

Assessing Reference Service Quality: A Chat Transcript Analysis

*Greta Valentine and Brian D. Moss**

Introduction

In 2016, members of the University of Kansas Libraries' Reference Services unit participated in a peer review of 60 chat transcripts from the Fall 2015 and 2016 semesters. This project grew out of the need to begin assessing the level of reference service quality provided at the University of Kansas (KU) Libraries.

In-person and virtual reference assistance is provided at the two main libraries up to 12 hours per day, six days per week, when classes are in session. Two main libraries plus four smaller branches make up the KU Libraries system. Staffing of reference services at the KU Libraries has evolved over time, from a mix of faculty-equivalent librarians and library staff to the current staffing model, which employs three full-time staff members and two half-time staff members, complemented by graduate student employees. While the current desk staff are all well qualified, very few have degrees in library science. Therefore, extra care must be taken to train reference staff on how to conduct a proper reference interview, and additional time is required to monitor reference stats to ensure that questions are being answered properly.

Previous local studies have been conducted in this vein. In 2007, librarians rated 2,300 chat transcripts from a two-year period to determine whether teaching moments were being taken advantage of, using ACRL's Information Literacy Standards for Higher Education as criteria.¹ A follow-up study took place in 2008, examining 50 of the previous 2,300 transcripts in detail to identify opportunities and best practices for incorporating instruction in virtual reference services.²

Since 2008, KU Libraries have not conducted additional formal studies to examine the quality of reference services provided. During the Fall 2015 semester, chat questions made up approximately 43 percent of overall reference questions asked at KU Libraries reference desks. This increased to 47 percent during the Fall 2016 semester. As chat becomes an increasingly common venue for reference questions at KU, this transcript analysis is a first step toward measuring the quality of reference service quality. Chat transcripts are attractive as a starting point since they provide a comprehensive snapshot of an interaction with a patron, and can be examined in a variety of ways to inform best practices in virtual reference. This paper describes the study's findings, and presents the project methodology alongside referenced literature as a resource for other librarians who may be looking for a starting point in establishing broader evaluation of reference services in general.

Literature Review

Virtual reference services (VRS) encapsulate library services which employ "human mediation to answer questions in a digital context."³ Beginning with e-mail and evolving to embedded web forms and synchronous chat, these services have been used for over 20 years as of 2017, and are no longer emerging but ubiquitous.⁴

* *Greta Valentine is Data & Research Analyst at University of Kansas, greta.valentine@ku.edu; Brian D. Moss is Reference Services Manager at University of Kansas, bdmoss@ku.edu.*

Evaluation of VRS has caught on more slowly. As late as 2011, assessment of VRS was in its early stages.⁵ One of the first comprehensive reviews of VRS effectiveness published at that time primarily referenced evaluations published between 2003 and 2008.⁶ Many of these studies used chat transcripts as the basis for some type of evaluation, since a transcript provides a comprehensive record of a virtual interaction with a patron.⁷ Many studies at this time focused primarily on the objectives, cost, and feasibility of implementing a chat service while quality control of chat interactions made up a smaller portion of the literature.

After 2008, a number of case studies used established guidelines as criteria for determining VRS quality. The American Library Association's Reference and User Services Association (RUSA) behavioral guidelines were most frequently cited, though some articles referred to guidelines established by the Internet Public Library,⁸ as well as locally developed lists of competencies.^{9,10} These guidelines were employed primarily as lists of behaviors that could be looked for in a chat transcript to determine the quality of the interaction.¹¹⁻¹⁵

Much of the literature referenced for this study indicated that VRS interactions should include components found in the reference interview, drawing in large part from established guidelines such as RUSA and IFLA.¹⁶ Likewise, evaluation of VRS sometimes centers on the assertion that the reference interview is central to any reference transaction, and critical to its success.¹⁷ For this reason, some studies examine transcripts for evidence that a reference interview, or elements of an interview occurred.¹⁸ However, chat presents some issues not present during the in-person reference interview. Not only must librarians know how to facilitate an effective reference interview, but they must also be competent in online communications skills, web search strategies, database usage, and knowledge of chat service policies and procedures.¹⁹

Most of the literature acknowledged additional issues that may arise in a VRS transaction (such as the lack of nonverbal cues²⁰ and technical issues like sudden disconnections or difficulties co-browsing²¹). However, few studies design their evaluations with these differences in mind. Recent articles indicate that the immediacy of patrons' needs when using VRS, as well as the fast pace of conversation pose new areas that are not accounted for in current behavioral guidelines for providing reference services in a virtual environment.²² This immediacy may necessitate sacrificing some established recommended behaviors in order to provide quick answers. Current guidelines do not always take this kind of patron behavior into account. Consequently, evaluating librarians using these guidelines as a rubric may lead to inflated negative ratings on categories such as instruction.

Methodology

The KU Libraries Reference unit began this project with the aim of gauging current service levels and moving towards more holistic assessment, rather than contributing a new case study to the literature. However, there are several unique aspects of this study which distinguish it from other case studies. First, this study was participatory, allowing current members of the reference team to rate themselves so the process did not seem punitive. Second, rather than noting the presence or absence of a suggested practice in the transcript, this study allowed participants to rate each transaction on a list of five characteristics derived from RUSA guidelines for a deeper picture of overall quality—Correctness, Completeness, Grammar, Courtesy and Instruction. Both practices were suggested by Logan and Lewis's 2011 study.²³

Chat transcripts functioned as the unit of analysis, since they present a comprehensive record of an interaction with a patron.²⁴ Additionally, this method is less obtrusive than a "mystery shopping" approach, which can "make service providers uncomfortable and increase tension among library employees, which is almost certain to translate negatively to the user experience."²⁵ Those rating the transcripts did so on a four-point scale, with a rating of one being a very poor response and a four being exemplary. The four-point scale was used to force a response that falls toward either a positive or negative end of the spectrum; fewer points on

the scale were also intended to make it relatively simple for participants to decide where a given response falls on the scale.

Literature examining VRS quality distinguishes between methodologies which assess from the user perspective and those which examine transcripts from a service perspective.²⁶ This study was designed from the service perspective, in that it examined the quality of the service provided. The authors wanted to focus on the content of answers themselves not only because an accurate answer is central to reference quality,²⁷ but also because KU Libraries patrons may well be satisfied with an answer that is incomplete or contains incorrect information. The authors wanted to take first steps towards measuring response quality by looking at what the librarians themselves could do to provide quality service—an approach that comprises a smaller portion of the literature. The authors borrowed a question from Pomerantz, Luo and McClure that allowed participants to maintain a mindset of improvement with each transcript they evaluated: “What could the librarian have done to improve this answer?”²⁸

The study was divided into two rounds. In the first round, participants examined 30 transcripts from the Fall 2015 semester, and the second round, an additional 30 transcripts from Fall 2016. Participants included both staff of the current Reference Services unit—six in the first round and ten in the second—as well as three former reference librarians with experience on the desk providing a more objective perspective. During the first round, transcripts were gathered from LibraryH3lp—KU Libraries’ chat client at the time—and anonymized. By Fall 2016, the reference unit had begun using Springshare’s LibAnswers for chat inquiries. LibAnswers provides additional metadata categories not available in LibraryH3lp, such as the Reference Effort Assessment Data (READ) Scale value assigned to the transcript by the librarian who answered the question. This scale assigns the values one through six to reference questions, indicating the effort, knowledge and skills required to adequately address the question.²⁹ During the first round of analysis, participants often disagreed about whether a given question required instruction or not, resulting in too few ratings to meaningfully examine. The authors wanted to find a balance between obtaining a sample that represented the range of questions that typically comes in via chat, but where instruction could also be evaluated. Consequently, the Fall 2016 sample was limited to questions rated a three or higher on the READ scale.

Borrowing from Logan and Lewis’s methodology, the first round of the study combined Correctness and Completeness as a single category. However, after discussing the results of the first round and consulting additional literature,³⁰ the authors determined that these attributes are distinct and should be analyzed separately. The second round of the study included separate categories for Correctness and Completeness.

There are a number of potential limitations to this study that should be noted. The small sample size, for example, was intended to provide participants with a manageable number of transcripts that could be rated in depth. However, this may limit the strength and statistical significance of statements about the study’s findings. The original intent was to compare patterns between current and former reference providers, but due to the small number in each group, the ratings were combined to assess overall patterns of strengths and weaknesses in the answers provided to KU patrons via VRS. Additionally, the literature notes a tendency for participants in similar rating-type exercises to give higher ratings because they know other participants, even if an answer is incomplete or incorrect.³¹ It is possible the effect may have been amplified because participants knew they were rating their own answers as well.

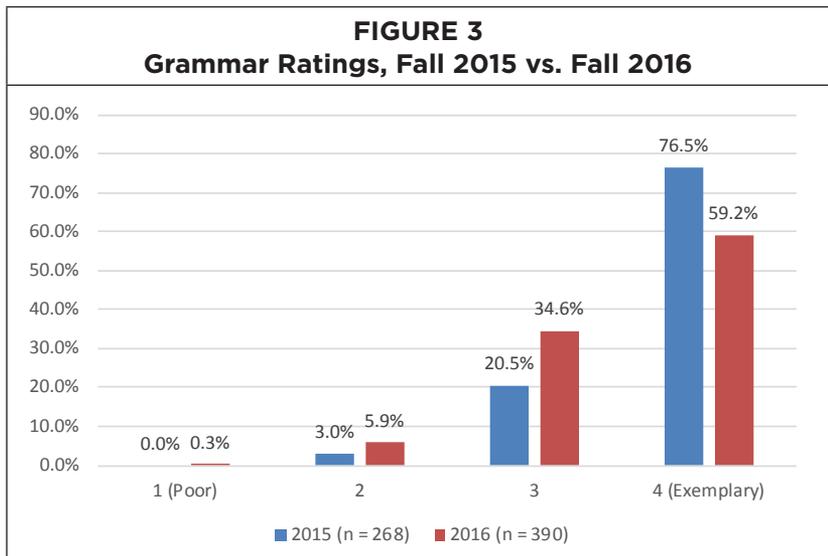
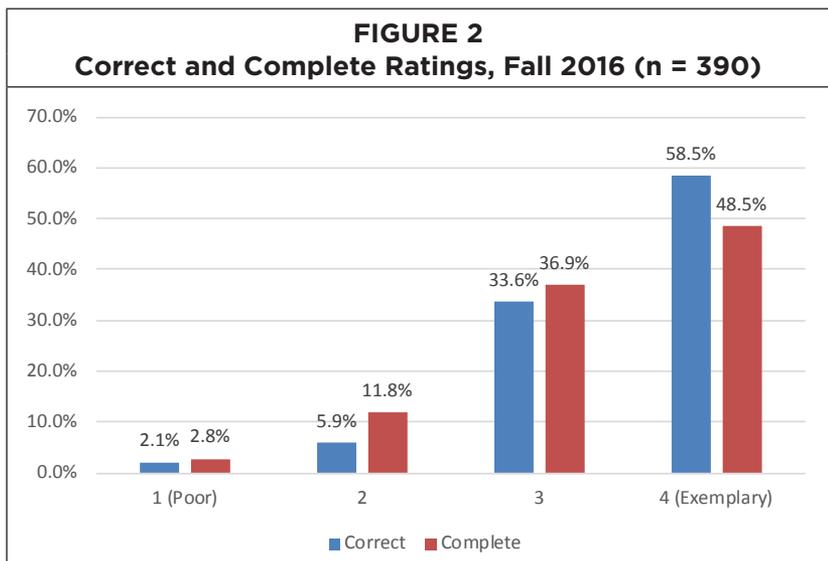
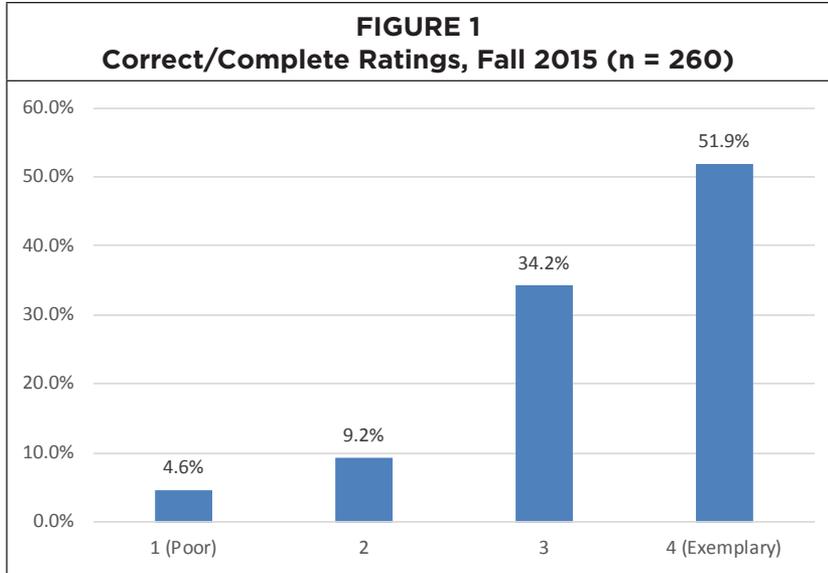
Findings

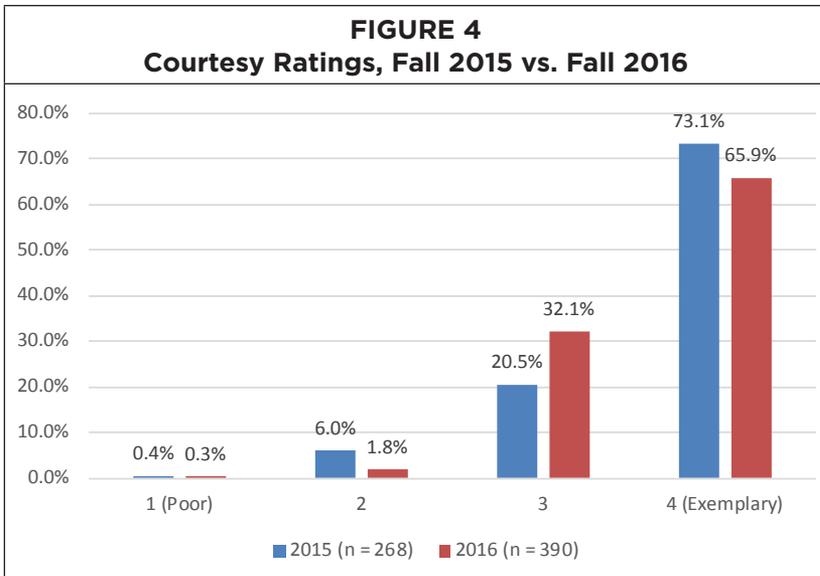
At the outset, this study aimed to examine transcripts for quality in the areas of Correctness, Completeness, Courtesy, Grammar, and Instruction.

The categories of Correctness and Completeness were among the most complex to diagnose, and received lower scores than did Grammar and Courtesy. Figure 1 shows the Fall 2015 sample, in which participants graded transcripts on a combined category of Correctness/Completeness.

When the scores were split into two separate categories in the Fall 2016 sample (Figure 2), participants rated all transcripts as being slightly more correct than complete. It is not surprising that the category of Completeness would be rated lower when the analysis question during Fall 2016 was framed around improving the answer. If participants were thinking about additional resources or instructions the librarian could have provided, they may have been more apt to rate the question as less complete. Many of the comments during the Fall 2016 round focused on specific resources the librarian could have provided such as links to articles, and directions on locating or using the Interlibrary Loan request page on the Libraries' website.

Grammar (Figure 3) was one of the highest-rated categories in the study. This is likely because it is a less nuanced category than some of the others, and a simple scan of the transcript can indicate whether the overall quality of grammar is appropriate or not. This category prompted a discussion among participants on whether grammar is an important focus in chat interactions, since this is not a consideration in face-to-face inquiries. Some wondered whether emulating patrons' "chat speak" including emoticons might be a way of building





rapport. Best practices resulting from these discussions encouraged reference providers to make the patron feel comfortable and to build a casual rapport while maintaining a professional use of grammar and punctuation.

Courtesy (Figure 4) was the highest-rated category in the study, with less than three percent of transcripts overall falling into an unacceptable range (rated a one or a two on the four-point scale). A small number of notably discourteous transcripts prompted a critical discussion of this as a result of the study. Common indicators of a discourteous response in participants’

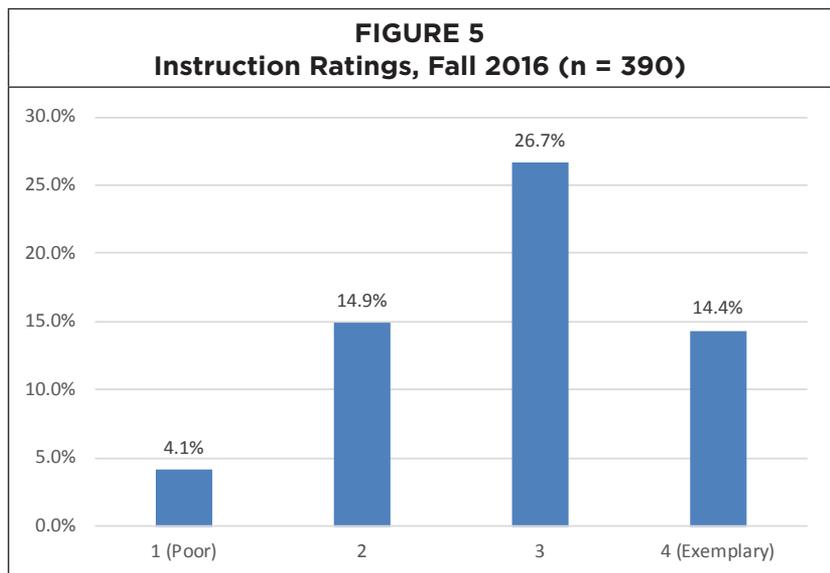
comments included lack of a greeting, short remarks that could be interpreted as curt in an online environment, and cutting the patron off before they had finished asking a question.

Participants suggested these alternative behaviors to better demonstrate courtesy in chat interactions:

- Break up lengthy responses, so as not to overwhelm the patron
- Inform the patron when another activity was keeping librarians from responding (such as searching for resources or helping another patron)
- List consulted resources while answering a patron’s question, so the patron is aware how thoroughly the librarian has searched.
- Ask the patron whether they need help with anything else, especially at the end of a transaction, to avoid appearing aloof or eager to send patrons on their way

Instruction was the most difficult category to rate in this study (see Figure 5). Largely because discretion of what qualified as “good” instruction versus “poor” instruction was left entirely up to the participants rating the Fall 2015 sample, there was much disagreement over which questions merited instruction to begin with. Consequently, each transcript received such a low number of ratings for this category that definitive reporting was impossible. The ratings below are from Fall 2016, which included only questions rated a three or higher on the READ scale, since this is the rating at and above which librarians may be required to provide instruction.³²

Two notable issues arose during the study regarding instruction via VRS. One was related to the Complete-



ness category, for which librarians sometimes missed an opportunity to instruct a patron. This was consistent with KU's 2008 study, which indicated "Librarians often asked students if they were finding the information they needed, but were not informing them about the quality of the information or helping them question the information's validity."³³ While that study focused on opportunities for information literacy instruction via chat, instruction was notably absent from some of the transcripts despite an opportunity to provide it.

The second issue relates to the first, and deals with reasons a patron may or may not be open to receiving instruction via chat. Participants commented that in some cases that instruction was unnecessary either because the patron needed their question answered immediately, or because chat was not a suitable medium for addressing the question in depth. In 15 percent of transcripts from both rounds of the study, the librarian suggested that the patron may benefit from coming to the reference desk in person. In one case, the participant commented, "The patron needs in-person instruction. Chat isn't like fast food!!" This theme is worth exploring: when is instruction appropriate, and how persistent should a librarian be in providing it? The 2008 study conducted at KU supports research indicating that patrons respond best to librarians who are persistent and approachable, especially in terms of instruction.³⁴

However, patrons often that the immediacy of their need prevented them from coming to the library in person, or from spending a long time working through a complex question via chat. This aligns with findings from van Duinkerken, Stephens, and MacDonald indicating that needs of patrons who choose VRS as a means to obtain information from the library may differ from those who approach the desk in person, and that current virtual reference guidelines do not account for these differences in behavior and expectation.³⁵ Some studies indicated that patrons want instruction in most cases, and stay engaged as long as librarians are providing it.³⁶ However, this study highlighted the reality that while an in-person consultation may be the best means of addressing a patron's need, patrons who use VRS often want quick answers, and are unwilling to come to the library in person or wait through a lengthy exchange where their question can be addressed in depth. While very few patrons in the transcripts for this study indicated an urgent need for information, literature referenced indicated that in a similar study, as many as 77 percent of patrons indicated they needed the requested information quickly.³⁷ Despite the relatively few patrons who had an urgent request in this study, the authors suspect that KU patrons' attitudes may mirror those reported by Luo, who indicated that not only do patrons prefer immediate answers over detailed instruction, but that some patrons were less satisfied with virtual reference when referred to another resource without an immediate answer.³⁸

Participant Comments: Looking towards Improvement

Several themes stood out in participants' comments about how various transactions in the study might be improved. For the Fall 2015 sample, participants commented generally about how they evaluated a given transcript. For transcripts in the Fall 2016 sample, and based on additional survey of the literature, the authors posed a more specific question, "How could this answer be improved?" Therefore, comments in the Fall 2016 sample focused more on changing or improving the librarians' behavior than on justifying a given rating.

The main themes from participants' comments involved issues of completeness and instruction. These comments suggested that the librarian could have improved the patron's experience by helping them understand the process for retrieving what was requested for future reference. Some participants interpreted incomplete responses as missed opportunities for instruction:

[D]idn't really give any instruction on how to do this next time—just typing in the article title isn't always going to give you what you need—lost opportunity to talk about citations, etc.

Normally preferable to walk the patron through the ILL process, or at least stay online with them to make sure they don't have any problems.

Could have offered to walk patron through interlibrary loan request.

One participant indicated that the librarian should communicate their process not only as a means of instruction, but as an indication that the patron was not being ignored:

The librarian could have shown the patron how they determined that a specialist was necessary; from the beginning, it appears that the librarian is already admitting defeat. Where did the librarian look? Where had the patron looked already?

Librarians also sometimes sacrificed a complete answer for a timely one, again reflecting the complexity of providing in-depth assistance via chat. Echoing the literature's claim that patrons will not always stick around for a lengthy exchange, participants noted reasons why instruction may not have been feasible in the responses they reviewed:

[T]he instruction was minimal, but this would have been difficult to explain [even] in person.

[T]his is a complex question that would be hard, and very long, to give instruction on, and the patron probably wouldn't pay that much attention—for most people, they just want the article and don't care how the library staff member found it. If the person using chat asked about how it was found, the person answering could have suggested that they set up a consultation with a librarian and told the person how to do that.

[T]he reference person did a pretty good job of trying to tease out what the student needed to think about to narrow down the question, but that's hard to do in chat. Would the reference person have had the option to suggest that the student call so they could have tried to do the reference interview portion in an easier and quicker way?

In some cases, the transcript itself indicated that librarians had suggested that chat was not an appropriate way to solve the patron's need. In 15 percent of transcripts overall, librarians suggested the patron come to the library in person, or explained that resources would be more readily available in person. This further necessitates considering ways to evaluate in-person reference, as well as taking into account the limitations of chat when evaluating reference provided there.

Lessons Learned: Assembling Your Toolkit

Participation

Conducting this analysis benefited the current Reference Service unit in two main ways: participation, and applying study results. Participation in the design and execution of the project allowed the unit to think about, discuss, and critique their own collective work, which encouraged a quality-centric mindset towards their current work at the desk. It is worth mentioning that this study builds on exercises the current Reference Services staff already conducts. Members gather a few transactions from LibAnswers to discuss at bi-weekly meetings or via e-mail, suggesting alternate ways to approach and answer the question.

Overall, this study yielded mostly positive outcomes, not only in terms of the transcript ratings themselves, but also in terms of the buy-in from reference staff. Initially there was some concern that staff would resist the study, but for the most part they came to recognize it as an opportunity to learn how to provide better service to library patrons. It may have helped that the reasoning behind the study was introduced early on as more of a training opportunity, rather than an intent to be punitive in any way.

As a result of this study, current desk staff have already proven to be much more likely to consider the quality of service they provide via chat, including the appropriate level of instruction required for each interaction. Therefore, the Reference Services unit plans to incorporate similar ongoing chat evaluations into regular departmental practice on an annual or semi-annual basis. As the first two rounds have demonstrated, having reference staff evaluate transcripts leads to valuable discussions about how best to provide virtual reference assistance to patrons, and causes staff to think about quality while answering future chat questions.

Application

In addition to buy-in from Reference Services staff, the added benefit of conducting this study is the ability to pinpoint specific themes in participants' comments and ratings to consider for training or discussion. For example, instruction bears closer scrutiny based on the complexity raised by this study, and the realization that guidelines do not always mesh with reality. When is it ok to give "fast-food" answers? When should the librarian persist in offering instruction? Finding ways to evaluate and discuss these issues can be the focus of future evaluations.

Anyone planning a similar study should make every effort to be transparent and provide frequent updates to desk staff throughout the process. Some degree of pushback will be inevitable, and some staff may simply fail to improve their service quality. However, having transcripts with relatively high or low ratings provided the group with examples around which to frame discussions and training activities. Several participants even mentioned that they recognized LibAnswers as a source of chat transcripts that could function as a training aid as a result of this study.

The timing of rating the transcripts played a potentially bigger role than initially expected. For example, evaluations of the second study took place at the very end of the Fall 2016 semester, with results presented to staff at the beginning of the Spring 2017 semester. Future assessments may have more impact if transcripts are evaluated early in a given semester, thus causing desk staff to consider service quality while answering chats going forward; results can then be presented mid-semester as a refresher. However, regardless of the timing, the analyses always provided an excellent topic of discussion during regular departmental meetings.

This study provided the KU Libraries' Reference Services unit with a basis on which to build ongoing evaluation. While the methodology changed between Fall 2015 and Fall 2016, the larger participation as well as the iterative process have provided a basis for comprehensive evaluation of reference service quality. The findings regarding what worked well as well as what went wrong will allow the group to move towards an improved service quality informed by specific areas on which to focus.

Notes

1. Devlin, Frances, John Stratton and Lea Currie, "Not Just Q&A: Teaching through Digital Reference" (paper presented at the Brick and Click Libraries Academic Library Symposium, Maryville, Missouri, November 2, 2007).
2. Devlin, Frances, Lea Currie and John Stratton, "Successful Approaches to Teaching Through Chat," *New Library World* 109 (2008): 223–234, accessed February 11, 2016, doi: 10.1108/00907320110399547.
3. Pinto, María and Ramón A. Manso, "Virtual References Services: Defining the Criteria and Indicators to Evaluate Them," *The Electronic Library* 30 (2012): 51–69.
4. Kern, Kathleen M, "Looking at the Bigger Picture: An Integrated Approach to Evaluation of Chat Reference Services," *The Refer-*

- ence Librarian* 46(2008):95–96,99–112, accessed December 22, 2016, doi: 10.1300/J120v46n95_07.
5. Logan, Firouzeh F. and Krystal Lewis, “Quality Control: A Necessary Good for Improving Service,” *The Reference Librarian*, 52(2011):218–230, accessed October 20, 2015, doi: 10.1080/02763877.2011.557314.
 6. Matteson, Miriam L., Jennifer Salamon, and Lindy Brewster, “A Systematic Review of Research on Live Chat Service,” *Reference & User Services Quarterly* 51(2011):82–100, accessed September 20, 2016, doi: 10.5860/rusq.51n2.172.
 7. Burger, Amy, Jung-ran Park and Guisu Li, “Application of Reference Guidelines for Assessing the Quality of the Internet Public Library’s Virtual Reference Services,” *Internet References Services Quarterly* 15(2010):209–226, accessed June 21, 2016, doi: 10.1080/10875301.2010.526479.
 8. Ibid, 215.
 9. Pomerantz, Jeffrey, Lili Luo, and Charles R. McClure, “Peer Review of Chat Reference Transcripts: Approaches and Strategies,” *Library & Information Science Research* 28(2006):24–48.
 10. Luo, Lili, “Chat Reference Competencies: Identification from a Literature Review and Librarian Interviews,” *Reference Services Review* 35(2007):195–209.
 11. Matteson, Salamon and Brewster, “A Systematic Review of Research on Live Chat Service,” 182.
 12. Logan, and Lewis, “Quality Control: A Necessary Good for Improving Service,” 225.
 13. Schwartz, Howard R., “The Application of RUSA Standards to the Virtual Reference Interview,” *Reference & User Services Quarterly* 54(2014): 8–11.
 14. Van Duinkerken, Wyoma, Jane Stephens and Karen I. MacDonald, “The Chat Reference Interview: Seeking Evidence Based on RUSA’s Guidelines: A Case Study at Texas A&M University Libraries,” *New Library World* 110(2009):107–121, accessed September 20, 2016, doi: 10.1108/03074800910941310.
 15. Maness, Jack M., Sarah Naper and Jayati Chaudhuri, “The Good, the Bad, but Mostly the Ugly: Adherence to RUSA Guidelines during Encounters with Inappropriate Behavior Online,” *Reference & User Services Quarterly* 49(2009):151–162.
 16. Burger, Park and Li, “Application of Reference Guidelines,” 216.
 17. Schwartz, “The Application of RUSA Standards,” 8.
 18. Pomerantz, Luo and McClure, “Peer Review of Chat Reference Transcripts,” 41.
 19. Luo, Lili, “Effective Training for Chat Reference Personnel: An Exploratory Study,” *Library & Information Science Research* 31(2009):210–224, accessed September 29, 2016, doi: 10.1016/j.lisr.2009.04.004
 20. Schwartz, “The Application of RUSA Standards to the Virtual Reference Interview,” 8.
 21. Devlin, Currie, and Stratton, “Successful Approaches to Teaching through Chat,” 229.
 22. Van Duinkerken, Stephens and MacDonald, “The Chat Reference Interview,” 117.
 23. Logan and Lewis, “Quality Control,” 221 and 224.
 24. Burger, Park and Li, “Application of Reference Guidelines,” 213.
 25. Todorinova, Lily, and Matt Torrence, “Implementing and Assessing Library Reference Training Programs,” *The Reference Librarian* 55(2014):37–48, accessed October 20, 2015, doi: 10.1080/02763877.2014.853277.
 26. Luo, Lili, “Chat Reference Evaluation: A Framework of Perspectives and Measures,” *Reference Services Review* 36(2008):71–85, accessed August 17, 2016, doi: 10.1108/00907320810852041.
 27. Kern, “Looking at the Bigger Picture,” 109.
 28. Pomerantz, Luo and McClure, “Peer Review of Chat Reference Transcripts,” 34.
 29. Gerlich, Bella Karr and G. Lynn Berard, “Testing the Viability of the READ Scale (Reference Effort Assessment Data): Qualitative Statistics for Academic Reference Services,” *College & Research Libraries* 71(2010):116–137, accessed December 29, 2016, doi: 10.5860/0710116.
 30. Luo, “Chat Reference Evaluation,” 75.
 31. Pomerantz, Luo & McClure, “Peer Review of Chat Reference Transcripts,” 43.
 32. Gerlich and Berard, “Testing the Viability of the READ Scale,” 117.
 33. Devlin, Currie, and Stratton, “Successful Approaches to Teaching through Chat,” 226.
 34. Ibid, 225.
 35. Van Duinkerken, Stephens, and MacDonald, “The Chat Reference Interview,” 117.
 36. Matteson, Salamon and Brewster, “A Systematic Review of Research on Live Chat Service,” 181.
 37. Van Duinkerken, Stephens, and MacDonald, “The Chat Reference Interview,” 117.
 38. Luo, “Chat Reference Evaluation,” 79.