Title: The Use of Transaction Logs to Study the Effectives of Librarian Behaviors on User Satisfaction in a Virtual Setting: A Mixed-Method Approach

Today, I'm going to share with you a research project that looked into the Use of Transaction Logs to Study the Effectives of Librarian Behaviors on User Satisfaction in a Virtual Setting. I will use this study that the two of us conducted to illustrate a combined qualitative and quantitative research methodology.
Outline
In presenting my research, first, we'll start with the background of this research. Then, we'll delineate specific research questions of the study followed by research method that tells you how the data were collected to answer the research questions. Then, we'll present major findings. At the end, we'll summarize the findings and discuss their implications in both research and practice and take questions.
Research Background

- Patron/staff interaction during reference interviews as a strong predictor of reference success (Gers & Seward, 1985; Whitlatch, 1990)
- RUSA behaviors as consistent predictors of reference success → Suggested the RUSA Guidelines as a useful reference service evaluation tool (Gatten & Radcliff, 2001; Saxton, 2002)
- Few assessed virtual reference success using a behavioral standard
- Methodological benefit of using available data

Research Background

(1) Successful reference interview between patrons and reference staff has been emphasized consistently in previous literature as an important factor of reference success.

In 1985, Gers & Seward (1985) demonstrated that reference librarians' verbal and nonverbal communication skills are very important in delivering effective reference services. This has brought a great attention on reference interview and staff training to teach effective verbal and nonverbal communication skills.

(2) From this background, RUSA has developed the behavioral Guidelines, known as the Guidelines for Behavioral Performance of Reference and Information Service Providers. It delineated that behaviors that can lead to effective reference interview. A handful of research studies consistently showed that the use of the behaviors in the Guidelines are positively associated with reference success (Gatten & Radcliff, 2001; Saxton, 2002). Thus, the Guideline has been recognized as a useful staff training and performance evaluation tool. Very recently in June 2004, RUSA has revised the original 1996 Guidelines to help librarians who provide virtual reference interview.

So far, very little is known about the nature of virtual reference interview. For example,
- how do reference librarians interact with their patrons in virtual reference setting,
- are they similar to traditional reference services or different,
- are the verbal and nonverbal cues in virtual references similar to physical interactions or not,
- what are the model behaviors that could lead the virtual reference service more successful.

This revised Guideline has not been utilized in virtual reference research.

(4) Among studies that looked into the Guidelines, Gatten & Radcliff (2001) evaluated the effectiveness of the staff training of the 1996 RUSA Guideline in physical reference service settings. Their study was conducted using an unobtrusive observational method where proxies approached the librarians and recorded the occurrence of the behavior and rated their satisfaction with the answer. In fact, this type of research can be much easily done in virtual reference setting. Librarians can easily access the interview transaction transcripts that recorded reference interview word by word along with co-browsing process. So far, to my best knowledge, there has been little research that assessed the virtual reference interview using this easily available observational data. Thus, we designed a study to investigate it.

Perhaps, the best reference that help us to understand the nature and effectiveness of virtual reference would be physical reference interviews.
RUSA Behavioral Guidelines: Before I go any further, I’ll briefly mention the basic structure of the 2004 Guideline.

It’s basically an extended version of the 1996 edition. Recognizing the increasing need for a guideline that help the staff who provide digital reference, RUSA revised the guideline. Maintaining the original five-component structure, the revised Guidelines categorized each of the five areas into three aspects: general setting, physical setting, in remote access settings.

First, **Approachability** includes behaviors that ensure easy access to the reference staff by removing/lowering the barrier to personal assistance. (Examples: Making signs visible, making the presence of virtual reference visible, making the eye-contact or word contact with the patrons)

Second, **Interest** includes behaviors that show librarians’ interest to the patron’s reference questions so that patrons could state their questions without hesitance. (Examples: Focus attention to the patron, maintaining and re-establishing “word” contact, making question scope visible and clear)

Third, **Listening / Inquiring** includes behaviors that ensure good listening and questioning skills so that librarians can find out patrons’ real information needs, which sometimes are buried or unclear. (Examples: communicating receptive cordial ways, use proper written language, using adequate probing, rephrasing questions for adequate understanding of the questions)

Fourth, **Searching** includes effective search skills and related behaviors during information searches to maximize searching effectiveness and accurate answer (Examples: explaining search strategies, escorting patrons in the search process, provide pointers and information sources).

Finally, **Follow-up** includes behaviors that ensures proper ending of the reference transactions, ensuring patron satisfaction with the answers, and referring to the alternative (Examples: “Asking if their questions have been completely answered, if they need additional information, refers to alternative sources if not answered, ask to come back if they need further assistance”)

This revised Guideline was approved in June 2004 for use. While the revised guideline has incorporated some issues that are relevant for virtual reference, the core component and idea remain the same.
Research Purpose

- To investigate the extent of the presence of behaviors prescribed in *the RUSA Behavioral Guidelines* in virtual reference interviews

- To assess the impact of the RUSA behaviors on user satisfaction in a virtual setting

- To investigate the effectiveness of *the Guidelines* as an assessment tool for virtual reference services

Research Purposes

In the current research, we, first of all, wanted to investigate to what extent each of the behaviors prescribed in the RUSA Behavioral Guidelines was present in virtual reference interviews in general.

The second purpose of the study was to assess whether or not the presence of RUSA behaviors was related to user satisfaction with virtual reference service. Are the RUSA behaviors make the users of virtual reference service more satisfied?

Finally, the third purpose was to examine the use of transaction log analysis combined with user survey an a useful research tool to study virtual services.
Research Questions

RQ1. To what extent is each RUSA behavior observed in virtual reference interviews?

RQ2. Do RUSA behaviors make virtual reference more successful?

RQ3. How good is the Guidelines as a virtual reference service assessment tool?

For each research purpose, a corresponding research question was developed:

RQ1. To what extent is each RUSA behavior observed in virtual reference interviews?
RQ2. Do RUSA behaviors make virtual reference successful?
RQ3. How good is the Guidelines as a virtual reference service assessment tool?
Research Methods:
The present study examined chat reference services delivered through the Broward County public library system in Florida. As the largest library system in Florida, it has 33 regional and branch libraries. Since August 2002, the system has used the 24/7 Reference chat reference service delivered by the Metropolitan Cooperative Library System (MCLS), an association of libraries located in the greater Los Angeles area, funded by a Federal Library Services and Technology Act (LSTA) grant.

The data analyzed for the present study were on-line chat reference transactions initiated by the patrons of the Broward Country library system, along with online survey questionnaires, which the service users completed voluntarily.

The questionnaire was devised and supplied by 24/7 Reference and popped up upon completion of each transaction. While the patrons were mostly the users of the library system, the reference staff who provided the service were from forty-nine library systems across the United States participating in the MCLS 24/7 Reference program.

During the six-month research period between January and June 2004, a total of 1,667 transactions took place and 1,387 were analyzable as being either completed or transferred transactions. In order to analyze the influence of referrals on user satisfaction, all 420 transactions that had a corresponding questionnaire that chat service users had submitted immediately after their session to report their satisfaction with the service were chosen. This amounted to 30.1% of the total analyzable transactions.

By having a chat reference session transcript with a matching user satisfaction survey questionnaire, the researcher was able to examine what question was asked in a reference session, whether the question was answered completely or referred, and to what extent the patron was satisfied with the answer or referrals she or he received.
1. Independent Variable: Librarian behaviors

- Data source: VR transaction transcripts
- 9 RUSA behaviors & 1 IFLA behavior
  - Selection: Easily observable behaviors
  - Coding: Observed=1; Unobserved=0

The Independent variable of the study is Librarian behaviors

Librarian Behaviors were measured in terms of the behaviors prescribed in the RUSA Guidelines. The behaviors were observed from reference session transcripts. The transcripts recorded all verbal interactions between reference librarians and patrons, including URLs from co-browsing activities where librarians guide users to see what the librarians show.

Initially, the behaviors in the transcripts were coded for all items in the Guidelines by assigning “1” for the appearance of the behavior and “0” for the absence of such behavior. Basically, we coded only those RUSA behaviors that can be readily observable. For example, we can code a behavior such as “Maintains or re-establishes "word contact" with the patron”. However, behaviors like “Focuses attention on the patrons” are not easily observable in text-transcript. So, we took a more conservative approach by including only apparent behaviors for the content analysis.

As the coding process progresses, 9 types of behaviors were emerged from this preliminary coding scheme. In the next slide, I’ll talk a little details of the coding.
Example: Reference Transaction Transcript and Co-browsing (cont’d)
Example: Reference Transaction Transcript and Co-browsing
Selected RUSA Behaviors

RUSA guidelines have a total of 49 items across the 5 areas. While some items were easily observable, some items were not because they were thinking process rather than actual behavior, some items were overlapped with other items, some items were not applicable in virtual reference setting. The final 9 behaviors were mostly derived by combining similar behaviors together.

**Area 1 (8 items)**: general concept of “welcoming” was opted (e.g.: Hi, this is reference librarian); Excluded: most of items were not coded because they are applicable to pre-reference interview stage (1.8), or to in-person situation. Items most for pre-interview stage.

**Area 2 (8 items)**: only two were applicable (2.2 & 2.6). Excluded: Physical (2.1,2.3-2.5), out of scope (2.7-email; 2.8 pre-transaction info).

**Area 3 (10 items)**: receptive communication (3.1 & 3.2); probing by rephrasing, clarifying, asking questions (3.5-3.8). Excluded: not easily observable (3.3, 3.4, 3.9); overlap (3.9).

**Area 4 (11 items)**: In fact many items are overlapped and thus one behavior could be coded across multiple behaviors. Thus, we decided to code Searching with two aspects: “Searching for or with patrons” which include 4 items (4.3, 4.6, 4.11) and “Providing information sources” which combine 3 items (4.5, 4.9, and 4.11). Excluded: falling in a better category (reference interview (4.1, 4.2-verifying words), follow-up (4.7)), cognitive process rather than explicit behavior (4.2 in part)

**Area 5 (10 items)**: has three distinctive behaviors: (1) proper closing (asking if answered and needs more info-4.7, 4.1, and 5.9); (2) offer alternatives or refer (5.4-5.8, and 5.9 remote); and 5.2 (comeback). Excluded: inapplicable(5.3)

Finally, I added one area from IFLA: “Use the client's name” (from 2.4 Guidelines for Chat Sessions : IFLA Digital Reference Guidelines) to “Area 1” approachability. We thought it’s worth looking into because people have been pondering about its effectiveness.
Inter-coder reliability test for content analysis

- 20% of transactions were coded by two-independent coders to test the level of agreement between two independent coders on the RUSA behavior coding.

- Result: Generally, they were above the threshold value of 80% -- Ranged between 96.4% and 76.2% across all 10 behaviors.

Inter-coder reliability

RUSA behaviors were coded by two independent coders for intercoder reliability in order to ensure the consistency. The tests show the level of agreement between two independent coders.

As the first step for this purpose, the primary researcher coded the entire 422 transactions. Then, the second coder, a reference librarian who received training for coding the RUSA behaviors, coded every fifth transaction (n = 84). This sample for intercoder reliability comprises 20% of the total transactions, which is the recommended percentage for social science research. Finally, the percentage agreement between the two coders was calculated for each behavior, which informs the level of agreement between two coders. The agreements were Welcoming (91.7%), Use of patrons’ name (96.4%), Interest (78.5%), Listening (85.6%), Inquiring (78.6%), Searching with or for patrons (76.2%), Offering pointers and information sources (78.6%), Asking if answered completely (91.7%), Offering referrals (85.0%), and Asking to come back for further assistance (90.5%).

All these intercoder percentage agreements were either above or close to the generally acceptable threshold value of 80%.
Dependent Variable: User Satisfaction

- Data source: Post transaction user survey

- User satisfaction (a composite of 1-3)
  1: Satisfaction with the answer
  2: Perceived staff quality
  3: Intention to return

* Didn’t use open-ended comments in the survey due to many missing responses

**Dependent Variable User satisfaction**

The dependent variable is User satisfaction. It was measured using patrons’ answer to a short pop-up survey that was submitted immediately after each reference transaction.

This survey has 8 questions that ask patrons’ experience of the reference service. The survey has three items that relate different aspects reflecting user satisfaction: (1) Satisfaction with the ANSWER (not just overall service), (2) user perception of the staff quality, (3) intention to use the service again (frequently used item since Joan Durrance proposed to use in 1986).

Each of these three items is a unique indicator of reference service evaluation criteria but each of them represents the service effectiveness only partially. Therefore, we created a more comprehensive, new variable by combining all three question items and named it as “user satisfaction.” This composite variable reflects overall user satisfaction with the virtual reference service.

At first, we considered to analyze the contents of the open-ended comments, but later decided not to because only one third of the survey respondents made comments, which means a lot of missing cases.
# User Satisfaction Composite Variable based on three User Satisfaction Items on a Five-Point Scale

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1. Satisfaction with the answer</td>
<td>Unsatisfied</td>
<td>Not sure</td>
</tr>
<tr>
<td>2. Perceived staff quality</td>
<td>Poor</td>
<td>Average</td>
</tr>
<tr>
<td>3. Willingness to use the service again</td>
<td>Never</td>
<td>Maybe</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>Excellent</td>
</tr>
<tr>
<td></td>
<td>Very likely</td>
<td></td>
</tr>
</tbody>
</table>

Mean = 12.7  SD=3.4  
Min=3 (Highly dissatisfied); Max=15 (Highly Satisfied)

- The overall user satisfaction was computed by summing up the above three indicators ([Footnote 1](#)). Measurement research literature indicates that composite variables yield scores that are generally more valid and reliable than does a single item ([DeVellis 1991](#)).

- In creating a composite variable, a series of computations was conducted because the three items in the survey questionnaire were measured either on a three-point or on a four-point ordinal scale (e.g., 'satisfied,' 'not sure,' and 'not satisfied' for the 'satisfaction with the answer' item). The three items were rescaled on the same five-point scale, and then summed as a single composite variable. This procedure of data management allowed the researcher to conduct the necessary inferential statistical tests as shown in the chart.

- As a result, the mean of the overall user satisfaction was 12.69 with a standard deviation of 3.44 in the range between a minimum value of 3 for 'highly dissatisfied' and the maximum value of 15 for 'highly satisfied.'

- We did one additional checking to minimize the arbitrariness in assigning numeric values to the verbal descriptors. That is, we tried two alternative scales (i.e., four-point and six-point scales) in addition to the selected five-point scale by making slight variations in the value assignment. When the test results from a series of statistical analyses for Research Question 2 and Research Question 3 were compared for all three scales to examine the stability of the assigned values, the test results were identical across all three scales, indicating the value assignment on the five-point scale is stable and reliable.
Validity & Reliability of User Satisfaction Composite Variable

Reliability. Cronbach's Alpha coefficient = .85

Validity. Conducted Factor analysis

Three items’ structure-pattern coefficients = .868, .916, and .876, respectively.

It explained 78.64% of the total variance.

In order to determine the reliability and validity of the scores yielded by the composite variable, a reliability test and a factor analysis were conducted, respectively.

For reliability: a Cronbach's alpha coefficient of .85 was obtained, indicating that the composite variable generated scores were reliable.

For validity: the factor analysis revealed that the three items were represented by one factor, with structure-pattern coefficients of .868, .916, and .876. This factor explained 78.64% of the total variance. This high validity score confirms that the composite variable is measuring a single construct, suggesting that the composite variable is a valid measure of user satisfaction.
Findings
Presence of RUSA Behaviors

(1) First, Findings to the RQ 1: To what extent is each of the behaviors in the Guidelines present in virtual reference interviews? (N = 422)

The presence of the selected behaviors was ranged between 28.7% and 63.7%.

The three most frequently observed behaviors appeared over 50% of the reference interviews analyzed:

1. Approachability (53.8%): “Hi, this is reference librarian from California.”
2. Interest (63.0%): “I’m reading your question now. Just a moment please…”
3. Come back (63.7%): “If you need further assistance, please contact us again.”

As you may recognize, the reason these behaviors were more frequently appear is that they came from scripted words.

(2) Comparison with previous studies: Since little research was done using the Guideline, comparison with previous literature is not possible across all areas. However, one possible area is “follow-up” area. In Dewdney & Ross, follow-up behavior appeared about 1/3. In Gatten & Radcliff, “Asking if question was answered was 29%;

(3) I should also mention this to help your interpretation of this graph. Among these ten behavioral categories, there are behaviors that are desirable to have in any type of RQ. But, some behaviors are not expected to appear in all transactions. For example, approachability, word contact, follow-up (Answered, Comeback) can be expected in any type of queries.

In contrast, listening, probing, pointers, offer alternatives, first name may not, depending on the type of questions (ready-reference vs. subject) or the information given to the librarians (some users used their names but others didn’t). Therefore, a low frequency in this graph doesn’t necessary reflect a failure.

We can interpret this simple frequency of the behaviors more meaningfully when we look at these behaviors in relation with reference success. This relates to Research question 2.
RQ2. Were patrons more satisfied when a RUSA behavior was observed than not?

User satisfaction (Mean=12.7; SD=3.4; Max=15; Min=3)

RQ 2: Were patrons more satisfied when a RUSA behavior was observed than not?

To the research question 2, the graph on the screen shows the answer. Let me remind you that user satisfaction score was obtained by combining three responses from user survey. The mean score of user satisfaction was 12.7.

The violet bar represents the average user satisfaction score when a behavior recommended by RUSA Guidelines was observed in the reference interview.

The purple Line represents the average user satisfaction score when the behavior was NOT observed in the reference interview.

The general pattern is that user satisfaction was higher when the behavior was observed from the virtual reference interviews, except two behaviors (welcoming & referrals). Now, we need to make a much concrete finding about this difference. Is each of these differences a statistically significant difference or is it just by chance?
### How significant was the difference? (t-test)

<table>
<thead>
<tr>
<th>Librarian Behaviors</th>
<th>User satisfaction</th>
<th>Difference</th>
<th>Sig. of Diff. (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>When Present</td>
<td>When Absent</td>
<td></td>
</tr>
<tr>
<td>Welcoming</td>
<td>12.59</td>
<td>12.80</td>
<td>-0.21</td>
</tr>
<tr>
<td>Patron name</td>
<td>13.29</td>
<td>12.45</td>
<td>0.84</td>
</tr>
<tr>
<td>Interest</td>
<td>12.78</td>
<td>12.54</td>
<td>0.24</td>
</tr>
<tr>
<td>Listening</td>
<td>13.47</td>
<td>12.27</td>
<td>1.2</td>
</tr>
<tr>
<td>Inquiring</td>
<td>13.04</td>
<td>12.51</td>
<td>0.53</td>
</tr>
<tr>
<td>Searching</td>
<td>13.51</td>
<td>12.12</td>
<td>1.39</td>
</tr>
<tr>
<td>Provide pointer</td>
<td>13.59</td>
<td>12.05</td>
<td>1.54</td>
</tr>
<tr>
<td>Answered?</td>
<td>13.58</td>
<td>11.90</td>
<td>1.68</td>
</tr>
<tr>
<td>Referrals</td>
<td>12.61</td>
<td>12.76</td>
<td>-0.15</td>
</tr>
<tr>
<td>Come back</td>
<td>13.00</td>
<td>12.15</td>
<td>0.85</td>
</tr>
</tbody>
</table>

(Vicki, we may not need this slide because it's just a statistical test result of the earlier slide. But, depending on the conference audience, we may need this technical data.)

How significant was the difference?

To answer this question, t-test was used to compare user satisfaction between when the behavior was present and when the behavior was not present. When there was a statistically significant difference, I highlighted it in bright green color in this table.

Let’s look at the behavior of Patron name, as an example. Among the transactions where reference librarians called patron’s name, their average user satisfaction was 13.29. Among the transactions where reference librarians didn’t call patron’s name, the average reference success was 12.45. The average reference success was 0.84 point higher when librarians called the patrons’ first name. Was the difference statistically significant? Yes. The number in the last column indicates the result of statistical test. The probability that this difference is just by chance is as low as .014, that is, 1.4% of the time. It’s a customary to use 5% error margin in statistics. Any result with the significance level below .05 can be interpreted as a significant difference. So, across six out of the ten observed behavior, there was a statistically greater user satisfaction when librarians showed the behaviors than not.

In other words, reference service tends to be more successful when librarians:
1) Used patron’s first name during the reference interview
2) Communicated more receptively and listened more carefully
3) Searched with or for the patrons
4) Provided pointers
5) Asked whether the question was completely answered
6) Asked to come back if patrons needed further assistance

This finding confirms the earlier literature that showed the importance of the RUSA Guideline behaviors to successful reference interview transactions. Furthermore, this...
RQ3. Which librarian behaviors in virtual reference interview can help us to predict user satisfaction?

- Multivariable regression analysis identified consistent predictors of user satisfaction (order of importance):

1. Asking whether the question was answered completely ($p < .001$)
2. Providing pointers, search paths, URLs ($p = .014$)
3. Asking to come back for further assistance ($p = .019$)
4. Searching with/for patrons by explaining search strategies and help with reformulating questions ($p = .023$)
5. Receptive, cordial listening ($p = .044$)

RQ3. Which Librarian behaviors in virtual reference interview can help us to predict user satisfaction?

To answer this question, we ran another statistical test, called “Multivariable regression analysis.” Through this test, we were able to identify four behaviors that allow us to predict virtual reference success.

1. The strongest behavioral predictor was “asking whether their question was answered completely,” followed by “providing pointers,” “asking patrons to come back when they have further assistance,” “searching with/for patrons, and “receptive, cordial listening.

The strongest predictor was asking a very simple follow-up question, like (“Did you find what you needed?, Does this completely answer your question?, and Is there anything else I can help you with?).

It’s pleasant surprise that this study confirms Gers and Seward (1985) contention that “the single most important behavior because it has the potential for allowing one to remedy lapses ion other desirable behaviors.” It is a chance to “repair the interaction or to formulate a new, more promising search strategy.

2. Let me talk a little about three behaviors that were not found as significant predictors.

Te three behaviors had no association with reference success: Approachability, Interest, and Providing alternative. We pondered about this outcome and have a couple of possible explanations.

(1) Behavioral indicator of Approachability was only measured by initial word contact in this study. The welcoming “Hi, this is the reference librarian” because I could not identify many other behavioral indicators that clearly represent the behavior. Thus, this initial contact word can be considered to be just customary word that doesn’t have any positive or negative impact on use.

(2) Likewise, behavioral indicators measuring Interest were not rich because it was not easy to identify such word when the librarian was not fully engaged and explicitly...
Wrap-up

Having said that all, let me wrap up this presentation by pointing out some of the implications of our study.
Implications:

- **For research:**
  Demonstrated a way to conduct a mixed-method research using easily available, unobtrusively collected reference transactions and user survey in studying user satisfaction in a virtual reference setting.

- **For practice:**
  Help training the reference staff.

- **For professional organization:**
  Help further promote the Revised Guidelines as an effective behavioral standard of virtual reference interviews.

### For the research community:
One of the advantages in studying virtual reference is the availability of reference transaction scripts. It allows the researcher to observe reference interview unobtrusively in natural setting. Methodologically, this study demonstrated a relatively easy way to utilize available, unobtrusively collected real reference transactions in assessing reference service performances. Furthermore, using the accompanying pop-up user survey, we were able to determine the causal relationship between librarian behavior and immediate user satisfaction. This mixed method approach is very powerful yet easily applicable research method by both researchers and practitioners.

### For the practitioner community:
The finding could help reference staff training in practice. As the previous literature consistently demonstrated (Gers & Seward, 1985; Dewdney & Ross, 1998), this study also showed that the simple questions like, "Did it completely answer your question?" "Do you need anything else?" could enhance reference success significantly than any other efforts. This finding should be reassured in reference staff training to enhance the performance.

### For RUSA:
we believe the findings of this study is very encouraging for RUSA. This study was the first study that demonstrated the value of the Guideline by assessing its real items in 5 areas in virtual reference setting. The findings demonstrated that the RUSA Guideline can continue to serve the reference community as a great training guideline. It needs some adjustment in order to be used as a direct assessment tool. Some items in the Guideline need further clarifications if it were to be used as a concrete evaluation tool without...
This ends the presentation. Thank you very much for your interest and listening. Now, let me take your questions.
Contact Information

• Nahyun Kwon  nkwon@cas.usf.edu

• Vicki L. Gregory  gregory@shell.cas.usf.edu