2010

User Satisfaction with Electronic Reference: A Systematic REview

Elaine M. Lasda Bergman
University at Albany, State University of New York, elasdabergman@albany.edu

Irina I. Holden
University at Albany, State University of New York, iholden@albany.edu

Using the Systematic Review Methodology to Assess User Satisfaction with Electronic Reference

Abstract

Purpose: This paper systematically reviews research that analyzes satisfaction with electronic reference services, paying particular attention to how user satisfaction is measured. The application and value of evidence-based methodologies for Library and Information Science (LIS) research is explored.

Design/methodology/approach: Database searches identified research concerned with electronic reference. Articles with a variable of user satisfaction were extracted and subjected to a critical appraisal. The remaining research was analyzed for similarities, differences and consistency.

Findings: A wide variety of methods are used to measure user satisfaction. There was almost no overlap in specific questions considered although there were some similarities in methodologies used. The results of this analysis show a lack of standardization in LIS research on this topic.

Research limitations/implications (if applicable): There may be some bias in the selection of research in that the reviewers were only able to obtain published findings. The lack of consistency in reporting results further limited the articles eligible for review and precluded a meta-analysis.

Practical implications (if applicable):
By synthesizing the research conducted on this topic, practicing librarians should be able to see patterns in user satisfaction with electronic reference, and become aware of common pitfalls in undertaking user satisfaction assessment. Those conducting or planning LIS research will be able to identify the characteristics of sound research and thorough reporting of results.

Originality/value: Systematic review is an underutilized methodology in LIS research. As evidence-based librarianship gains traction, it will become a more important tool for LIS researchers. The synthesis and analysis of previous research brings together disparate findings and shows patterns and/or differences in providing these services, and brings into focus the lack of consistency in LIS research on this topic.

Introduction

Digital reference, particularly instant message, chat, and e-mail reference, is becoming an increasingly important facet of library services, and consequently, library science research. Lankes (2005) defines digital reference as "the use of human intermediation to answer questions in a digital environment (p. 321)." Digital reference has been the subject of various types of evaluation and research in recent years. This paper analyzes patterns in research articles that are written about user satisfaction with digital reference

using systematic review methodology. Systematic reviews differ from narrative literature reviews: while a narrative review provides a summary and overview of the content available on a given topic, a systematic review is more narrowly focused and seeks to "assemble, critically appraise or evaluate and synthesize the results of primary studies in an integrative approach" (McKibbon, 2006, pp. 203-204).

Systematic review is primarily used in the medical field, and medical librarians were among the first in the LIS field to adopt this methodology and apply it to library research (Ankem, 2008; Brettle, 2003; Grant, 2007; Koufogiannakis and Wiebe, 2006; Ranking, Grefsheim and Canto, 2008; Weightman and Williamson, 2005; Winning and Beverly, 2003; Zhang, Watson, and Banfield, 2007). Several authors have written about the importance of systematic reviews to librarianship (DeLuca et al., 2008; McKibbon, 2006; Saxton, 2006; Sampson et al., 2008). Incorporating the systematic review methodology into library science research has two benefits for librarians: (1) systematic reviews provide a greater understanding of our collective body of library science research and (2) librarians can sharpen their literature searching skills and better assist patrons who undertake systematic reviews in their own disciplines. One situation where systematic reviews are helpful is when there are numerous studies comparing similar variables. Reading a systematic review provides a researcher not only with time saving information on the studies in question, but also a comparison, synthesis and analysis of the findings (McKibbon, 2006, p. 205). Systematic reviews also provide valuable information for future researchers by showing the important criteria essential to successfully conducted and presented research. As Booth and Brice (2004) note, systematic reviews are a tool for evidence based librarianship (EBL) because they make "evidence" more accessible and facilitate "sharing of research across international boundaries," providing a social role in information exchange and a utilitarian role in generating research that works to improve circumstances rather than simply understand them (p. 5-6).

User satisfaction is one method of evaluating the effectiveness of library services (Cullen, 2001; Andaleeb and Simmonds, 1998; Niyonsenga and Bizimana, 1996). Receiving data from the user's point of view may provide insights for service providers and point to changes in the service to make it easier to use, more welcoming, or otherwise enhance the user's experience. Thus, user satisfaction can be a key variable in assessing and evaluating library services. As may be expected there have been a large number of studies that wholly or partially address user satisfaction in the digital reference environment, making it a topic ripe for systematic review. This systematic review will examine the quality of studies, consistency of measures, and the importance of user satisfaction with online reference (e-mail, chat, and instant message), analyze the context in which user satisfaction is evaluated and consider other measured variables in user satisfaction survey instruments.

Methodology

The methodology for performing a systematic review typically consists of six parts: formulation of a research question, a database search and retrieval of research to be considered for the review, applying pre-set inclusion and exclusion criteria for choosing

which studies to assess, quality evaluation and extraction of data, presentation of results and analysis (McKibbon 2006, pp. 210-212).

- Research Question

The questions the authors wish to address with this research are: (1) What is the level of satisfaction of patrons who utilize digital reference? and (2) What are the measures researchers use to quantify user satisfaction and how do they compare? The authors hypothesize that while digital reference has received a generally positive response from users, there is no consistent measure of satisfaction and that user satisfaction provides only a small piece of the total picture when assessing the effectiveness of any library service.

- Studies for Consideration

The reviewers searched three databases to undertake the literature search: Library, Information Science, and Technology Abstracts (LISTA), Library and Information Science Abstracts (LISA), and the Education Resources Information Center (ERIC).

The reviewers searched the database Library, Information Science and Technology Abstracts (LISTA) using a multi-field search on the EBSCO platform. These results were then limited to peer-reviewed articles, providing a final result list of 123 articles from this search. Library and Information Science Abstracts (LISA) was searched using a multi-field search on the CSA platform, A result list of 209 articles was obtained from this search. A similar search was also performed on the Education Resources Information Center (ERIC); however, no unique studies matching our criteria were retrieved. Eliminating duplicate articles from the database searches left a total of 279 results. The results obtained happened to all be journal articles, so no conference proceedings or other materials were gleaned from this search. The authors did not include books in this search because the authors felt that longer monographs might not be directly comparable to journal articles using our assessment tools and methodologies.

- Inclusion and Exclusion Criteria and Full Text Review

The reviewers then applied the inclusion and exclusion criteria identified in Table 1.[Take in Table 1]

INCLUDE:	EXCLUDE:
Articles that describe the transactions of e-reference service through surveys, analysis of transcripts and other types of studies.	Articles discussing or describing how to implement digital reference programs
Peer reviewed journals.	Articles discussing the continuation or discontinuation of such programs
Articles on 24-7 (consortial) reference	Articles providing only quantitative

	analyses, (eg: demographic analyses identifying only criteria such as gender, groups, ethnic groups)
Articles comparing e-reference to face to face reference	Reviews, editorials and commentary
Articles on academic, public and special libraries	Non-English articles
Articles on e-mail, IM, and "chat" reference	

Table 1. Inclusion and Exclusion criteria for selecting articles to be subject to review

The abstracts of the articles were reviewed under these inclusion and exclusion criteria to determine their applicability. This evaluation left a total of 93 articles. The full text of the 93 articles was obtained and read by both authors to see if they also met the criteria of having one or more variables pertaining to user satisfaction.

Of this number, 32 articles studied electronic reference transactions but did not analyze user satisfaction, 18 articles provided advice or commentary and did not utilize any type of research methodology, 5 articles were about electronic reference but not about the nature of transactions or interactions, 5 only discussed the resources used in providing electronic reference services, 3 showcased examples of electronic reference services at specific libraries or library systems, 3 provided best practices on how to do electronic reference, 1 was a literature review, 1 was a review of another study, 1 was not scholarly, and 1 was not about electronic reference at all. This left 23 articles to be included in the quality review and possibly subject to further data analysis. In reviewing the full text of these articles, a citation to the Ruppel and Fagan (2002) article was discovered, and seemed relevant. The full text of that article was then obtained and found to have sufficient content to be included in the critical appraisal step, making a total number of 24 articles to be considered for critical appraisal [Take in Table 2].

No variable on user satisfaction	32
Advice or commentary	18
Not about e-ref transactions	5
About resources used	5
Showcases library experience	3
Literature review	2
Review of another study	1
Not Scholarly	1
Not about electronic Reference	1
Selected for critical appraisal	23
Found during citation search	1
Total	94

Table 2 Articles Reviewed For User Satisfaction Variable

[Take in Fig. 1]

Quality Review

The remaining 24 articles were subjected to a critical appraisal process developed by Lindsay Glynn (2006) for evidence based librarianship (EBL). [Take in Table 3] The critical appraisal tool consisted of a rubric with four sections: population, data collection, study design and results. It seemed the appropriate choice when compared with similar tools developed earlier, such as CriSTAL, mostly used for the studies related to health and medical librarianship (Booth, 2003, Booth and Brice, 2003). Glynn developed her tool a few years later, noting "the need for a more detailed, functional EBL tool has arisen" (Glynn, 2006, p. 388). Glynn's detailed list of 22 questions and a mathematical formula for determining a precise cutoff for articles that do not meet quality standards seemed most rigorous and relevant.

The reviewers noticed that there is some overlap between certain questions in this tool and some questions were irrelevant to the overwhelming majority of the studies. However, the level of detail in the tool provided a substantive assessment of the collected research. The collected research provided varying levels of detail pertaining to the fourmain areas of review in the Glynn tool. For example, Koufogiannakis and Wiebe noticed in their systematic review (2006) that there are some "gaps in reporting." Some common gaps in each section are detailed below.

- Population

Criteria related to population often made the difference between an article's inclusion or exclusion in the final review. Important criteria include: representation of the entire possible population of users, inclusion/exclusion criteria, bias, sample size, response rate, and informed consent. In general, the articles considered for this review were lacking sufficient attention to these issues, often providing minimal detail about the selection process, or using a weak selection process. Also in many studies, users were sent a follow-up survey at the end of each transaction but it was unclear if an informed consent letter was also sent or if prior ethics approval to survey human subjects was obtained.

The response rate reported in the results often was quite low, but many authors justified it by referencing other studies in regard to response rates or citing statistics. The general consensus was that with the proliferation of web surveys the response rate dropped drastically, as reported by several studies (Chang and Holland, 2005; Lankes and Shostack, 2002). Kloss and Zhang (2003) reported that a low response rate could be attributed to the fact that the survey was pushed at the very end of the transaction when some of the users already managed to log off before the survey could be sent.

The inclusion and exclusion criteria for selecting study participants often were hard to determine, making it difficult to assess bias in the population studied.

Sample size also varied and affected the quality assessment. If the sample size was too

small, it was very difficult to include the studies into the group for consideration (Kloss and Zhang, 2003; Maness, 2008). While no fixed methodology for determining adequate sample size was adopted in this process, the authors used a "best judgment" assessment of this metric.

Articles using both obtrusive and unobtrusive methodologies were included in the research sample, which presented difficulties assessing population choice. Unobtrusive studies are considered as one of the best ways to determine the quality of reference services, hence user satisfaction. As some studies point out, the better the quality of the reference answer the more users are satisfied with the reference service in general (Nilsen, 2004, Nilsen 2006, Nilsen and Ross, 2006).

The choice of population included in the studies was often biased even when the users were sent automatic surveys upon the completion of the transaction. Many researchers commenting on surveys note that it is unknown what makes a subject answer a survey versus subjects who choose not to fill out the questionnaire. In Ruppel and Fagan (2002), it was noted that patrons whose transaction ended unexpectedly, or who could not connect to the reference service due to technical problems didn't get the surveys and therefore couldn't participate in the survey (2002).

- Data Collection

Data collection was described with clarity and attention to detail by the majority of the studies. Most of the studies measured outcomes at an appropriate time, as per the requirement of Glynn's critical appraisal tool. Most of the articles included either the entire or part of the survey. Inclusion of the survey instrument is an indication of adherence to good research practice and also allows replicability. Unfortunately, it was not present in every article. For example, Wilson and Keys (2004) and Ware, et al. (2002) did not include the survey instrument or questions that indicated what the survey looked like but rather presented the results in their own words.

It was not always clear if the survey instrument was validated, i.e., subjected to a pre-testing phase prior to administering it to users, in order to assure the instrument measures what it is intended to measure. Even articles that made it to the final review did not always validate their instruments, for example Desai and Graves (2008), Nilsen (2006), Sugimoto (2008).

Although not always validated, generally the survey questions, when included, seemed clear and understandable. Glynn (2006, p. 394) makes an emphasis on the importance of this particular component ("Are questions posed clearly enough to be able to elicit precise answers?").

Another question in the critical appraisal tool was concerned with possible bias: "Were those involved in collection of the data also involved in delivering a service to the target population?" (Glynn, 2006). In the case of unobtrusive studies the answer was negative;

the obtrusive studies were more difficult to analyze in this regard. If the authors were affiliated with the location that was being studied it was assumed that they were most likely providing services to the surveyed population.

The timeliness of the surveys was assessed positively in general. The majority of the user satisfaction surveys were sent immediately upon completion of the reference transaction. Two of the studies reported sending the surveys later: a week later after the reference transaction occurred (Stoffel and Tucker, 2004) and three days later (Lankes and Shostack, 2002). This was so patrons could have time to utilize the information in the transaction and determine its helpfulness. The reviewers felt that both of these were reasonable time frames.

In the unobtrusive studies, often there was an electronic record of the exact transaction as it occurred, rendering the time frame question less important – because the exact transaction is preserved and the researchers are not losing any information over time (Nilsen, 2006; Nilsen and Ross, 2006; Sugimoto, 2008).

The majority of the articles used results of their own research. However, a few reported using previously conducted and reported research (Lankes and Shostack, 2002; Maness, 2008).

- Study Design

The study design requirements from the appraisal tool focused on five questions: was the methodology utilized appropriate, was there face validity, would the replication be possible based on the description, did the researchers obtain ethics approval, and did the discussion clearly reflect the outcomes in relation to the data collection (Glynn, 2006).

Both reviewers agreed that all the appraised articles used an appropriate methodology. Multi-center studies, i.e. involving several libraries, like Chang and Holland, Kwon (2006, 2007), Kwon and Gregory (2007), and several studies by Nilsen (2004, 2006, Nilsen and Ross, 2006), gave researchers a better chance to reduce internal bias and increase external validity (Glynn, 2006).

Another important component was face validity. Face validity should be judged on the grounds of common sense - is the approach undertaken by the researchers logical? The reviewers agreed that face validity was present in all of the articles subjected to critical appraisal.

Ethics approval was not addressed in a majority of the articles. It is unclear whether this is due to the researchers' failing to obtain ethics approval or simply taking this step for granted and failing to report that the approval was obtained. Two studies, Desai and Graves (2008), Stoffel and Tucker (2004) report about this issue clearly and in detail. Future library and information science library and information science (LIS) researchers need to be more careful in addressing this issue.

Most of the studies provided and clearly described the outcomes in relation to data collection. This certainly appears to be a strong side of the reviewed research articles.

- Results

Glynn focused on the following aspects: the clarity of results, confounding variables, analysis reflected in conclusions, subset analysis, and their focus in the article, suggestions for future research and external validity.

The first question related to the clarity in reporting the results. Most of the articles adhered to this requirement and reported their discoveries clearly; the conclusions thus accurately reflected the analysis.

Another important issue to be appraised was that of confounding variables. Confounding variables as defined by Glynn are "other things that could have had an effect on the outcome aside from the identified intervention" (Glynn, 2006). For example, if the intervention is electronic reference service, other things that could have an effect on user satisfaction could be technological problems, librarian adherence to RUSA guidelines, delay in librarian response, lack of communication cues in an electronic format, etc. For example, Chang and Holland (2005) discuss self-selection bias as a possible implication on the measure of user satisfaction, those more likely respond to the survey could be those who were most happy with the service. On the other hand, some authors did not discuss such factors: Jerant and Firestein (2003) report about positive feedback presenting the data, but the discussion stops at this point. Similar reporting can be observed in Kipnis and Kaplan (2008), Kloss and Zhang (2003), Kwon (2006).

Accuracy of the results reflected in the discussion was addressed appropriately in majority of the articles. The reviewers noted that the articles as a rule correctly interpreted the data and results that they reported. Glynn (2006) intended that there should not be any discrepancies between the data and its analysis.

It was also important to determine if the articles discussed results in proportion to their importance in the study, or as Glynn described it, was "subset analysis a minor, rather than a major focus of the article" (Glynn, 2006). Kloss and Zhang (2003) only briefly mention patron evaluations in their conclusions and discussion, though the question that addressed user satisfaction seemed to be of the same importance as other research questions posed. Far greater detail was included about the two other research questions in their study.

Most of the studies provided suggestions for further research (Broughton, 2002; Kloss and Zhang, 2003; Kipnis and Kaplan, 2008). However, some publications concentrated on the practical applications of their studies rather than on hypothesizing about future research. Jerant and Firestein (2003) write about positive experience brought by virtual reference service and elaborate about future plans for service expansion and promotional activities, did not propose directions for future research. Wilson and Keys (2003)

demonstrated a similar approach by evaluating the success of their virtual reference pilot project and lessons learned, and also practical applications. Ware, et al. (2002) provided a detailed description of the pilot virtual reference service (VRS) at their academic institution, but the measurement of user satisfaction seemed under-reported. The authors emphasized the practical side of the issues related to the VRS. However, little or nothing was said about theoretical aspects of the issue and the authors' ideas for future research.

External validity was adequately addressed in every article. External validity refers to the applicability of the research results to a broader population than that studied (Glynn, 2006). The reviewers concluded that both the obtrusive and unobtrusive studies selected for the critical appraisal had external validity. Each of the studies described in the articles seemed to have a practical approach if applied to a larger population. [Take in Table 4]

Broughton, K.M. (2002), "Usage and user analysis of a real-time digital reference service", *Reference Librarian*, Vol. 38 No.79/80, pp.183-200.

Chang, H.R. and Holland, M.P. (2005), "User satisfaction survey of ask-a-question service at the Internet Public Library", *Internet Reference Services Quarterly*, Vol. 10 No. 2, pp.61-73.

Croft, R. and Eichenlaub, N. (2006), "E-mail reference in a distributed learning environment: best practices, user satisfaction and the reference services continuum", *Journal of Library Administration*, Vol. 45 No. 1, pp.117-147.

Desai, C.M. and Graves, S.J.(2008), "Cyberspace or fact to face: the teachable moment and changing reference mediums", *Reference and User Services Quarterly*, Vol. 47 No. 3, pp. 242-254.

Jerant, L.L. and Firestein, K. (2003), "Not virtual, but a real, live, online, interactive reference service", *Medical Reference Services Quarterly*, Vol. 22 No. 2, pp. 57-68.

Kipnis, D.G. and Kaplan, G.E. (2008), "Analysis and lessons learned instituting an instant messaging reference service at an academic health sciences library: first year", *Medical Reference Services Quarterly*, Vol. 27 No. 1, pp. 33-51.

Kloss, L. and Zhang, Y. (2003), "An evaluative case study of a real-time online reference service", *The Electronic Library*, Vol. 21 No. 6, pp. 565-575.

Kwon, N. (2006), "User satisfaction with referrals at a collaborative virtual reference service", *Information Research*, Vol. 11 No. 2, pp. 1-8.

Kwon, N. (2007), "Public library patrons' use of collaborative chat reference service: the effectiveness of question answering by question type", *Library and Information Science Research*, Vol. 29, pp.70-91.

Kwon, N. and Gregory, V.L. (2007), "The effects of librarians' behavioral performance on user satisfaction in chat reference services", *Reference and User Services Quarterly*, Vol. 47 No. 2, pp. 137-148.

Lankes, R.D. and Shostack, P. (2002), "The necessity of real-time: fact and fiction in digital reference systems", *Reference and User Services Quarterly*, Vol. 41 No. 4, pp. 350-355.

Maness, J.M. (2008), "A linguistic analysis of chat reference conversations with 18-24 year-old college students", *Journal of Academic Librarianship*, Vol. 34, No. 1, pp. 31-38.

Marsteller, M.R., and Mizzy, D. (2003), "Exploring the synchronous digital reference interaction for query types, question negotiation and patron response", *Internet Reference Services Quarterly*, Vol. 8 No. 1/2, pp. 149-165.

Nilsen, K. (2004), "The library visit study: user experiences at the virtual reference desk", *Information Research*, Vol. 9 No. 2, paper 171 [available at http:InformationR.net/ir/9-2/paper171.html].

Nilsen, K. (2006), "Comparing users' perspectives of in-person and virtual reference", *New Library World*, Vol. 107 No. 1222/1223, pp. 91-104.

Nilsen, K. and Ross, C.S. (2006), "Evaluating virtual reference from the users' perspective", *The Reference Librarian*, No. 95/96, pp. 53-79.

Ruppel, M., and Fagan, J.C. (2002), "Instant messaging reference: users' evaluation of library chat", *Reference Services Review*, Vol. 30 No. 3, pp. 183-197.

Schachaf, P., and Horowitz, S.M. (2008), "Virtual reference service evaluation: adherence to RUSA behavioral guidelines and IFLA digital reference guidelines", *Library and Information Science Research*, Vol. 30, pp. 122-137.

Smyth, J.B., and MacKenzie, J.C. (2006), "Comparing virtual reference exit survey results and transcript analysis: a model for service evaluation", *Public Services Quarterly*, Vol. 2 No. 2/3, pp. 85-105.

Stoffel, B. and Tucker, T. (2004), "E-mail and chat reference: assessing patron satisfaction", *Reference Services Review*, Vol. 32 No. 2, pp. 120-140.

Sugimoto, C.R. (2008), "Evaluating reference transactions in academic music libraries", *Music Reference Services Quarterly*, Vol. 11 No. 1, pp.1-32.

Vandecreek, L.M., (2006), "E-mail reference evaluation", *The Reference Librarian*, Vol. 45 No. 93, pp. 99-108.

Ware, S.A., Moyo, Fennewald, J., Moyo, L.M., and Probst, L.K. (2003), "Ask a Penn State librarian, live", *The Reference Librarian*, Vol. 38 No. 79, pp. 281-295.

Wilson, F. and Keys, J. (2004), "AskNow! -- evaluating an Australian collaborative chat reference service: a project manager's perspective", *Australian Academic and Research Libraries*, Vol. 35 No. 2, pp. 81-94.

Table 4 Articles Reviewed for Quality (those selected for analysis in bold type)

Data Analysis and Results

- Setting and General Characteristics

Twelve research papers about satisfaction with electronic reference met the standards described above and were selected for systematic review. All of these articles originated from the United States. Six of the articles detailed multiple studies on digital reference (Desai and Graves, 2008; Nilsen, 2004; Nilsen and Ross, 2006; Nilsen, 2006; Stoffel and Tucker, 2004; Sugimoto, 2008; Ruppel and Fagan, 2002). Three of these also compared electronic reference to face-to-face reference. In addition, Kwon published two articles with the same data, so the article that contained less data about satisfaction was dropped from the review (2007). The reviewers looked at each individual survey or instrument separately and did not include research related to traditional reference desk services, providing a total of eighteen unique surveys or instruments that were included in the review. For the purposes of clarity, each time a survey or instrument was used will heretofore be referred to as "studies" as opposed to the "articles" which are the complete research paper which may have contained one or more "studies".

The articles appeared in nine unique journals, with two articles appearing in *Reference and User Services Quarterly* (Desai and Graves, 2008; Kwon and Gregory, 2007) and two in *Reference Services Review* (Stoffel and Tucker, 2004; Ruppel and Fagan, 2002). Six articles included a study on e-mail reference (Nilsen, 2004; Nilsen, 2006; Nilsen and Ross 2006; Stoffel and Tucker (2004); Sugimoto, 2008; Vandecreek, 2006). Nine articles discussed chat reference or instant message reference. While chat reference often refers to the use of software that allows for co-browsing and other advanced features, the reviewers were aware that sometimes the use of instant messaging (IM) technology is also called “chat” so these were handled together. Schachaf and Horowitz (2008) and Vandecreek (2006) were the only articles that solely analyzed e-mail reference.

The lack of articles related solely to e-mail reference was a surprising result of this analysis, as the literature search had no date restriction and should have included e-mail reference as well as instant message or chat reference. Since e-mail reference has been around longer than chat reference, it was presumed that most of the satisfaction studies would relate to e-mail. Five articles compared e-mail reference and chat reference (Nilsen, 2004; Nilsen, 2006; Nilsen and Ross, 2006; Stoffel and Tucker, 2004; Sugimoto, 2008). The authors suspect that because e-mail reference was at first the only form of electronic reference, there was not sufficient impetus to study satisfaction with this modality. However, when chat and IM reference became an alternative electronic reference modality, libraries began to examine the most effective ways of providing reference service, comparing one modality to another.

- Methods and Timing of Data Collection

Three of the eighteen distinct studies contained in the articles utilized a pop-up survey sent to users immediately after the reference transaction was completed (Kwon and Gregory, 2007; Vandecreek, 2006; Ruppel and Fagan, 2002). Five studies used e-mails to patrons some time after the reference transaction was completed (Desai and Graves, 2008; Kwon and Gregory, 2007; Stoffel and Tucker, 2004; Vandecreek, 2006). One study suggested that this was preferable to the pop-up surveys because it would give the patron time to assess whether the information received was useful, but found that it may not be as effective with chat reference (Stoffel and Tucker, 2004, p.130). Seven distinct studies were done with paper surveys (Nilsen, 2004; Nilsen, 2006; Nilsen and Ross, 2006; Ruppel and Fagan, 2002). There were several unobtrusive studies that used transcript analysis or saved e-mails as the basis for their data collection. (Marsteller and Mizzy, 2003; Schachaf and Horowitz, 2008; Sugimoto, 2008). It was again surprising to the authors that many of the studies employed paper surveys instead of pop-ups or e-mails; however this had to do in some cases with the timing of the survey. Those employing paper surveys gave them to specific groups of students to complete after the reference transaction took place (Nilsen, 2004; Nilsen, 2006; Nilsen and Ross, 2006; Ruppel and Fagan, 2002). Thus use of a paper format may have been simply because handing the paper survey out in class was more expeditious than e-mailing each student a link.

A broad range of questions was used to capture the qualitative variable of satisfaction in the studies. As noted, there were both obtrusive (i.e., patron surveys) and unobtrusive (i.e., transcript analysis or “fake” patrons asking questions and reporting on the experience) methods for capturing satisfaction with electronic reference. There were five articles which used obtrusive methods to obtain their data (Desai and Graves, 2008; Kwon and Gregory, 2007; Stoffel and Tucker, 2004; Vandecreek, 2006; Ruppel and Fagan, 2002); while six articles used unobtrusive methods (Marsteller, 2003; Nilsen, 2004; Nilsen, 2006, Nilsen and Ross, 2006, Schachaf and Horowitz, 2008; Sugimoto, 2008).

- Comparison of Variables

There were no consistent patterns in how satisfaction was measured by the studies included in this review. In particular, there was not sufficient consistency in variables in order to undertake a meta-analysis. However a few themes did emerge that will be discussed below.

"Willingness to Return"

A theme among many articles was the concept of determining whether or not a patron would use the service again. The three Nilsen articles discuss Durrance's use of the "willingness to return" concept in her 1989 article in Library Journal. Durrance was specifically looking at proxy patrons' willingness to return to the same staff member, as opposed to repeat use of the service in general (1989, 35). This "willingness to return" variable was addressed in various forms in six articles, totaling 11 unique studies (Desai and Graves, 2008; Kwon and Gregory, 2007; Nilsen 2004; Nilsen 2006; Nilsen and Ross, 2006; Stoffel and Tucker, 2004). Desai and Graves (2008), Kwon and Gregory (2007) and Stoffel and Tucker (2004) inquired about williness to use the service again, whereas the three Nilsen articles, like Durrance, focused on return to a specific staff person. Respondents who expressed a willingness to return ranged from 55% in Nilsen and Ross (2006) to 98% in Desai and Graves (2008). The numbers in the Nilsen articles pertaining to returning to specific staff people were significantly lower, ranging from 58% for e-mail reference (2006), to 68% for chat reference (Nilsen and Ross, 2006). The reviewers suspect that this difference has to do with the specificity of the question. It is possible that patrons in the Nilsen studies who were not satisfied with the staff person with whom they interacted, may have favorably answered a question about using the service again if they thought they might encounter a more knowledgeable staff person.

"Have You Used It Before?"

Four unique studies in two articles asked if the patrons had used the service before (Desai and Graves, 2008; Stoffel and Tucker, 2004). These ranged from 30% in Desai and Graves' survey of instant message patrons, to 69% in Stoffel and Tucker's survey of e-mail patrons. As would be expected, more patrons had used e-mail reference than IM or chat reference. This is due to the fact that real time electronic communication (chat) has been employed more recently than asynchronous (e-mail) electronic communication.

Positivity of Experience

Five unique studies in three articles asked patrons to assess the positivity or negativity of their experience in various ways (Stoffel and Tucker, 2004; Vandecreek, 2006; Ruppel and Fagan 2002). As with the "willingness to return" variable, this was not measured exactly the same way in any of the studies. Ruppel and Fagan (2002) asked the question "how much did you like the service as a way of getting help?" They used a 7 point scale in their longer survey ranging from "hated it" to "loved it". This unobtrusive study had the lowest result for this question: 69% of 51 respondents rated the service a 1, 2, or 3

(those numbers closest to "loved it"). Their short survey asked if the service was "a good way to get help" and used a 3 point scale (not good, fairly good, very good) to measure this variable. They found 82% of 340 were "very good". Stoffel and Tucker asked how satisfied patrons were with the e-mail or chat reference experience and had a four point scale (very satisfied, satisfied, dissatisfied, very dissatisfied) in both their e-mail and chat surveys (2004). They collapsed their results into 2 categories: "satisfied" and "dissatisfied" and combined results. They found 98.2% of 55 responses to the e-mail were in the "satisfied" category -- the highest percentage in this category, while 85% of their 14 chat responses fell into the "satisfied" category. Vandecreek simply asked respondents to rate the service on a four point scale (excellent, good, only fair, poor) and found that out of 167 respondents, 92% rated it excellent or good (2006).

As stated above, the lowest results were the 69% rating in Ruppel and Fagan's unobtrusive long survey on IM reference. Again, one can infer that patrons may have been more comfortable with e-mail communication because e-mail is more pervasive and has been widely used for a longer period of time than IM and chat reference. Since it was not a random population chosen for this unobtrusive study, but a group of students in a library skills class, who likely could havebeen schooled in what types of qualities constitute a satisfactory reference transaction, there may be bias in the results.

Staff Quality

Four unique studies in three articles asked a question about staff quality (Kwon and Gregory, 2007; Stoffel and Tucker, 2004; Ruppel and Fagan, 2002). Staff quality in Kwon and Gregory's (2007) study was measured using a 3 point scale (excellent, good or poor) and the metric received a 68.2% "excellent" rating. Stoffel and Tucker used a four point scale (very satisfied, satisfied, dissatisfied, no response), with 76.4% selecting "very satisfied" in the e-mail reference survey and 92.8% selecting "very satisfied" for the chat reference survey (2004). Ruppel and Fagan used a seven point scale (ranging from one: fantastic to seven: terrible) and 86% of their respondents ranked the service with a one, two, or three. It is very difficult to compare these figures because of the range of the scales used and the numbers of respondents: Kwon had 417 respondents in her chat reference survey, Stoffel and Tucker had 55 for their e-mail survey and a mere 14 for the chat reference survey, and Ruppel and Fagan had 51 respondents in their long survey of IM reference. The reviewers suspect that the very high satisfaction from Stoffel and Tucker's chat reference survey is directly related to the significantly low number of respondents.

Other Questions in Obtrusive Studies

Only one patron survey asked the direct question: "Were you satisfied with the answer?" (Kwon and Gregory, 2007) (although as noted above, others asked direct or indirect questions about satisfaction with the service in general). The reviewers wonder if most researchers felt asking such an obvious and direct question would not return accurate results. Furthermore, only one patron survey asked the question: "Would you

recommend this service to a colleague?" (Stoffel and Tucker, 2004). Certainly if "willingness to return" is a measure of satisfaction, willingness to recommend the service seems to be a natural follow-up measure for gauging user satisfaction. Other questions about ease of use, helpfulness of service, whether it is a "good way to find information", and response quality all seemed to provide useful indicators for measuring satisfaction. However, they were not used in more than one article. One difference in these articles has to do with the context in which satisfaction was measured. For example, several articles measured satisfaction in an attempt to correlate it with or draw inferences about adherence to RUSA or IFLA guidelines (Desai and Graves, 2008; Kwon and Gregory, 2007; Schachaf and Horowitz, 2008). Stoffel and Tucker (2004) studied the technical aspects of using an electronic reference medium as much or more than they were studying the nature of the transaction itself. Nilsen's Library Visit Studies collected anecdotal information about the frequency with which reference interviews take place, factors affecting the success of reference transactions, and behaviors of librarians causing patron dissatisfaction (Nilsen, 2004; Nilsen, 2006; Nilsen and Ross, 2006). Very few authors studied various aspects of satisfaction as the sole goal of the research; more often, satisfaction was tied to some other aspect of providing electronic reference and trying to assess the relationship of that aspect to achieving a satisfactory electronic reference transaction.

Transcript analysis of unobtrusive studies

Three of the articles reviewed used an unobtrusive methodology based on evaluation of chat transcripts or e-mail records of reference transactions (Marsteller and Mizzy, 2003; Schachaf and Horowitz, 2008; Sugimoto, 2008). Of these, the articles by Schachaf and Horowowitz and Sugimoto used the satisfaction criteria of “completeness” of the answer to predetermined reference questions posed by “proxy” patrons. Schachaf and Horowitz used two reference questions and addressed them separately, one being 75% complete, one being 68.9% complete; whereas Sugimoto provided composite scores for each of her studies finding that 64% of instant message reference questions were “mostly incomplete” and e-mail reference questions were 78.4% “mostly complete.”

Schachaf and Horowitz (2008) and Marsteller and Mizzy (2003) on the other hand, used independent evaluators to determine the relative satisfaction provided by the transaction. Each used coders to analyze the transcript. Schachaf and Horowitz’s coders rated each transaction on a 3-2-1 scale (good-fair-poor) and Marsteller and Mizzy’s coders assessed the positivity or negativity of patron responses in chat transcripts. The first study found a satisfaction rating of 2.24 out of 3, and Marsteller and Mizzy found that only five of 200 transactions contained an obviously negative response. Using coders who did not participate in the transaction is the most unobtrusive method for determining the success of a reference transaction, but to determine the usefulness of having an arbitrary judge of satisfaction, more research is needed which compares a patron satisfaction survey with an independently coded assessment relating to the same transaction to learn how well these types of unobtrusive studies truly capture the success of a reference transaction.

Discussion

According to the research reviewed for this study, satisfaction remains high for all types of electronic reference regardless of the methodology being used to measure it. The one significant outlier brought forth in this review was Sugimoto's unobtrusive study on the completeness of IM reference transactions in music libraries (2008). In this case, because it was one reviewer who both constructed the mock reference questionsand evaluated the transactions, the possibility exists that there is bias in this result. In this case, the measurement was not a pure "user satisfaction" assessment because the user was not a true patron but rather a librarian with clear intentions and presuppositions of what information should be prescribed in the transaction. In the other unobtrusive studies that used students as proxy patrons (who, although likely instructed on how to evaluate the transaction, do not have the same level of judgment as a degree-holding librarian), a majority of responses indicated satisfaction. This corroborates the results from the obtrusive patron studies, which again might hold some bias due to the self selection of respondents.

Probably the most important result of this systematic review is understanding that the variable of user satisfaction is contingent upon the context of other variables being considered as a part of a study. While a few authors had the single goal of understanding satisfaction, others tried to correlate satisfaction to variables related to librarian behavior or to the quality of the answer provided. This analysis has led the reviewers to ponder the best way to measure user satisfaction. Does the researcher choose a randomized follow up survey when there is bias in that the respondents are self selecting? Unobtrusive studies with pools of proxy patrons may be biased if they are given instruction or coaching in how to evaluate a reference transaction. Whose satisfaction is measured when independent coders are used to evaluate reference transcripts? Further research might well focus on establishing "best practices" for measuring satisfaction.

A benefit of using systematic review methodology is achieving a greater understanding of the improvements that can be made in the quality of library and information science (LIS) research. Future studies would do well to review critical appraisal tools when developing a research project to strengthen their case. While this study used a very specific, quantitative appraisal tool (Glynn, 2006), there are other appraisal tools both in LIS and health sciences that can be utilized (Booth, 2003; Booth, 2007; Downs and Black, 1998; Eldredge, 2002; Greenhalgh and Taylor, 1997, Moher, et.al., 1999). In addition, LIS research can be improved by standardizing methodologies for user surveys, transcript analysis, and unobtrusive studies, and providing suggestions for reducing the bias inherent in each of these methodologies. There are simple things that can be done to improve the quality of published research, mainly to do with completeness of reporting. Authors should do whatever they can to assure the replicability of their methodology by others. This would include providing a copy of the instrument and clearly describing all steps taken to gather and analyze data. Full reporting of statistics should not be overlooked, such as the total response rate and the total counts of responses to a given

question, rather than simply reporting percentages. Researchers should mention that they have obtained informed consent and ethical clearance. Pretesting a survey instrument would help to assure that the questions are understood by respondents and that the instrument measures what the researchers intend it to measure. If researchers made efforts in these directions it would improve significantly the quality of LIS research.

Conclusion

Each of the articles considered in this review chose to evaluate user satisfaction differently -- sometimes the differences were slight, sometimes they were significant. This made direct comparisons of differing studies very difficult, however we have seen some general patterns showing that electronic reference is, on the whole, a satisfactory experience regardless of the questions asked or the methodology chosen to measure this variable. Many of the articles draw inferences about characteristics of successful electronic reference transactions, however these aspects were not considered in this review. It is not often that researchers have made user satisfaction the sole focus of a research study -- instead there are a variety of contexts and circumstances that are included in these studies.

The biggest concern with the quality of the research considered for this study is obtaining a representative population. It is often not clear if the factors that have an impact on population, such as recruiting participants, clearly outlining inclusion/exclusion criteria, and protecting the rights of participants, are taken into consideration and reported. Since sample population has a significant impact on data collection, it is crucial to provide a full report about it in the research article. Booth and Brice (2004) identified problems with LIS research that hamper the ability to conduct systematic reviews like this one, such as heterogeneous samples, poor research designs, and inadequate identification and retrieval of relevant research (p.112). To various extents, each of these issues has been present in this systematic review as well. Therefore the authors echo the vital need for standards that will help LIS researchers better structure their research construction and writing processes.

In writing about the practical applications of research, Booth (2004) described the following possibilities: direct application of results, conditional application of results, or derivation and enlightenment. The last of these best describes the research reviewed for this study; librarians effectively highlight new knowledge and implications of their work. However, further standardization of research protocols and reporting would benefit the LIS research community and enhance the quality of research in our field. Such standardization could facilitate more thorough systematic reviews and meta-analyses in the future. The library world needs research; librarians can benefit from learning about research quality and best practices. By incorporating methodologies which synthesize disparate but similar research, such as the systematic review, LIS researchers can carry the movement towards evidence-based practice beyond the medical field, and serve as a model for incorporating evidence-based research in other social science fields.

Bibliography

Andaleeb, S. S. and Simmonds, P. L. (1998), "Explaining user satisfaction with academic libraries: strategic implications", *College & Research Libraries*, Vol. 59 No. 2, pp.156-68.

Ankem, K. (2008), "Evaluation of method in systematic reviews and meta-analyses published in LIS", *Library and Information Research,* Vol. 32 No. 101, pp. 91-104.

Booth, A. (2003), "CriSTAL online", available at: http://www.shef.ac.uk/scharr/eblib/ (accessed 5 February, 2010).

Booth, A. and Brice, A. (2003), "Evidence based librarianship: the first steps", available at http://www.shef.ac.uk/scharr/eblib/conf2003.htm (accessed February 10, 2010).

Booth, A. and Brice, A. (2004), "Why evidence-based information practice?", in Booth, A. and Brice, A. (Eds.) *Evidence-Based Practice for Information Professionals: a Handbook,* Facet Pub., London, pp. 1-13.

Booth, A. (2007), "Who will appraise the appraisers?—the paper, the instrument and the user", *Health Information and Libraries Journal*, Vol. 24 No. 1, pp. 72-76.

Brettle, A. (2003), "Information skills training: a systematic review of the literature", *Health Information and Libraries Journal*, Vol. 20 Suppl. 1, pp. 3-9.

Brice, A. and Hill, A. (2004), "A brief history of evidence-based practice", in Booth, A. and Brice, A. (Eds.) *Evidence-Based Practice for Information Professional : a Handbook,* Facet Pub., London, pp. 13-23.

Broughton, K.M. (2002), "Usage and user analysis of a real-time digital reference service", *Reference Librarian,* Vol. 38 No.79/80, pp.183-200.

Chang, H.R. and Holland, M.P. (2005), "User satisfaction survey of ask-a-question service at the Internet Public Library", *Internet Reference Services Quarterly*, Vol. 10 No. 2, pp.61-73.

Croft, R. and Eichenlaub, N. (2006), "E-mail reference in a distributed learning environment: best practices, user satisfaction and the reference services continuum", *Journal of Library Administration*, Vol. 45 No. 1, pp.117-147.

Cullen, R. (2001), "Perspectives on user satisfaction surveys", *Library Trends,* Vol. 49 No. 4, pp. 665-87.

DeLuca, J.B., Mullins, M.M., Lyles, C. M., Crepaz, N., Kay, L. and Thadiparthi, S. (2008), Developing a comprehensive search strategy for evidence based systematic reviews", *Evidence Based Library and Information Practice,* Vol. 3 No. 1, pp. 3-32.

Desai, C.M. and Graves, S.J. (2008), "Cyberspace or face to face: the teachable moment and changing reference mediums", *Reference and User Services Quarterly*, Vol. 47 No. 3, pp. 242-254.

Downs, S.H. and Black, N. (1998), "The feasibility of creating a checklist for the assessment of the methodological quality both of randomized and non-randomized studies of health care interventions", *Journal of Epidemiology Community Health*, Vol. 52, pp. 377-384.

Eldredge, J.D. (2002), "Evidence-based librarianship: what might we expect in the years ahead", *Health Information and Libraries Journal*, Vol. 19 No. 2, pp. 71-8.

Glynn, L. (2006), "A critical appraisal tool for library and information research", *Library Hi Tech*, Vol. 24 No. 3, pp. 387-99.

Grant, M. (2007), "The role of reflection in the library and information sector: a systematic review", *Health Information and Libraries Journal*, Vol. 24 No. 3, pp. 155-66.

Greenhalgh, T. and Taylor, R. (1997), "Papers that go beyond numbers (qualitative research)", *BMJ*, Vol. 315 No. 711, pp. 740-3.

Jerant, L.L. and Firestein, K. (2003), "Not virtual, but a real, live, online, interactive reference service", *Medical Reference Services Quarterly*, Vol. 22 No. 2, pp. 57-68.

Kipnis, D.G. and Kaplan, G.E. (2008), "Analysis and lessons learned instituting an instant messaging reference service at an academic health sciences library: first year", *Medical Reference Services Quarterly*, Vol. 27 No. 1, pp. 33-51.

Kloss, L. and Zhang, Y. (2003), "An evaluative case study of a real-time online reference service", *The Electronic Library*, Vol. 21 No. 6, pp. 565-575.

Koufogiannakis, D. and Wiebe, N. (2006), "Effective methods for teaching information literacy skills to undergraduate students: a systematic review and meta-analysis", *Evidence Based Library and Information Practice*, Vol. 1 No. 3, pp. 3-43.

Kwon, N. (2006), "User satisfaction with referrals at a collaborative virtual reference service", *Information Research*, Vol. 11 No. 2, pp. 1-8.

Kwon, N. (2007), "Public library patrons' use of collaborative chat reference service: the effectiveness of question answering by question type", *Library and Information Science Research*, Vol. 29, pp.70-91.

Kwon, N. and Gregory, V.L. (2007), "The effects of librarians' behavioral performance on user satisfaction in chat reference services", *Reference and User Services Quarterly*, Vol. 47 No. 2, pp. 137-148.

Lankes, R.D. (2005), “Digital reference research: fusing research and practice”, *Reference & User Services Quarterly*, Vol. 44 No. 4, pp. 320-26.

Lankes, R.D. and Shostack, P. (2002), "The necessity of real-time: fact and fiction in digital reference systems", *Reference and User Services Quarterly*, Vol. 41 No. 4, pp. 350-355.

Maness, J.M. (2008), "A linguistic analysis of chat reference conversations with 18-24 year-old college students", *Journal of Academic Librarianship*, Vol. 34, No. 1, pp. 31-38.

Marsteller, M.R., and Mizzy, D. (2003), "Exploring the synchronous digital reference interaction for query types, question negotiation and patron response", *Internet Reference Services Quarterly*, Vol. 8 No. 1/2, pp. 149-165.

McKibbon, K.A. (2006), “Systematic reviews and librarians”, *Library Trends*, Vol. 55 No. 1, pp. 202-15.

Moher, D., Cook, D. J., Eastwood, S., Olkin, I., Rennie, D. and Stroup, D. F. (1999), “Improving the quality of reports of meta-analyses of randomized controlled trials: the QUORUM statement”, *Lancet*, Vol. 354 No. 9193, pp. 1896–1900.

Nilsen, K. (2004), "The library visit study: user experiences at the virtual reference desk", *Information Research*, Vol. 9 No. 2, paper 171 [available at http:InformationR.net/ir/9-2/paper171.html].

Nilsen, K. (2006), "Comparing users' perspectives of in-person and virtual reference", *New Library World*, Vol. 107 No. 1222/1223, pp. 91-104.

Nilsen, K. and Ross, C.S. (2006), "Evaluating virtual reference from the users' perspective", *The Reference Librarian*, No. 95/96, pp. 53-79.

Niyonsenga, T. and Bizimana, B. (1996), "Measures of library use and user satisfaction with academic library services", *Library & Information Science Research*, Vol. 18, pp. 225 - 240.

Rankin, J.A., Grefsheim, S.F. and Canto, C.C. (2008), “The emerging informationist specialty: a systematic review of the literature”, *Journal of the Medical Library Association*, Vol. 96 No. 3, pp. 194-206.

Ruppel, M., and Fagan, J.C. (2002), "Instant messaging reference: users' evaluation of library chat", *Reference Services Review*, Vol. 30 No. 3, pp. 183-197.

Sampson, M., Shojania, K.G., Carrity, C., Horsley, T., Ocampo, M. and Moher, D. (2008), “Systematic reviews can be produced and published faster”, *Journal of Clinical Epidemiology*, Vol. 61 No. 6, pp. 531-536

Saxton, M.L. (1997), “Reference service evaluation and meta-analysis: findings and methodological issues”, *Library Quarterly*, Vol. 67 No. 3, pp. 267-89.

Schachaf, P., and Horowitz, S.M. (2008), "Virtual reference service evaluation: adherence to RUSA behavioral guidelines and IFLA digital reference guidelines", *Library and Information Science Research*, Vol. 30, pp. 122-137.

Smyth, J.B., and MacKenzie, J.C. (2006), "Comparing virtual reference exit survey results and transcript analysis: a model for service evaluation", *Public Services Quarterly*, Vol. 2 No. 2/3, pp. 85-105.

Stoffel, B. and Tucker, T. (2004), "E-mail and chat reference: assessing patron satisfaction", *Reference Services Review*, Vol. 32 No. 2, pp. 120-140.

Sugimoto, C.R. (2008), "Evaluating reference transactions in academic music libraries", *Music Reference Services Quarterly*, Vol. 11 No. 1, pp.1-32.

Vandecreek, L.M., (2006), "E-mail reference evaluation", *The Reference Librarian*, Vol. 45 No. 93, pp. 99-108.

Ware, S.A., Moyo, Fennewald, J., Moyo, L.M., and Probst, L.K. (2003), "Ask a Penn State librarian, live", *The Reference Librarian*, Vol. 38 No. 79, pp. 281-295.

Weightman, A.L. and Williamson, J. (2005), “The value and impact of information provided through library services for patient care: a systematic review”, *Health Information and Libraries Journal*, Vol. 22 No. 1, pp.4-25.

Wilson, F. and Keys, J. (2004), "AskNow! -- evaluating an Australian collaborative chat reference service: a project manager's perspective", *Australian Academic and Research Libraries*, Vol. 35 No. 2, pp. 81-94.

Winning, M.A. and Beverley, C.A. (2003), “Clinical librarianship: a systematic review of the literature”, *Health Information and Libraries Journal*, Vol. 20 Suppl. 1, pp. 10-21.

Zhang, L., Watson, E.M. and Banfield, L. (2007), “The efficacy of computer-assisted instruction versus face-to-face instruction in academic libraries: a systematic review”, *Journal of Academic Librarianship*, Vol. 33 No. 4, pp. 478-84.