

Designing Information Savvy Societies: An Introduction to Assessability

Andrea Forte, Nazanin Andalibi, Thomas Park, Heather Willever-Farr

College of Computing and Informatics

Drexel University, Philadelphia, PA, USA

{aforte, na477, tp352}@drexel.edu, hwilleverfarr@gmail.com

ABSTRACT

This paper provides first steps toward an empirically grounded design vocabulary for *assessable design* as an HCI response to the global need for better information literacy skills. We present a framework for synthesizing literatures called the Interdisciplinary Literacy Framework and use it to highlight gaps in our understanding of information literacy that HCI as a field is particularly well suited to fill. We report on two studies that lay a foundation for developing guidelines for assessable information system design. The first is a study of Wikipedians', librarians', and laypersons' information assessment practices from which we derive two important features of assessable designs: information provenance and stewardship. The second is an experimental study in which we operationalize these concepts in designs and test them using Amazon Mechanical Turk (MTurk).

Author Keywords

Credibility; information literacy; assessability; Wikipedia

ACM Classification Keywords

H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction

INTRODUCTION

Technologies can be designed to support human needs in different ways; for example, they can be designed using principles of usability [25], accessibility [30], ergonomics [29], or, as we suggest in this paper, *assessability*. We define assessability as the extent to which an interface helps people make appropriate assessments of information quality in a particular context of use. Thinking in terms of assessable design could change how designers approach the problem of creating environments where people search, produce, and access information.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI 2014 April 26 - May 01 2014, Toronto, ON, Canada. Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2473-1/14/04...\$15.00.
<http://dx.doi.org/10.1145/2556288.2557072>

Terms like “information overload,” “drinking from the firehose,” and “information pollution” have surfaced in scholarly and popular press to describe problems associated with unprecedented access to information. Barack Obama issued a presidential proclamation about the need for U.S. citizens to learn sophisticated information literacy skills to function in an information-rich democratic society [26] and global organizations such as the International Federation of Library Associations and Institutions (IFLA) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) compile and publish information literacy materials from around the world [16].

Supporting the development of information literacy skills has largely been the domain of education and librarianship communities. This emphasis on education implicitly suggests that information production systems are features of the world within which people must adapt and learn. Yet, how people learn to act within systems is only half of the story. From a sociotechnical HCI perspective, it is clear that the technologies we use to find, access, and produce information are designed environments that can be modified and extended in response to and anticipation of social, cultural, and cognitive needs.

The need for an integrated HCI response to the problem of information literacy is particularly salient because of the proliferation of participatory information sources such as Wikipedia, Wikia sites, Reddit, and Ancestry.com, which are information sources produced and curated by numerous contributors using collaboration or aggregation platforms. Such systems mediate information production and access for millions of users.

In this paper, we introduce a framework to integrate HCI literature with literature from other disciplines that address information literacy to highlight complementary approaches and gaps. We then report on two studies that advance our understanding of designing systems to support information literacy practices: an interview study from which we derived potential features of assessable designs, and an experimental study in which we operationalized these proposed features as design elements in Wikipedia interface modifications and tested them using MTurk.

THE INTERDISCIPLINARY LITERACY FRAMEWORK

Innovations in information production, search, and access have attracted many researchers to problems associated

with information literacy. It's an area that is important, complex, and ubiquitous, which makes it attractive to researchers of all stripes. Not surprisingly, information literacy research extends across many literatures, draws on diverse theoretical and disciplinary traditions, and encompasses overlapping vocabularies, some of which identify as research on media literacy, search, credibility, persuasion, and trust.

We synthesize these diverse vocabularies by introducing a framework that highlights two critical dimensions of information literacy: assumptions about the skills involved in literate practices, and approaches to transforming practice (See Figure 1).

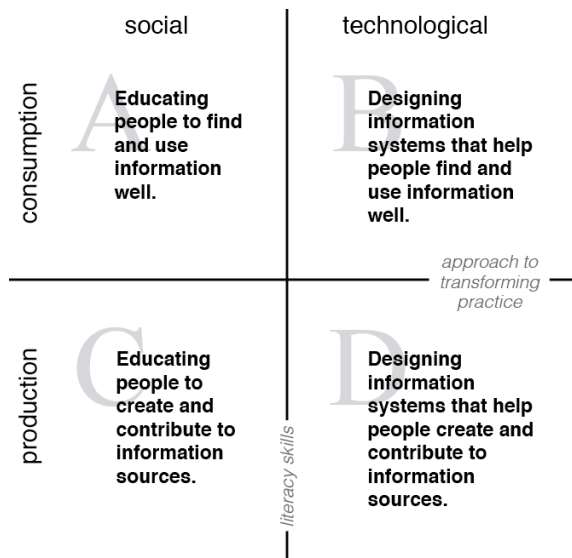


Figure 1: The Interdisciplinary Literacy Framework applied to the domain of information literacy.

Skills involved in literate practice. In order to measure literacy, researchers and educators must define what skills are common to literate individuals and develop methods of assessing these skills. These definitions and methods are based on a set of culturally informed assumptions about what literacy entails. We suggest that literacy entails both consumption and production skills; in traditional terms: both reading *and* writing. In terms of information literacy, on one end of the axis lie practices associated with consumption such as searching, browsing, accessing, and assessing information. Increasingly information literacy research also addresses the other end of the axis: producing and contributing to information resources.

Approach to transforming practice. A second set of assumptions concerns the ways in which these practices can be transformed. Information literacy research often includes a transformative agenda: researchers seek to understand practice in order to improve people's abilities to use and participate in systems of information production. On one end of this axis, social solutions to information literacy

propose innovative education standards and interventions to help people develop more sophisticated literacy skills. On the other end, technological solutions aim to design information environments to make them more amenable to user participation and assessment.

By juxtaposing these dimensions, we can map the literature on information literacy onto four quadrants. What follows is not an exhaustive review, but integrates some of the most relevant HCI work with other disciplines to exemplify four types of approaches to information literacy research.

- A. **Social approaches to encouraging better information consumption practices.** These approaches are by far the most abundant, for example, library instruction classes in which students are taught how to find and assess information resources or specialized programs such as Google search education [15]. Research that supports these approaches has sought to identify existing practices and ways of interpreting technologies [9, 22, 24], to evaluate educational approaches to modifying these practices [20], and to offer standards and guidelines for implementing such activities [2]. (Social approaches may include using educational technologies.)
- B. **Technological approaches to designing for better information consumption practices.** Examples of these approaches include novel search engines designed to support credibility judgments [32, 41] and use of concepts like "information foraging" [31] to predict online information seeking behaviors [3]. In the domain of participatory information sources, WikiTrust [1] is computes a rating of reliability for Wikipedia text based on revision histories and Wikidashboard [18] is a system that exposes editing activity to readers.
- C. **Social approaches to helping people become good contributors to information sources.** These approaches are often initiated to help communities attract contributors as socializing newcomers is a challenge for open collaboration communities [13]. People need to learn not only that they can contribute to information resources, but how to do so well. For example, through efforts like Wikipedia Academies¹, individuals from under-represented groups in the Wikipedia community can receive training. Similarly, projects have been launched to involve university students in the production of articles [8, 21]. Others have evaluated socialization tactics to identify factors that contribute to success [4]. The efforts above are usually framed as projects to support Wikipedia, not would-be learners of production skills.
- D. **Technological approaches to designing for higher quality participation in information production.**

¹ <http://en.wikipedia.org/wiki/Wikipedia:Academy>

Many tools have been developed to support experts as they negotiate complex interactions and production activities. For example, SuggestBot helps Wikipedia editors find tasks they might enjoy and that need to be done [5]. Re:Flex is a system that creates visualizations of interactions among Wikipedia editors and content to support social translucence [23]. Despite research that documents newcomer practices in online production communities [27, 34] and a pervasive belief that interface features such as lack of WYSIWYG can hamper newcomer participation [40], design experimentation to support newcomers in participatory information systems is scarce.

A Critical Gap: Making People Smarter, Not Computers

The quadrants in the Interdisciplinary Literacy Framework do not represent mutually exclusive agendas; most research implicitly or explicitly involve elements of at least two even if it only emphasizes one. Many researchers have discussed the traversal of these axes, particularly as people move between consumption and production using participatory forms of media [17, 28]. Sormunen et al. used writing on a wiki as a context for learning both critical consumption and production skills in Swedish schools [35]. St. Jean et al. examined how content producers assess information they find [36]. Forte and Bruckman developed a wiki for high schools with specialized tools to support sharing citations in a bibliographic database; they suggested that producing an information resource for others not only supports practicing writing, but also transforms strategies for finding and assessing information [12].

Looking ahead, we see a continued need for work that spans these quadrants, in particular where participatory information sources are concerned. Moreover, we see an important gap in quadrant B—an area that HCI is well-suited to fill. Although algorithmic approaches such as those behind credibility-based search results [32, 41] or content ratings in WikiTrust [1] can help deliver important information that support critical consumption, these methods rely on the designers of algorithms to define credibility. Our vision of assessable design is one in which designers support people in learning how information is produced so they can make informed assessments.

Kittur et al.'s early work with Wikidashboard is one of the few projects that has addressed this problem directly [18]. They used Amazon Mechanical Turk to demonstrate that showing people editing activity can affect their judgments of Wikipedia articles: more perceived volatility led to lowered trust and more perceived stability led to higher trust. More recently, Towne et al. [39] tested whether exposing content-related discussions would engender less trust in Wikipedia content by making the messiness of the process more visible; in fact, they discovered that viewing discussions generally lowered trust in a given article (especially when conflict was perceived), but participants *reported* that it raised their assessment both of the article

and Wikipedia as a source. Towne et al. speculate this might be because viewing discussions engenders more confidence in one's ability to assess Wikipedia articles.

These findings suggest that understanding how information is produced informs not only how people assess information (information literacy skills) but also their self-efficacy (propensity to use and confidence in information literacy skills). This raises two questions. First, the findings that people feel more critical when they are exposed to conflict and that viewing visualizations of edit histories affect how people judge content both raise the question of how expert Wikipedians, who should be deeply familiar with production on the site, differ from others' in interpreting such information to inform their assessment strategies:

RQ1. What strategies do expert Wikipedians use to assess articles and how do they differ from the strategies used by people with less knowledge of how information in Wikipedia is produced?

Second, if we can identify features of the site that give people insight into article quality, we want to capitalize on these and begin to develop a vocabulary for talking about assessable designs, using Wikipedia as a testbed.

RQ2. Can we use expert Wikipedia strategies to help inform the design of assessable interface features that help inform non-Wikipedians' strategies?

We address these questions in the following two studies.

STUDY 1: DIARY AND INTERVIEW STUDY

In order to answer RQ1 and better understand interactions between different forms of expertise and assessment of Wikipedia articles, we recruited a strategic sample of individuals that represent a broad swath of information consumers for a diary and interview study:

- *Amateur experts*. Highly experienced Wikipedians, expert producers of a participatory information resource who are not professionally trained. Wikipedians were recruited through Wikimedia Foundation contacts and using snowball recruitment.
- *Librarians*. Librarians are formally trained to be discerning about the selection and assessment of information. We wanted to understand how this formal training influences information assessment strategies. Information professionals were recruited from local libraries using flyers and mailing lists.
- *Novices*. These are individuals who do not participate in information production and have no special training in assessing information, but who use Wikipedia. These were recruited from local undergraduate student populations using flyers and mailing lists.

Methods

We conducted phone or in-person interviews with 6 librarians, 3 Wikipedians, and 3 students for a total of 12

interviewees. Before the interview, participants completed a search diary for up to 7 days (average 5.2 days) in which they recorded details about Internet searches they had conducted. Interviews began with open-ended explorations of participants' experiences with participatory media, then included a semi-structured protocol [33] in which participants responded to specific questions about their search practices grounded in data from their diaries. Finally, in a think-aloud protocol [7], the participants were asked to examine two Wikipedia articles. The first article, on Apollo 9, was relatively short, tagged as needing more citations, and included primarily technical information; whereas the second article, about the moon, had attained the level of featured article by undergoing a rigorous process of community review. We asked participants to provide an assessment of them and verbally explain how they made judgments about the information. Both articles were saved to the first author's website before the interview study in order to ensure that all participants saw the same version.

Interview data were transcribed and coded using the software Dedoose to identify patterns of information assessment strategies within and across participant groups. After open coding yielded a catalog of assessment strategies used by participants, we examined, compared, and grouped these to identify higher-level concepts that describe important features of assessable participatory information sources. This process of open, axial, and selective coding is typical of a grounded theory approach to data analysis [14].

Findings

Our analysis yielded two concepts that describe important features of assessable designs: provenance and stewardship.

Information Provenance. Provenance is a concept that is familiar in disciplines such as history, archival studies, and the sciences and relates to metadata that make it possible to trace the ownership or origins of an artifact or dataset. This is a critical concept for people dealing with participatory information sources. *Where did this information come from?* The most simple technique for sourcing is a citation. In all populations, the source of information surfaced as an important element of assessment, although the ways people used that information varied. Knowing where information came from is related to the concept of verifiability.

All our participants demonstrated assessment strategies that relied on establishing provenance, but these strategies were informed by different understandings of Wikipedia and its interface. For example, compare the responses of a student, a librarian, and a Wikipedian to the presence of a warning posted at the top of the Apollo 9 article that suggested it was lacking citations:

Student: It says it needs more citations for verification. That's important. So I'd probably want something that... Is this the only Apollo 9 article on Wikipedia?

Interviewer: Yeah, this is the official one.

Student: That's the official one? Well someone needs to work on this one, because I'm sure there's more information on it.

Librarian: This article needs additional sources for verification. My response is, of course it does! It's crowd sourced, or it's user-generated content. It always probably will.

Wikipedian: You know, the citation needed template at the top shows up too many places to actually be a good indicator of whether or not there is stuff that actually needs citations... One reason people are not good at judging the quality of Wikipedia articles is that Wikipedians are bad about tagging poor quality articles and I mean that to say that we tag too many of them.

The student took the warning at face value and it influenced his perception of the article's quality; once warned that the article did not sufficiently establish where its information came from he reflected that someone needed to work on it. He interpreted the warning as a reader and acknowledged it is a signal for attention from editors. The librarian had already committed to an assessment of Wikipedia as an information resource and, thus, the warning didn't affect her. She views Wikipedia articles as eternally in need of better sourcing. The Wikipedian uses his knowledge about the practice of tagging articles and suggests that the warning isn't reliable; he went on to offer his own assessment of the citations.

All participants mentioned citations as important features of source quality, not only in Wikipedia articles, but also in blogs or other information sources they encounter online. They most often noted the quantity of citations (the featured article, *Moon*, included over 150 citations); however, librarians and Wikipedians also noted characteristics like type, recency, publisher, or length of the resources cited.

Stewardship. It is often difficult to attribute authorship of—and thus responsibility for—any particular passage in participatory information sources to a specific author; however, the idea that many individuals share responsibility for maintaining the resource is a rationale that people used for trusting Wikipedia. Librarians and student participants understood that many people maintain Wikipedia; however, only Wikipedians used knowledge of stewardship processes to inform their assessments.

Expert Wikipedians use their knowledge of tools and practices on the site to determine how likely an article is to be well maintained; for example, by assessing whether an article is likely to have been worked on by many editors. One clear indication of this is the "featured article" status, which is signified by an unobtrusive star in the corner of the article. Wikipedian interviewees immediately noted that the moon article was featured and, when asked, explained that featured articles have been rigorously reviewed and are the best content on Wikipedia. Explained one Wikipedian, "it's very hard to get an article to featured status. I've got several

of them, and I'm proud of them, and they're on my resume.” None of the other participants noticed or could explain what the star meant without clicking on it.

Casual Wikipedia users likewise commented on the potential for many editors to create a robust article and for many eyes to catch mistakes:

Student: There's a lot of subcategories and everything, which really leads me to think that a lot of different people went and looked at this. Because if you get one guy from NASA on the Apollo 9, but you don't have a physicist, then you're lacking some information. Like different fields get different information.

Like a Wikipedian, this student uses his general understanding of how Wikipedia works to assess the article at hand; however, later he explains that he doesn't know much about Wikipedia's rules. He knows *that* many people create and maintain Wikipedia, but he has never done it and doesn't know *how* this process is managed. In a couple of cases, casual Wikipedia users referred to the process of continuous refinement by multiple editors as “peer review,” a term reserved by our librarian participants for traditional information production and stewardship processes that they typically held in contrast to that of Wikipedia.

In some cases, librarians' professional training was evident. Some librarians explained their assessments in terms of traditional information production processes that place trained professionals in the role of gatekeeper. For example, when asked about students using Wikipedia:

Librarian: The web is great for students but at the same time they have to do so much more work. They have to do the selecting and then the verifying instead of some editorial board doing it for them... an editorial board has looked at this journal but then also the librarian who knows the most about the subject has also curated or looked through as well as looking at reviews.

Another librarian stated that Wikipedia is “a bibliographic aid or portal”—not a credible information source itself, but a list of places to look for information. “One day,” she reflected, “maybe it could be considered more of a scholarly source, because it does have a lot of information in one place.” At times librarians seemed to struggle with the conflicting sense that the articles were useful and a belief that stewardship on the site is non-existent or inferior to traditionally authored reference works.

Outcome: Proposed Features of Assessable Designs

Based on these findings, we propose that assessable designs will expose features of provenance and stewardship.

Although information provenance can be documented in multiple ways, for Wikipedia and other kinds of participatory information sources discussed by participants such as blogs, citations play a central role in informing their assessments. In particular where the citation list exceeded

150 items, people who wanted to assess the *quality* of citations struggled to do so. The “wall of citations” appeared to overwhelm many participants. It may be that a standard bibliographic list of citations, appropriate for scholarly publications whose readers are experts at reading and interpreting citations, is not the ideal presentation for a lay population. Our first design exploration of provenance will be an alternate representation of what's cited in a Wikipedia article.

Our data on how people interpret stewardship on the site suggest that people who haven't participated in editing articles aren't even aware that Wikipedia has processes for managing article quality. Thus, our first design exploration of stewardship will be a representation of Wikipedia's quality review process.

STUDY 2

In a follow-up study, we used Amazon Mechanical Turk (MTurk) to test our design explorations by measuring whether they affect assessments of Wikipedia articles and how participants interpret representations of information provenance and stewardship.

Experimental Design

To develop the different interface conditions, we selected three articles from Wikipedia that were of roughly comparable quality by searching for biology-related articles rated ‘B-class’ on the Wikipedia quality scale: Dengue Virus, Domestication, and Root. We lightly edited a local copy of these articles to be of similar length and include similar numbers of citations and images. We asked three experienced Wikipedians to compare the modified articles; all three rated them of similar quality.

We developed two types of visualizations to represent features of information provenance and stewardship (see Figure 2). To create different experimental conditions, we included one at the top of the otherwise unchanged articles. In the information provenance visualization, participants were shown a pie chart that gave information about what kinds of resources were cited in the article. This is intended not only to give people a heuristic for making assessments,

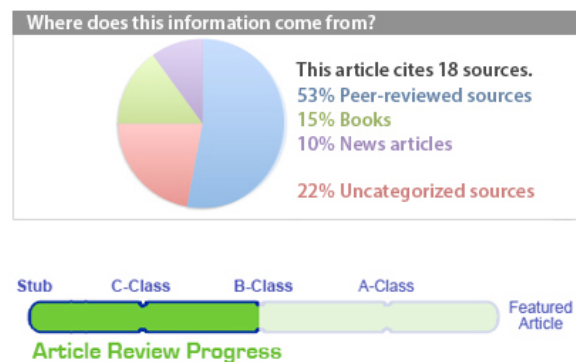


Figure 2: Top - Representation of Information Provenance
Bottom - Representation of Stewardship

but also to suggest the importance of paying attention to where the information comes from. In the information stewardship visualization, participants were shown a thermometer chart that depicts how far the article has progressed from a stub-class article toward featured article. This is intended, again, not only to alert people to the article's status, but also to communicate the fact that Wikipedia has a review process and articles may be at different levels of quality.

Two versions of each representation were created: one for a lower quality condition and one for a higher quality condition. This yielded 5 total conditions in which the visualization depicted:

- REF: 75% uncategorized sources, 15% news, 10% peer-reviewed, and 5% books.
- REFA: 85% peer-reviewed sources, 10% books, and 5% uncategorized sources.
- REVIEW: article had reached B-class.
- REVIEWA: article had reached A-class.
- PLAIN: Control group with no visualization

We followed best practices for using Mechanical Turk for user studies [19, 39]. After completing a short demographic survey, each participant assessed a plain article, then two of the four visualizations. Participants were offered \$0.50 for each article they assessed with an additional \$0.10 for completing all three. We varied the order in which visualizations were presented so that half the participants saw REF before REVIEW, and half saw REVIEW before REF. Higher and lower quality conditions were paired. We included verification questions to ensure participants attended to articles. To control the order and number of tasks available to participants, we provided a link to a custom survey form on the first author's website where participants were asked to enter their MTurk worker ID. At the end of each task, participants were given a code to enter in MTurk to receive payment. Statistical analysis of responses was done using SPSS.

Measures

Quality. To measure quality, we included five survey items that asked to what extent participants agreed that each article was written well, accurate, a bad source of information, unbiased, and trustworthy on a 4-point scale (strongly disagree, somewhat disagree, somewhat agree, strongly agree). Because we were interested in comparing differences, not in measuring quality on an absolute scale, a symmetric four-point continuum was chosen as a forced measure. One item was presented negatively to ensure that the participants were attending to the task and was inverted for data analysis. A principle component analysis was conducted to verify these variables measured one construct. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.78, above the recommended value of 0.6, and Bartlett's test of sphericity was significant ($p < .01$). One

component was extracted and the percentage of variance accounted for by that was 56.56%. The component loadings for writing, trustworthiness, accuracy, neutrality, and good information were 0.76, 0.86, 0.85, 0.67, and 0.56 respectively. Because these variables are measuring a single construct, the sum of these five variables was used as the overall quality variable for each article assessment.

Visualization Effect. This construct refers to the extent participants believed the visualizations of provenance and stewardship affected their assessment of the article. This survey item was included as a five-point scale on which participants indicated whether the visualizations [strongly lowered, somewhat lowered, did not affect, somewhat raised, strongly raised] their assessment of the article.

Global Visualization Effect. This construct refers to the extent participants believed the visualizations affected their assessment of Wikipedia as an information source. This survey item included the same five-point scale on which participants indicated whether the visualizations affected their assessment of Wikipedia as a source.

Visualization Interpretation. To support our interpretations and discussion of quality assessments, we included two prompts with free response answers: "Briefly describe what the infographic at the top of the article represents" and "Did you use the information in the infographic to answer the survey questions? If so, how? What did it mean to you?"

Participants

We conducted a pilot study with 20 participants to ensure that the plain article conditions were comparable and that the survey was understandable. With minor modifications, the final version of the study was run with 244 English-speaking participants, which yielded 682 article assessments. See Table 1 for participant demographics.

Using responses to verification questions in combination with free text responses and task completion time, we checked whether any participant had entered "junk" data and should be excluded. All participants were found to have participated in earnest.

Analysis and Findings

Before comparing the visualization conditions, we first confirmed that participants assessed PLAIN articles on root, dengue fever, and domestication to be of equal quality. A one-way ANOVA was conducted on article quality to compare each of the PLAIN conditions and demonstrated that there was no significant difference between them [$F(2, 230) = 2.718, p = .07$].

Comparing the Visualizations

Next we compared quality in the four visualization conditions. To compare each of the paired conditions, we used the Mann-Whitney U test, which is a non-parametric test that compares distributions of ranked data, making it appropriate for ordinal data. A Mann-Whitney test

		#	%
	Male	149	61%
	Female	93	38%
Education	High school	31	13%
	Some college	95	39%
	College degree	95	39%
	Graduate degree	22	9%
Reads Wikipedia	Never/Don't know	4	2%
	< Once per month	21	9%
	1-3 times per month	54	22%
	1-3 times per week	91	37%
	Most days	53	22%
	Multiple times per day	21	9%
Edits Wikipedia	Never/Don't know	185	76%
	> 10 times	54	22%
	10-300 times	3	1%
	300-5000 times	1	<1%
	Over 5000 times	0	-

Table 1: Participant Demographics

Not all sections add up to 244 because of four missing values.

confirmed that quality assessments were significantly higher for the REFA condition (Mean=17.6) than for the REF condition (Mean=16.4) [U=4209, p<.01, r=.25]. Likewise, participants rated REVIEWA (Mean=17.6) higher than REVIEW (Mean=16.1) [U=4022, p<.01, r=.3]. The *r* value (effect size) for the REVIEW visualization is larger than for the REF visualization, and therefore we can conclude that the REVIEW visualizations had a more powerful effect on quality ratings.

An analysis of Visualization Effect and Global Visualization Effect indicated that participants' perceptions of the visualizations' effects on their rating behaviors are aligned with their actual rating behaviors (See Figures 3 and 4). In the case of REF and REVIEW, nearly 35% of participants believed that the presence of the visualization lowered their assessment of the article. In the case of REFA and REVIEWA, perceptions shifted dramatically toward a belief that the visualizations raised assessments.

Comparing the Participants

We were interested in comparing Wikipedia editors and non-editors to see if editing experience affected perceptions of the visualizations. We hypothesized that experienced Wikipedians would be less affected by our visualizations since they represented features of Wikipedia that should already be familiar to frequent editors. Only four participants had significant Wikipedia editing experience, which made it impossible to carry out a meaningful statistical analysis. Although we believe editing fewer than 10 times is a weak signal of expertise, by collapsing participants who had ever edited into one group, we were able to run Mann-Whitney U tests on Visualization Effect

in each condition and compare the results to those who had and had not edited Wikipedia.

For both editors and non-editors we found significant differences in all conditions. Among people who had edited Wikipedia, the difference between Visualization Effect ratings in the REVIEWA (Median=4.5) and REVIEW conditions (Median=4) was significant. [U=232.5, p=.01, r=.34]. Similarly, there were significantly higher effects for REVIEWA (Median=5) than for REVIEW (Median=4) [U=1807, p<.01, r=.43] among non editors. There were significantly higher effects for the REFA condition (Median=5) than for the REF condition (Median=4) [U=170.5, p<.01, r=.48] among those who had edited Wikipedia and those who had not edited Wikipedia also reported significantly higher Visualization Effects in the REFA condition (Median=5) than the REF condition (Median=4) [U=1592, p<.01, r=.46].

Table 2 provides a comparison of the effect sizes for these findings. The effect size for REF conditions is similar among editors and non-editors. This is not surprising since paying attention to citations is not something unique to Wikipedia editors. The effect sizes for REVIEW

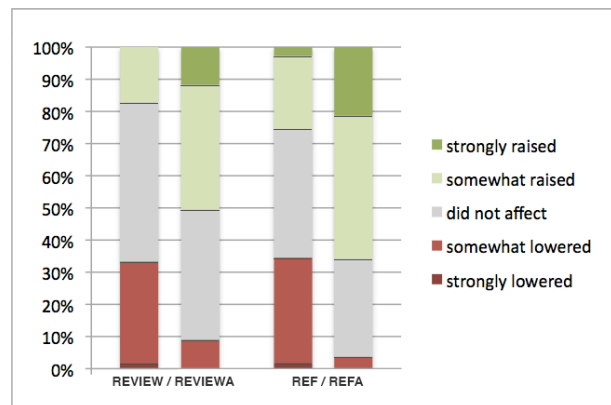


Figure 3: Visualization Effect as measured by the question "The infographic at the top of the article ___ my assessment of this article."

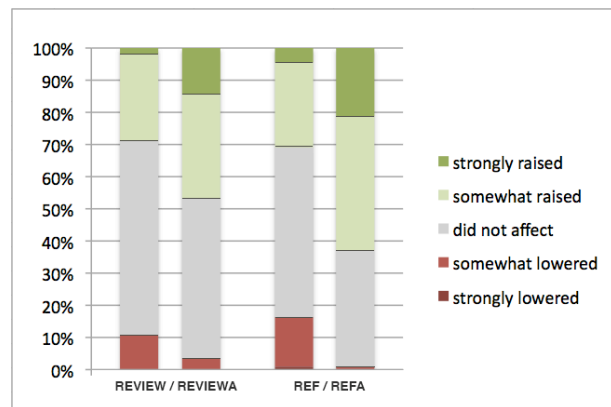


Figure 3: Global Visualization Effect as measured by the question "The infographic at the top of the article ___ my assessment of Wikipedia as an information resource."

conditions, however, are quite different. The effect size of the difference between REVIEW and REVIEWA is much higher for non-editors. This suggests that non-editors perceived information about information stewardship on Wikipedia as a stronger influence on their assessments than editors.

Effect size of difference between:	Edited	Not Edited
REF and REFA	.48	.46
REVIEW and REVIEWA	.34	.43

Table 2: Effect sizes for differences between VIZ and VIZA conditions for participants who had/had not edited Wikipedia

The visibility of formal training on librarians’ assessment strategies in Study 1 prompted us to wonder whether participants with higher levels of education would be better able to use the information in our visualizations to inform their assessments. We hypothesized that people who had spent more years in school would be better trained to take note of things like the quality of sources or indication of a review process. To investigate this, we used a Mann-Whitney U test to compare the differences between visualization effect, global visualization effect, and quality ratings of REF and REFA, and REVIEW and REVIEWA by education level. With one exception, the mean values for REFA and REVIEWA conditions were equal or higher than for REF and REVIEW conditions respectively. Table 3 highlights statistically significant results and provides the difference between mean ratings to help readers interpret the data. (Quality is a composite represented by the sum of five ranked items whereas Viz Effect and Global Viz Effect are each measured using a single ranked item.)

		Graduate Degree	College	Some College	High School
Quality	REF	4.1 p<.01	1.85 p=.01	.78 p=.04	.06 p=.96
	REVIEW	.59 p=.5	2.31 p<.01	1.16 p=.01	.79 p=.86
Viz Effect	REF	1.32 p<.01	.96 p<.01	.77 p<.01	.76 p=.02
	REVIEW	.29 p=.44	1.01 p<.01	.58 p<.01	.41 p=.33
Global Viz Effect	REF	1.23 p<.01	.61 p<.01	.52 p<.01	.6 p=.04
	REVIEW	-.19 p=.28	.41 p=.01	.42 p=.01	.44 p=.26

Table 3: Differences between mean ratings for VIZ and VIZA conditions. p values < .05 and .01 are highlighted.

We found that the highest and lowest levels of education were associated with reduced impact of visualizations, particularly in the REVIEW conditions. REF visualizations consistently had an effect across education levels, with reduced impact among high school students. We speculate

that advanced scholarship may yield effects similar to that observed in interviews with librarians: although people with advanced degrees are very familiar with and use citations to assess the articles, they are more likely to make their own judgments about the quality of Wikipedia, regardless of what the Wikipedia quality scale says.

We turned to free text Visualization Interpretation data to help us further understand these findings. These data confirmed that graduate students were able to correctly interpret the REVIEW visualizations but that they were making their own judgments independent of it. Wrote one, “This seems like a trustworthy and accurate article. The infographic suggests it is only average. That rating indicates Wikipedia has very high quality standards.” His interpretation of the visualization was flawed; however, this participant separated his own assessment of the article (trustworthy and accurate) from the one he was given on the Wikipedia quality scale.

DISCUSSION

We have completed a few first steps toward an empirically grounded design vocabulary for assessability. So far, we have established that assessable designs help people understand something about:

- information provenance, where information came from,
- and information stewardship, how it is maintained.

The concepts we’ve identified and tested resonate with literature on credibility, which dates back millennia to Aristotelian logos (logic of content), pathos (appeal to emotion), and ethos (reputation of speaker/writer). Contemporary scholars in many fields use related concepts like trust, believability and expertise in sometimes inconsistent and overlapping ways to describe subjective and objective features of credibility. We define credibility as a perceptual, subjective experience, not a characteristic of information itself; provenance and stewardship are important features of information content and production insofar as they are perceived and understood by people.

Perceptions of content characteristics (such as the presence of citations) appear frequently in the literature on web credibility. Fogg’s credibility grid [11] is a prime example of identifying online content features that can serve as credibility cues to information consumers. Stewardship surfaces less frequently. Sundar’s MAIN model of credibility assessment [38] accounts for something like stewardship in its treatment of “agency.” Provenance and stewardship are also closely related to but not precisely congruent with Flanagin and Metzger’s message credibility and media credibility [10]. They differ in that media credibility refers to perceived credibility of a communication channel rather than the process by which information is produced and maintained.

Our dimensions of assessability also complement two important bodies of work in the HCI literature that were

introduced earlier in this paper. By developing a visualization that allows readers to quickly see editing activity, WikiDashboard was developed to enhance editor accountability and “help users to identify interesting edit patterns in Wikipedia pages” [37] as an effort to improve social translucence and support assessment of “trustworthiness” [18]. This is a step toward communicating stewardship in that it helps readers perceive editing activity; however, although WikiDashboard exposes the activity itself in “interesting” ways, it does not provide guidance on interpreting these data to help readers perceive policy/quality control mechanisms familiar to Wikipedians.

More recently, Re:Flex is a set of interface components that was developed for Wikipedia atop a general system architecture (Re:Arch) that supports designing for social translucence [23]. This infrastructure has not been used to support assessability; however, McDonald et al. identify dimensions of social translucence including content, interaction, relations, and systemic states that could be used to expose important features of provenance and stewardship and develop robust assessable design prototypes.

Provenance and stewardship are not the only two concepts that will be important for designing assessable information environments, but they are a starting point grounded in user experiences. We can imagine explorations of assessable designs that help people know how the information has been used by others and in what contexts. References to reputation and social networks ties are absent in our interview data. One possible reason for this is that social information is not typically used for the tasks we studied and people neither expect it nor have developed a repertoire of strategies for using it for information assessment. Another explanation is that most people do not have ready access in their networks to the expertise necessary to vet the kind of information they need for the tasks we’ve studied.

We think it is important to note that Wikipedia is a remarkably conservative resource given its reputation as a renegade reference. Policies surrounding citation defer to well-established publishing processes like scientific peer review and traditional journalism and prohibit the production of personalized content. Studying informational sites with different characteristics, such as Twitter, Amazon, or Ancestry.com will broaden our understanding of assessability and help with next steps toward a robust design framework and guidelines.

We began this paper by suggesting that a vocabulary for assessability could change how designers think about information environments. Our inspiration for this vision is the literature on social translucence [6, 23, 37]. Erickson and Kellogg suggested that common forms of computer-mediated communication (CMC) provide little social information that can help people make choices about how to behave. The physical world is rich with social information that we use continuously, but “[i]n the digital world, we are socially blind” [6]. By analogy, Internet users

suffer from information blindness. The content we encounter online is not often accompanied by clues about how it was created or where it came from that can help inform information behaviors. Search engines and aggregators exacerbate the problem by divorcing information from its original context of production.

The system properties of visibility, awareness, and accountability articulated in the social translucence literature do not dictate the form that social norms might take; rather, they provide a basis for making choices about how to behave. Likewise, our vision of assessable design is not to dictate what information is considered relevant, credible, useful or junk. Instead, designs should make features like information provenance and stewardship visible so that people can learn to make assessments themselves. Assessable design does not mean filtering information, sifting the wheat from the chaff, the credible from the suspicious; it is a vision for design that facilitates the development of sophisticated understandings about how information is produced in participatory environments and, by extension, perhaps how one might contribute.

ACKNOWLEDGMENTS

This work is supported by NSF (CISE IIS-1253302). We also thank members of Wikimedia Foundation and community for their feedback and assistance.

REFERENCES

1. Adler, B.T., Chatterjee, K., Alfaro, L.d., Faella, M., Pye, I., and Raman, V. Assigning trust to Wikipedia content. *WikiSym 2008*. ACM Press (2008), 1-12.
2. American Association of School Librarians, *Standards for the 21st-century learner in action*. American Association of School Librarians, Chicago, 2009.
3. Chi, E., Pirolli, P., and Pitkow, J. The scent of a site: a system for analyzing and predicting information scent, usage, and usability of a Web site. *CHI* (2000), 161-168.
4. Choi, B., Alexander, K., Kraut, R.E., and Levine, J.M. Socialization tactics in wikipedia and their effects. *CSCW*, ACM (2010), 107-116.
5. Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: using intelligent task routing to help people find work in wikipedia. in *Proceedings of the 12th international conference on Intelligent user interfaces*. ACM (2007), 32-41.
6. Erickson, T. and Kellogg, W.A., Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, 7, 1, (2000), 59-83.
7. Ericsson, K. and Simon, H., *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge, 1993.
8. Farzan, R. and Kraut, R.E. Wikipedia classroom experiment: bidirectional benefits of students' engagement in online production communities. *CHI*. ACM (2013), 783-792.

9. Flanagin, A. and Metzger, M., The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media and Society*, 9, 2, (2007), 319-342.
10. Flanagin, A. and Metzger, M., Digital media and youth: unparalleled opportunity and unprecedented responsibility, in *Digital Media, Youth, and Credibility*, M. Metzger and A. Flanagin, Editors., MIT Press, Cambridge, MA, 2008, 5-28.
11. Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., and Tauber, E.R. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. in *Proceedings of the 2003 conference on Designing for user experiences*. ACM, 2003, 1-15.
12. Forte, A. and Bruckman, A., Citing, writing and participatory media: wikis as learning environments in the high school classroom. *International Journal of Learning and Media*, 1, 4, (2009), 23-44.
13. Forte, A. and Lampe, C., Defining, Understanding and Supporting Open Collaboration: Lessons from the Literature. *Am Behav Scientist*, 57, 5, (2013), 535-547.
14. Glaser, B. and Strauss, A., *The Discovery of Grounded Theory: strategies for qualitative research*. Transaction Publishers, New Brunswick, 1967.
15. Google. *Search Education*. [cited Jan 4, 2014]; <http://www.google.com/insidesearch/searcheducation>.
16. IFLA and UNESCO. *Information Literacy International Resources Dir*. <http://infolitglobal.net/directory/en/>.
17. Jenkins, H., Purushotma, R., Weigel, M., K., C., and Robinson, A.J., *Confronting the challenges of participatory culture: Media education for the 21st century*. MIT Press, Cambridge, MA, 2009.
18. Kittur, A., Chi, E., and Suh, B. Can you ever trust a wiki? Impacting perceived trustworthiness in Wikipedia. in *CSCW*, ACM (2008), 477-480.
19. Kittur, A., Chi, E.H., and Suh, B. Crowdsourcing user studies with Mechanical Turk. *CHI*, ACM, 2008, 453-456.
20. Kuiper, E., Volman, M., and Terwel, J., The Web as an Information Resource in K-12 Education: Strategies for Supporting Students in Searching and Processing Information. *Rev of Ed Research*, 75, 3, (2005), 285-328.
21. Lampe, C., Obar, J., Ozkaya, E., Zube, P., and Velasquez, A. Classroom Wikipedia participation effects on future intentions to contribute. *CSCW*, ACM, 2012, 403-406.
22. Lucassen, T. and Schraagen, J.M. Trust in wikipedia: how users trust information from an unknown source. in *Proceedings of the 4th workshop on Information credibility*. ACM (2010), 19-26.
23. McDonald, D.W., Gokhman, S., and Zachry, M. Building for social translucence: a domain analysis and prototype system. *CSCW*, ACM (2012), 637-646.
24. Menchen-Trevino, E. and Hargittai, E., Young Adults' Credibility Assessment of Wikipedia. *Information, Communication & Society*, 14, 1, (2011), 24-51.
25. Nielsen, J., *Usability Engineering*. Morgan Kaufmann, San Diego, CA, 1993.
26. Obama, B. *National Information Literacy Awareness Month, A Proclamation*. 2009; Available from: http://www.whitehouse.gov/the_press_office/Presidential-Proclamation-National-Information-Literacy-Awareness-Month.
27. Panciera, K., Priedhorsky, R., Erickson, T., and Terveen, L. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. *CHI*, ACM, 2010, 1917-1926.
28. Peppler, K. and Kafai, Y., From SuperGoo to Scratch: Exploring creative digital media production in informal learning. *Learning, Media and Tech*, 32, 2, (2007), 17.
29. Peres, S.C., Nguyen, V., Kortum, P.T., Akladios, M., Wood, S.B., and Muddimer, A. Software ergonomics: relating subjective and objective measures. *CHI*, ACM (2009), 3949-3954.
30. Petrie, H. and Kheir, O. The relationship between accessibility and usability of websites. *CHI*, ACM (2007), 397-406.
31. Pirolli, P. and Card, S., Information Foraging. *Psychological Review*, 106, 4, (1999), 32.
32. Schwarz, J. and Morris, M. Augmenting web pages and search results to support credibility assessment. *CHI*, ACM (2011), 1245-1254.
33. Seidman, I., *Interviewing as Qualitative Research: a guide for researchers in education and the social sciences*. Teachers College Press, New York, 2006.
34. Singh, V., Johri, A., and Mitra, R. Types of newcomers in an online developer community. in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 2011, 717-720.
35. Sormunen, E., Eriksson, H., and Kurkipää, T., Wikipedia and wikis as forums of information literacy instruction in schools, in *The Road to Information Literacy: Librarians as facilitators of learning*, R. Gwyer, R. Stubbings, and G. Walton, Editors., International Federation of Library Associations and Institutions, 2012, 311-330.
36. St Jean, B., Rieh, S.Y., Yang, J.Y., and Kim, Y.Ä., How content contributors assess and establish credibility on the web. *ASIS&T*, 48, 1, (2011), 1-11.
37. Suh, B., Chi, E., Kittur, A., and Pendleton, B.A. Lifting the veil: improving accountability and social transparency in Wikipedia with WikiDashboard. *CHI*, ACM (2008), 1037-1040.
38. Sundar, S.S., The MAIN model: A heuristic approach to understanding technology effects on credibility. *Digital media, youth, and credibility*, (2008), 73-100.
39. Towne, W.B., Kittur, A., Kinnaird, P., and Herbsleb, J. Your process is showing: controversy management and perceived quality in wikipedia. *CSCW*, ACM (2013), 1059-1068.
40. Wales, J. Opening Remarks. *Wikimania* (2006).
41. Yamamoto, Y. and Tanaka, K. Enhancing credibility judgment of web search results. *CHI*, ACM, 2011, 1235-1244.