Data Collection Interfaces in Online Communities: The Impact of Data Structuredness and Nature of Shared Content on Perceived Information Quality

Mahed Maddah
Suffolk University
mmaddah@suffolk.edu

Roman Lukyanenko
HEC Montréal
roman.lukyanenko@hec.ca

Debra VanderMeer
Florida International University
vanderd@fiu.edu

Binny Samuel
University of Cincinnati
samuelby@uc.edu

Abstract

The growth of online communities has resulted in an increased availability of user-generated content (UGC). Given the varied sources of UGC, the quality of information it provides is a growing challenge. While many aspects of UGC have been studied, the role of data structures in gathering UGC and nature of to-be-shared content has yet to receive attention. UGC is created in online platforms with varying degrees of data structure, ranging from unstructured to highly-structured formats. These platforms are often designed without regard to how the structure of the input format impacts the quality of outcome. In this study, we investigate the impact of the degree of data structure on the perceived quality of information from the novel perspective of data creators. We also propose and evaluate a novel moderating effect due to the nature of content online users wish to share. The preliminary findings support our claims of the importance of these factors for information quality. We conclude the paper with directions for future research and expected contributions for theory and practice.

1. Introduction

Information quality (IQ) has been a core topic in Information Systems for many decades [1]. Traditionally, this topic has been studied in the context of organizations, where data creation involved employees or customers [2]. However, starting with the explosion of online content, especially in social media and crowdsourcing, organizations are attempting to harness the large volume of user-generated content (UGC) to enhance their decision making [3]. UGC is quite different from traditional content and requires updated approaches to information quality. UGC is created by members of the general public, who are often casual content contributors (the crowd) with weak or no formal ties to the organizations wanting to use their data [4]. This feature of UGC makes data consumers skeptical about the quality of UGC. This growing concern – information quality of UGC or crowd IQ [5] – has attracted much attention in recent years [6], [7]. We add to this growing body of work by considering data creators' perception of the quality of the data they produce – thus far an under-explored research area.

Data collection is conducted in different ways in online platforms. Especially in UGC context, the information collection interfaces are designed in different structures, formats and features. Despite significant growth in online services, there has been little research on conceptual modeling of IS built to collect online content [8] (i.e., data collection platforms). Given the popularity of UGC platforms, we argue that investigating the design of data collection platforms could benefit both data creators and data consumers. [1] As UGC participation is voluntary, individuals may be dissuaded if the process of contributing is difficult [12], [13].

While usually IQ is measured from the point of view of data consumers, perceptions of IQ by those who create the data also matters as they represent the end result of the data creation transaction. Once the users are finished creating data (e.g., click on a "submit" button), they may reflect on the process and wonder whether the data they submitted accurately and fully reflected their perceived

1 Following a widely accepted convention in the IS research community we use the terms information and data interchangeably [4], [9]–[11].

URI: https://hdl.handle.net/10125/64253
978-0-9981331-3-3

HICSS

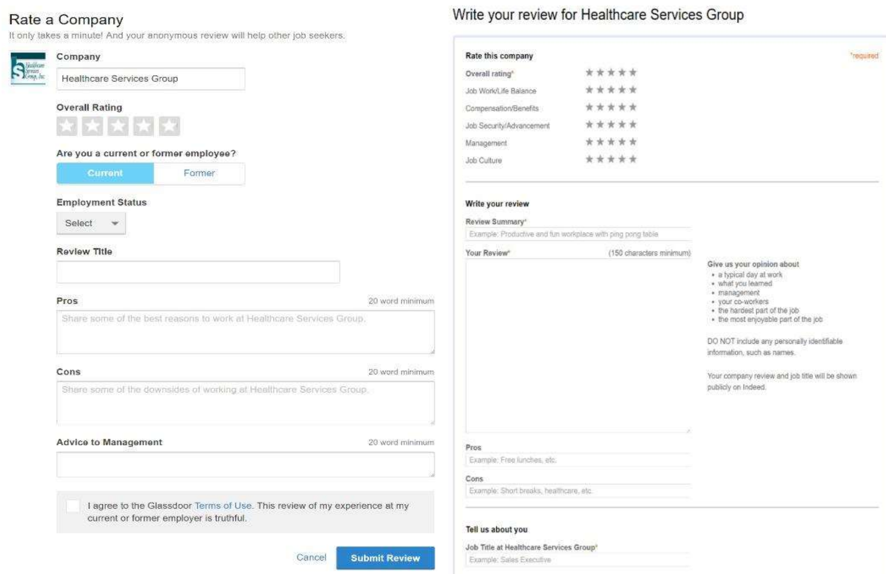

Figure 1. Screenshot of two similar-purpose websites with different data-entry structure [Glassdoor.com on the left, and Indeed.com on the right]

experience. Notwithstanding the commonsense nature of this issue, virtually no research to date investigated information quality from the perception of data consumers *who provide these data* (as opposed to evaluate the data created by others).

Previous findings imply that data creators are sensitive to the quality of data they produce. Studies show that inadequate data entry choices may discourage data creators from entering data [14]. Users faced with inadequate choices may resort to guessing or intentionally sabotage or provide erroneous information [4], [5]. This may also result in dissatisfaction and a loss of immediate users, as well as potential users since a discouraged user will not promote the platform to others.

Finally, such incomplete or inaccurate data, brings about serious problems for data consumers hoping to leverage UGC in their decision making and analysis [15]–[17]. These issues strongly suggest the need for continued research on improvements in designing data collection platforms to allow users to fully and accurately express themselves (reflected in higher perceptions of IQ by the users). We posit that two critical aspects have yet to receive scrutiny: (1) Data structuredness - varying degrees of the structure of data collection interfaces, and (2) Nature of shared content - the type of data that creators wish to share.

First, our model aims to improve IQ by focusing on data structures. We argue that this aspect should be examined, because IS design can affect users' behavior [18]. Currently, UGC platforms have varying degrees of data structure, ranging from unstructured (e.g., text area) to highly structured format (e.g., rigid and specific forms) [19]. The existence of diverse UGC platforms and different formats show that a unified design for UGC lacks consensus [20], [21]. Although some recent research has been done on this issue, research on UGC indicates that we have little understanding of the appropriate degree of data structure for UGC data gathering and its downstream impact on the quality of information [22]. There are many websites, where we can see similar data is collected with different structures. For example, Glassdoor (www.Glassdoor.com), which is a recruiting website, lets users review companies by providing a form asking for overall rating (1 to 5 stars), employee status (two-choice question), employment status (Drop-down list), review title (blank field), Pros, Cons, and advice to management (previous three use text areas). On the other hand, a similar website, Indeed (www.Indeed.com) asks for overall rating in addition to several specific dimensions about the company (1 to 5 stars), review summary (blank field), a review (open box), and pros/cons (each a blank field) (see Fig. 1).

Admittedly, designers use different interfaces to better serve different purposes. For example, Lukyanenko et al 2019 [4] found that both structured and unstructured (in their case – instance-based) data collection interfaces have their advantages and disadvantages [5] and serve different purposes. However, every interface design decision constrains the data received by the platform. Typically, when data is created and stored in a structured form, the process of information creation is relatively transparent and well-controlled [16]. On the other hand, the resulting data from an unstructured data collection interface may be more detailed and nuanced, but harder to analyze and understand [4], [5], [23]. Without a structure to guide them, users have autonomy in providing what they consider helpful or important in unstructured interfaces. While this autonomy generally leads to greater richness and detail with the potential for novel insights, the inherent heterogeneity in the data received makes its interpretation and analysis challenging. Incomplete data is also a greater possibility with unstructured data collection interfaces. Without being asked or prompted, data creators are unlikely to consider providing all important dimensions of the data desired by data consumers.

Second, previous research shows that the design of data-collection interfaces is typically based on users' memory [27], [28]. Briefly, semantic memory is described as the memory of facts and general knowledge, while episodic memory is related to personal experiences. For example, one's memory of having a cup of coffee with a friend at a specific coffee shop on X street at a specific time is related to episodic memory. However, knowing that there is a coffee shop on X street is associated with semantic memory [29].

In summary, we believe that conceptual modeling of data structruedness and nature of shared content are important factors to consider in the design of data collection interfaces. These factors are prevalent in online interfaces in use today, yet little thought has been given to them by prior research. We believe both of these factors will improve the perceived quality of data

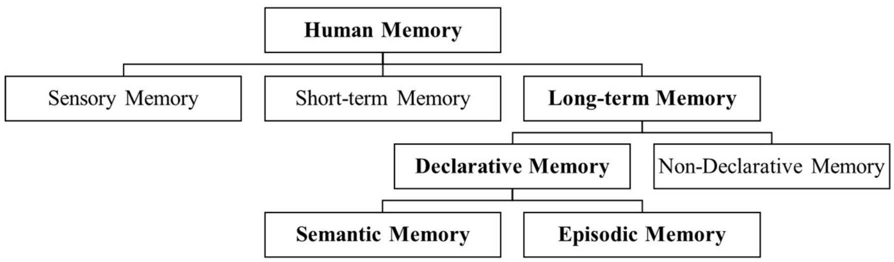

Figure 2. Memory hierarchy

preferences and factors that encourage participation [24]. However, we contend that interface designers should consider designing the data collection interface in alignment with the nature of shared content that is supposed to be recorded as UGC. Prior research in non-UGC contexts has shown that users are frustrated by this misalignment. For instance, an employed doctor in a hospital, with a level of high incentive for participation, has little option for recourse except to vent frustrations [25]. However, a voluntary participant in a UGC context, with a much lower incentive to participate, might simply walk away when presented with an interface which does not provide appropriate alignment. The interface is designed to collect data creators' opinions which stem from their cognition. We argue that considering this cognition will result in better interface design. Studies of human memory have suggested a generic hierarchy (Fig. 2 – memory hierarchy, adapted from [26]). Data creators draw from long-term memory to provide content. Long-term memory is usually divided into two types: declarative memory and non-declarative. UGC in online communities is associated with declarative memory, i.e., expressions of beliefs and perceptions about and stories of experiences with some phenomenon. The generated content is a way for the users as data creators to assert themselves. Declarative memory, which is referred to as explicit or conscious memory, can be categorized into semantic and episodic provided by data creators.

We performed several experiments to assess the data creators' perceived quality of information resulted from differently designed interfaces and nature of shared content. Our results show that using interfaces with different degrees of structure does indeed result in noticeable differences in data creators' perception of information quality.

The following section provides an overview of our theoretical background. Then, we briefly go through our experiment. In the final section, we provide a short description of the preliminary results and conclusion.

2. Theoretical background

In Fig. 3, we illustrate the theoretical model used to explore the impact of data collection interface on the perceived quality of data in UGC platforms.

As discussed earlier, we refer to our conceptualization of the dependent variable as Perceived Information Quality (perceived IQ) - defined here as the assessment of quality of *own data* by a data creator.

As prior research has already studied the perceptions of IQ (i.e., perceptions of data consumers of the data created by others) [30], we adapted the measurement model for our construct from prior research [31]. Consistent with prior research (focused on consumers) we evaluate perceived IQ in terms of content, process, and structure of the data creation. We evaluate *content* in terms of completeness and accuracy of the recorded content. We use time and speed of data entry to assess the *process* and also flexibility and utility of structure to assess *structure* of the given interface.

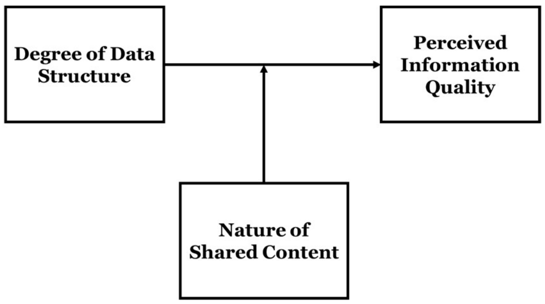

Figure 3. Theoretical model

Next, data structure (see the box on the left in Fig. 3) has always been a concern in designing the information collection platforms [32], [33]. Some studies show that the interface can influence the system outcome and users' perception of the system. The user interface in UGC setting, that we refer to as the information collection interface, can be designed in different ways. There are a wide variety of factors that can be considered in designing the interface. Interface simplicity, generality, convenience, ease of navigation, functionality are some of the factors that have been studied in previous research [34], [35]. One of the factors that has yet to be studied is data structuredness. In this research, data structuredness refers to the degree an interface allows for various representations of data [36]. Data structuredness illustrates the extent to which the collected data is organized within pre-defined fields and can thereby be directly processed [37]. Losee [36] used three categories for interface design in terms of degree of structure. A format with the lowest degree of data structure is unstructured (flexible) format in which data fields are text areas, and users can write their answers in sentences or paragraphs. On the other hand, structured (fixed) format has higher degree of structure in which users are allowed to select the preferred answer among the pre-defined options. In this format, data is collected through drop-down menus, multiple-choice options, or some kind of predefined categories [36], [38]. In this research, the flexibility and design of data collection interface is referred to as data structuredness. As we already mentioned, UGC data are collected using different formats. Some online platforms provide text areas and ask users to write their opinion, status, comment or feedback, while others provide rigid forms with drop-down menus. Data consumers expect unstructured and flexible information when they provide text areas. On the other hand, they expect fixed and structured information when they give users rigid forms. This concept has been studied, showing the importance of this factor in collecting data from crowds [39]. Knowing that a well-designed interface can improve the performance of a system [40], we posit that there could be an impact of the degree of structure on perceived IQ. Given the freedom that unstructured interface brings to data creators, we hypothesize:

Hypothesis 1: *In UGC settings, data creators perceive that using less-structured information collection interfaces results in recording data with higher quality than using more-structured information collection interfaces.*

We also take into account another factor that could affect the perceived quality of data collected in UGC. User-centered design, introduced by Norman and Draper [41], considers a design process where end-users can influence how a design takes shape. This perspective has been utilized in terms of consulting users about their needs, considering their expectations [42] or users' physical abilities [43]. Simply, user-centered design seeks an output design where the task fits users' abilities. While this is important, it does not consider whether the task fits how users think. In other words, users' thoughts that originated from their mind will be turned into the resulting UGC through the medium of the data collection interface [44]. We suggest that the nature of these thoughts that are going to be the UGC should be considered in designing the data collection interface. To the best of our knowledge, there is no previous study in the IS discipline that considers the nature of shared content (see bottom box in Fig. 3) in designing data collection interfaces. We borrow from prior research on the importance of designing an interface for use in accordance with users' information processing capabilities [45], [46].

According to human information-processing approach in psychology, the humans' mind performance is a function of several processing stages, with the central metaphor of "a human is like a computer" [46], [47]. These studies imply the potential similarities between human mind and IS. We posit that not only the design of interface [46] but also the process of data collection in online platforms should be in accordance with users' information processing capabilities. However, the nature of shared content does not influence the information quality of recorded outcome directly. In other words, we argue that the nature of shared content can moderate the relationship between data collection interface and IQ.

In UGC settings, we deal with data provided by online users (i.e., our data creators) from their declarative, long-term memory given that this is where much of the produced UGC originates. We posit that designing interfaces with sensitivity to the human memory structure may result in improvements in IQ of UGC. The two sub-categories of declarative memory are episodic and semantic human memory. Episodic memory is about temporally dated events and the temporal-spatial relations of the events [27], [48]. Episodic data in human mind is stored as a perceptual event in terms of its autobiographical reference to other similar contents and is quite susceptible to transformation and loss of information. This kind of memory is describable in terms of their perceptible dimensions, and its retrieval is related to the knowledge of the individual of his or her personal identity [49], [50]. As Tulving [27] stated: "since information in episodic memory is always temporally dated, and since it can only be retrieved if its temporal date is sufficiently accurately specified by the retrieval cue, interference with temporal coding may render access to the to-be-retrieved material difficult or impossible." Tulving's description of episodic information features and its comparison to semantic information implies less structure and organization in storing the information for episodic type of content. To be able to enhance the quality of episodic information in UGC setting, we hypothesize the following:

Hypothesis 2: *While recording episodic content, data creators perceive that using less-structured information collection interfaces results in recording data with higher quality than using more-structured information collection interfaces.*

On the other hand, Tulving [50], [51] referred to semantic memory as an organized knowledge that an individual possesses about verbal symbols and their meaning or referents. Semantic memory is much less vulnerable to be changed or be forgotten than episodic memory [48], [52]. According to Tulving [27], semantic information is "always referred to an existing cognitive structure, that is, they always have some cognitive reference…". Semantic memory information usually represents objects, concepts, facts and so on, which all are detached from their autobiographical reference. Also, Tulving mentioned: "information in semantic memory, on the other hand, is usually encoded as part of, or assimilated into, a rich multi-dimensional structure of concepts and their relations, and such embeddedness protects the stored information from interference by other inputs." Considering the features of semantic memory and how semantic information is stored and treated in the human mind, we examine the way it should be stored as UGC in online platforms. We hypothesize the following about the appropriate degree of structure to store semantic information.

Hypothesis 3: *While recording semantic content, data creators perceive that using more-structured information collection interfaces results in recording data with higher quality than using less-structured information collection interfaces*.

3. Methodology

We conducted multiple experiments to test our hypotheses. Considering the moderating effect of nature of shared content, we provided participants with different interfaces and asked them to record the requested data. Finally, we asked data creators to rate the perception of the quality of information they recorded. In the experiments, three online information collection interfaces were designed with different degrees of structure. We used the following popular and commonly used degrees of data structure [36], [38]:

- Structured (fixed) format: data fields in this format are either drop-down menus or multiple-choice options. Users are allowed to select the preferred answer among the pre-defined options.
- Semi-structured format: data fields in this format let users write down their answers. However, they are asked to provide at least two hashtags. They can create a hashtag by typing # followed by a term. This term can be a keyword or an important point of the answer (to the best of users' understanding). In this condition, the generated hashtag adds some structure to the unstructured text-area-driven content.
- Unstructured (flexible) format: data fields in this format is text area, and users can write down whatever they want in sentences or paragraphs.

To consider different types of human memory, the participants were asked to record either semantic or episodic types of information. We conducted two experiments in two different domains: asking participants to provide their experiences on their recent flight, and also on their recent medical/health issues. Using psychology literature on semantic and episodic memory, we developed several scenarios that are looking for content representing these two memory types. To validate the scenarios, we consulted with two psychology experts and conducted a classification task among independent judges.

Fig. 4 shows the detailed version of our theoretical model that includes all the studied aspects of each variable. We used Amazon Mechanical Turk (MTurk) to run the experiments. MTurk is a crowdsourcing platform that offers access to large numbers of job requesters (people who post tasks) and crowd workers (people who perform the tasks [53]). It allows crowd workers to perform tasks in exchange for monetary rewards [54], [55]. Previous studies show various applications of MTurk in research. It has been widely used for different purposes becoming widely accepted in business research [53], [60], [61], transcriptions [56], experiment designs [57], qualitative designs [58] and user evaluation studies [59]. As our theory is based on general memory systems, we conducted the experiments with the general population. We attracted MTurk workers ages 18 to 60 living in the United States, to participate in the experiments.

We followed between-subject design where participants were assigned to only one of the interfaces and one of the scenarios. We collected total of 181 responses. 58 responses were collected through unstructured interface, 61 responses were collected through semi-structured interface, and 62 responses were collected through structured format[2]. Based on the collected demographic information, our participants include 123 (68 %) female and 58 (32 %) male MTurk workers, and they represent a wide range of age from 18 to over 55 years old. More than 80 percent of the participants hold at least a college degree. And more than 96 percent were native English speakers, so they did not have a language proficiency barrier to respond to the questions.

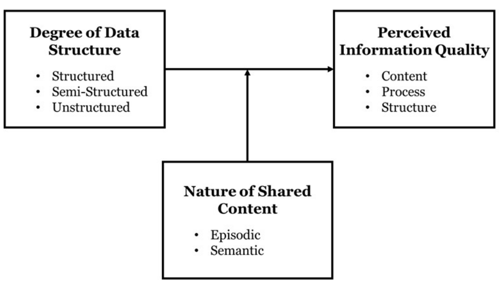

Figure 4. Detailed theoretical model

After recording their responses in the provided information collection interfaces, each MTurk participant was asked to take a survey to evaluate the perceived quality of recorded information. The survey is adapted from previous research on IQ [30], [31], [62], [63] and aims to assess the perceived quality of content, process, and structure of recorded information. The survey contains several items for each category. The participants were asked to indicate their assessment on a seven-point Likert scale.

4. Results

In this section, we briefly present the initial results of the first experiment. We collected data from users who use differently structured interfaces to share their experience or knowledge. We asked users to assess their perception of information quality of their recorded responses. Our measurement instrument evaluates this perception of information quality in terms of completeness, accuracy, ease of data entry, the speed of data entry, flexibility and utility of structure. Now, we have data sets that represent users' perceived IQ based on the three different interfaces they used. To compare the interfaces, we run analyses of variance test. This test is usually used to compare three or more group means for statistical significance.

First, we run an ANOVA test to assess the differences among the three categories of data structure: structured, semi-structured, and unstructured (table 1). The preliminary findings show that using differently-designed interfaces result in noticeable differences in perceptions of IQ. This result is consistent with hypothesis 1, where we proposed that collecting data through interfaces with different degrees of structure affects perceived information quality.

We have evaluated the recorded information in terms of completeness and accuracy. Also, the results on perceived flexibility and utility of structure are significant. Specifically, data creators' perception of accuracy, completeness, flexibility, and utility of structure decreases for interfaces with highly structured formats compared to those with unstructured and semi-structured formats.

The insignificant findings for the ease of data entry and speed of data entry suggest that people perceive structured and unstructured interfaces to be generally comparable in the amount of expanded effort. This is a notable finding that implies that designers should not sacrifice usability while tailoring the interfaces to the nature of to-be-shared content.

Table 1. ANOVA results of perceived IQ measures (unstructured, structured and semi-structured)

Dimension	**DF**	**F-value**	**Sig.**
Completeness	2	5.055	0.007^{**}
Accuracy	2	4.873	0.009^{**}
Ease of data entry	2	0.383	0.682
Speed of data entry	2	1.409	0.247
Flexibility	2	2.822	0.062
Utility of structure	2	3.223	0.042^{*}

** $p < 0.01$; * $p < 0.05$

[2] Additionally, as many real-world websites implement character limits, for ecological validity, we also had unstructured conditions with 700 character limit. For brevity, we do not discuss the results of these conditions here, but note that the results are fully consistent (i.e., having the same direction and effect significance) as the conditions we report. Having multiple variants of the same underlying condition behave in the same way strengthens the validity of our findings.

Table 2. Post hoc analyses

Dimension	Cond.	US	SS	HS
Completeness	US		0.03	0.46
	SS	-0.03		0.43
	HS	-0.46	-0.43	
Accuracy	US		0.05	0.42
	SS	-0.05		0.37
	HS	-0.42	-0.37	
Ease of data entry	US		0.11	0.09
	SS	-0.11		-0.02
	HS	-0.09	0.02	
Speed of data entry	US		0.27	0.08
	SS	-0.27		-0.20
	HS	-0.08	0.20	
Flexibility	US		-0.01	0.33
	SS	0.01		0.34
	HS	-0.33	-0.34	
Utility of structure	US		0.17	0.41
	SS	-0.17		0.24
	HS	-0.41	-0.24	

- US: unstructured, SS: semi-structured, HS: highly structured
- The cell values indicate the mean difference (condition in a row - condition in a column)
- The shadowed cells indicate that the mean difference is significant at the 0.05 level.

Second, we run another test to evaluate the episodic and semantic content separately. We run two separate ANOVA tests for each group and assess the difference in data structure. First, we only consider episodic content. These contents were recorded using either unstructured, semi-structured or structured interface. So, we run the test to investigate any significant difference among the interfaces. The findings are very similar to the situation where we did not consider the nature of content. The unstructured format was able to collect episodic content with significantly higher completeness (mean score of the 7-point Likert scale = 6.30, $p < 0.001$) and accuracy (6.15, $p < 0.01$). The results for perceived flexibility (5.63, $p < 0.05$) and utility of structure (6.17, $p < 0.01$) are also significant. However, there is no significant difference detected for perceived ease ($p = 0.445$) and speed of data entry ($p=0.616$). Then, we run the same test for semantic content. These contents were recorded using either unstructured, semi-structured or structured interface. The findings are very different from episodic content. We did not detect any significant differences between groups for any dimension.

Our findings indicate that using different degrees of structure in designing information collection interfaces results in different levels of perceived IQ. Online users perceive that they are able to provide more complete and accurate information while using less-structured interfaces. Also, the results indicate that the highest results for IQ dimensions are consistently obtained when the data is entered in an unstructured format, which demonstrates peoples' appreciation of the value of unstructured data collection.

The preliminary findings are of great significance to theory and practice of UGC and information quality. They show the value of unstructured data, especially for storing episodic information. In particular, they suggest that when more rigid formats are used to capture UGC, people may feel that they were not able to fully express themselves. This may have a negative impact on the overall experience of people with the platform where data collection took place. Equally important, considering that users perceive some deficiencies for completeness and accuracy when the wrong format is chosen, it means that when organizations begin to analyze such UGC, these deficiencies may result in inappropriate decisions being made based on such data. Further, our initial findings support our hypotheses on the importance of nature of shared content. In certain conditions, our results were significant, when we compared the data sets based on the type of recorded information. This finding encourages us to consider more detailed future work to see the users' response to record different types of information.

Our work contributes to theory by showing the importance of distinguishing episodic and semantic type of information. That human memory is organized differently, reflecting different aspects of human experience, may have vast ramifications for the information systems discipline. For example, episodic information appears to be more prone to external interferences. This means that in applications such as customer service, Q&A platforms or crowdsourcing apps having the aim is to understand the experiences of users, need to be especially careful in asking probing or leading questions (which may frame and even distort memory) when dealing with episodic information. Likewise, the platforms collecting episodic data (e.g., social media), need to continue exploring innovative designs for flexible data collection (e.g., voice, video interfaces) for the episodic content to be fully captured.

Our research demonstrates a connection between information system design and human memory – an obvious, but thus far neglected IQ factor. Indeed, especially in UGC settings, much of the content produced originated in the memory of the contributors. Consequently, it is reasonable to posit that a greater sensitivity to the memory structures of humans, may result in improvements in IQ - a proposition that has already been supported by our preliminary findings. Further, this study will extend the literature by enhancing our understanding of the structure-degree of different kinds of UGC. The results of this study can be used by researchers to further discuss the role of an appropriate degree of structure to collect data effectively and efficiently. Furthermore, findings of this research enrich the IQ literature by empirically comparing perceived IQ by data creators and data consumers– another common issue that has been consistently understudied in previous research.

5. Implications and Future Studies

In this study, we investigate the impact of the degree of data structure on the perceived quality of information from the novel perspective of data creators. Our work highlights the importance of considering type of memory in interface design. Data creators perceive that they are able to record more complete and accurate content, while using less-structured interface format. However, this is not always true. Our findings show that data creators' perception depends on the nature of shared content. In other words, data creators do perceive that they are able to record more complete and accurate content, while using less-structured interface format, but just for episodic content. On the other hand, data creators' perception of IQ for an interface intended to collect semantic content does not depend on the interface's degree of structure.

Our work can help practitioners improve the design of UGC platforms and support the adoption of UGC by organizations. Providing an appropriate data-entry interface helps to enhance users' self-expression. The findings of this study will lead to new design principles that improve users' ability to express themselves online. At this stage, our proposed principles could include:

Assigning a suitable interface to collect the information. Perhaps the most notable design guideline originating in our findings is that different kinds of information should be collected using different interfaces. This principle is driven by the clear perception among subjects that there is a difference in ideal levels of structured-ness to support episodic and semantic content sharing, impacting the completeness and accuracy of the information they hope to share. Specifically, we see that episodic information is best collected using less-structured interfaces, whereas semantic content requires more structured ones.

Identifying the nature of to-be-shared content. Given the differences noted in perceptions of IQ for episodic and semantic content, it is possible to dynamically configure an interface based on a user's desired content form. Prior to recording content, the data creators could be given a very simple question that asks about the nature of content: whether the to-be shared content is associated to data creators' personal experience or his/her knowledge about something, and the interface modified accordingly.

Level of structure is clearly a defining factor for capturing different types of shared content. In particular, episodic content requires less structure to ensure high perceived completeness and accuracy; however, interface designs with some additional structure may help to focus users on characteristics of their experience that might be relevant. We suggest the following:

Specifying the topic of to-be-shared content. Depending on the context, usually the possible topics are predictable. For example, in a specific forum of an online healthcare community that is meant to discuss a particular disease, symptoms, drugs, side effects, treatments, lifestyles, and physicians are the possible topics. Data creators could help specifying the topic of to-be shared content, by responding to a single multiple-choice question. This structure allows the user to self-label the content, which may improve accuracy of IQ, since this would capture the user's intention.

Recognizing notable possible points in a specific topic. If all of the most-rated previous comments on the topic of side effects have talked about a severe headache, the system could ask data creator about the possible headache or suggest talking about headache. Using text mining, designers could use previous content to create a potential structured data collection interface for different topics. This structure directs users to common themes in the data, which may help users to greater completeness IQ.

Our study is not without limitations. For instance, different prompts might have elicited different content, leading to different results. Further, we asked about personal experiences with the flu, and it is possible that data creators may have opted to not share full details of health-related experiences.

Overall, we are encouraged by our preliminary findings, and are working on a follow-up study to measure data consumers' perceived quality of the semantic versus episodic UGC data collected using the structured, semi-structured and unstructured interfaces. The recorded content from the current study will be shown to the independent participants of a new study, who will be asked to assess the quality of recorded information and make inferences about the person who provided this information based on the information provided. Using this method, we want to empirically compare data creators versus data consumers' perception of IQ and determine if people's perceptions of quality agree with objective measures. We expect the findings to be consistent with those of the data creators' perceived IQ. The comparison between the two sides of perceived IQ stands to further contribute to the theory of information quality. Our results may indicate the need to measure both data creators and consumers point of view on perceived IQ to gain a comprehensive perspective on the impact of the variables of interest on information quality.

6. References

[1] S. Petter, W. DeLone, and E. R. McLean, "Information Systems Success: The Quest for the Independent Variables," *Journal of Management Information Systems*, vol. 29, no. 4, pp. 7–62, 2013.

[2] N. Agmon and N. Ahituv, "Assessing data reliability in an information system," *Journal of Management Information Systems*, pp. 34–44, 1987.

[3] C. Wyrwoll, "User-Generated Content," in *Social Media*, Springer, 2014, pp. 11–45.

[4] R. Lukyanenko, J. Parsons, Y. Wiersma, and M. Maddah, "Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-generated Content.," *MIS Quarterly*, vol. 43, no. 2, 2019.

[5] R. Lukyanenko, J. Parsons, and Y. Wiersma, "The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-generated Content," *Information Systems Research*, vol. 25, no. 4, pp. 669–689, 2014.

[6] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, "Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions.," *ACM Computing Surveys*, vol. 51, no. 1, p. 7, 2018.

[7] M. Kosmala, A. Wiggins, A. Swanson, and B. Simmons, "Assessing Data Quality in Citizen Science," *Frontiers in Ecology and the Environment*, vol. 14, no. 10, pp. 551–560, 2016.

[8] R. Lukyanenko and J. Parsons, "Conceptual modeling principles for crowdsourcing," in *International Workshop on Multimodal Crowdsensing*, Maui, Hawaii, USA, 2012, pp. 3–6.

[9] Y. W. Lee, L. L. Pipino, J. D. Funk, and R. Y. Wang, *Journey to data quality*. Cambridge, Mass: MIT Press, 2006.

[10] A. Parssian, S. Sarkar, and V. S. Jacob, "Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product," *Management Science*, vol. 50, no. 7, pp. 967–982, Jul. 2004.

[11] R. Y. Wang, "A product perspective on total data quality management," *Communications of the ACM*, vol. 41, no. 2, pp. 58–65, 1998.

[12] O. Nov, O. Arazy, and D. Anderson, "Dusting for science: motivation and participation of digital citizen science volunteers," *2011 iConference*, pp. 68–74, 2011.

[13] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk," in *Proceedings of the ACM SIGKDD workshop on human computation*, 2010, pp. 64–67.

[14] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[15] R. Y. Wang, V. C. Storey, and C. P. Firth, "A framework for analysis of data quality research," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 7, no. 4, pp. 623–640, 1995.

[16] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data quality in context," *Communications of the ACM*, vol. 40, no. 5, pp. 103–110, 1997.

[17] H. Xu, J. Horn Nord, N. Brown, and G. Daryl Nord, "Data quality issues in implementing an ERP," *Industrial Management & Data Systems*, vol. 102, no. 1, pp. 47–58, 2002.

[18] M. Hibbeln, J. L. Jenkins, C. Schneider, J. S. Valacich, and M. Weinmann, "How Is Your User Feeling? Inferring Emotion Through Human–Computer Interaction Devices," *MIS Quarterly*, vol. 41, no. 1, pp. 1-A7, Mar. 2017.

[19] M. Maddah, R. Lukyanenko, and M. Chiarini Tremblay, "Toward Theory and Method of Hybrid Data Collection," presented at the International Conference on Design Science Research in Information Systems and Technology (DESRIST), St. John's, NL Canada, 2016, pp. 93–95.

[20] Y. Jin, Y. Wen, G. Shi, G. Wang, and A. V. Vasilakos, "CoDaaS: An experimental cloud-centric content delivery platform for user-generated contents," in *2012 International Conference on Computing, Networking and Communications (ICNC)*, 2012, pp. 934–938.

[21] M. Obrist, D. Geerts, P. B. Brandtz\a eg, and M. Tscheligi, "Design for creating, uploading and sharing user generated content," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 2008, pp. 2391–2394.

[22] R. Lukyanenko, "Information Quality Research Challenge: Information Quality in the Age of Ubiquitous Digital Intermediation," *J. Data and Information Quality*, vol. 7, no. 1–2, pp. 3:1–3:3, 2016.

[23] P. Buneman, S. Davidson, M. Fernandez, and D. Suciu, "Adding structure to unstructured data," in *Database Theory—ICDT'97*, Springer, 1997, pp. 336–350.

[24] R. Stoeckl, P. Rohrmeier, and T. Hess, "Motivations to produce user generated content: Differences between webloggers and videobloggers," *BLED 2007 Proceedings*, p. 30, 2007.

[25] M. Green, "25 Quotes That Show Just How Fed Up Physicians Are With EHRs," 2015.

[26] M. Herz and K. H. Brunk, "Conceptual advances in consumers' semantic and episodic brand memories: A mixed methods exploration," *Psychology & Marketing*, vol. 34, no. 1, pp. 70–91, 2017.

[27] E. Tulving, "Episodic and semantic memory 1," *Organization of Memory*, vol. 381, no. 4, 1972.

[28] L. R. Squire and S. M. Zola, "Episodic memory, semantic memory, and amnesia," *Hippocampus*, vol. 8, no. 3, pp. 205–211, 1998.

[29] E. Goldstein, *Cognitive psychology: Connecting mind, research and everyday experience*. Cengage Learning, 2010.

[30] B. H. Wixom and P. A. Todd, "A theoretical integration of user satisfaction and technology acceptance," *Information Systems Research*, vol. 16, no. 1, pp. 85–102, 2005.

[31] J. D. Xu, I. Benbasat, and R. T. Cenfetelli, "Integrating service quality with system and information quality: An empirical test in the e-service context.," *Mis Quarterly*, vol. 37, no. 3, 2013.

[32] B. Laurel and S. J. Mountford, *The art of human-computer interface design*. Addison-Wesley Longman Publishing Co., Inc., 1990.
[33] B. Shneiderman, *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India, 2010.
[34] R. T. Fielding and R. N. Taylor, "Principled design of the modern Web architecture," *ACM Transactions on Internet Technology (TOIT)*, vol. 2, no. 2, pp. 115–150, 2002.
[35] E. Kim, S. Eom, and S. Uoo, "Effective user interface design for online stores in the Asia Pacific region: A survey study," *AMCIS 2001 Proceedings*, p. 168, 2001.
[36] R. M. Losee, "Browsing mixed structured and unstructured data," *Information processing & management*, vol. 42, no. 2, pp. 440–452, 2006.
[37] H. Baars and H.-G. Kemper, "Management support with structured and unstructured data—an integrated business intelligence framework," *Information Systems Management*, vol. 25, no. 2, pp. 132–148, 2008.
[38] X. Zhang, Y. Yu, H. Li, and Z. Lin, "Sentimental Interplay between Structured and Unstructured User-Generated Contents-An Empirical Study on Online Hotel Reviews," *Online Information Review*, vol. 40, no. 1, 2016.
[39] M. Turoff, M. Chumer, B. V. de Walle, and X. Yao, "The design of a dynamic emergency response management information system (DERMIS)," *Journal of Information Technology Theory and Application (JITTA)*, vol. 5, no. 4, p. 3, 2004.
[40] B. Shneiderman, *Designing the user interface*. Boston, MA: Pearson Education, 2003.
[41] D. A. Norman and S. W. Draper, *User centered system design: New perspectives on human-computer interaction*. CRC Press, 1986.
[42] C. Abras, D. Maloney-Krichmar, and J. Preece, "User-centered design," *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, vol. 37, no. 4, pp. 445–456, 2004.
[43] A. Holzinger, "User-Centered Interface Design for disabled and elderly people: First experiences with designing a patient communication system (PACOSY)," in *International Conference on Computers for Handicapped Persons*, 2002, pp. 33–40.
[44] M. Maddah, R. Lukyanenko, and D. VanderMeer, "Designing UGC Interfaces: Considering Level of Structure and Type of Data," presented at the AIS SIGSAND Symposium, Syracuse, NY, 2018.
[45] A. Dix, "Human-computer interaction: A stable discipline, a nascent science, and the growth of the long tail," *Interacting with Computers*, vol. 22, no. 1, pp. 13–27, 2010.
[46] J. A. Jacko, *Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. CRC press, 2012.
[47] R. Lachman, J. L. Lachman, and E. C. Butterfield, *Cognitive psychology and information processing: An introduction*. Psychology Press, 2015.
[48] H. Burianova, A. R. McIntosh, and C. L. Grady, "A common functional brain network for autobiographical, episodic, and semantic memory retrieval," *Neuroimage*, vol. 49, no. 1, pp. 865–874, 2010.
[49] D. B. Mitchell, "How many memory systems? Evidence from aging.," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 15, no. 1, p. 31, 1989.
[50] E. Tulving, "How many memory systems are there?," *American psychologist*, vol. 40, no. 4, p. 385, 1985.
[51] E. Tulving, "Episodic memory: from mind to brain," *Annual review of psychology*, vol. 53, no. 1, pp. 1–25, 2002.
[52] D. L. Greenberg and M. Verfaellie, "Interdependence of episodic and semantic memory: evidence from neuropsychology," *Journal of the International Neuropsychological society*, vol. 16, no. 5, pp. 748–753, 2010.
[53] X. Deng, K. D. Joshi, and R. D. Galliers, "The duality of empowerment and marginalization in microtask crowdsourcing: Giving voice to the less powerful through value sensitive design," *Mis Quarterly*, vol. 40, no. 2, pp. 279–302, 2016.
[54] M. Buhrmester, T. Kwang, and S. D. Gosling, "Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data?," *Perspectives on psychological science*, vol. 6, no. 1, pp. 3–5, 2011.
[55] M. B. O'Leary, J. M. Wilson, and A. Metiu, "Beyond being there: The symbolic role of communication and identification in perceptions of proximity to geographically dispersed colleagues.," *MIS quarterly*, vol. 38, no. 4, pp. 1219–1243, 2014.
[56] M. Marge, S. Banerjee, and A. I. Rudnicky, "Using the Amazon Mechanical Turk for transcription of spoken language," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 5270–5273.
[57] E. Bursztein, S. Bethard, C. Fabry, J. C. Mitchell, and D. Jurafsky, "How good are humans at solving CAPTCHAs? A large scale evaluation," in *2010 IEEE symposium on security and privacy*, 2010, pp. 399–413.
[58] M. K. Ward and S. M. Broniarczyk, "It's not me, it's you: How gift giving creates giver identity threat as a function of social closeness," *Journal of Consumer Research*, vol. 38, no. 1, pp. 164–181, 2011.
[59] A. Sorokin, D. Berenson, S. S. Srinivasa, and M. Hebert, "People helping robots helping people: Crowdsourcing for grasping novel objects," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 2117–2122.
[60] J. Kees, C. Berry, S. Burton, and K. Sheehan, "An analysis of data quality: professional panels, student subject pools, and Amazon's mechanical turk," *Journal of Advertising*, vol. 46, no. 1, pp. 141–155, 2017.
[61] C. Schneider and T. Cheung, "The Power of the Crowd: Performing Usability Testing Using an On-Demand Workforce," in *Information Systems Development*, 2013, pp. 551–560.
[62] J. E. Collier and D. C. Barnes, "Self-service delight: Exploring the hedonic aspects of self-service," *Journal of Business Research*, vol. 68, no. 5, pp. 986–993, 2015.
[63] T. Pikkarainen, K. Pikkarainen, H. Karjaluoto, and S. Pahnila, "Consumer acceptance of online banking: an extension of the technology acceptance model," *Internet research*, vol. 14, no. 3, pp. 224–235, 2004.