short-listed or named as finalists but did not win tended to decline precipitously following the announcement of the award-winning books.

Thelwall's (2017a) investigation of reviews of books in 50 major English book genres found gender-based differences in reviewers' attitudes toward the gender of authors in almost all categories and in the level of interest and ratings of books. However, the study did not show a clear relationship between an author's gender and his/her success or prevalence within a genre. In a subsequent study, Thelwall's (2017b) quantitative analysis of the extent to which a writer's gender influences the attitudes of male and female readers toward books within specific genres has demonstrated a strong gender differences in reviewers' ratings

within genres, such as higher ratings given by female reviewers to contemporary romance and by male reviewers to short stories.

Evidence of the influence of such factors as genre, literary awards and gender of both reviewers and authors suggests the need to investigate a common aspect across reviews. Of particular salience in this regard is what the language of users' reviews reveals about the reviewers and the nature of the reviewing enterprise as a vehicle of cultural production. Commenting on individuals' language as a valuable resource for gaining insight into their psychological characteristics, Boyd and Pennebaker (2017) described it as "reliable over time, internally consistent, predictive of a wide range of behaviors and even biological activity" (p. 64), a conviction famously promoted by Freud (1901). Language-based measures of personality can be useful than traditional personality measures for capturing/modeling lower-level personality processes that are more closely associated with important objective behavioral outcomes. In this regard, Biber (1991) argues that if "certain [linguistic] features consistently co-occur, then it is reasonable to look for an underlying functional influence that encourages their use" (p. 13). Comparing powerful and powerless speech, Lakoff (1975) proposed that the latter was characterized by the frequent use of tag questions (e.g. "isn't it?"), intensifiers (e.g. "really?") and hedges (e.g. "maybe"). In this vein, Tausczik and Pennebaker (2010) argued that patterns of attention, behaviors and sentiments are profoundly entrenched in an individual's language. From a computational linguistic perspective, Biber (1991, 1993), in a series of studies, observed significant differences across various types of texts, such as use of first person pronouns and present tense verbs in romance novels, which made them seem more personal than mysteries or any other genres.

Boyd and Pennebaker (2017) argue that in the current climate of big data, "many of the logistical properties for which self-report measures are often lauded ring even truer for language-based measures of personality" (p. 64), particularly as digital data from social platforms are "inherently ecologically valid" because they "originated from thoughts and behaviors that occur in the absence of research intervention" (p. 64). In this regard, the Linguistic Inquiry and Word Count (LIWC) text analysis application (Pennebaker *et al.*, 2015; Pennebaker, 2011) was developed to investigate how individuals' formal and informal written and spoken language reflects their psychological states. A number of studies that have utilized LIWC (e.g. Fernández-Cabana *et al.*, 2013; Ireland and Pennebaker, 2010; Schwartz *et al.*, 2013) have found, for instance, that function words, which constitute approximately of 55 percent of spoken and written language (Rochon *et al.*, 2000), not only clarify syntactical meaning, but also reflect personal and emotional features of social relationships (e.g. Chung and Pennebaker, 2007). The use of we-words as well as concrete nouns were found to be highly associated with individuals' positive experiences, whereas I-words and past- and future tense verbs often signaled individuals' attention to their feelings, especially sadness as well as implied a sense of ownership (Pennebaker, 2011). The use of second- and third-person pronouns and use of present tense was found to be associated with the negative emotions such as anger and writing about real experiences with greater length and detail, although traumatic experiences elicited fewer emotional and cognitive words (Pennebaker, 2011). Studies have also associated the frequent use of pronouns with one's greater attention to oneself or one's social world (Tausczik and Pennebaker, 2010), use of auxiliary verbs with a narrative style (Robinson *et al.*, 2013) and article use with the writer's cognitive complexity (Pennebaker and King, 1999).

Language-based measures of personality can be useful for capturing/modeling lower-level personality processes that are closely associated with important objective behavioral outcomes than traditional personality measures. We propose a complementary model that draws on a big data solution: the analysis of the words people use. Language use is

relatively reliable over time, internally consistent, and differs considerably among individuals. However, there has been little analysis of users' reviews from a linguistic perspective for insights into psychological aspects of reviewers' perceptions and behaviors. Pennebaker and King (1999) state that the "ability to isolate reliable language patterns from person to person allows some insight into the ways people perceive the world and naturally express themselves" (p. 1309). Moreover, taking into consideration the inherently textual nature of representation of resources, understanding the users' language and perspectives may shed light on the role and value of user-generated reviews in complementing traditional representation of resources and facilitating the discoverability of cultural objects.

This study, therefore, aims to investigate characteristics of Goodreads' reviews through the lens of Bourdieu's (1985) theory of cultural production, guided by the broad question: How may linguistic characteristics of Goodreads' reviews faciiltate findability, production and consumption of cultural objects?

Cultural production and cultural intermediaries

Bourdieu (1985) views the restricted and large-scale cultural productions, as complementary:

> One should beware of seeing anything more than limiting parameter construction in the opposition between the two modes of production of symbolic goods, which can only be defined in terms of their relations with each other. (p. 29)

Moreover, he claims that both modes of cultural production are subject to attribution of "*public* meaning of […] [the] work," which originates in the process of dissemination and consumption and in the achievement of cultural recognition and is dominated by various agents or cultural intermediaries, who not only endow works with certain characteristics and rankings of legitimacy, but also provide consumers with the capability and instruments for "adopting the posture [that is] designated as specific aesthetic" and "appropriation of these legitimized symbolic goods" (p. 13). Discussing the effects of audience ratings of cultural productions, Bourdieu (1985) argues that they "can and should be contested in the name of democracy" (p. 66) and that:

> […] everyone is fixated on ratings. Wherever you look, people are thinking in terms of market success. Only 30 years ago, and since the middle of the 19th century […] writers [were] acknowledged by other writers or even artists [were] acknowledged by other artists – immediate market success was suspect. It was taken as a sign of compromise with the times, with money. Today, on the contrary, the market is accepted increasingly as [a] legitimate means of legitimation. (Bourdieu, 2001, p. 251)

Viewing critics as cultural intermediaries between cultural production and consumption, Bourdieu differentiates between those of the pre-mass media age, who were the critics of serious and legitimized culture, and the new cultural intermediaries of the mass media age, who represent a "new petite bourgeoisie" with idiosyncratic cultural tastes and practices (Hesmondhalgh, 2006). Bourdieu (2001) claims that while the audience ratings "impose the sales model on cultural products" […] , "historically, all of the cultural productions […] the highest human products – maths, poetry, literature, philosophy – were all produced against market imperatives" (p. 251).

Viewed through the lens of Bourdieu's (1992) theory of practice, the language of Goodreads' reviewers represents "situated encounters between agents endowed with socially structured resources and competencies, in such a way that every linguistic interaction […] bears the traces of the social structure that it both expresses and helps to reproduce" (p. 2). Language as a form of practice is an integral part of individuals' habitus, i.e., "a set of *dispositions* which

JD

incline agents to act and react in certain ways" (p. 12), or more specifically, a linguistic habitus that denotes a sub-set of the dispositions that agents' attain in the course of learning to communicate in certain contexts (p. 17). Along these lines, Wittgenstein (1953/1963) contends that because language, particularly, the meaning it conveys, is manifested in the forms of life in which the individual participates, it is intimately interwoven with the individual's particular world view.

This study explores Goodreads' reviewers' linguistic habitus to understand how the language of reviewers of top-rated Goodreads' books may facilitate the discoverability and consumption of cultural objects.

Methodology

Sample

To ensure a representative sample of popular and commonly reviewed titles, a sample of reviews of 20 first titles was drawn from Goodreads' Best Books of 2014 during the period of December 2015–January 2016. The sample included 18 Goodreads' genres: Business, Debut Goodreads' Authors, Fantasy, Fiction, Food and Cookbooks, Graphic Novels and Comics, Historical Fiction, History and Biography, Horror, Humor, Memoir and Autobiography, Mystery and Thriller, Non-Fiction, Poetry, Romance, Science Fiction, Young Adult Fantasy and Science Fiction, Young Adult Fiction. In total, the sample comprised of 474,803 unique reviews generated by 9,335 Goodreads' reviewers.

Analysis

This exploratory study utilized both human and machine textual analysis of Goodreads-generated reviews. LIWC software was utilized for textual analysis of the reviews. LIWC has been widely applied (e.g. McHaney *et al.*, 2018; Tausczik and Pennebaker, 2010) and is considered an effective tool for measuring the psychological aspects of language (e.g. Iliev *et al.*, 2015; Kahn *et al.*, 2007). The LIWC application includes a processing component, which analyzes text files and dictionaries, is comprised of collections of words that define particular categories that are used to compare processed text files (Tausczik and Pennebaker, 2010). For each given text, LIWC analysis yields approximately 90 output variables, including word count, four summary language variables (analytical thinking, clout, authenticity and emotional tone), three general descriptor categories (words per sentence, percent of target words captured by the dictionary and percent of words in the text that are longer than six letters), 21 standard linguistic dimensions (such as percentage of words in the text that are pronouns, articles, auxiliary verbs), 41 word categories tapping psychological constructs (such as affect, cognition, biological processes, drives), six personal concern categories (such as work, home, leisure activities), five informal language markers (such as assents, fillers, swear words, netspeak) and 12 punctuation categories (such as periods, commas). A *LIWC Dictionary* is composed of almost 6,400 words, word stems and select emoticons. Each dictionary entry additionally defines one or more word categories or subdictionaries. For example, the word "mourned" is included in five word categories: sadness, negative emotion, overall affect, verbs and past focus. A *LIWC Dictionary* also captures word stems. The LIWC application assigns each word into defined linguistic categories and calculates the total number of words in each category in relation to the total number of words in the given text. For example, the denotation of 4.50 for positive emotions indicates the percentage of positive emotion words among all the words in the analyzed text (Pennebaker *et al.*, 2015).

Having obtained a large number of variables with LIWC analysis, was applied factor analysis, which estimates the correlations among all the variables and provides a set of factor loadings. The correlation of an individual variable with a factor was considered meaningful if it had a loading above 0.30 (Aron *et al.*, 2006).

Because of the popularity of the fiction genre, coupled with its limited representation in traditional systems of representation and organization (Mikkonen and Vakkari, 2016) the constructivist grounded theory (Charmaz, 2014) was applied to closely analyze the top 500 of 2,500 user-generated reviews of Goodreads' five top fiction titles for 2014: *Big Little Lies* by Liane Moriarty, *Landline* by Rainbow Rowell, *Leaving Time* by Jodi Picoult, *One Plus One* by Jojo Moyes and *The Storied Life of A.J. Fikry* by Gabrielle Zevin). A coding scheme was developed by two coders, one of whom was predominantly engaged in the coding process and met weekly with the second coder over a three-month period to discuss emerging themes and refine categories as needed. This approach allowed the development of thematic codes and formulation of connections between various socio-technical elements that make up the reviewer experience of reading and commenting on Goodreads (see Table I).

Fitting each review into one category was challenging; therefore, all reviews were coded across all relevant categories. In order to assess agreement between two coders, Cohen's (1960) κ inter-rater agreement coefficient was applied to coding of the top 100 reviews of each of the five titles (500 or 20 percent of the sample). The overall κ coefficient for inter-rater agreement was $k = 0.75$. According to Landis and Koch's (1977) standards for assessing strength of agreement of the κ coefficient: from 0.40–0.59 indicates moderate agreement, 0.60–.79 indicates substantial agreement, and over 0.80 indicates outstanding agreement. Thus, the κ coefficient for inter-rater agreement was substantial ($k = 0.75$).

Results and discussions

The analysis of 474,803 reviews revealed an average of 45 words per review ($M = 45$, $SD = 162.89$). In keeping with overall trends, users preferred using the social button to convey "like" sentiments ($SD = 1,144$) to posting comments ($SD = 355$).

Reviews for the Young Adult Fiction genre tended to have the highest number of words ($M = 64$) and those for the Mystery genre the least number ($M = 37$). Pennebaker (2011) observed that the older the individuals the more words they use (such as in blogs); thus, it is likely that reviewers of Young Adult Fiction were relatively older individuals and Mystery reviewers relatively younger. In line with previous studies (e.g. Ye *et al.*, 2011), the analysis of reviews revealed the affective nature of user-generated reviews ($M = 7.69$, $SD = 9.3$), with

Codes	Code definition and example
Emotions	Affective tone of the reviews
Positive	Positive affective terms, e.g., "enjoyable", "sweet", "loved it"
Negative	Negative affective terms, e.g., "dislike", "I hated this book"
Neutral	Neutral tone of the review, e.g., "[…] a story that uses class differences as a plot device", "Here is what about this story"
Mixed	Mix expression of positive, negative and/or neutral emotions, e.g., "enjoyable read […] half of the book tedious"
Evaluation of author's work	Judgment of author's writing, such as "author is incredibly accurate in […]", "author's snarky commentary about"
Personal experiences/ associations	Personal expressions associated with reading, such as "their journey feels as if it was my own," "when I finished reading […] I didn't know what to do with myself"
Perceived Problems	Expression of problems with the writing, such as "corny dialogue and predictable tropes", "insanely overhyped fluff"
Storyline	Description of storyline, such as "The story evolves around […]," "centers on the lives of a group of […]"
Rating	Provision of rating or explaining the thinking behind rating, such as "I usually reserve my five star ratings for […]," "I'm torn between a 3 and a 4 star for this book"
Character Discussion	Reference to and discussion of characters, such as "Moriarty handles these […]", "[…] brings Fikry and Amelia closer […]"

Table I.
Coding schema

positive emotions ($M = 5.59$, SD $= 8.6$) expressed more often than negative ones ($M = 0.83$, SD $= 4.0$). The most positive reviews were of titles in the Humor genre ($M = 7.63$), whereas the most negative reviews were of titles in Horror genre ($M = 1.44$).

Taking into consideration the importance of the eight types of function words (i.e. personal and impersonal pronouns, articles, prepositions, auxiliary verbs, common adverbs, conjunctions and negations) in understanding social psychological process (Tausczik and Pennebaker, 2010), the LIWC analysis found a total of 231,496 function words or 48.8 percent of all words used in the reviews. Young Adult Fiction reviews had the highest percentage of function words ($M = 54.17$) and relatively high use of pronouns (such as I, them, itself) ($M = 16.3$). Personal pronouns had a relatively high use in reviews of Humor ($M = 8.82$) and Romance ($M = 8.67$) titles, whereas I-words were highly associated with Romance and Young Adult Fiction genres ($M = 5.36$ and $M = 5.26$, respectively). Given that high use of function words and especially pronouns are associated with sincerity, we may infer that reviews of Young Adult Fiction titles were likely to be candid, reflecting Pennebaker's (2011) assertion that "when telling the truth, we tend to "own" what we say" (p. 294).

Pennebaker (2011) argues that I-words have particular social and psychological significance as they signify statements of identity, in which the speaker announces his/her self-awareness and self-focus, an orientation more characteristic of women than of men. Wicklund (1975) associates such self-referencing with honesty and humbleness, and Kacewicz *et al.* (2014) link it with lower status, suggesting that "those in a higher position in the hierarchy are more other- focused, whereas those lower in the social hierarchy are more self-focused as gauged by the use of personal pronouns" (p. 138). Thus, the findings of this study imply that the sampled reviews associated of top-ranked titles were provided predominantly by women who honestly shared their personal reading experiences.

In particular, the relatively high use of "she/he" references in the Memoir and Autobiography ($M = 1.26$) and Humor ($M = 0.58$) genre reviews suggest reviewers' immersion in the task and/or expression of something that made them emotional (such as happy, furious or sad). Reviewers also utilized the impersonal pronouns (such it, its, those) ($M = 6.25$, SD $= 5.11$), which could be associated with self-deception as there is greater narrative distance between the self that narrates and the self that is narrated about in deceptive narratives (Pennebaker, 2011). Moreover, narrative selves are constituted not as autonomous selves, but as subject to processes (e.g. psychological, social) that are most likely operating on a subconscious level (Bedwell *et al.*, 2011). Across the sampled genres, impersonal pronouns were relatively highly associated with Horror and Romance genres ($M = 6.59$ and $M = 6.57$, respectively).

Using LIWC, the rates of the eight functions (personal pronouns, impersonal pronouns, articles, prepositions, auxiliary verbs, adverbs, conjunctions, and negations) and the affective (positive and negative emotions) word categories were computed separately for each review, and the analysis revealed that the rates of use of each of the function and affective word categories were positively correlated across all reviews. A diagnostic test indicated that a factor model was appropriate for the data (KMO $= 0.56$, Bartlett's test of sphericity $= 293,701$, $p < 0.001$).

A principal component analysis on ten dimensions yielded four factors (eigenvalues above 1), and varimax rotation was used to facilitate the interpretation of factors (Pennebaker and King, 1999) (see Table I). The first factor (eigenvalue $= 1.725$) included three high power sources, i.e., personal pronouns (0.589), impersonal pronouns (0.627), auxiliary verbs (0.733), negations (0.494) and a small loading of positive emotions (−0.301); and the second factor (eigenvalue $= 1.507$) included two high loadings of prepositions (0.736) and articles (0.598), and a small loading of positive emotions (−0.686) (see Table II). Correlation analysis revealed strong correlations between factors ($r = 0.539$, $p < 0.01$).

The first factor of high loading of personal and impersonal pronouns, auxiliary verbs and negations reflects the overall extensive use of function words in the English language (Chung and Pennebaker, 2007). The high rate of function words in the reviewer's vocabulary is also associated with younger women of a lower status, and use of pronouns and auxiliary verbs suggests individuals' social orientation (Pennebaker, 2011). The overly enthusiastic writings are associated with many forms of deception (Pennebaker, 2011), an outcome which has also been reported in the studies of online reviews (e.g. Zhou *et al.*, 2004). Yoo and Gretzel's (2009) analysis of travel reviews, for instance, found that frequent use of first person pronouns and positive sentiments signal deception in reviews. Relatively high use of function words coupled with a positive tone may also imply reviewers' attempts to promote their reviews and influence one's reading choices.

Although the use of articles may not reveal which objects or events a person is talking about, it may provide a sense of his/her general approach to the world, that is, serve as a reference to concrete and impersonal objects and/or events (Pennebaker and King, 1999). Thus, the second factor implies some association between positive emotions that emerge and the object, that is, the reviewed titles. The third factor represents users' causal thinking as such conjunction words as because and why connote attempts to explain causes and effects (e.g. because and why) (Pennebaker, 2011). The high loading of negative emotions may imply that some of the reviews were provided by relatively young users and most likely by women, who tend to use negative emotion words more often (Pennebaker, 2011).

Content analysis
In order to understand the nuances of user-generated reviews, a close analysis of the first 500 reviews of each of the five top-ranked titles (*Big Little Lies* by Liane Moriarty, *Landline* by Rainbow Rowell, *Leaving Time* by Jodi Picoult, *One Plus One* by Jojo Moyes and *The Storied Life of A.J. Fikry* by Gabrielle Zevin) in Goodreads' Fiction Literature genre category was conducted. The analysis has demonstrated the prevalence of positive reviews (2,002, 80.08 percent) with some negative (175, 7 percent), neutral (106, 4.24 percent) or mixed (217, 8.68 percent) reviews of the titles. However, while the sample reviews most often reflected positive emotions, negative and blended positive and negative emotions associated with the reviewers' reading experiences were also found, as exemplified in the following comment: "3.5 stars […] As always, […] created another great, realistic cast of characters. But this plot I felt was a little bit dry and in the end, […] there were several loose ends. My least favorite […]."

Power sources by factor	Loadings			
	Factor 1	Factor 2	Factor 3	Factor 4
Auxiliary verbs	0.733*	0.130	0.059	−0.047
Impersonal pronouns	0.627*	−0.177	−0.076	−0.231
Personal pronouns	0.589*	0.041	0.091	0.009
Negations	0.494*	−0.035	−0.027	0.207
Prepositions	0.023	0.736*	0.024	−0.099
Positive emotions	−0.301*	−0.686*	0.012	−0.271
Article	−0.291	0.598*	−0.094	−0.143
Conjunctions	0.015	0.147	0.803*	0.042
Adverbs	0.043	−0.212	0.753*	−0.028
Negative emotions	−0.035	−0.065	0.013	0.921*
Note: *Denotes significant loadings				

Table II.
Factor analysis loadings for the ten power sources

Emotions associated with and/or emerging from the reading constituted one of the main themes of the reviews, followed by evaluation of author's current and/or previous writings, personal experiences and associations with other titles or movies that seemed relevant to the reviewer, and storyline or "spoiler alerts," and perceived problems with the title and ratings (see Table III).

However, often reviews were blends of these themes. For instance, one reviewer's discussion of an author's work referred to the author's writing style, her previous publications and the reviewer's own evaluation, feelings and associations with other relevant artistic expressions: "Fabulous book by […] this is an author that can really create engaging characters […] a lot of writers seem to have a discernible style […] but this author just writes books that are believable, witty, sad and always leave me thinking about the book […]."

As observed in the machine analysis of reviews, users' personal experiences were reflected in the extensive application of I-words: "I'm not usually one for books about […], but I am highly susceptible to books everyone seems to love […] And I'm so glad I was curious about this book, because I really ended up loving it […]."

Goodreads' users have also pursued their own interest by directing readers either to their own blog or YouTube channel ("There will be a more in-depth review on my YouTube channel!) as well as publisher's interests when a review was provided in return for access to an electronic copy ("An ecopy of the book was provided by the publisher through […] in return for a review"). There were also more direct provisions of ratings and recommendations to read (or not to read) a title: "4.5 loved the story and characters […]."

Taken together, both human and machine analyses of user-generated reviews suggest that Goodreads' users' language provides useful insights into their psychological experiences as readers and members of a broadly defined reading community. The high rate of function words utilized, in particular I-words, coupled with positive emotion words, suggests that reviewers tended to convey their opinions in order to influence other individuals' reading choices or in Bourdieu's (1985) terms influence cultural consumption. In line with the previous studies of user-generated reviews (e.g. Yoo and Gretzel, 2009), the prevalence of positive reviews may also imply the unreliable nature of such reviews.

Overall, Goodreads affordances seemed to be effectively utilized to serve individuals' interests by enabling them to not only to express their reading experiences but also contribute to the public perception of a cultural work. A characteristic blend of personal sentiments and experiences makes user-generated reviews a good complementation to traditional, rather formal, representation of cultural expressions. However, the relatively high use of function words coupled with predominantly positive tone of reviews found in this study can be interpreted as individuals' attempts to "sell" their reviews and influence the overall process of dissemination and consumption of cultural goods.

Themes	Title No. 1[a]	Title No. 2[b]	Title No. 3[c]	Title No. 4[d]	Title No. 5[e]
Emotions	317 (63.4%)	295 (59%)	274 (54.8%)	307 (61.4%)	289 (57.8%)
Evaluation of author's work	477 (95.4%)	287 (57.4%)	102 (20.4%)	241 (48.2%)	183 (36.6%)
Personal experiences/associations	165 (33%)	172 (34.4%)	188 (37.6%)	204 (40.8%)	279 (55.8%)
Perceived problems	98 (19.6%)	77 (15.4%)	36 (7.2%)	18 (3.6%)	66 (13.2%)
Storyline	98 (19.6%)	119 (23.8%)	128 (25.6%)	144 (28.8%)	125 (25%)
Rating	98 (19.6%)	77 (15.4%)	76 (15.2%)	64 (12.8%)	123 (24.6%)
Character discussion	54 (10.8%)	0	0	21 (4.2%)	96 (19.2%)

Notes: $n = 2,500$. [a]*Landline*; [b]*Big Little Lies*; [c]*The Storied Life of A.J. Fikry*; [d]*One Plus One*; [e]*Leaving Time*

Table III.
Main themes
across five titles

This study suggests that algorithmic transparency regarding the inclusion of user-generated reviews in traditional systems of information representation and organization, such as WorldCat, as well the overall trend toward persuasion in public meanings or rankings of cultural products, in Bourdieu's (1985) words, "can and should be contested in the name of democracy" (p. 66).

Conclusions

The Goodreads' architecture provides the socio-technical framing that supports the enterprise of making meaning of users' reading and reviewing culture. Both leisure and labor are framed in Goodreads' architecture, and their meanings and value are shaped by users' subjective positions, the kinds of communications afforded, the communities supported, the algorithms in place to account for social or monetary capital and the social practices engaged (Postigo, 2016).

This paper demonstrates that Goodreads' users' language is quite stable across individuals and may serve as a window into their psychological perspectives as expressed in their reviews. For instance, the presence of a negative tone in reviews may imply that they were composed by relatively young users and most likely young women, who as a group have been found to favor of using negative emotion words (Pennebaker, 2011). The observed high rate of function words coupled with copious use of pronouns and auxiliary verbs may suggest reviewers' social orientation (Pennebaker, 2011), whereas the use of function words and overall positive tone of the reviews might be interpreted as reviewers' attempts to market their reviews and influence readers' cultural consumption and overall cultural production. As Goodreads' reviews have become a part of individuals' daily interactions with knowledge representation, organization and discovery systems, the patterns that emerged in this study scale up the potential influence of Goodreads' reviews on cultural consumption and production. Moreover, they raise concerns about an algorithmically reinforced "sales model on cultural products" (Bourdieu, 2001) as well as the cultural prestige of Goodreads' reviews and their value in providing the capability and instruments for enacting Bourdieu's (1985) notion of "adopting the posture [that is] designated as specific aesthetic" and "appropriation of these legitimized symbolic goods" (p. 13). On a positive note, we hope that one day that having one's work subject to a Goodreads' review may also enter the lexicon as a verb – to be "Goodreads" – to signify not only public ranking, but, most importantly, the ultimate cultural prestige that might, as quoted from Schwartz (2017) in the introduction of this paper, be "scary to writers" but reliable to readers, as one cannot "easily predict where […][the] favor would fall" (para. 1); however, for now, we call for readers and professionals to be mindful of a potentially deceptive aspect of user-generated reviews.

This study has a number of limitations that should be addressed in further research. First, machine algorithms for detecting sentiments might not be very accurate, particularly when it comes to individuals' language use, such as the algorithms' failure to detect contexts in which terms and expressions are used ironically or sarcastically (Wang *et al.*, 2014). For example, such methods of semantic analysis might not interpret as ironic such expressions as "I am fired, great!" (Wang *et al.*, 2014). There might be also limitations in detecting all variations of emoticons, abbreviations and fashionably misspelled words that are widely used in the online environment (Wang *et al.*, 2014).

This study is also subject to the limitations of exploratory factor analysis as utilized in this study. For example, while a factor many accurately reflect the variables within the factor, they could be difficult to interpret due to the fact that they may load onto more than one factor and may correlate with each another to produce a factor despite sharing little underlying meaning for the factor (Yong and Pearce, 2013). Therefore, future studies should consider variations in factor loading across and within genres.

JD

This exploratory study presents analysis of a one-year period of Goodreads' reviews and close analysis of 2,500 reviews associated with only five titles. Further studies involving a longitudinal analysis of Goodreads' data to provide a comprehensive understanding of Goodreads' reviews' patterns and users' behaviors across a longer span of time are needed to better understand trends and patterns in user-generated book reviews and their influence on the process of cultural production.

Notes

1. Amazon.com: www.amazon.com

2. Goodreads.com: www.goodreads.com

3. Facebook.com: www.facebook.com

4. WorldCat: www.worldcat.org

References

Aerts, G., Smits, T. and Verlegh, P.W.J. (2017), "How online consumer reviews are influenced by the language and valence of prior reviews: a construal level perspective", *Computer in Human Behavior*, Vol. 75, pp. 855-864, available at: www.sciencedirect.com/science/article/pii/S07475632 17303941

Archak, N., Ghose, A. and Ipeirotis, P.G. (2011), "Deriving the pricing power of product features by mining consumer reviews", *Management Science*, Vol. 57 No. 8, pp. 1485-1509.

Aron, A., Aron, E.N. and Coups, E.J. (2006), *Statistics for Psychology*, Pearson, Upper Saddle River, NJ.

Bedwell, J.S., Gallagher, S., Whitten, S.N. and Fiore, S.M. (2011), "Linguistic correlates of self in deceptive oral autobiographical narratives", *Consciousness and Cognition*, Vol. 20 No. 3, pp. 547-555.

Biber, D. (1991), *Variations across Speech and Writing*, Cambridge University Press, Cambridge.

Biber, D. (1993), "Using register-diversified corpora for general language studies", *Journal of Computational Linguistics*, Vol. 19 No. 2, pp. 219-241.

Bourdieu, P. (1985), "The market of symbolic goods", *Poetics*, Vol. 14 Nos 1/2, pp. 13-44.

Bourdieu, P. (1992), *Language and Symbolic Power*, Harvard University Press, Cambridge, MA.

Bourdieu, P. (2001), "Television", *European Review*, Vol. 9 No. 3, pp. 245-256.

Boyd, R.L. and Pennebaker, J.W. (2017), "Language-based personality: a new approach to personality in a digital world", *Current Opinion in Behavioral Sciences*, Vol. 18, pp. 63-68, available at: www.sciencedirect.com/science/article/pii/S2352154617300487

Charmaz, K. (2014), *Constructing Grounded Theory*, Sage, Thousand Oaks, CA.

Chevalier, J.A. and Mayzlin, D. (2006), "The effect of word of mouth on sales: online book reviews", *Journal of Marketing Research*, Vol. 43 No. 3, pp. 345-354.

Chung, C. and Pennebaker, J. (2007), "The psychological functions of function words", in Fiedler, K. (Ed.), *Social Communication*, Psychology Press, New York, NY, pp. 343-360.

Cohen, J. (1960), "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, Vol. 20 No. 1, pp. 37-46.

Csikszentmihalyi, M. (1975), *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*, Jossey-Bass, San Francisco, CA.

Csikszentmihalyi, M. (1990), *Flow: The Psychology of Optimal Experience*, Harper & Row, New York, NY.

Dimitrov, S., Zamal, F., Piper, A. and Ruths, D. (2015), "Goodreads versus Amazon: the effect of decoupling book reviewing and book selling", *Proceedings of the Ninth International AAAI Conference on Web and Social Media in Oxford, England*, AAAI Press, Palo Alto, CA, pp. 602-605, available at: www.aaai.org/ocs/index.php/ICWSM/ICWSM15/rt/captureCite/10557/0

Fernández-Cabana, M., García-Caballero, A., Alves-Pérez, M.T., García-García, M.J. and Mateos, R. (2013), "Suicidal traits in Marilyn Monroe's Fragments: an LIWC analysis", *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, Vol. 34 No. 2, pp. 124-130.

Fish, A. and Srinivasan, R. (2012), "Digital labor is the new killer app", *New Media & Society*, Vol. 14 No. 1, pp. 137-152.

Freud, S. (1901), *The Psychopathology of Everyday Life*, White Press, Freiberg, available at: Amazon.com (accessed June 1, 2018).

Goodreads (n.d.), "About Goodreads", available at: www.goodreads.com/about/us?rel=nofollow (accessed June 1, 2018).

Hesmondhalgh, D. (2006), "Bourdieu, the media and cultural production", *Media, Culture & Society*, Vol. 28 No. 2, pp. 211-231.

Iliev, R., Dehghani, M. and Sagi, E. (2015), "Automated text analysis in psychology: methods, applications, and future developments", *Language and Cognition*, Vol. 7 No. 2, pp. 265-290.

Ireland, M.E. and Pennebaker, J.W. (2010), "Language style matching in writing: synchrony in essays, correspondence, and poetry", *Journal of Personality and Social Psychology*, Vol. 99 No. 3, pp. 549-571.

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K. and Robinson, A. (2009), *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, MIT Press, Cambridge, MA.

Kacewicz, E., Pennebaker, J.W., Davis, M., Jeon, M. and Graesser, A.C. (2014), "Pronoun use reflects standings in social hierarchies", *Journal of Language and Social Psychology*, Vol. 33 No. 2, pp. 125-143.

Kahn, J.H., Tobin, R.M., Massey, A.E. and Anderson, J.A. (2007), "Measuring emotional with the linguistic inquiry and word count", *The American Journal of Psychology*, Vol. 120 No. 2, pp. 263-286.

Katz, J.E. and Rice, E.R. (2002), *Social Consequences of Internet Use: Access, Involvement, and Interaction*, MIT Press, Cambridge, MA.

Kousha, K., Thelwall, M. and Abdoli, M. (2017), "Goodreads reviews to assess the wider impacts of books", *Journal of the Association for Information Science & Technology*, Vol. 68 No. 8, pp. 2004-2016.

Kovács, B. and Sharkey, A.J. (2014), "The paradox of publicity: how awards can negatively affect the evaluation of quality", *Administrative Science Quarterly*, Vol. 59 No. 1, pp. 1-33.

Lakoff, R.T. (1975), *Language and A Woman's Place*, Harper & Row, New York, NY.

Landis, J. and Koch, G. (1977), "The measurement of observer agreement for categorical data", *Biometrics*, Vol. 33 No. 1, pp. 159-174.

Lee, K.-T. and Koo, D.-M. (2015), "Evaluating right versus just evaluating online consumer reviews", *Computers in Human Behavior*, Vol. 45, pp. 316-327, available at: www.sciencedirect.com/science/article/pii/S0747563214007535

Leonardi, P.M. (2011), "When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies", *MIS Quarterly*, Vol. 35 No. 1, pp. 147-167.

McHaney, R., Tako, A. and Robinson, S. (2018), "Using LIWC to choose simulation approaches. A feasibility study", *Decision Support Systems*, Vol. 111, pp. 1-12, available at: www.sciencedirect.com/science/article/pii/S0167923618300691

McWilliam, G. (2000), "Building stronger brands through online communities", *Sloan Management Review*, Vol. 41 No. 3, pp. 43-54.

Mikkonen, A. and Vakkari, P. (2016), "Readers' interest criteria in fiction book search in library catalogs", *Journal of Documentation*, Vol. 72 No. 4, pp. 696-715.

Mudambi, S.M. and Schuff, D. (2010), "What makes a helpful online review? A study of customer reviews on Amazon.com", *MIS Quarterly*, Vol. 34 No. 1, pp. 185-200.

Nakamura, J. and Csikszentmihalyi, M. (2009), "Flow theory and research", in Snyder, C.R. and Lopez, S.J. (Eds), *Oxford Handbook of Positive Psychology*, Oxford University Press, Oxford, pp. 195-206.

Otterbacher, J. (2011), "Being heard in review communities: communication tactics and review prominence", *Journal of Computer-Mediate Communication*, Vol. 16 No. 3, pp. 424-444.

Pennebaker, J. (2011), *The Secret Life of Pronouns: What Our Words Say About Us*, Bloomsbury Press, New York, NY.

Pennebaker, J.W. and King, L.A. (1999), "Linguistic styles: language use as an individual difference", *Journal of Personality and Social Psychology*, Vol. 77 No. 6, pp. 1296-1312.

Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. (2015), *The Development and Psychometric Properties of LIWC*, University of Texas at Austin, Austin, TX.

Postigo, H. (2016), "The socio-technical architecture of digital labor: converting play into YouTube money", *New Media & Society*, Vol. 18 No. 2, pp. 332-349.

Quaschning, S., Pandelaere, M. and Vermeir, I. (2015), "When consistency matters: the effect of valence consistency on review helpfulness", *Journal of Computer-Mediated Communication*, Vol. 20 No. 2, pp. 136-152.

Robinson, R.L., Navea, R. and Ickes, W. (2013), "Predicting final course performance from students' written self-introductions: a LIWC analysis", *Journal of Language and Social Psychology*, Vol. 32 No. 4, pp. 481-491.

Rochon, E., Saffran, E.M., Berndt, R.S. and Schwartz, M.F. (2000), "Quantitative analysis of aphasic sentence production: further development and new data", *Brain and Language*, Vol. 72 No. 3, pp. 193-218.

Schellekens, G.A.C., Verlegh, P.W.J. and Smidts, A. (2010), "Language abstraction in world of mouth", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 207-223.

Schwartz, A. (2017), "Farewell, Michiko Kakutani!", *The New Yorker*, available at: www.newyorker.com/books/page-turner/farewell-michiko-kakutani (accessed December 1, 2018).

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P. and Ungar, L.H. (2013), "Personality, gender, and age in the language of social media: the open-vocabulary approach", *PLoS ONE*, Vol. 8 No. 9, e73791, available at: https://doi.org/10.1371/journal.pone.0073791 (accessed June 10, 2018).

Shirky, C. (2010), *Cognitive Surplus: Creativity and Generosity in a Connected Age*, Penguin, London.

Spink, A., Robins, D. and Schamber, L. (1998), "Use of scholarly book reviews: implications for electronic publishing and scholarly communication", *Journal of the Association for Information Science and Technology*, Vol. 49 No. 4, pp. 364-374.

Steiner, A. (2008), "Private criticism in the public space: personal writing on literature in readers' reviews on Amazon", *Particip@ations*, Vol. 5 No. 2, available at: www.participations.org/Volume%205/Issue%202/5_02_steiner.htm (accessed June 10, 2018).

Suleman, K. and Vechtomova, O. (2015), "Discovering aspects of online consumer reviews", *Journal of Information Science*, Vol. 42 No. 4, pp. 492-506.

Tausczik, Y.R. and Pennebaker, J.W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54.

Thelwall, M. (2017a), "Book genre and author gender: romance > paranormal–romance to autobiography > memoir", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 5, pp. 1212-1223.

Thelwall, M. (2017b), "Reader and author gender and genre in Goodreads", *Journal of Librarianship and Information Science*, available at: https://journals.sagepub.com/doi/10.1177/0961000617709061 (accessed June 1, 2018).

Thelwall, M. and Kousha, K. (2017), "Goodreads: a social network site for book readers", *Journal of the Association for Information Science & Technology*, Vol. 68 No. 4, pp. 972-983.

van Dijck, J. (2009), "Users like you? Theorizing agency in user-generated content", *Media, culture & society*, Vol. 31 No. 1, pp. 41-58.

Wang, N., Kolsinski, M., Stillwell, D.J. and Rust, J. (2014), "Can well-being be measured using Facebook status updates? Validation of Facebook's Gross National Happiness Index", *Social Indicators Research*, Vol. 15 No. 115, pp. 483-491.

Wicklund, R.A. (1975), "Objective self-awareness", *Advances in Experimental Social Psychology*, Vol. 8, pp. 233-275, available at: www.sciencedirect.com/science/article/pii/S006526010860252X

Willemsen, L.M., Neijens, P.C. and de Ridder, J.A. (2011), "'Highly recommended!' the content characteristics and perceived usefulness of online consumer reviews", *Journal of Computer-Mediated Communication*, Vol. 17 No. 1, pp. 19-38.

Wittgenstein, L. (1953/1963), *Philosophical Investigations*, (translated by Anscombe, G.E.M.), Macmillan Publishing, New York, NY.

Wu, J., Huang, L., Zhao, J.L. and Hua, Z. (2015), "The deeper, the better? Effect of online brand community activity on customer purchase frequency", *Information & Management*, Vol. 52 No. 7, pp. 813-823.

Ye, Q., Law, R., Gu, B. and Chen, W. (2011), "The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings", *Computers in Human Behavior*, Vol. 27 No. 2, pp. 634-639.

Yong, A.G. and Pearce, S. (2013), "A beginner's guide to factor analysis: focusing on exploratory factor analysis", *Tutorials in Quantitative Methods for Psychology*, Vol. 9 No. 2, pp. 79-94.

Yoo, K.-H. and Gretzel, U. (2009), "Comparison of deceptive and truthful travel reviews", in Höpken, W., Gretzel, U. and Law, R. (Eds), *Proceedings of Sixteenth International Conference on Information and Communication Technologies in Travel and Tourism in Amsterdam*, Springer-Verlag, Wien, Amsterdam, pp. 37-47.

Zhang, Z. and Varadarajan, B. (2006), "Utility scoring of product reviews", *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*, ACM, New York, Arlington, VA, pp. 51-57.

Zhou, L., Burgoon, J.K., Nunamaker, J.F. and Twitchell, D. (2004), "Automating linguistic-based cues for detecting deception in text-based asynchronous computer-mediated communication", *Group Decision and Negotiation*, Vol. 13 No. 1, pp. 81-106.

Zhu, D.H., Ye, Z.Q. and Chang, Y.P. (2017), "Understanding the textual content of online customer reviews in B2C websites: a cross-cultural comparison between the US and China", *Computers in Human Behavior*, Vol. 76, pp. 483-493, available at: www.sciencedirect.com/science/article/pii/S0 74756321730465X

Zuccala, A., Verleysen, F.T., Cornacchia, R. and Engels, T.C. (2015), "Altmetrics for the humanities: comparing Goodreads reader ratings with citations to history books", *Aslib Journal of Information Management*, Vol. 67 No. 3, pp. 320-336.

Further reading

Verboord, M. (2010), "The legitimacy of book critics in the age of the internet and omnivorousness: expert critics, internet critics and peer critics in Flanders and the Netherlands", *European Sociological Review*, Vol. 26 No. 6, pp. 623-637.

Corresponding author
Lala Hajibayova can be contacted at: lhajibay@kent.edu