Summer 6-15-2016

INFORMATION QUALITY NEEDS THROUGHOUT THE PURCHASE PROCESS

Stefan Hirschmeier
University of Cologne, hirschmeier@wim.uni-koeln.de

Roman Tilly
University of Cologne, tilly@wim.uni-koeln.de

Detlef Schoder
University of Cologne, schoder@wim.uni-koeln.de

INFORMATION QUALITY NEEDS THROUGHOUT THE PURCHASE PROCESS

Research

Hirschmeier, Stefan, University of Cologne, Köln, Germany, hirschmeier@wim.uni-koeln.de

Tilly, Roman, University of Cologne, Köln, Germany, tilly@wim.uni-koeln.de

Schoder, Detlef, University of Cologne, Köln, Germany, schoder@wim.uni-koeln.de

Abstract

Information quality (IQ) plays a critical role in ecommerce consumers' purchase decisions, and product-related user-generated content (PUGC) is an increasingly important source of information. Nevertheless, vendors currently do not take full advantage of PUGC since it is mainly used in later phases of the purchase process to evaluate single products. Integrating PUGC into earlier phases of the purchase process might be promising. However, little is known about which dimensions of IQ consumers evaluate as being most important in different phases of the purchase process. To close this gap, this study comprises an exploratory survey to investigate the perceived importance of various IQ dimensions in each of the three phases of an idealized purchase process (i.e., screening, filtering, and evaluation). Conceptually, this study thus extends the concept of IQ to a dynamic perspective. The results indicated that IQ needs are stable throughout the process and that users strongly favor accuracy and believability when using aggregated PUGC extracted from reviews in contrast to timeliness, completeness, and amount of data, which were assessed as the least important IQ dimensions. Practically, the findings can inform the design of ecommerce websites, specifically how to optimize the presentation and integration of PUGC.

Keywords: Information quality, User-generated content, Consumer decision making, Consumer purchase funnel

1 Introduction

Information quality (IQ) plays a critical role in ecommerce consumers' purchase decisions (Chen, Shang, and Kao, 2009; Napitupulu and Kartavianus, 2014; Narwal and Kant, 2014; Tsai and Chuang, 2011) as there is no opportunity to physically interact with products when shopping online. User-generated content (UGC) is an increasingly important source of information when it comes to product-selection and decision processes. With its advantages of richness, authenticity, and independency, UGC also comes with challenges like fraud, subjectivity, and diverse forms of heterogeneity. Of special interest is a subset of UGC that deals with product-related information. For example, for laptops, this would be statements like "bright display," "annoying fan," or "long-lasting battery." Although the presence of product-related UGC (PUGC[1]) has become common on ecommerce sites, it is questionable whether vendors and intermediaries take full advantage of PUGC. Today, PUGC is heavily used for evaluation purposes. That is, PUGC is typically not accessible on ecommerce sites before the corresponding product has been found (Lee, Li, and Wei, 2008). PUGC is mostly treated like an extension to the vendor's product description, and individuals must access the product site before accessing the associated PUGC.

Nevertheless, decision making entails several phases (e.g., screening of opportunities, reduction of a consideration set, selection of candidates, confirmation, evaluation), but the vast majority of PUGC that is written down in texts has rarely found its way into decision phases other than evaluation. When consumers do use PUGC for evaluation purposes, it turns out to be a time-consuming process for them as they have to read the full texts of several reviews and iteratively refine their search according to the insights obtained. In a lot of ecommerce applications, there is a need to integrate PUGC in earlier phases of the purchase process. As more and more product reviews are written, the need to organize and rearrange PUGC becomes more important (Hu and Bing, 2004), getting away from vast full-text displays of reviews. Issues of questionable IQ, however, constrain the use of PUGC in earlier search phases. Using PUGC in an early search phase (e.g., screening of opportunities or reduction of a consideration set) implies pulling relevant PUGC information "in front of the product" and presenting the content in an aggregated and consolidated way. However, it is challenging for information system designers to provide filters or condensed information of appropriate quality in a screening or filtering phase for some obvious reasons: PUGC mostly covers only a few aspects of a product, and single reviews are like illuminating a product in the dark with a spotlight—they never give a full picture. As completeness of information is one problem, proper aggregation and consolidation of PUGC turns out to be another. Therefore, this study examines the IQ needs of users throughout the purchase.

A widely accepted definition of IQ is "fitness for use" of data/information for data consumers (Ballou, Madnick, and Wang, 2003; Madnick et al., 2009; Wang and Strong, 1996). Fitness for use implies a user-centric view of IQ issues that is related to a specific task, and it reflects that "use" can be different depending on the situation. Thus, fitness for use is an appropriate IQ definition to use when studying potentially changing IQ evaluations throughout the purchase process, during which the user is confronted with different situations and tasks.

IQ research has shown that IQ not only comprises accuracy, believability, and other intrinsic dimensions but also contextual, representational, and accessibility dimensions (Wang and Strong, 1996). This multidimensional conceptualization emphasizes that changing the representation, context, or accessibility of information will not only result in a different assessment of IQ but might also impose different IQ needs on the user side.

However, there is little knowledge about users' IQ preferences throughout the phases of the purchase process. Researchers have not yet revealed in which contexts PUGC becomes most valuable on product platforms or how to build information systems that can leverage the value of PUGC. To close this gap, this study enhances the understanding of PUGC IQ by combining IQ with purchase-process phases in order to obtain a more differentiated view of IQ for the same PUGC content over time or, more concisely, throughout process phases. The IQ needs that arise during the purchase process are to be determined, so the questions are which IQ dimensions are perceived as most important and do these IQ needs remain stable or change throughout the process? Therefore, the idea of this study is to pursue a dynamic approach to IQ to optimally support the purchase process in different phases. Hence, the research questions are formulated as follows:

[1] The creation of product-related content by users is often referred to as electronic word of mouth (eWOM). Whereas the term eWOM is sometimes used in the literature to describe content, in this study, eWOM is understood as a process, not the textual product of the process, which is denoted with PUGC.

RQ: How do users perceive the importance of different IQ dimensions compared to each other, and do the levels of importance change depending on the phase of the purchase process in which PUGC is integrated?

The remainder of this paper is organized as follows. First, related work from three fields is introduced: IQ frameworks, purchase-process models, and approaches to integrating PUGC into early purchase-process phases by aggregating PUGC. Then, this study's research approach of a dynamic IQ observation combining an IQ framework with a purchase phase model is depicted, which is followed by the presentation of the conceptual model. Afterwards, the survey instrument and data analysis are explained. The study concludes with its findings and implications.

The research approach underlying this study, specifically the research question, related work, conceptual model, and research method, has already been accepted for publication as a research-in-progress paper (Hirschmeier et al., 2015). Besides minor changes, the research method was substantially revised and the results, discussion, and conclusion were added.

2 Related Work

First, existing literature in the field of IQ and purchase processes are presented as both concepts will be combined in this research. Further, an overview of corresponding approaches and methods for aggregating PUGC as a basis for its use in early phases of the purchase process is given.

2.1 Information quality

Typical research questions related to IQ are (1) how can IQ be defined; (2) how can IQ be measured; and (3) how does IQ relate to other concepts, such as user satisfaction and information system success.

Relating to the first point, different authors have proposed a variety of definitions. As already mentioned, the fitness-for-use definition was introduced by Wang and Strong (1996) and has since been the conceptual basis for various IQ frameworks (e.g., Kahn, Strong, and Wang, 2002; Lee et al., 2002) and research studies (e.g., Nelson, Todd, and Wixom, 2005; Otto, 2011), including applications in online contexts (Arazy et al., 2011; Scholz and Dorner, 2013). Another common definition proposed, for example, by Orr (1998) and Wand and Wang (1996) is that of IQ[2] being the degree of correspondence of information with external phenomena (i.e., accurately representing real-world entities and their attributes). Shanks and Darke (1998) and Price and Shanks (2005) proposed a definition of IQ based on semiotic theory, which integrates different perspectives and defines IQ on three semiotic levels: syntactic (conforming to rules), semantic (corresponding to external phenomena), and pragmatic (corresponding to a user-centric perspective).

Regarding the second point—measuring—IQ definitions are usually operationalized through a set of dimensions like accuracy, timeliness, completeness, consistency, and ease of understanding (for an overview of definitions, see, for example, Knight and Burn (2005); Jayawardene et al. (2013); Lee et al. (2002)). The dimensions chosen to be relevant in order to assess IQ in specific contexts depend on the task and formal specifications and expectations of the consumer (Kahn, Strong, and Wang, 2002; Lee et al., 2002). Dimensions are then measured through technical means (e.g., structural and textual features of UGC (Vir Singh, Sahoo, and Mukhopadhyay, 2014; Wang, Liu, and Fan, 2011)) or via user surveys (e.g., Blanco, Sarasa, and Sanclemente, 2010; Liang and Xue, 2013)).

[2] Though Orr (1998) and Wand and Wang (1996) actually spoke of "data quality," "data (quality)" and "information (quality)" are used synonymously in this study as it is common practice in IQ research (Madnick et al., 2009). In some contexts, it is necessary to differentiate between data and information, or, respectively, data quality and information quality. Some studies have done so, but there is no shared common understanding regarding these and other terms (Kahn, Strong, and Wang, 2002; Madnick et al., 2009; Price and Shanks, 2005; Wand and Wang, 1996). For the research objective of this study, no differentiation between data quality and information quality is needed.

Finally, regarding the third point, IQ has been shown to positively influence information system success (DeLone and McLean, 2003; DeLone and McLean, 1992; Petter, DeLone, and McLean, 2013) and user satisfaction (Doll and Torkzadeh, 1988; Wixom and Todd, 2005).

However, the *assessment of the importance* of different IQ dimensions has received less attention. From the perspective of information users, this problem can be defined as follows. Given a set of IQ dimensions (e.g., accuracy and timeliness), there are several tradeoffs (e.g., the more accurate information is supposed to be, the less timely it can be provided and vice versa) as well as different contexts and tasks (e.g., phases in the purchase process). As such, which IQ dimension should be prioritized first, second, third, and so forth?

In fact, there are some studies on the tradeoffs between IQ dimensions and their relative importance. Fehrenbacher and Helfert (2012) investigated tradeoffs between the perceived importance of selected IQ dimensions, and Seethamraju (2005) studied the relative importance of web quality dimensions. However, to the best of our knowledge, this problem has not been investigated in the context of PUGC in the purchase process.

2.2 Purchase-decision process models

Information processing theory identified three decision-making phases (Cook, 1993; Simon, 2001) in pre-internet times—intelligence, design, and choice—which have since been adapted to online purchase decision making (Gao et al., 2012). In general, when talking about the purchase-decision process, the consumer purchase funnel model (Evans, 2008) is widely used in various forms and under differing names (e.g., the consumer decision journey (Court et al., 2009)) or in diverse "funnel models," such as the ecommerce funnel, sales funnel, or conversion funnel. The consumer decision model (Engel, Blackwell, and Miniard, 1995; Jobber, 1995) proposes seven phases: need recognition, search for information, pre-purchase alternative evaluation, purchase, consumption, post-purchase alternative evaluation, and divestment. Several other theories have been applied to the decision-making process, such as mental accounting theory (Gupta, 2006). Vázquez et al. (2014) presented a novel analysis and classification of (P)UGC in terms of how it is involved in the phases of the consumer decision journey. Depending on the perspective, purchase-decision models have a different number of phases (three to seven) with different names. Throughout this study, "purchase process" and "purchase funnel" are used to identify the funnel model and the underlying purchase process.

2.3 Techniques for integrating PUGC into earlier phases of the purchase process

Integrating PUGC into earlier phases of the purchase process requires techniques for aggregating it. Knowing which IQ needs arise during the purchase process is important for generating value-added by aggregating PUGC. For instance, one important way to aggregate PUGC is to perform product feature extractions. However, when extracting features from texts, uncertainty plays an important role, and IQ needs come into play. Information about which IQ dimensions users prefer in a specific purchase-process stage may distinguish a value-adding feature extraction from a useless one. In the following, some approaches for aggregating PUGC and integrating it into earlier phases of the purchase process are presented.

When the IQ preferences of individuals or a customer segment are known, integrating PUGC into early search phases is one of the aims of content-based recommender systems. While, in general, "the recommendation problem is reduced to the problem of estimating ratings for the items that have not been seen by a user" (Adomavicius and Tuzhilin, 2005, p. 2), a recommender system's goal in a broader view might not only be to estimate ratings for *items* that a user has not seen yet but also to provide ratings for *filters* or *information* about items that a user has not seen yet, for instance, PUGC filters that match individuals' needs.

Content-based recommender systems do not take user behavior or collaborative filtering into consideration but only rely on content. They are rooted in information-filtering (Belkin and Croft, 1992) and information-retrieval research (Salton, 1989) and are typically connected to automatic feature-extraction techniques for textual data. An overview of text-extraction techniques is given by Jiang (2012). Further, Lee et al. (2008) integrated traditional information-retrieval relevance rankings with database aggregation to model the knowledge within online product reviews and product descriptions in order to provide a needs-centric search wherein users can input free-text queries. Dave et al. (2003) proposed a classifier that draws on information-retrieval techniques for feature extraction and scoring in order to generate a list of product attributes (e.g., quality, features, etc.) and aggregate opinions about each of them. Aggregating subjective opinions is a challenge but leads to more intersubjective information, referred to as "objectivity by averaging" (Parameswaran and Whinston, 2007). Approaches for aggregating sentiments using ontologies have been proposed by Mukherjee and Joshi (2013; also see Mukherjee and Joshi, 2014) and Grandi et al. (2014).

Social business intelligence (BI) approaches (Francia, Golfarelli, and Rizzi, 2014) make up an interdisciplinary research area and combine data-mining technologies, natural language processing, and other promising techniques to summarize (P)UGC. Gallinucci et al. (2013) proposed a way to aggregate topics for social BI, and Hu and Bing (2004) proposed an approach to mine and summarize customer product reviews, focusing on product features only and generating feature-based summaries. Further, Popescu and Etzioni (2005) constructed an unsupervised information-extraction system, which mines reviews in order to build a model of important product features.

Especially for ecommerce websites, it is vital to process and present PUGC in an appropriate way. Nevertheless, although diverse techniques exist to extract product features from full-text data and approaches have been proposed to integrate PUGC into early purchase-process phases, in practice, these approaches have rarely been applied. Most importantly, it remains unclear which IQ dimensions—from the perspective of the consumer—PUGC should be adapted to when designing ecommerce sites in order to correspond to the importance of IQ on behalf of users. In other words, while techniques to extract, aggregate, filter, and present PUGC exist, it is still unclear *which aim* information aggregation, filtering, and presentation should be tuned to so that they are most suitable to users' requirements in each stage of the purchase process.

3 Research Approach

In this study's research approach, two models are combined: a purchase funnel model and an IQ framework. The following depicts how both models apply to the research objective.

3.1 Purchase funnel

The consumer purchase funnel model has been widely used in various forms and under different names, with different numbers of and naming for phases. In nearly all models, the first phase is referred to as "awareness," referring to consumers' becoming aware of relevant product features. The last phase is mostly referred to as "action" or "purchase" and is preceded by "evaluation" in some models. The middle part of the funnel has diverse names and granularities. It is mainly about consumers' considering whether to include the product in their consideration set (i.e., narrowing down the set). As such, this phase is usually referred to as "consideration." Basic models distinguish between three phases that are common to all purchase funnel models: the beginning (awareness), the end (evaluation), and the middle (narrowing down the consideration set). In this research approach, a parsimonious model of a purchase funnel was used that consists of three phases (see Figure 1): (1) a screening phase, during which the user gathers an overview about products; (2) a filtering phase, during which the user narrows down the consideration set; and (3) an evaluation phase, during which the user reads product information (i.e., descriptions and reviews) in detail.

When dealing with the fitness-for-use approach for IQ dimensions, all content is in a specific use context. The research object of fitness for use is therefore a combination of content and use context. The following gives an overview of what PUGC means in the three different fitness-for-use situations.

PUGC in the screening phase. Screening PUGC is the first step in the purchase process. "In the beginning phases of purchase, a buyer lacks experience, his choice criteria is not well-developed and he doesn't have any knowledge of various brands and their potential" (Gupta 2006, p. 27). PUGC must be aggregated, consolidated, and condensed to be presentable in the screening phase.

PUGC in the filtering phase. Filtering by PUGC is the second step in the purchase process. If PUGC is used for filters, it must be aggregated, consolidated, and condensed just like PUGC in the screening phase. The big difference between PUGC for screening and filtering is not the presentation of the content but what users want to do with this content (i.e., the use context).

PUGC in the evaluation phase. Evaluating PUGC is the last step in the purchase process. In this phase, the user reads detailed information about the product (i.e., the full text of the review).

Examples of PUGC in each phase are presented in Figure 1.

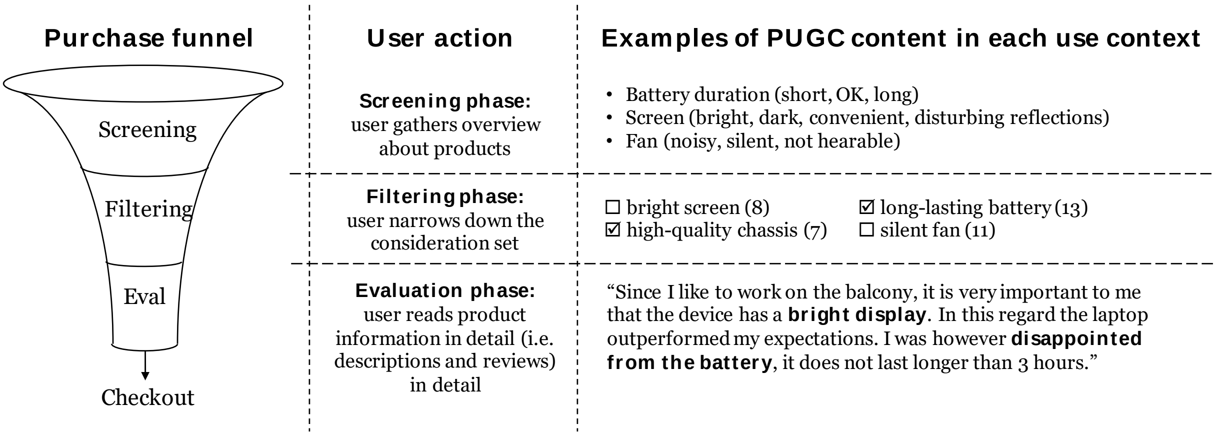

Figure 1. Purchase funnel model and corresponding examples of PUGC

3.2 Conceptual model

In order to discern the importance of IQ dimensions during the purchase process in the context of PUGC, not only are assessments of users' IQ needs of major interest but also determining whether levels of importance remain stable throughout the search process. The process phase might have an effect on levels of importance depending on how close users are to checkout and the completed purchase.

Since little is known about how levels of importance of IQ dimensions might change during the purchase process, the design of this research is exploratory. That is, users' perceived importance regarding multiple IQ dimensions in each of the three phases of the purchase funnel model are investigated, but a specific proposition regarding the influence of phases on perceived importance is not postulated. Figure 2 summarizes the conceptual model of the research.

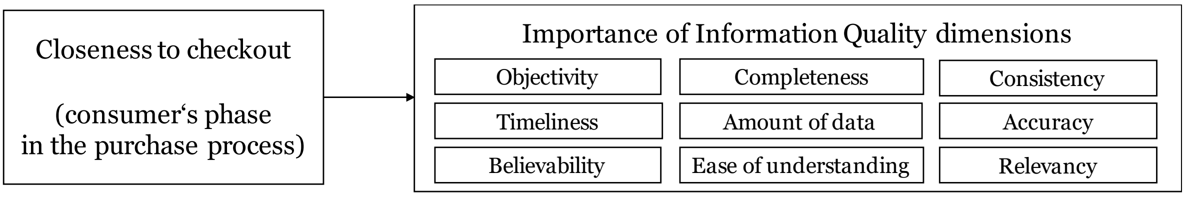

Figure 2. Conceptual model

4 Research Method

This study pursues an empirical-exploratory approach to IQ as it explicitly focuses on the subjective perspective of the information consumer. Therefore, this study stands in the tradition of empirical-exploratory approaches to IQ and its dimensions, which date back to (at least) the work of Wang and Strong (1996). This section describes which IQ dimensions were selected for empirical investigation, how the survey instrument was designed, and how data were collected.

4.1 Selection of IQ dimensions

The set of IQ dimensions used in this study is based on Wang and Strong's framework of 15 IQ dimensions due to its wide acceptance and fit with the empirical consumer-centric view (Wang and Strong, 1996). Their definition of IQ and its dimensions explicitly take into account the subjective perspective of information users (and their tasks and contexts) on information and related quality. IQ as fitness for use is conceptually well suited to this study's research question, which asks for subjective evaluations of IQ on behalf of information users. It is particularly better suited than, for example, defining IQ as the degree of correspondence of information with external phenomena—as proposed by Orr (1998) and Wand and Wang (1996)—because the latter definition excludes the user perspective.

From Wang and Strong's framework of 15 IQ dimensions, five dimensions were dropped for this study, namely, accessibility, access security, value-added, interpretability, and concise representation, for the following reasons. Accessibility was dropped because accessibility is a prerequisite in all phases. As there is no non-accessible PUGC, there is also no sense in assessing the importance of accessibility. Also, there is no access security for PUGC. Regarding value-added, its importance is expected to be high in all phases, and no meaningful results were hoped for. Interpretability was dropped because of its interference with "ease of understanding" (also see the explanation below). Since the representation of data in the screening and filtering phase is condensed, conciseness of representation does not seem to be comparable among the phases. The survey therefore included an initial set of 10 IQ dimensions.

Before the questionnaire was sent out to participants, a pre-test was conducted with seven participants. They answered the questionnaire and provided verbal feedback on how to improve it. Accordingly, parts of the questionnaire were rephrased to improve understandability. Some participants complained about there being too much text in the questionnaire, but the decision was made to keep detailed descriptions of the three purchase phases to ensure a clear understanding on the part of the participants rather than running the risk of misunderstandings regarding the purchase phases. Further, six pre-test participants found it difficult to apply the dimension of reputation of information source to the scenario and thus to assess its importance. Therefore, reputation was excluded from the set IQ dimensions, which made up a final set of nine IQ dimensions for the final questionnaire.

The following provides definitions for the nine IQ dimensions in the final set. Based on the framework by Wang and Strong (1996), several IQ frameworks have emerged, and various names and definitions for the IQ dimensions have been proposed in the literature. Only a few papers give detailed definitions of the IQ dimensions, and definitions differ across the frameworks. The questionnaire in this study mainly sticks to the definitions of Wang and Strong (1996). Only occasionally were definitions from other authors used when those definitions were perceived to be (1) better suited for this study's research context or (2) easier to understand for survey participants.

Believability. It is hard to distinguish believability from similar terms, like confidence (McGilvray, 2008). The best definition for this study's purpose is the one from Wang and Strong (1996, p. 31): "Data are accepted or regarded as true, real, and credible."

Timeliness. Timeliness (Byrne et al., 2008; Loshin, 2006; McGilvray, 2008; Price and Shanks, 2005) is also often referred to as currency (English, 2009; Eppler, 2006; Redman, 1997; Stvilia et al., 2007). This study's questionnaire refers to the definition of Price and Shanks (2005, p. 10): "The currency (age) of the data is appropriate to its use."

Accuracy. Accuracy is included in a lot of studies (Eppler, 2006; Gatling et al., 2012; Loshin, 2001; McGilvray, 2008; Redman, 1997; Stvilia et al., 2007). The following definition was considered to be most appropriate: “Determines the extent to which data objects correctly represent the real-world values for which they were designed” (Gatling et al., 2012, p. 334).

Objectivity. Price and Shanks (2005) discussed several understandings of objectivity. Nevertheless, in this study, the definition of Wang and Strong (1996, p. 32) was applied: “Data are unbiased and impartial.”

Completeness. The term completeness can be found in many studies (Byrne et al., 2008; Gatling et al., 2012; Gomes, Farinha, and Trigueiros, 2007; Loshin, 2006; Redman, 1997; Wang and Strong, 1996) but also under different names with similar meanings (e.g., data coverage) (McGilvray, 2008). For the survey, the definition of Gomes et al. was selected: “Data is complete if no piece of information is missing” (Gomes, Farinha, and Trigueiros, 2007, p. 17).

Relevancy. From relevancy definitions (English, 2009; Lyon, 2008; Stvilia et al., 2007), the one from Wang and Strong (1996, p. 31) was picked: “Data are applicable and useful for the task at hand.”

Consistency. Several definitions for consistency can be found (Eppler, 2006; Gatling et al., 2012; Kimball and Caserta, 2004; Loshin, 2001), out of which the following was used: “Is the information free of contradictions or convention breaks?” (Eppler, 2006, p. 83)

Ease of understanding. Ease of understanding is sometimes used synonymously with interpretability (Redman, 1997). Interpretability was dropped from the set of IQ dimensions, and the definition of Wang and Strong (1996, p. 32) was used for ease of understanding: “Data are clear without ambiguity and easily comprehended.”

Amount of data. “The quantity or volume of available data is appropriate” (Wang and Strong, 1996, p. 32).

4.2 Survey instrument

Several methods exist to gather data about the priorities/importance of different options—in this case IQ dimensions. A common and straightforward approach is to use Likert scales for all options, which is generally suitable to assess user perceptions (Smith, 1997). However, it does allow the participant to respond that all options (i.e., dimensions) are important to him or her, which may be rather unrealistic, instead of making him or her think about what is really important. More sophisticated approaches apply resource limitations, for example, by asking participants to assign points from a given “budget” to the options. This approach is referred to as comparative scaling with constant sum, which allows for better discrimination among options without being too time consuming (Malhotra, 2009). However, participants have to calculate and sum up points, which makes this method difficult and error prone. As an advantage, both approaches allow operations on a metric scale.

Working with ordinal scales, a simple way to collect data about preferences is to ask participants to rank options. However, the results neglect how big differences in perceptions are. It has been proven useful to ask participants to choose between IQ dimensions (Fehrenbacher and Palit, 2013) in pairwise comparisons. However, given the relatively large number of items to compare in this study, this is less feasible as the number of comparisons increases nonlinearly.

Accepting the risk that users might rate all IQ dimensions as very important, the straightforward method of using Likert scales was chosen in order to assess the importance of IQ dimensions in each purchase phase (i.e., screening, filtering, and evaluation).

In the questionnaire, users were put in a scenario of searching for a laptop on the internet. First, they were introduced to the fictitious scenario and to the three purchase phases (i.e., screening, filtering, and evaluation), including illustrative examples, so they could gain a good understanding of the context. Following this, participants consecutively received a more detailed description of each purchase phase and were asked to indicate the importance they perceived regarding each of the IQ dimensions using a seven-point Likert scale. To prevent a systematic bias arising from the sequence in which the IQ dimensions were presented, the order of the IQ dimensions was randomized across all participants and phases. Appendix A shows a questionnaire example for the scenario screening (Figure 5). Questions for the filtering and evaluation phases were assembled accordingly. Further, participants were asked to indicate their sex and age.

4.3 Data collection

The link to the online survey was given to students of an undergraduate (approximately 120 students) and a graduate (approximately 50 students) information systems course in a large German university in November 2015. Students were asked to answer the questionnaire in class (but on their own) using their laptops, smartphones, or tablets. In total, 132 students participated in the survey, which equals a response rate of approximately 78%. Six records had to be excluded from further analysis because the respective participants only used the Likert scales up to the value of 3 (across all three purchase phases and nine dimensions, i.e., 27 times). This pattern occurred because they opened the questionnaire on their smartphones in portrait mode and were thus not able to see the full scale from 1 to 7, but only saw 1 to 3. All of the remaining 126 records were used for further analysis.

4.4 Data analysis

To investigate the effect that different purchase-process phases possibly have on each IQ dimension's perceived importance, boxplots were computed for each IQ dimension's importance across each of the three phases to explore the results visually. Further, non-parametric Friedman tests (Friedman 1937) were applied to each of the nine IQ dimensions to test for differences between the three measurement points—that is, the three purchase-process phases. The Friedman test is well suited for repeated measurements (i.e., paired samples) with more than two measurement points (i.e., process phases) and an ordinal level of measurement as in the case at hand.

5 Results

Owing to the large share of male students in the university's information systems program, the sample was unbalanced in terms of sex (90 male and 36 female, which equals 71.4% and 28.6%, respectively). Figure 3 provides a breakdown of respondents' age and sex.

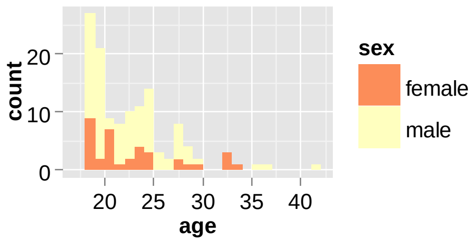

Figure 3. Breakdown of respondents' age and sex

Regarding the assessment of IQ dimensions' importance, all IQ dimensions were assessed to be important to a certain extent, and most assessments were in the upper half of the scale. Figure 4 presents the boxplots for each of the nine IQ dimensions across all three purchase phases. Table 1 provides the respective descriptive statistics.

Comparing IQ dimensions against each other, respondents reported two dimensions to be of paramount importance: namely, believability and accuracy. These two dimensions are not only the two most important IQ dimensions across all phases, but their medians are also located at the upper end of the scale (7) with a very small deviation. This is strong evidence that these two IQ dimensions were favored above all others. Completeness, timeliness, and appropriate amount of data, in contrast, seem to be among the least important IQ dimensions to users.

Regarding the question of whether the importance of an IQ dimension changes throughout the purchase process, there are surprisingly few changes between the measures for screening, filtering, and evaluation for each of the dimensions. Participants' perceived importance seems to be stable throughout the purchase process. Friedman tests for differences between the levels of importance in the three phases for each of the nine dimensions also did not indicate significant differences except for completeness (see Table 1).

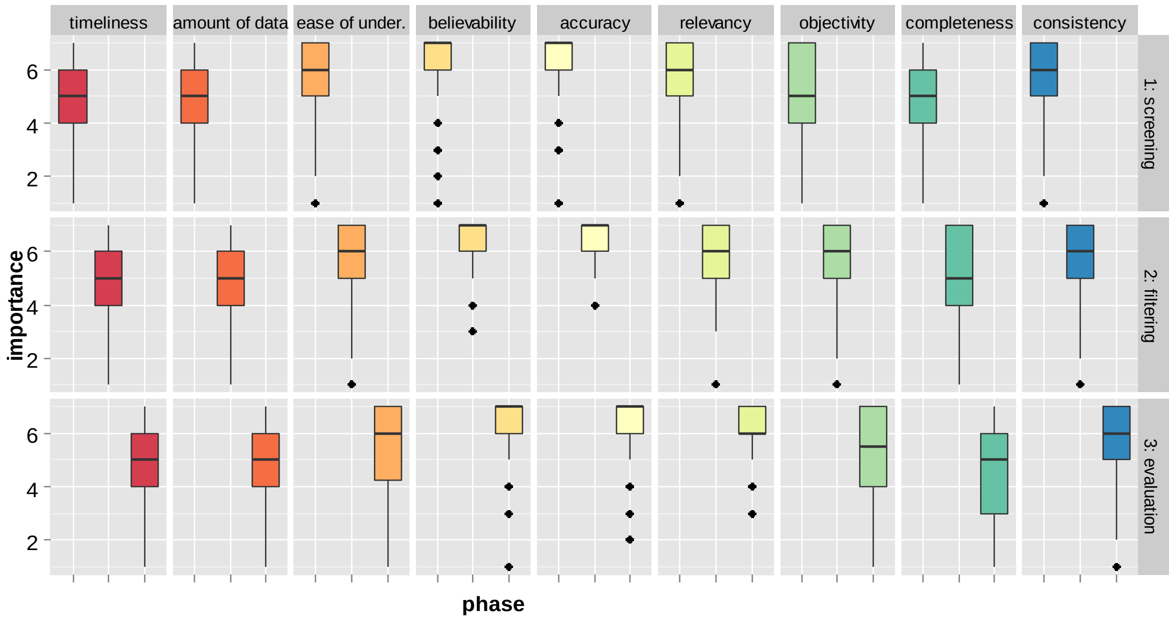

Figure 4. Boxplots for all nine IQ dimensions and all three phases

6 Discussion and Limitations

The results provide some interesting insights on how to aggregate PUGC in order to manage the vast amount of PUGC that arises on ecommerce platforms today. If, for example, features are extracted from PUGC, one always faces quality problems related to UGC in general. Depending on perceived importance, different IQ dimensions could be emphasized differently. The following are some examples illustrating how PUGC aggregation can be conducted to emphasize different IQ dimensions: emphasizing believability by including how many distinct reviewers have contributed, by including origin of PUGC, or by including only reviewers that were actual buyers; emphasizing timeliness by including only the most recent PUGC; emphasizing accuracy by aggregating either every distinct value (e.g., weight 1.2 kg, 1.25 kg, 1.3 kg) or allowing truncation (e.g., "about" 1 kg, 1.5 kg, 2 kg); emphasizing completeness by including all PUGC no matter what the quality and amount is; emphasizing consistency by excluding conflicting evaluations (and only using the majority opinion); emphasizing ease of understanding by replacing/merging numeric values with "fuzzy" values (e.g., "light," "medium," "heavy" for weight); and emphasizing amount of data by including only the most commonly mentioned product features. This study provides initial insights into which IQ dimensions should be emphasized in a given scenario.

According to the results, when viewing product features that have been extracted from PUGC (instead of being presented dozens of reviews, which would be very time-consuming to read), users seem to prefer primarily seeing product information that is highly accurate and believable. Conversely, users seem to be willing to accept information that is not complete and exhaustive (indicated by lower importance values for completeness and amount of data) or somewhat outdated (timeliness).

Dimension		**Phase**			**Friedman test**		
		Screening	**Filtering**	**Evaluation**	**χ^2**	**df**	**p**
Timeliness	Minimum	1.00	1.00	1.00	2.734	2	.255
	1st Quartile	4.00	4.00	4.00			
	Mean	5.00	5.00	5.00			
	3rd Quartile	6.00	6.00	6.00			
	Maximum	7.00	7.00	7.00			
Amount of data	Minimum	1.00	1.00	1.00	2.340	2	.311
	1st Quartile	4.00	4.00	4.00			
	Mean	5.00	5.00	5.00			
	3rd Quartile	6.00	6.00	6.00			
	Maximum	7.00	7.00	7.00			
Ease of understanding	Minimum	1.00	1.00	1.00	.193	2	.908
	1st Quartile	5.00	5.00	4.25			
	Mean	6.00	6.00	6.00			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			
Believability	Minimum	1.00	3.00	1.00	3.164	2	.206
	1st Quartile	6.00	6.00	6.00			
	Mean	7.00	7.00	7.00			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			
Accuracy	Minimum	1.00	4.00	2.00	3.529	2	.171
	1st Quartile	6.00	6.00	6.00			
	Mean	7.00	7.00	7.00			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			
Relevancy	Minimum	1.00	1.00	3.00	5.427	2	.066
	1st Quartile	5.00	5.00	6.00			
	Mean	6.00	6.00	6.00			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			
Objectivity	Minimum	1.00	1.00	1.00	1.052	2	.591
	1st Quartile	4.00	5.00	4.00			
	Mean	5.00	6.00	5.50			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			
Completeness	Minimum	1.00	1.00	1.00	10.625	2	.005
	1st Quartile	4.00	4.00	3.00			
	Mean	5.00	5.00	5.00			
	3rd Quartile	6.00	7.00	6.00			
	Maximum	7.00	7.00	7.00			
Consistency	Minimum	1.00	1.00	1.00	0.370	2	.831
	1st Quartile	5.00	5.00	5.00			
	Mean	6.00	6.00	6.00			
	3rd Quartile	7.00	7.00	7.00			
	Maximum	7.00	7.00	7.00			

Table 1. Descriptive statistics of participants' assessments of the IQ dimensions and Friedman tests for differences

Practical suggestions on how to implement the relevant IQ dimensions in the purchase process can be derived from these results. The tradeoff between little information with high accuracy and more information with low accuracy should clearly be decided in favor of accuracy. An implication for ecommerce managers is that less can be more when it comes to displaying extracted product features: if PUGC is to be aggregated by feature-extraction techniques, a smaller set of derived features with higher accuracy would be more valuable than a larger set of derived product features with lower accuracy and believability. This holds for the evaluation phase of this study's purchase-process model as well as for earlier phases. If filters are offered in the filtering phase on the basis of extracted product features, only those filters that can be applied with high accuracy should be shown. Accordingly, in the screening phase, only an overview of those product features/information that can be accurately extracted should be selected.

These findings underpin the common conceptualization of IQ as a multidimensional concept (Knight and Burn, 2005; Lee et al., 2002; Zhang et al., 2014). Confronted with nine different dimensions of IQ, respondents indicated different levels of perceived importance. However, while the fitness-for-use perspective might have proposed changing evaluations of IQ depending on the task and situation, IQ evaluations appeared to be stable throughout the purchase-process phases (i.e., screening, filtering, and evaluation). As an implication for future research, more attention should be devoted to ensuring believability and accuracy in PUGC rather than testing finer-grained purchase-process models (i.e., more than three phases, as done in this study).

However, it is possible that participants were not sufficiently able to put themselves in the three different purchase-process phases. One possible reason for this could be that today's ecommerce websites still mainly present PUGC along with product descriptions at a rather late stage in the purchase process (i.e., the evaluation phase using this paper's terminology). This issue could indicate that participants might not be able to imagine using PUGC in phases other than the evaluation phase. Using mockups or prototypes of (fictitious) ecommerce sites during the questionnaire that provide visual examples integrating PUGC into earlier phases might help overcome this chicken-and-the-egg problem. Mockups and prototypes, however, bear the risk of priming participants toward certain design solutions. This is why a survey approach was chosen for this early-stage research on dynamic IQ throughout the purchase process.

Regarding limitations, the sample of 126 student participants was unbalanced in terms of gender (i.e., the majority of the respondents were male). To address this issue, the analysis of importance of IQ dimensions was repeated for female and male participants separately. Since the results for females and males were very similar, the findings seem to be robust to gender bias; at least there is no direct indication for the existence of a gender bias. Thus, a more balanced sample in terms of gender is not likely to show different results, but the existence of a gender bias cannot be eliminated for sure. Further, the sample is surely not representative of the whole online shopping population and is also probably biased in terms of age, affinity to shopping online, education level, background, setting, and other factors.

Regarding the validity of the results, only one item per question was used in order to keep the survey instrument short and prevent users from terminating the survey; hence, internal consistency and validity could not be calculated.

7 Conclusion and Future Research

This study's new dynamic perspective on IQ perceptions throughout the purchase process and the exploratory research approach revealed interesting results. The patterns of relative importance of the nine IQ dimensions were remarkably stable across the three phases of an idealized purchase process (i.e., screening, filtering, and evaluation of products). Further, two IQ dimensions—accuracy and believability—clearly stood out from the rest. These results can help in the design of ecommerce site and recommendation techniques that employ PUGC when it comes to decisions on which aspects of IQ to prioritize at the cost of others. As future ecommerce sites will surely have to be able to process more and more PUGC and transform the information into product properties (i.e., features) that serve the information needs of customers in each phase of the purchase process, this research may help companies design ecommerce sites and understand users' IQ needs.

Insights into users' IQ needs throughout the purchase process are key to the construction of new product search mechanisms, tomorrow's ecommerce platforms, and the treatment of PUGC in general. Ecommerce site managers will benefit from these insights when they need to decide how to integrate aggregated PUGC into the search process. By gaining more knowledge about information needs throughout the purchase process, ecommerce managers will also be able to better design future content-based recommendation systems according to the fitness-for-use paradigm.

The survey instrument of this study did not take tradeoffs between IQ dimensions into account. However, such tradeoffs are very likely to exist in practice (e.g., completeness versus timeliness of PUGC) and may limit the degree to which users' demand for IQ dimensions can be incorporated in PUGC-based search and recommendation mechanisms. Hence, future research on the design of such mechanisms should take into account these and similar constraints.

In future studies on the evaluation of PUGC IQ dimensions in the purchase process, one could also (1) employ qualitative methods to better control and elicit participants' understanding of IQ dimensions, (2) conduct more focused surveys (i.e., fewer dimensions, more items per dimension) or scenario-based experiments, (3) investigate users' IQ needs for different types of products (e.g., search versus experience goods), and (4) conduct the study with different types of user communities.

With the knowledge of users' IQ needs comes the potential to aggregate PUGC more effectively and integrate it into earlier phases of the purchase process. However, further research needs to measure the quality of the aggregation on the one hand and the efficiency and effectiveness of the purchase process on the other, for which the findings of this study may serve as a basis.

Appendix A: Questionnaire example

Phase "Screening"

In the phase "Screening" you want to get an overview over product features which could be relevant to your purchase decision for a laptop. You focus less on single products, but rather try to grasp which features of a laptop would be relevant to you. You are especially interested in information about what has been important to other users,

The goal of the phase "screening" is to find out a set of relevant product features for your purchase.
The website you visit shows you some features of laptops, so you can get a better image for your decision.

An example for a product feature of laptops is "battery duration" and its properties are "short", "OK", "all day long", "enough for a four-hour train ride". Some more examples are:
- "Screen": "bright", "dark", "convenient", "disturbing reflections"
- "Fan": "noisy", "silent", "not hearable", "like a jet"

In the screening phase you can find out the features that are relevant to you from the features presented.

In the phase "Screening" it is important to me, that ...	1 not important at all	2	3	4	5	6	7 very important
... the information is regarded as true, real, and credible	○	○	○	○	○	○	○
... the information is unbiased and impartial	○	○	○	○	○	○	○
... the information is applicable and useful for the task at hand	○	○	○	○	○	○	○
... no piece of information is missing	○	○	○	○	○	○	○
... the information is free of contradictions or convention breaks	○	○	○	○	○	○	○
... the information is clear without ambiguity and easily comprehended	○	○	○	○	○	○	○
... the quantity or volume of available data is appropriate	○	○	○	○	○	○	○
... the age of the data is appropriate to my use	○	○	○	○	○	○	○
... the information correctly represents real-world properties	○	○	○	○	○	○	○

Figure 5. Sample questionnaire with Likert scales for the screening use context

References

Adomavicius, G., and A. Tuzhilin (2005). "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions." *IEEE Transactions on Knowledge and Data Engineering*, 17 (6), 734–749.

Arazy, O., O. Nov, R. Patterson, and L. Yeo (2011). "Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict." *Journal of Management Information Systems*, 27 (4), 71–98.

Ballou, D., S. Madnick, and R. Wang (2003). "Special Section: Assuring Information Quality." *Journal of Management Information Systems*, 20 (3), 9–11.

Belkin, N., and B. Croft (1992). "Information filtering and information retrieval." *Communications of the ACM*, 35 (12), 29–37.

Blanco, C. F., R. G. Sarasa, and C. O. Sanclemente (2010). "Effects of visual and textual information in online product presentations: Looking for the best combination in website design." *European Journal of Information Systems*, 19 (6), 668–686.

Byrne, B., J. Kling, D. McCarty, G. Sauter, H. Smith, and P. Worcester (2008). *The information perspective of SOA design, Part 6: The value of applying the data quality analysis pattern in SOA*. IBM corporation.

Chen, Y.-C., R.-A. Shang, and C.-Y. Kao (2009). "The Effects of Information Overload on Consumers' Subjective State towards Buying Decision in The Internet Shopping Environment." *Electronic Commerce Research and Applications*, 8, 48–58.

Cook, G. (1993). "An empirical investigation of information search strategies with implications for decision support system design." *Decision Sciences*, 24 (3), 683–698.

Court, D., D. Elzinga, S. Mulder, and O. Vetvik (2009). *The consumer decision journey*. McKinsey Quarterly.

Dave, K., S. Lawrence, and M. Pennock (2003). "Mining the peanut gallery: opinion extraction and semantic classification of product reviews." In: *Proceedings of the 12th international conference on World Wide Web*.

DeLone, W. H., and E. R. McLean (1992). "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research*, 3 (1), 60–95.

DeLone, W. H., and E. R. McLean (2003). "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update." *Journal of Management Information Systems*, 19 (4), 9–30.

Doll, W. J., and G. Torkzadeh (1988). "The Measurement of End-User Computing Satisfaction." *MIS Quarterly*, 12 (2), 259–274.

Engel, J. F., R. D. Blackwell, and P. W. Miniard (1995). *Consumer Behavior* (8th Edition). The Dryden Press, Harcourt Brace College Publishers.

English, L. P. (2009). *Information quality applied: Best practices for improving business information, processes and systems.* Wiley Publishing.

Eppler, M. J. (2006). *Managing information quality: increasing the value of information in knowledge-intensive products and processes.* Springer.

Evans, D. (2008). *Social Media Marketing: An Hour a Day.* Indianapolis: Wiley Publishing.

Fehrenbacher, D. D., and M. Helfert (2012). "Contextual Factors Influencing Perceived Importance and Trade-offs of Information Quality." *Communications of the Association for Information Systems*, 30 (1).

Fehrenbacher, D. D., and I. Palit (2013). "A Quasi-Experimental Analysis on the Influence of Satisfaction and Complexity on Information Quality Outcomes." In: *CONF-IRM Proceedings*.

Francia, M., M. Golfarelli, and S. Rizzi (2014). "A methodology for social BI." *International Database Engineering & Applications Symposium*, 207–216.

Friedman, M. (137). "The use of ranks to avoid the assumption of normality implicit in the analysis of variance." *Journal of the American Statistical Association*, 32 (200).

Gallinucci, E., M. Golfarelli, and S. Rizzi (2013). "Meta-stars: multidimensional modeling for social business intelligence." In: *Proceedings of the sixteenth international workshop on Data warehousing and OLAP*.

Gao, J., C. Zhang, K. Wang, and S. Ba (2012). "Understanding online purchase decision making: The effects of unconscious thought, information quality, and information quantity." *Decision Support Systems*, 53 (4), 772–781.

Gatling, G., C. Brague, R. Champlin, H. Stefani, N. Weigel, G. Bryce, I. Felsheim, A. Engel, W. Gardella, S. Grewal (2012). *Enterprise Information Management with SAP*. SAP Press.

Gomes, P., J. Farinha, and M. J. Trigueiros (2007). "A Data Quality Metamodel Extension to CWM." In: *PCCM '07 Proceedings of the 4th Asia-Pacific conference on Conceptual Modelling*.

Grandi, U., A. Loreggia, F. Rossi, and V. A. Saraswat (2014). "From Sentiment Analysis to Preference Aggregation." In: *Proceedings of the 2014 International Symposium on Artificial Intelligence and Mathematics.*

Gupta, S. (2006). "Examining Online Purchase Decision-calculus: The Mental Accounting Theory Perspective." Thesis. National University of Singapore.

Hirschmeier, S., R. Tilly, M. Hirschmeier, and D. Schoder (2015). "A Dynamic Approach to Information Quality in User-Generated Content. " In: *ICIS 2015 Proceedings*.

Hu, M., and L. Bing (2004). "Mining and Summarizing Customer Reviews." In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Jayawardene, V., S. Sadiq, and M. Indulska (2013). "An analysis of data quality dimensions." Technical Report. University of Queensland.

Jiang, J. (2012). "Information Extraction from Text." In: *Mining Text Data* (pp. 11–41). Ed. by C. C. Aggarwal and C. Zhai. Springer US.

Jobber, D. (1995). *Principles and practice of marketing.* Berkshire: McGraw-Hill.

Kahn, B. K., D. M. Strong, and R. Y. Wang (2002). "Information Quality Benchmarks: Product and Service Performance." *Communications of the ACM*, 45 (4), 184–192.

Kimball, R., and J. Caserta (2004). *The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data*. Indianapolis: Wiley Publishing.

Knight, S.-A., and J. Burn (2005). "Developing a Framework for Assessing Information Quality on the World Wide Web." *Informing Science Journal*, 8, 159–172.

Lee, T. Y., S. Li, and R. Wei (2008). "Needs-Centric Searching and Ranking Based on Customer Reviews." In: *IEEE Conference on E-Commerce Technology*.

Lee, Y. W., D. M. Strong, B. K. Kahn, and R. Y. Wang (2002). "AIMQ: a methodology for information quality assessment." *Information & Management*, 40 (2), 133–146.

Liang, H., and Y. Xue (2013). "Online Health Information Use by Disabled People: the Moderating Role of Disability." In: *ICIS 2013 Proceedings*.

Loshin, D. (2001). *Enterprise knowledge management: The data quality approach*. Morgan Kaufmann Academic Press.

Loshin, D. (2006). *Monitoring Data Quality Performance using Data Quality Metrics*. Informatica Corporation.

Lyon, M. (2008). *Assessing Data Quality, Monetary and Financial Statistics*. Bank of England.

Madnick, S. E., R. Y. Wang, Y. W. Lee, and H. Zhu (2009). "Overview and Framework for Data and Information Quality Research." *ACM Journal of Data and Information Quality*, 1 (1), 1–22.

Malhotra, N. K. (2009). *Marketing Research: An Applied Orientation*. London: Prentice-Hall.

McGilvray, D. (2008). *Executing data quality projects: Ten steps to quality data and trusted information*. Morgan Kaufmann.

Mukherjee, S., and S. Joshi (2013). "Sentiment Aggregation using ConceptNet Ontology." In: *International Joint Conference on Natural Language Processing*. Nagoya, Japan.

Mukherjee, S., and S. Joshi (2014). "Author-Specific Sentiment Aggregation for Polarity Prediction of Reviews." In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation*. Reykjavik, Iceland: ELRA.

Napitupulu, T. A., and O. Kartavianus (2014). "A structural equations modeling of purchase decision through e-commerce." *Journal of Theoretical and Applied Information Technology*, 60 (2).

Narwal, M., and R. Kant (2014). "Role of e-commerce in consumer decision making." *International Journal of Research in Business Management*, 2 (6), 77–88.

Nelson, R. R., P. A. Todd, and B. H. Wixom (2005). "Antecedents of Information and System Quality: An Empirical Examination within the Context of Data Warehousing." *Journal of Management Information Systems*, 21 (4), 199–235.

Orr, K. (1998). "Data Quality and Systems Theory." *Communications of the ACM*, 41 (2), 66–71.

Otto, B. (2011). "Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers." *Communications of the Association for Information Systems*, 29 (1).

Parameswaran, M., and A. B. Whinston (2007). "Research Issues in Social computing." *Journal of the Association for Information Systems*, 8 (6), 336–350.

Petter, S., W. DeLone, and E. R. McLean (2013). "Information Systems Success: The Quest for the Independent Variables." *Journal of Management Information Systems*, 29 (4), 7–62.

Popescu, A.-M., and O. Etzioni (2005). "Extracting Product Features and Opinions from Reviews." In: *Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*.

Price, R., and G. Shanks (2005). "A semiotic information quality framework: development and comparative analysis." *Journal of Information technology*, 20 (2), 88–102.

Redman, T. C. (1997). *Data quality for the information age*. Artech House, Inc.

Salton, G. (1989). *Automatic Text Processing: the transformation, analysis, and retrieval of information by computer*. Boston: MA: Addison-Wesley.

Scholz, M., and V. Dorner (2013). "The Recipe for the Perfect Review? - An Investigation into the Determinants of Review Helpfulness." *Business & Information Systems Engineering*, 5 (3), 141–151.

Seethamraju, R. (2005). "Relative Importance of Web Quality Dimensions." In: *AMCIS Proceedings*.

Shanks, G. G., and P. Darke (1998). "Understanding Data Quality in Data Warehousing: A Semiotic Approach." In: *Proceedings of the International Conference on Information Quality*.

Simon, H. (2001). "A behavioral model of rational choice." *The quarterly journal of economics*, 69 (1), 99–118.

Smith, A. (1997). *Human Computer Factors: A Study of Users and Information Systems*. McGraw-Hill.

Stvilia, B., L. Gasser, M. B. Twidale, and L. C. Smith (2007). "A framework for information quality assessment." *Journal of the American Society for Information Science and Technology*, 58, 1720–1733.

Tsai, C.-C., and H.-M. Chuang (2011). "The Role of Cognitive Decision Effort in Electronic Commerce Recommendation System." *World Academy of Science, Engineering and Technology*, 58, 66–70.

Vázquez, S., I. Muñoz-García, I. Campanella, M. Poch, B. Fisas, N. Bel, and G. Andreu (2014). "A classification of user-generated content into consumer decision journey stages." *Neural Networks*, 58, 68–81.

Vir Singh, P., N. Sahoo, and T. Mukhopadhyay (2014). "How to Attract and Retain Readers in Enterprise Blogging?" *Information Systems Research*, 25 (1), 35–52.

Wand, Y., and R. Y. Wang (1996). "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM*, 39 (11), 86–95.

Wang, G., X. Liu, and W. Fan (2011). "A Knowledge Adoption Model Based Framework for Finding Helpful User-Generated Contents in Online Communities." In: *ICIS 2011 Proceedings*.

Wang, R. Y., and D. M. Strong (1996). "Beyond Accuracy: What Data Quality Means to data Consumers." *Journal of Management Information Systems*, 12 (4), 5–33.

Wixom, B. H., and P. A. Todd (2005). "A Theoretical Integration of User Satisfaction and Technology Acceptance." *Information Systems Research*, 16 (1), 85–102.

Zhang, R., V. Jayawardene, M. Indulska, S. Sadiq, and X. Zhou (2014). "A Data Driven Approach for Discovering Data Quality Requirements." In: *ICIS 2014 Proceedings*.