

# University of Washington Libraries Chat Reference Transcript Assessment

Jackie Belanger, Kathleen Collins, Alyssa Deutschler, Rebecca Greer, Nancy Huling, Caitlan Maxwell, Ekaterini Papadopoulou, Lauren Ray, Robin Chin Roemer

## **Executive Summary**

This report provides a summation of a project undertaken in 2014-2016 that sought to assess the UW Libraries' Ask a Librarian chat reference service. Since 2002, the UW Libraries has provided live, online chat help to patrons, advancing our vision of meeting the information needs of our diverse communities at any time, and in any place. This project represents the first systematic assessment of transcripts from these chat interactions with our patrons, and is intended to help us learn more about who is using the service and the types of questions we receive.

A team of librarians performed a content analysis on one academic quarter's worth of transcripts using codes that included a modified READ (Reference Effort Assessment Data) Scale. Coding was intended to help us explore the level of effort staff expend on chat reference questions, as well as patterns of use in terms of subject areas, e-book related questions, types of users, and question frequency per quarter. Team members met regularly to discuss their observations throughout the coding process.

The goal of this project was to help us identify optimal staffing models and areas for improvement for the chat service, as well as to understand the chat transcripts' potential in helping us improve other Libraries services (websites, instruction, etc). Finally, qualitative feedback gathered from team discussions was intended to help us identify models for continued assessment of the chat service going forward. Project limitations included a lack of demographic information provided by the QuestionPoint system, applicability of the READ Scale to the online reference environment, and lack of information regarding librarian effort if that effort was not visible in the transcripts.

Results of the coding process showed that:

- Undergraduate students are most likely to ask questions that a complex response, followed by graduate students and then faculty/staff.
- Peak period for chat reference activity is in week six, with the majority of questions during that time being higher in complexity.

- Subject area and e-book categories proved of limited use in revealing new, specific information about the chat service.

Observational feedback showed that:

- There is wide variation in the level of instruction and service provided to users, including how much help is provided during the chat vs in an email follow-up.
- There are missed opportunities for following up with patrons post-chat with additional guidance and information.
- Patrons may have expectations regarding who they are chatting with that do not match reality.

Results suggest areas of opportunity for enhancing and improving current staff training, contributing to the development of chat reference best practices, and ensuring that staffing is adequate during peak times. Project findings have also highlighted new questions about patron expectations, and how peak question times impact chat service quality. This process has also provided direction for improving our assessment of chat reference in the future by utilizing a coding process better matched to the online environment, and coordinating the coding process with other Libraries assessment processes.

## **Aims and questions**

The central aim of this 2014-15 project was to assess the UW Libraries chat reference service in order to learn more about patterns of use and to explore potential opportunities for service improvements. We were interested in learning more about who was using our service, the types of questions submitted, and the degree of effort expended on the service by staff. By investigating these questions, we hoped to revisit staffing decisions for chat reference, as well as to gain a better understanding of how the service is used at our institution in general. As the first systematic assessment of the chat reference service undertaken at the UW Libraries, we also wished to explore the best methods for implementing a more sustainable, ongoing assessment of this heavily-used service.

The team assessed 3721 chat reference transcripts using the Reference Effort Assessment Data (READ) Scale, identified patterns of use, and examined the post-chat user survey tool with the aim of exploring the following questions:

- What is the level of staff effort expended on chat reference questions at various times of the quarter, and what are the optimal staffing models to match the types of questions received?
- Are there patterns of use in terms of questions related to particular subject areas, types of questions, and types of users (undergrads, graduate students, faculty) and, if so, how can this information be used to improve chat staffing and other Libraries services (e.g., instruction, liaison outreach, websites/guides)?
- What is the best model to build in more formal continuous assessment of the service, either in the form of changes to the follow-up user survey or ongoing chat transcript analysis?

## **Background**

Use of the 24/7 chat reference service has increased steadily over the past few years, while reference traffic at physical service points has continued to decline over the same period. Chat reference is growing both in numbers as well as percentage of reference questions. In 2011-12 there were 8500 chat reference sessions reported, representing about 13% of total reference transactions. In 2014-15 there were 20,000 chat questions, or about 31% of total reference questions. The total number of reference questions for those years was about the same. The

physical reference desk service was assessed using the READ Scale in 2011, with the result that librarian staffing of in-person service points at selected libraries was reduced. The READ Scale, developed by Bella Karr Gerlich, Ph.D., Professor and Dean of Libraries at Texas Tech University, is a six-point scale tool used for coding qualitative statistics around “level of effort, knowledge, skills and teaching utilized by staff during a reference transaction” (Gerlich & Berard 2010). The 2014-15 project enabled the team to explore a similar approach to assessing the virtual reference service as that used in the 2011 reference review.

UW Libraries chat service is staffed by both UW librarians, staff and MLIS graduate students, as well as consortium librarians, referred to here as “QP backups”. During the day, UW chat staff volunteer their time to cover chat reference hours, typically in one-hour shifts. This coverage then extends to backups in the evening hours to enable a 24/7 service to our user population.

Chat transcripts have been regularly monitored for quality by the Head of Reference & Research Services and the Online Reference Services Coordinator, which has led to the development of training materials and continuous improvements to the service. However, no systematic assessment had yet been undertaken on a significant set of chat transcripts to learn more about who is using the service and the types of questions we receive via chat. Given the increasing importance of this service to Libraries users, and the hire of a new Libraries online services coordinator, the Head of Reference & Research Services, in consultation with the Libraries Assessment Coordinator, felt that 2014-15 was an opportune moment to undertake an in-depth, systematic exploration of chat reference.

## **Literature Review**

Prior research on chat reference assessment has often focused primarily on staffing models (Bravender, Lyon, & Molaro 2011), patron satisfaction (Lasda Bergman & Holden 2010), or quality of chat service (Arnold & Kaske 2005). Over recent years, a growing body of research has focused on qualitative content analysis of chat transcripts in order to gain a deeper and more nuanced understanding of the types of questions handled via virtual reference (Youngbar 2012; Armann-Keown, Cooke, & Matheson, 2015).

In 2011, a systematic review of virtual reference service literature published between 1995-2010 was conducted by Matteson, Salamon and Brewster. Their findings highlight that the predominant types of data analysis employed by those researching chat reference are: content

analysis of chat transcripts or of survey and interview questions, and quantitative analysis from survey data (2011). The top categories focused on in the studies have been patrons (user motivation, satisfaction, etc.), questions (type and outcome), the question-answering process (looking at the chat interaction as a communication event), and response guidelines (librarian behavior and the offering of instruction in chat).

In reviewing the literature, we looked for examples of chat reference assessment methodologies and approaches that might address our desire to establish a sustainable, ongoing evaluation model. Were there templates or ways of approaching chat assessment that could be reasonably replicated on an ongoing basis, and that integrated assessments of in-person service points? Unfortunately, while there has been a fair amount of research analysing chat reference in academic libraries, few address how assessment methodologies might be sustained over time. Most of the studies we found described one-time assessments that were done with the aim of expanded understanding of patron questions, and chat service improvement. Many describe the creation of coding schemes for the purpose of answering questions specific to that library's assessment needs. Often these coding schemes are based on Katz's *Introduction to Reference Work* (1997), ACRL Reference Transaction definitions (Fennewald, 2006) or build on the work done in other chat reference assessment studies.

While previous studies did not address our questions regarding sustaining chat reference assessment over time, a few studies did provide approaches to question coding that might help us address new questions raised during the course of this project. For instance, Marsteller & Mizzy's study of chat reference transcripts at Carnegie Mellon University involved the creation of coding categories that (in addition to coding for question type) addressed instruction (librarian asking of closed/open questions) and communication style (patron engagement with the chat dialogue) (2003). Meert & Given's study of chat reference transcripts at looked at why questions were referred, examining why patrons did not receive a complete real-time response during the chat interaction, and comparing these between university and consortia chat staff (2009). Other studies analyze chat reference transcripts for instruction elements provided by chat staff, and perceived desire on the part of patrons for receiving instructions in chat (Matteson, 2011).

The body of literature on use of the READ Scale for reference assessment focuses primarily on the use of the Scale to assess in-person reference services (Gerlich & Berard 2010; Gerlich & Whatley 2009; Vassady, Archer, & Ackermann 2015). Although the READ Scale has been

widely used in evaluating questions at physical reference desks, it has been employed less frequently in the realm of chat reference (Ward & Phetteplace 2012). The team decided to employ the READ Scale in order to explore its viability in the chat environment, and in order to provide some continuity between the previous 2011 study of the in-person service points.

## **Methods**

A team of six librarians volunteered to perform a content analysis on one academic quarter's worth of chat transcripts (Fall quarter 2014, September 24 – December 12, 2014). The team chose not to use a sampling method, but to examine all Fall quarter transcripts in order to gain a better understanding of the scope of the work involved for future assessments, and to understand better patterns across the whole of this typically busy quarter. A total of 3721 transcripts were coded, although some of these (n=226) were later removed from the set of results as they were duplicates (see below for additional details).

Coding of the transcripts was completed in two rounds. The first round involved categories including patron type, student type (where applicable), campus location, and whether the question was answered by a librarian from our institution or elsewhere. The two-stage approach enabled team to have an initial pass through the transcripts in order to resolve any logistical issues arising in terms of working with the transcripts, and also enabled us to familiarize ourselves with the content of the transcripts in order to best determine the most suitable codes for the second round. More detailed discussion of coding decisions are available in Appendix A.

For round two, we added categories for transaction complexity using a slightly modified READ Scale (5 levels instead of 6), as well as for subject area and to capture any reference to e-books. Four subject area descriptors were used: Health, Engineering, Business, and Law. Past anecdotal observations from chat coordinators and staff pointed to these four subjects as particularly challenging, as they often require (or are perceived to require) more specialized knowledge on the part of the librarian. The team was interested to learn about the volume of questions in these subject areas, if any relationship existed between the subject areas and level of effort expended by chat staff to answer these questions, and whether any staffing changes needed to be made in order to better serve patrons from these areas.

While the team discussed using a variety of additional codes in addition to the READ Scale and subject, ultimately we decided to only add a code for whether a question was e-book related. Questions of patron desire for electronic versus print format, as well as ease of access and use of electronic materials, are of significant interest to selectors, technical services, and collections staff. Focusing on the single category of e-books enabled the team to test our process for applying question type codes. Also, those staffing the chat reference service perceived a high frequency of questions from patrons confused about their ability to access e-books found through our online catalogs, particularly those owned by non-UW Libraries. We hope to provide a sub-set of transcripts to those directly involved in the selection and discoverability of our e-books, in order to inform their work with patrons.

All six raters “normed” the READ scale using a sample set of transcripts. Members of the team independently scored the sample transcripts and then met to discuss any differences between scores. Discrepancies between scores were resolved in order to reach a shared understanding of how to apply the READ scale and other codes. During this norming process, we determined that we would only code the live transcript rather than any follow up responses offered by a librarian. The team decided that this approach provided a more realistic picture of how a typical chat evolves within the temporal space of an active session, providing insight as to the degree of service offered at the user's point of need.

Librarians entered their data into Excel spreadsheets, which were then merged at the end of the project. Dedoose, a web tool used by the Libraries in the past for qualitative analysis, was tested for the purpose of coding and analyzing transcript data, but the barriers for regularly working with the transcripts in this tool were significant. The transcripts downloaded from the OCLC QuestionPoint system came to us in a single, extremely large xml file. Using Dedoose effectively would require us to enter each transcript into the system individually. Given our goal of project sustainability in the future, the team determined that Dedoose was not a viable solution for the ongoing assessment of transcripts. While a qualitative data analysis system such as Dedoose or NVivo might be useful for a single large scale research project, the team decided that a basic tool such as Excel requires less investment of time and staff training and is currently adequate for our needs.

There was an overlap of approximately 5% in the chat transcript assignments (that is, certain transcripts were coded by more than one person). While this project was not designed to be a

full-scale research study, this approach provided a greater opportunity to achieve inter-rater reliability and robust READ scale scores. Overlapping scores were examined and discrepancies resolved by the assessment coordinator. In addition, another small set of transcripts (an additional 5%) was “spot checked” – raters were encouraged to indicate any uncertainty about their READ scale scores (e.g., if they were uncertain about whether to code a chat as a 3 or a 4). These were examined and questions resolved by the assessment coordinator. Once all outstanding questions and coding discrepancies had been resolved, the coordinator removed duplicates and “lost calls” – those transcripts that contained no effort on the part of chat service staff (n= 226). If a patron accidentally disconnected and reconnected to ask the same question (sometimes multiple times), transcripts were retained if there was distinct librarian effort in each case. Results were calculated after duplicates and “lost call” transcripts were removed from the pool.

Team members also explored using the post-chat survey results in order to gain a better understanding of user satisfaction with the service. The survey results did not provide much detailed information, and so were not analyzed for this study. However, this prompted the team to decide that the survey should be revised significantly in light of transcript assessment results in order to gather more meaningful feedback from users.

In addition to data analysis, qualitative feedback from team members was gathered in order to assess the usefulness of the methodology, and record observations from the transcript assessment that might not have been captured through the codes themselves. These observations were used in conjunction with the data to formulate the recommendations in this report.

### Limitations

One limitation of this approach to chat transcript assessment was the lack of patron demographic data captured by our system. Approximately 72% of patrons enter the chat system via the Qwidget located on library web pages and the Libraries search tool. There is a long form available to patrons, where they can enter details about their campus affiliation and patron type (e.g., faculty, graduate student), but most patrons use the Qwidget option, which does not require any of this data. In order to maximize the amount of patron type data available to us, the team used information derived from the long form, where available, and also information gleaned from the transcripts themselves (e.g., where a patron identifies themselves in the

course of a chat as a student in a certain undergraduate class at the Tacoma campus). While we were interested to explore patterns of use among different patron types, we were only able to glean this information via these two approaches in 51% of the transcripts.

The central limitation of our methodology was in the applicability of the READ Scale to the virtual reference environment. The READ Scale may fit an in-person and perhaps slightly older model of reference transaction that relies more heavily on print resources. The origins of the Scale for use at in-person service points is reflected in many of the categories: for example, many questions at READ Scale 1 (“Answers that require the least amount of effort”) are related to physical use of spaces and services, such as directional inquiries or assistance using printers. These types of questions are rarely asked via our chat reference service, which rendered this code automatically less applicable to many of our online interactions. It is also the case that the examples of instruction given in the READ Scale tend to be focused more on mechanical use of search tools (e.g., “basic instruction in searching the online catalog”) rather than more conceptual aspects of instruction (e.g., how do I know if this is a peer-reviewed journal article?).

In addition, the transcripts often do not provide a complete picture of what the librarian is doing (for example, their search process, the number of sources they may have used for a search, etc.), which made accurate coding of effort level difficult. A chat librarian can undertake many steps in determining out how to get a patron to the right information, without describing this in the transcript: there were numerous examples of a librarian saying to a patron "hold on, I'm looking for a few things" without providing any detail about what they were doing. This is a known limitation of transcript analysis more generally, but was made more challenging by the use of a coding scheme that is based on the amount of effort taken to answer a question. See, for example, Question ID 9858143 (located in Appendix B), where the librarian clearly consulted multiple sources in order to find an item (which would point to a READ Scale score of 4), but that work was not visible to those coding the transcript. Unless the librarian indicated all the sources they used to answer the question, it may not have been coded as a 4 on the Scale.

It is also the case that there was often a misalignment in the online environment between the complexity of the question asked, the amount of effort expended, and the amount of instruction provided to the patron. Members of the team noted that the instruction evident in the transcript may not always correlate to the complexity of the question. One patron may need a great deal

of instruction on what we would consider a relatively simple concept (explaining to a patron less familiar with web searching how to request a book found in the catalog, for instance), so the question might require the librarian to provide a significant amount of guidance (which in turn resulted in a higher score on the READ Scale). Conversely, there were also multiple instances where complex questions were referred to a subject specialist quickly and therefore demonstrated minimal librarian effort (see, for example, Question ID 9965590, in Appendix B). Despite these particular limitations, however, such cases did highlight the potential need for better training for chat staff about how much instruction to provide and when to refer.

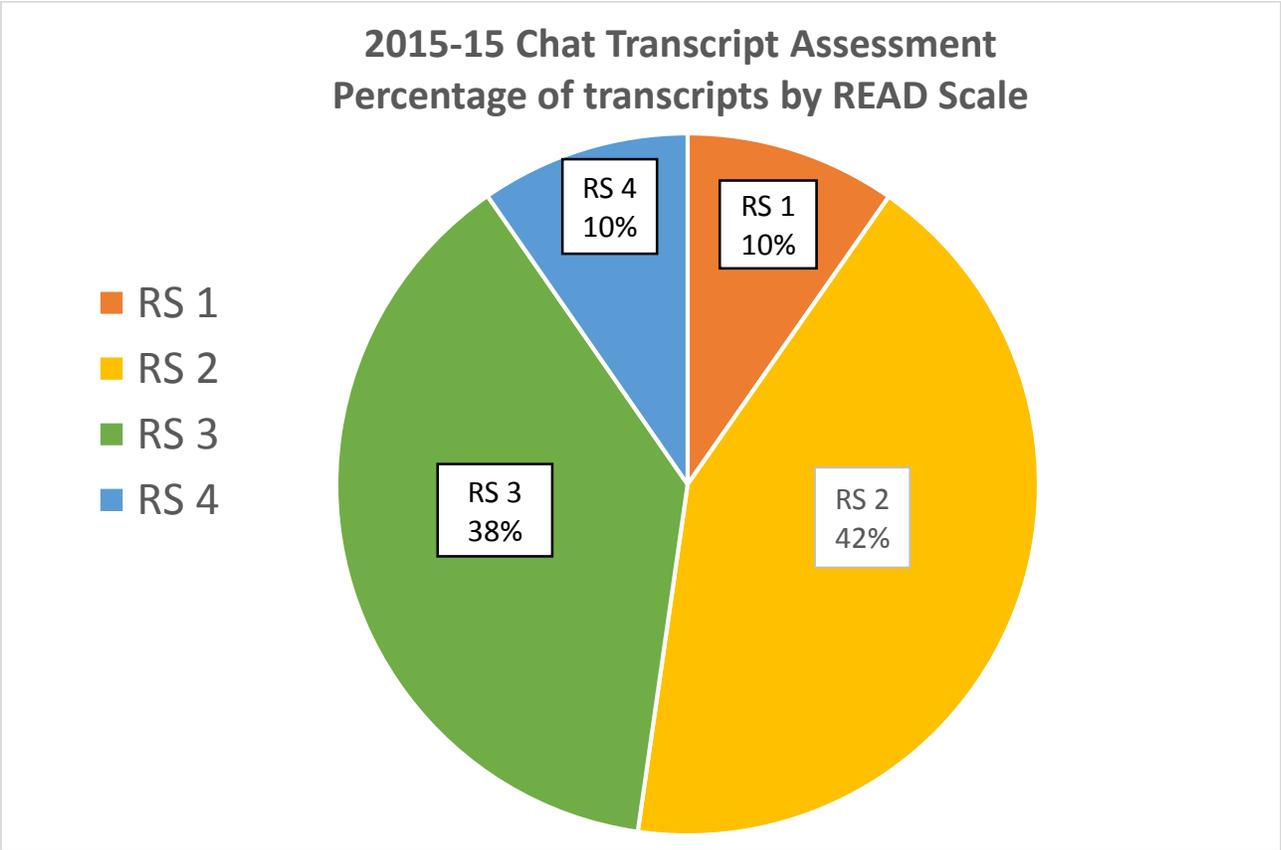
Separating questions of service quality from those of effort level complicated the coding process. While we do have general guidelines for chat customer service behavior at the UW Libraries, participation in the service is voluntary, and interpretations of what constitutes successful reference interaction can vary among individuals staffing the service. In addition, the UW Libraries share this service with consortium librarians who may have a different set of expectations and institutional cultures around customer service and instruction within reference work. What one librarian may consider an acceptable level of effort expended on a question, another may consider inadequate. Therefore, a staff member's own interpretation of "good service" will impact how much effort is taken, and in turn how their transcript is coded.

In order to address these limitations, future assessment efforts of virtual reference transactions might wish to code chats based on the level of difficulty of the patron question, instead of on librarian effort. Alternatively, a model could be used that coded for difficulty level of question and librarian effort, which would help us to see where there is a lack of alignment between the type of question asked and the depth of response. Overall, while the READ Scale was a useful trial model for gaining a deeper level of understanding of the nature of the chat questions we receive, the team determined that future assessment efforts would benefit from using an instrument more sensitive to the complexities and limitations of the online environment.

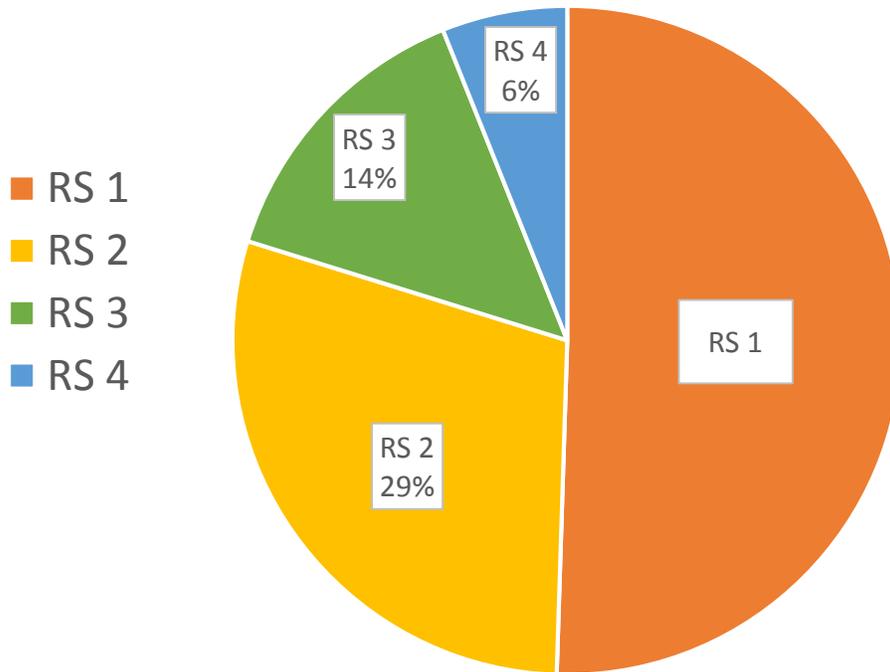
## **Results & Observations**

80% of the questions analyzed were considered by the coders to be READ Scale 2 or 3 level of complexity. Of the remaining 20%, an equal proportion were coded at READ Scale 1 and 4. This finding differed sharply from data collected by our institution during a 2011 assessment of physical reference desks, where 50% of questions were coded at READ Scale 1 (although the results from the two projects are not necessarily comparable, due to differences in the scope of

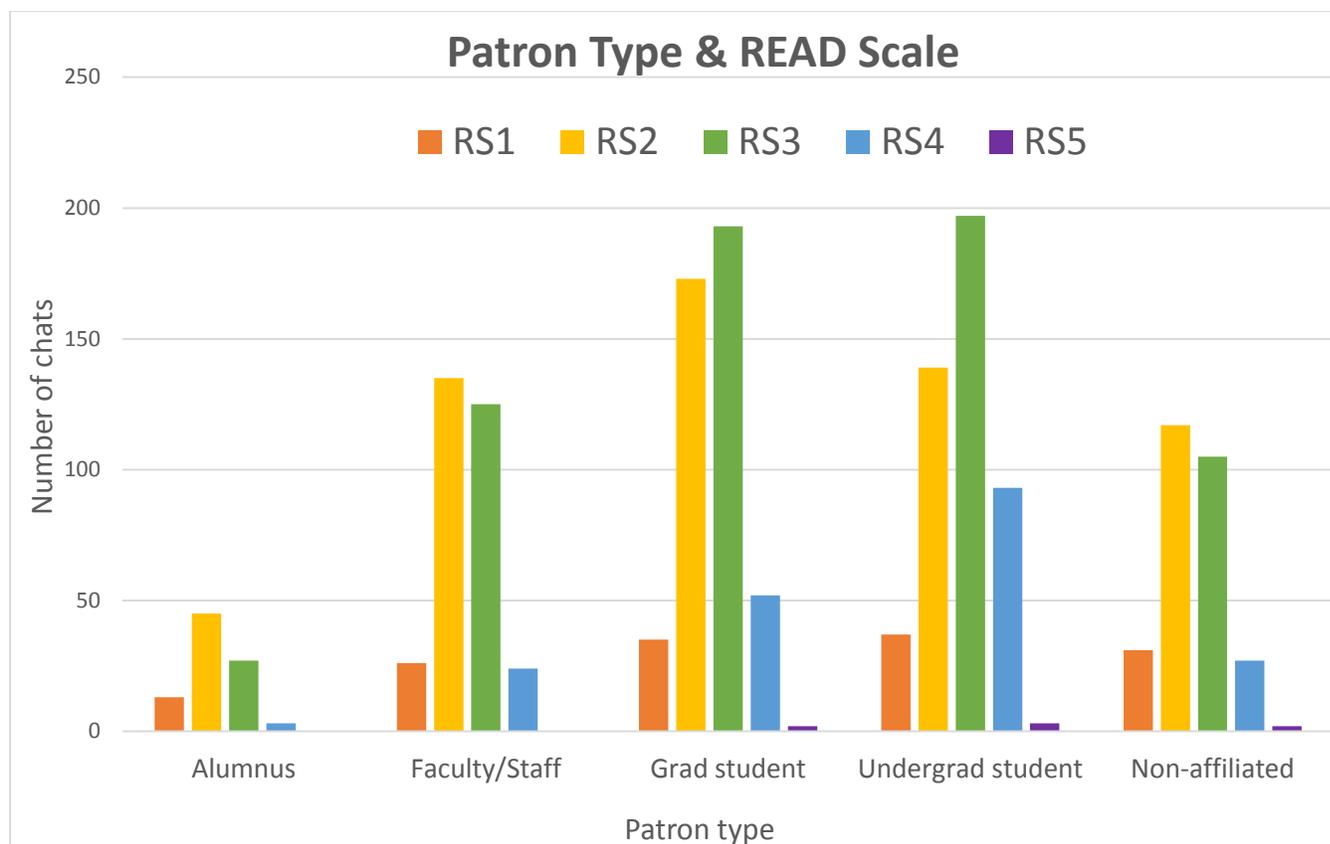
the two studies). It also contradicts the assumption held by some library staff that chat reference is not used as often by patrons for complex or “serious” research questions.



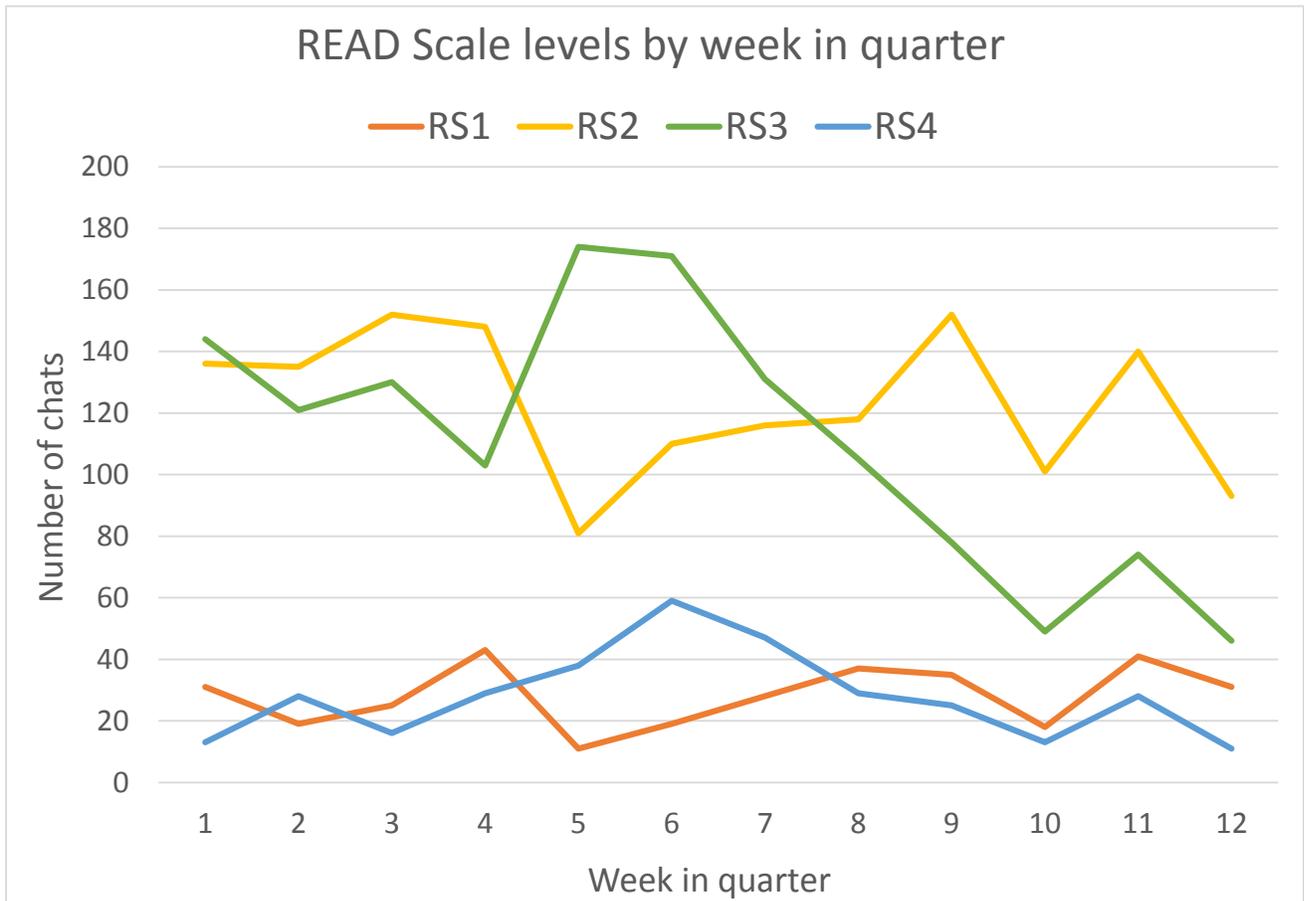
### Fall 2011 Reference Desk Assessment: Percentage of questions by READ Scale



Our analysis also found that of all the patron types, undergraduate students were the ones most likely to ask reference questions requiring a complex response, identified by the coders as READ Scale 3 or above. This patron type was followed in turn by graduate students, and then faculty/staff. These findings suggest that amount of effort expended by staff in a chat reference session drops in accordance with the patron's educational experience—a conclusion that may impact how our institution trains new chat staff, as well seasonal staffing patterns.



The peak period for chat reference for Fall quarter 2014 was in week six (n=359, or 10% of the total number of questions); the lowest volume of chats were in weeks ten and twelve of the quarter, (n=183 and 182, respectively). The majority of the questions in week six were at the higher end of the READ Scale (3 and 4), with the highest number of level 4 questions occurring during this week (n=59). The peak for READ Scale 3 questions was in week 5 (n=174), followed closely by week 6 (n=171). These findings are in line with the timing of many assignments at mid-quarter, and suggest that increased staffing during weeks five and six may be useful to provide the best service to patrons with more complex questions.



In terms of the four subject areas coded in this project (Health, Engineering, Business, and Law), only a small number of the transcripts (n=563, or 16%) were assigned to one of these four categories. Health-related subject questions made up the majority of these questions, with 58% of the questions where one of the above subject areas was identified. Subject area and e-book categories, however, proved to be of limited use in revealing new, specific information about our chat service. This was in part due to the fact that it was often difficult to determine the exact subject area from the transcript. More nuanced guidelines might help address this problem in future chat assessment projects, but it is also recommended that the team shares the subject-specific information with selected groups (such as Health Science Library staff) for further analysis.

All project team members noted in their observations that there were wide variations in the level of instruction and service provided to users: it was often the case that chat staff provided links to

resources users needed to find, without providing instructions on how to complete a task independently alongside these links. As noted above in the limitations section, there were also a number of instances where questions were referred quickly with little librarian effort. Team members observed that complex questions were handled by UW Libraries staff and QP backups in two ways: some staff immediately sent a complex question for follow up and others who attempted to answer the question. This was the case even in UW Libraries staff, pointing to the need for additional training. (Please see Appendix B for examples of chat transcripts that illustrate the difference in librarian effort expended on a reference question.)

Team members also noted that questions were often closed and not followed up on, even though there might have been additional assistance needed. There were many questions raised by these observations: is there a perception on the part of chat staff (possibly a misperception?) that users expect quick answers without instruction? Is it that multiple questions are being handled simultaneously and there is not enough time to provide instruction, or to engage at all with a seemingly complex question? There are many variables that impact how much instruction, and how many consulted resources, a librarian incorporates into the chat: the staff member's familiarity with the subject and search strategies, level of comfort instructing/communicating via chat/IM or "thinking aloud", personal beliefs regarding service expectations, and working with one versus multiple patrons.

Finally, team members noted that patrons often seemed to expect to be chatting to UW staff, even at night or on weekends. Patrons sometimes also seemed to expect a certain type of librarian to be available based on their chat entry point (e.g., a person coming in via HSL chat form might immediately begin the chat assuming that they are working with a medical librarian). These questions (in particular around the links between level of effort evidenced in a transcript and a single librarian handling multiple chats simultaneously) should not only be explored in further assessments, but used to inform continuing training of chat staff. Likewise, the relationship between chat transcripts and the post-chat surveys submitted by patrons of the service should be reexamined in future assessments. Our present study was unable to draw meaningful conclusions in this area, due to the unexpected discovery that different versions of the post-chat survey had been deployed to service users throughout the period of data collection.

## **Recommendations:**

## Improving the Chat Service

The findings revealed in the 2014-2015 Chat Reference Transcript Analysis Project have sparked a number of ideas that could enhance our current staff training, contribute to the development of best practices, and ensure staffing is adequate during peak times:

- Utilize transcripts of chat transactions in which staff used instructional techniques as jumping-off points for training. By providing these concrete examples of what robust instruction looks like in the virtual reference environment, we could demonstrate to our chat staff that it is possible to teach information literacy skills, as well as provide timely answers to their questions. This training could perhaps be done in collaboration with the Libraries Teaching & Learning Group:
  - Ask chat staff to read transcripts in which approaches to helping the patron varied widely, and ask them to discuss what they believe the outcomes would be for each patron. Use this exercise to help staff recognize teachable moments, and provide examples of how these openings might be addressed in the chat conversation. A similar exercise, called “Teachable Instants: Taking the Opportunity or Taking a Pass” was conducted during the 2010 Libraries Teaching & Learning Group’s training series, provided by Megan Oakleaf, and was well received.
  - Utilize the Libraries Patron Personas, in combination with sample chat transcripts, to guide a discussion around how patron needs and expectations might vary.
  - Encourage staff to “think aloud” and incorporate instructional best practices into chat, using strategies outlined in the literature.
- Provide additional guidance to chat staff regarding when a question should be referred to a subject specialist, and how they might provide some guidance to the patron before doing so.
- Create clearer standards for chat staff regarding when a question that has been partially answered needs additional follow-up, and when it can be closed.
- Encourage chat staff to refer patrons to the [recently developed Libraries FAQ page](#), which provides instruction on common patron questions, and reinforces strategies for problem solving in the future. Explore ways in which the FAQ page could be automatically “pushed” to patrons following each chat.
- Ensure that staffing levels during weeks five and six of the quarter, when question frequency and complexity is higher, are adequate for providing the best service.

- Clarify best practices/ service guidelines for UW chat staff, and explore ways in which we could communicate our (ideal) institutional approach of incorporating instruction into chat sessions with our the QP backup staff.

### Sharing Project Data

- Provide data and sample transcripts from this project with Libraries staff working in the four subject areas coded (Business, Health Sciences, Law and Engineering). Encourage them to explore this data in more depth, in order to inform their own service improvement and chat staffing levels.
- Communicate the results of this project more widely with Libraries staff, either at an In Service and/or in the Weekly Online News. Results regarding level of question complexity and patron type could contradict some staff assumptions that chat reference is not used as often by patrons for complex or “serious” research questions.

### Refining and Sustaining Chat Assessment

While this project has given us much to work with in terms of service improvement, it has also highlighted new assessment questions that we have about patron expectations, demographics, and the impact that simultaneously helping multiple chat patrons has on service quality. It has also allowed us to better understand the strengths and limitations of the READ coding process on our chat reference service, and to begin developing ideas for a more efficient, sustainable and precise assessment strategy.

This project represented a significant undertaking in terms of staff time and effort, and is not a sustainable long-term approach to ongoing assessment. While we will not be taking this approach every year, it has enabled us to see how we can scale this for ongoing assessment (e.g., using a small sample of transcripts assessed by a core team of librarians and graduate students).

Based on what we have learned over the course of this project, our recommendations are to:

- Explore further the correlation between the number of simultaneous chats handled by a single librarian and the level of instruction provided. Raters observed in many of transcripts that chat staff missed opportunities for “teachable moments”, and this often seemed to be the case when the chatter was working with multiple chat patrons at the same time. Having a better understanding of why staff are missing opportunities to

facilitate instruction could have important implications for training and staffing of the service.

- Revise the survey that is pushed to patrons at the end of each chat session, to capture actionable information beyond basic satisfaction.
  - Determine what is most important to UW chat staff from the current survey. What would they like to know, in terms of an evaluation of their own service, and eliminate things that aren't important to their own work.
  - Keep patron survey short - eliminate questions such as "did this technology work for you?" that aren't important and take up space.
  - Use the survey to get more demographic data, thus helping to fill in the gaps in demographic data that we currently have.
  - Use the survey to ask questions that align with questions in other ongoing Libraries assessments (about patron satisfaction, wishes, and needs). For instance, consider asking the "Which of the following services would be useful to your work" question that we ask in the Triennial Survey in the chat patron survey, but perhaps with options that fit an online environment (skype consultations, online tutorials etc).
- Use the revised survey or another assessment method to learn more about patron expectations for chat service. What do patrons expect (help from subject expert, 24/7 UW librarians, etc.)? Do they want a quick answer, or more guidance on how to do things themselves? Were their expectations met? Gaining a better understanding of patron expectations could help us market the service more accurately, and provide guidance for staff in addressing these during the chat conversation.
- Address the issue of lack of demographic information:
  - Encourage UW chat staffers to follow the example of QP backups non-UW chat staff and ask demographic questions of about patrons (are you a student? What is your primary campus? Are you an online student?) in order to provide better service. Possibly make this a best practice for UW chat staff and/or develop scripts.
- Assess transcripts on a regular basis, in line with Libraries User Query Sampling Weeks (which occur five times a year). Currently, staff categorize each of their chat transactions as "reference" or "non-reference" during User Query Sampling Weeks. Expanding on this basic coding would provide a predictable schedule for coordinated, targeted assessment of both in-person and virtual services.

- Use the data from this project to determine a more suitable coding scheme for virtual reference transactions in future assessments. Our experience using the READ Scale in this project has led us to conclude that the scale, while useful, requires adaptation for the online chat environment.
  - As part of this new coding scheme, consider which questions are of interest across the Libraries and code accordingly. We might wish to have a core set of codes that remain stable over time (for the purposes of collecting longitudinal data), combined with some variable coding that enables us to address specific, targeted questions on a yearly or quarterly basis.
  - Track results from ongoing assessments and resulting improvements (post on assessment page) so this activity is visible.
  - Consider coding chats based on the level of difficulty of the patron question, in addition to, or instead of, librarian effort.

In conclusion, this project has highlighted the potential that chat reference transcripts hold in helping us understand our users, improve library services, and communicate value. Chat transcripts provide a raw record of what patrons are seeking from us, as well as of our approaches to helping them in this “anytime, anyplace” environment. Over the course of this project, we have identified ways in which we could improve the efficiency and accuracy of our transcript coding and assessment process. Equally important, this project has highlighted that improved chat assessment data would be more powerful if viewed in combination with data taken from other service points in the libraries. Taking a more holistic approach to assessment across virtual and in-person services would allow us to get a fuller picture of our patrons, their needs, and satisfaction with the Libraries. Our patrons do not use the chat service in isolation. Their questions are diverse and touch on all aspects of our organization, from collections to facilities, discovery services to online teaching and learning, and more. Sharing assessment data captured from chat transcripts and other service points can help deepen our understanding of our community, in turn helping us improve a range of programs, including online and in-person instruction, and liaison services. Finally, ongoing assessment of chat has the potential to support the case for the institutional value of the service in supporting teaching, learning, and research.

## References

Arnold, J., & Kaske, N. (2005). Evaluating the Quality of a Chat Service. *Portal: Libraries and the Academy*, 5(2), 177-193.

Bravender, P., Lyon, C., & Molaro, A. (2011). Should chat reference be staffed by librarians? An assessment of chat reference at an academic library using LibStats. *Internet Reference Services Quarterly*, 16(3), 111-127.

Gerlich, Bella Karr, & Berard, G. Lynn. (2010). Testing the Viability of the READ Scale (Reference Effort Assessment Data)©: Qualitative Statistics for Academic Reference Services. *College & Research Libraries*, 71(2), 116-137.

Gerlich, B. K., & Whatley, E. (2009). Using the READ scale for staffing Strategies: The Georgia College and State University experience. *Library Leadership and Management*, 23(1), 26-30.

Lasda Bergman, E. M., & Holden, I. I. (2010). User satisfaction with electronic reference: a systematic review. *Reference Services Review*, 38(3), 493-509.

Marsteller, M., & Mizzy, D. (2003). Exploring the Synchronous Digital Reference Interaction for Query Types, Question Negotiation, and Patron Response. *Internet Reference Services Quarterly*, 8(1-2), 149-165.

Matteson, Miriam L., Salamon, Jennifer, & Brewster, Lindy. (2011). A systematic review of research on live chat service.(Report). *Reference & User Services Quarterly*, 51(2), 172.

Meert, D., & Given, L. (2009). Measuring Quality in Chat Reference Consortia: A Comparative Analysis of Responses to Users' Queries. *College & Research Libraries*, 70(1), 71.

Vassady, L., Archer, A., & Ackermann, E. (2015). READ-ing Our Way to Success: Using the READ Scale to Successfully Train Reference Student Assistants in the Referral Model. *Journal of Library Administration*, 55(7), 535-548.

Ward, D., & Phetteplace, E. (2012). Staffing by design: A methodology for staffing reference. *Public services quarterly*, 8(3), 193-207.

Youngbar, A. C. (2012). Questions by Keystroke: An Analysis of Chat Transcripts at Albert S. Cook Library. *Library Student Journal*, 7.

## **Appendix A: Codes**

### **Demographic information**

Librarian: handled by UW Librarian or QP backup

- UW
- QP Backup
  - Includes all non-UW librarians (those librarians identified as QP backup or, for example, "Librarian at Pacific Lutheran University")

Patron type:

- Alumnus
- Non-affiliated
- Faculty/staff (includes clinical)
- Student
- Unknown

Student type:

- Undergraduate
- Graduate
- Online
- Unknown
- Not applicable

Campus (the campus with which a patron is affiliated):

- Seattle
- Bothell (use for both Bothell and Cascadia)
- Tacoma
- Unknown
- Not applicable

**Effort level/subject**

- READ Scale (<http://readscale.org/>)
- E-books (includes technical issues, requesting an ebook if it is in print; can't access because it is Alliance/Summit)
- Subjects:
  - Business

- Health (medicine, health, public health, social work, nursing)
- Engineering (patents, proceedings, standards)
- Law (as subject, not visiting to Law library)

### **Coding descriptions/decisions**

- Only coded for chat itself, not follow up (this was considered a separate transaction).
- Did not look up patron information (e.g. status, campus), but coded based on what we can see in the transcript or patron-supplied information, e.g., if a patron self-identifies as faculty, or if they say “I’m looking for this book for my class” and gives a class code that is clearly Seattle, Tacoma, etc. If coders were not sure, “Unknown” was used. It would have been time-consuming to systematically look up patron information.
- All HSL locations coded as Seattle campus.
- Grouped faculty/staff together
- “Lost call” or discarded questions: in some cases, the question is not technically a “lost call”, but the patron disconnects quickly and there is no real exchange with a librarian, although there is sometimes later follow-up by a UW Librarian.
- We did not code formally for certain elements (Engl131, PCE, history day), but put these in the notes. These can help to inform QP descriptive codes.
- In cases where multiple questions were asked that required varying levels of staff effort (e.g., a basic policy question that might be coded at a 2, combined with a question that required more effort on the part of the librarian that might be coded at a 3 or 4), the entire question was assigned a single score, representing the score for the most difficult aspect of the question.

## **Appendix B: Sample Transcripts**

### **Chat Transcript #1**

**Patron:**  
12:12:31 2014/10/27 Chat Transcript: Is there a way to eliminate meta-analysis and medical hypothesis from search engines? I want to just find the actual studies done for my topic and not the meta-analysis that have already been done.

**Librarian:**  
12:12:31 2014/10/27 Note: Patron's screen name: XXXXX

**Librarian:**  
12:12:47 2014/10/27 Librarian XXXXX has joined the session.

**Librarian:**  
12:13:04 2014/10/27 HI

**Librarian:**  
12:13:46 2014/10/27 I think you'd better talk to the Health Sciences Library about how to do this. 206-543-3390

**Librarian:**  
12:17:26 2014/10/27 Bye for now.

**Librarian:**  
12:17:27 2014/10/27 Librarian ended chat session.

**Librarian:**  
12:17:31 2014/10/27 Note:  
Set Resolution: Answered

## Chat Transcript #2

**Patron:**  
13:56:15 2014/09/25 Chat Transcript: Evaluate widely used early childhood curriculums

**Librarian:**  
13:56:15 2014/09/25 Note: Patron's screen name: XXXXX

**Librarian:**  
13:56:33 2014/09/25 Librarian XXXXX has joined the session.

**Librarian:**  
13:56:52 2014/09/25 Hello, my name is XXXXX. How can I help you?

**Patron:**  
13:57:30 2014/09/25 Initially - i would like to search for what are the widely used curriculums in Early Childhood education

**Patron:**  
13:57:42 2014/09/25 and there effects on outcomes

**Librarian:**  
13:58:07 2014/09/25 Okay, have you done any research on this yet?

**Patron:**  
13:58:36 2014/09/25 not entirely sure how to navigate this search engine - i did background reading on google

**Librarian:**  
13:59:15 2014/09/25 Great, well let's start at the library homepage. Are you there?  
<http://www.lib.washington.edu/>

**Librarian:**  
13:59:37 2014/09/25 Click on Advanced Search under the main search box. This way you will be able to enter more than 1 search term.

**Patron:**  
13:59:52 2014/09/25 yes

**Librarian:**  
14:00:05 2014/09/25 Let's use keywords to help you find some resources that match your research interest.

**Librarian:**  
14:00:48 2014/09/25 Keywords are key concepts and terms that are the essence of your question. In this case I might start with "early childhood education" and "outcomes".

**Patron:**  
14:01:27 2014/09/25 i am interested in particular which curriculums used in early childhood education and their tie to outcomes

**Patron:**  
14:01:37 2014/09/25 widely used..

**Librarian:**  
14:02:44 2014/09/25 Let's try searching this at the same time. When you click Advanced Search, scroll down the page and you'll see 3 boxes to enter terms. Enter them here. You can define them as a type of search term. I would start with keyword as the type of term. Yes let's also use "curriculum" as a term. Leave out outcomes to start.

**Patron:**  
14:03:27 2014/09/25 ok

**Librarian:**  
14:04:11 2014/09/25 This looks like a pretty good search to start with. You can now refine the results along the lefthand bar.

**Patron:**  
14:04:41 2014/09/25 under advanced options?

**Librarian:**  
14:06:09 2014/09/25 You don't even need to go to advanced options. From your results page just scroll to the bottom. There are Topics that you can select to limit the results. But before that even, I think in the first 10 results many seem like they might be good matches. Why don't you browse those titles and see. Did you get this page of results?

**Librarian:**  
14:06:12 2014/09/25 <http://uwashington.worldcat.org.offcampus.lib.washington.edu/search?q=k%3Aearly+childhood+education+k%3Acurriculum&fq=&start=1&dblist=>

**Patron:**  
14:07:31 2014/09/25 right, however, I would like to evaluate widely used curriculums

**Patron:**  
14:07:47 2014/09/25 in early childhood education and their tie with outcomes

**Patron:**  
14:08:02 2014/09/25 perse, creative curriculum or Highscope

**Librarian:**  
14:09:53 2014/09/25 You can refine your keywords in your search if you have particular curricula of interest. I did this here for highscope. These were the results.

**Librarian:**  
14:09:54 2014/09/25 [http://uwashington.worldcat.org.offcampus.lib.washington.edu/search?q=k%3Aearly+childhood+education+k%3Ahigh+scope+curriculum&qt=results\\_page&dblist=](http://uwashington.worldcat.org.offcampus.lib.washington.edu/search?q=k%3Aearly+childhood+education+k%3Ahigh+scope+curriculum&qt=results_page&dblist=)

**Librarian:**  
14:10:07 2014/09/25 The more specific terms you give the more directed your search will be.

**Librarian:**  
14:11:01 2014/09/25 a comprehensive review of which are the widely used curriculums and why?

**Librarian:**  
14:11:07 2014/09/25 in the USA

**Librarian:**  
14:12:29 2014/09/25 The first link is an encyclopedia, but if you open the record and scroll to the bottom you can see that highscope is one of the terms tagged for this item. Presumably that curriculum is covered in the text. It looks like it covers quite a few techniques, so I think it would be comprehensive.

**Librarian:**  
14:14:12 2014/09/25 The second correlates the relationship between curricula type and early childhood education, speaking to the outcomes part of your question.

**Patron:**  
14:14:52 2014/09/25 which source are you viewing

**Librarian:**  
14:15:28 2014/09/25 To focus on the US I would add that as a search term. I was looking at Curriculum Models in Early Childhood Education:

**Librarian:**  
14:15:29 2014/09/25 <http://uwashington.worldcat.org.offcampus.lib.washington.edu/oclc/44265803>

**Patron:**  
14:17:01 2014/09/25 true, this is a great source. however, it does not cover some of the new up to date curriculums that are widely used

**Librarian:**  
14:17:40 2014/09/25 You can limit to dates on the left side too. You could only look at things extremely recent this way.

**Librarian:**  
14:18:04 2014/09/25 Try clicking on 2014.

**Patron:**  
14:18:10 2014/09/25 ok - I will play around a bit and see ways in which to narrow my search

**Patron:**  
14:18:11 2014/09/25 thank you

**Librarian:**  
14:19:26 2014/09/25 You can use the same techniques in education-related databases too if you want to go more in depth. Here is a research guide to find more places to search if you like:

**Librarian:**  
14:19:46 2014/09/25 <http://guides.lib.washington.edu.offcampus.lib.washington.edu/education>

**Librarian:**  
14:19:56 2014/09/25 Is there anything else today?

**Patron:**  
14:20:26 2014/09/25 Patron ended chat session.