

You Say Hello and I Say Goodbye: Methodology and Findings from An Academic Library's Virtual Chat Service Assessment Program

Helen Cawley

University of Notre Dame, United States

ABSTRACT

This *Report from the Field* outlines the implementation of a formal, comprehensive assessment program for an academic library's virtual chat service, designed to analyze quality and inform improvements to all reference interactions. The methodology features twice-yearly, team-based peer review among graduate student employees, utilizing an eight-metric rubric to generate both quantitative and qualitative data. Initial sessions found that most interactions were Proficient or Excellent, but highlighted the need for coaching on closing statements and providing updates during long searches. Recommendations include adding canned messages to the chat platform and refining training with tutorials and more reference documentation to reinforce consistent service quality.

Keywords: Virtual Reference; Chat Assessment; Peer Review; Reference Services; Academic Libraries

INTRODUCTION

Academic libraries rely on high-quality chat services to support all patrons, especially those with busy schedules, those studying or working abroad, and anyone who cannot visit the library in person. Well-run chat services ensure consistent, effective user support in a dynamic digital environment that is often more familiar to patrons than the in-person desk. Assessing library in-person reference interactions is often difficult without survey feedback, which is usually unavailable due to patron anonymity. Readily available chat transcripts, encompassing the entire exchange from the patron's first "Hello" to the reference desk's concluding "Goodbye," provide a valuable resource for evaluating reference interaction quality and customer service satisfaction.

While the reference desk in this report has operated its virtual chat service for many years, it has only recently begun a formal, comprehensive assessment of that service. Two complete chat

assessment sessions have been conducted since its inception, one in Spring and one in Fall, 2025. The chat assessment program focuses on two critical perspectives: the patron experience and the viewpoint of the chat support staff. Comprehensive quantitative and qualitative data about the library's chat service can be analyzed and interpreted through an ongoing assessment process, providing insights into improving all library interactions, including those in person.

This project report will outline the methodology (including the program design and structure of the assessment), the quantitative performance analysis schedule and procedures, and the qualitative insights and thematic analysis performed by the student and supervisor chat staff. Findings from the first two assessment sessions, recommendations from the student employees, and the future direction of the assessment program will be discussed.

BACKGROUND

The academic library discussed in this report is located at the University of Notre Dame, a mid-sized, private university in the United States Midwest. The system includes one main (Hesburgh Library) and three subject-specific branch libraries. Extensive collections, totaling approximately four million physical items, are maintained and provide research resources for nearly 20,000 students, faculty, and staff, in addition to thousands of global patrons through in-person, phone, email, and chat services.

The main reference (Ask Us) desk within the Hesburgh Library is operated by a full-time desk supervisor who hires, trains, and manages seven to ten graduate student employees yearly, and also supports, instructs, and schedules the same number of library colleagues to back up the desk. It is the only location within the library system that operates a virtual chat reference service.

METHODOLOGY

Before Spring Semester 2025, the Ask Us desk did not conduct any formal assessment of in-person or virtual interactions. Student employees performing reference interactions are supervised in person by librarians and the interactions are guided and afterwards discussed in person to instruct student employees about best practices for excellent service. Although the students undergo a lengthy online training program on conducting reference interviews and utilizing research resources before working at the Ask Us desk, a specific unit on online chat was only recently introduced.

The "Chatting at the Ask Us Desk" unit added in the Spring of 2025 includes several lessons, including those on online chat and in-person customer service, how to chat, chat etiquette, and chat evaluation. Student employees are tested on basic typing and grammar to meet online customer service standards, and practice specific scenarios taken from real chats received at the desk. They experience the online chat process from both a patron viewpoint on the library website and in the staff platform, receiving immediate feedback as they practice responding quickly and accurately to "mystery" questions.

To create the assessment portion of the unit, I surveyed other university librarians to find out what criteria they were examining, who participated in the evaluation process at their libraries, how they submitted feedback, and how improvements based on the feedback were incorporated

into daily practice. Based on this information, I decided that a structured, team-based chat assessment program would allow student employees to objectively evaluate peers conducting the chats, as well as foster continuous improvement to our patron service.

Since the student employees often work shifts alone or alongside a librarian, their limited interaction with each other makes it difficult for them to view themselves as part of a larger, cohesive desk team. By working together on this project, I wanted the student employees to gain an opportunity to connect with one another and recognize the importance of cooperation and collaboration in reference work at the library. To do that, I integrated teamwork into the assessment by grouping student employees into teams of two or three. Each team was given an animal mascot and picture, contributing to team camaraderie in a fun, engaging way.

The evaluation metrics are crucial to this project, and they were finessed after the first attempt at assessment. The final version of the rubric consisted of eight metrics, each having one or two submetrics, which examined several aspects of each chat (see Table 1).

Table 1

Chat Evaluation Metrics

Metric	Submetric 1	Submetric 2
Greeting	Chat staff member replies with correct canned message in a timely manner.	
Response Time	Chat staff member responds quickly to all incoming texts from chat patrons.	Chat staff member updates the patron frequently as they search for information.
Clarity	Responses are clear, concise, and free of jargon and errors.	Chat staff member responds in short, concise snippets of information rather than long explanations.
Accuracy	Chat staff member demonstrates the ability to find and provide the correct resources and answers the question asked by the patron.	Chat staff member shares screenshots and permalinks correctly, with explanations.
Problem-Solving	Chat staff member shares strategies to use different library resources and services.	Chat staff member refers the chat patron appropriately for the best help.
Professionalism	The overall language and	

	interaction is professional, courteous, and friendly.
Empathy	Clear effort is made to understand the patron's feelings or question.
Closing	Patron is given the opportunity to ask follow-up or additional questions. A professional or friendly closing of the chat is present.

Each submetric question is scored to determine the chat staff member's performance during the interaction. The student employees used the following assessment scale (see Table 2).

Table 2

Chat Evaluation Rating Scoring System

Score	Rating
4	Excellent or Not Applicable
3	Proficient
2	Developing
1	Needs Improvement
0	Not Observed

This scoring system aims to identify not only the highest performance in a submetric but also the complete absence of the required action. A score of 4 is assigned to actions that demonstrate a top example of performance for the submetric question, while zero is assigned to a required action that was missing in the chat. To avoid penalizing chats for actions that were not needed in the context of the chat, a score of 4 is also given to a submetric if it is not relevant to the chat. The scores are totaled to provide an overall chat Proficiency Rating (see Table 3).

Table 3*Chat Evaluation Proficiency Ratings*

Total Score	Proficiency Rating
47-52 points	Excellent
39-46 points	Proficient
26-38 points	Developing
13-25 points	Needs Improvement
0-12 points	Not a great chat!

Having devised a workable rubric to assess the chats, the next goal was to create the assessment format. I set several requirements for the format: it must be easy to input, identify, and assign the chats to team members; it must be easy to score each chat; and it must be easy to analyze the data to create reports that can be shared.

The initial assessment format used a Google Form linked to a separate tracking spreadsheet containing team information and the chats for evaluation. To ensure high-quality, actionable data, team leaders were specifically instructed to select only chats that were longer than three minutes and included significant back-and-forth interaction. Chats were sourced from the previous year, with one month's worth (averaging 150 chats) assigned to each team. All staff-identifying details were removed for anonymity. Over the three weeks, teams ultimately assessed between eighteen and twenty-four total chats, depending on the team size. The weekly workload was managed by an alternating team leader, who assigned three chats for evaluation to each member.

Before scheduling the first assessment, I needed to create an evaluation timeline. The schedule needed to allow teams to rotate leadership so all members could direct the process and also ensure they evaluated a sufficiently large sample of chat interactions. Additionally, teams required adequate time to collaborate on their findings and prepare final reports for presentation at an in-person Ask Us desk meeting. To achieve these goals, a four-week, twice-yearly cycle was constructed (Table 4).

Table 4*Chat Assessment Yearly Timeline*

Fall Semester		Feedback	Spring Semester		Feedback
Weeks 1-3:	Team leaders assign chats, teams evaluate, prepare summary reports; rotate team roles	Week 4: Supervisor reviews evaluations, prepares complete summary report, teams meet to give reports and discuss findings	Weeks 1-3:	Team leaders assign chats, teams evaluate, prepare summary reports; rotate team roles	Week 4: Supervisor reviews evaluations, prepares complete summary report, teams meet to give reports and discuss findings

At the end of the three-week evaluation period, the teams began their group reports and presentations, while I compiled the data from all teams to create an overarching report, charting metric averages, and closely examining the scores given for low and high scoring chats to confirm and comment on the assessment given.

At the meeting following the first assessment cycle, it was clear that a Google Form was not the ideal format for this process. While easy to use for scoring the individual submetrics, it did not automatically total the individual scores for student employees, making the process of assigning a Proficiency Rating onerous. Also, each team member's assessments were downloaded to a master spreadsheet that only I could access, and student employees were unable to refer back to their evaluations. Feedback from the student employees demonstrated that, while they liked the form itself and found it easy to use, they found it awkward to use separate documents for viewing and assessing the chats.

Based on this feedback, I completely revised the format and created a single Google Sheet for each team. The sheet includes the metric scores and links to an instruction document which outlines the entire process, including the rubric. Each metric is stated on the sheet with a box next to it for the score assignment. The sheet automatically provides a total using a formula, allowing the student employee to immediately identify and record the Proficiency Rating for each chat. The team leaders access the chat transcripts via a dedicated tab in the team sheet, copying each chat into a unique note within their team member's assessment section. They also record crucial metadata, including the unique chat identifier, the date and time the chat was received, and the duration of the chats in minutes and seconds. This allows me to quickly find the chats on the transcript tab in order to check for accuracy.

FINDINGS AND FUTURE DIRECTIONS

The two chat assessment cycles clarified several quantitative and qualitative data points. The student employees unanimously approved of the revised process, favoring the Google Sheet over the Google Form. They valued having the rubric visible in the sheet, the integrated chat transcripts for the submetric review, and the automatic scoring that simplified the Proficiency Ratings

assignment. Although I had to copy the individual scores to a new spreadsheet for analysis, this minor difficulty was offset by the benefit of having all team information consolidated in one spreadsheet file.

While the vast majority of chats were rated to be “Proficient” and “Excellent” in both sessions, the student employees felt that the Fall Semester chats were better in comparison to those of the Spring. Both sets of chats were pulled from those conducted before the additional training sessions were implemented, so this might be attributed to either sampling variability in the chat transcripts assigned by the team leaders, or more experience with the process. When analyzing the data, one team recorded a decline in performance across the Accuracy (-0.51), Problem-Solving (-0.55), and Professionalism (-0.44) in their second round of chats as compared to the Spring Semester.

To ensure a more robust data set for performance analysis, I propose to increase the number of evaluated chat transcripts in the next assessment cycle from three chats to six to ten chats per week per team member. I selected the initial number to evaluate to not overwhelm the students with the new assignment, and to ensure an adequate supply of qualifying chats (longer than three minutes in duration and including a significant back-and-forth interaction).

Coaching is needed across multiple areas, with chat closing identified as the lowest-performing metric. Failures in the closure process were frequent, even when the patron did not leave early. Staff members failed to follow up by asking whether the information provided was helpful to the patron’s question, or if they had additional questions. There was also a significant lack of explanation before and after sharing links. A few staff members offered a reassuring question to ensure that the link was accessible to the patron.

Other qualitative and professional shortcomings included the lack of frequent updates while working on a question, leaving long time gaps that could be misconstrued as a lack of empathy or a dropped connection. A tendency toward informality in interactions sometimes resulted in responses that were viewed as being less “professional” than appropriate in this context.

While high-performing interactions featured proactive investigations, assessments highlighted specific training needs. Student employees stated in their assessment that they felt that complex research questions should have been referred more frequently to a subject librarian. Also, the importance of providing correct and detailed instructions (e.g., VPN use, ILL procedures, library policies) and avoiding the dissemination of incorrect or incomplete information was repeatedly emphasized. We discussed the benefits of “teachable moments” in the chats, where describing how the information was located would have provided valuable future research skills to the patron.

The student employees offered several recommendations for the assessment program going forward. They suggested adding more canned messages to the chat platform, which would allow them to quickly insert updates during long searches and provide clear closing statements to ask patrons for any final questions. Other recommendations include creating internal procedural documentation, outward-facing subject guides to direct patrons to common questions about copyright issues, or instructions on accessing materials from off-campus or overseas. Using this feedback, I can refine training content further, allowing student employees to practice their chat

skills and quick retrieval of information. The students appreciated opportunities to reenact scenarios based on the transcripts and the common questions that the Ask Us desk receives.

The assessment process has already provided benefits to the patron service in reference interactions, both online and in person. Student employees' in-person and chat interactions have become more thoughtful and intentional as they incorporate more teaching moments and follow-up when sharing links and screenshots, helping patrons become more self-sufficient in their research. The chat proficiency ratings and charted metric averages can also be shared with the backup librarians and the library administration to demonstrate the quality and effectiveness of our reference service. This data can be vital for contributing to discussions on service improvement and library contributions.

By including the Hesburgh Library's Ask Us desk student employees in the assessment process, they feel more engaged and comfortable in their roles, and they have expressed a readiness to continue with the program. The opportunity presented in every chat interaction—from every patron “hello” to the Ask Us desk “goodbye”—is a golden opportunity to achieve several goals: to interact with patrons; to teach students, faculty, and visitors about library services; to increase the library's visibility and value; and to provide a solid foundation for academic library reference services. Because we are always looking for ways to improve the service to our patrons, virtual chat assessment has provided one more valuable tool in our user satisfaction toolkit.

About the author

Helen Cawley has supervised the reference service at the Ask Us Desk at the Hesburgh Library, University of Notre Dame since 2015. She is an active member of the ALA's RUSA Virtual Reference Services Committee. Her interests include chat reference, patron accessibility, and leisure reading.