Lab 1 - AskMissy Product Description

Matthew Stephens

Old Dominion University

CS 411W: Professional Workforce Development I

Professor Janet Brunelle

December 13, 2021

Version 1

Table of Contents

List of Figures

List of Tables

1.   Introduction

Students face two large problems in the domain of reading. The first problem is that they lack easy access to relevant resources in relation to assignments they receive in school, notably in grades 9-12. The second problem is a general downward trend in the interest in reading for students of all ages. Whether the students come from historically underprivileged backgrounds, or they are exploring concepts that are underrepresented in traditional literature, a weak beginning in literature leads to a lack of interest, trust, and reliance on reading skills. Skills that many professionals find necessary for their day to day lives after their school careers.

In a 2007 study, researchers from the National Endowment of the Arts found that American students are reading less (Fuglei, 2019). From 1992 to 2002 students showed a 5% decline in extracurricular reading. This decline grew more severe for adults in this time frame, increasing to 7%. Additionally, a study by the Pew Research Center found that adults who have lower levels of education-or are members of economically disadvantaged demographics-are less likely to read than adults with higher levels of education or economic advantage (Gelles-Watnick & Perrin, 2021). This shows that as people age, their tendency to read whether for scholastic or personal drive diminishes.

One way in which the harmful impact of a lower reading skill level is shown is in the Virginia Standards of Learning assessments, or SOLs. These assessments are the state's method of determining the learning and retention of the students in elementary, middle and high school for subjects such as math, reading, science and social studies. The curriculum of public schools is built around the SOL, and many students are taught that the SOL exam is the most important part of the school year as the culmination of all the lessons they have learned in that class. The

Department of Education uses SOL pass rates, among other statistics, to determine a school's accreditation. When students do poorly on SOLs, schools are at risk of losing accreditation. This can cost schools valuable resources necessary to maintain or improve the student's ability to learn.

When the COVID-19 pandemic shut down the in-person learning for schools in the 2019-2020 school year, students were thrust into an alien world of distance learning. They, and their parents, suddenly had to take on a much greater burden in terms of school organization and scheduling. A greater focus was put on independent study to keep up with the previous in-person learning standards. Simultaneously, teachers had to learn a completely new mode of teaching, disseminating lessons when they couldn't verify that students were paying attention. Even when students returned to school for the 2020-2021 school year, the impact was clear by the drop in the SOL exam pass rate (*Sol Test Pass Rates & other results*). Math SOL pass rates dropped an average of 30% between the 2018-2019 and 2020-2021 school year. For reading SOLs in grades 3-12 the drop was lower at around 11%, though this isn't cause for relief. Math skills can be buffered through memorization and practice of the same set of skills. Reading SOLs measure the test taker's comprehension, which requires the intake of a wide range of books. Reading comprehension is another practice-based skill, but requires different learning strategies than math, which can be difficult when learning by oneself.

The drop in SOL pass rates clarifies the need for all parties involved in primary education to restructure the learning process for students whose performance is non-optimized because of a lack of materials tailored to their learning preferences. Without this change, students lack

exposure to the critical reading comprehension skills that will benefit them after their scholastic careers.

To address both the symptoms and core problems, AskMissy is a software application designed for students, teachers and librarians in grades six through twelve to help users find the resources necessary to bridge the gap between conventional and digital learning for a more personalized experience. Machine learning algorithms will be implemented to personalize the experience for users and predict the types of resources that the user searches for.

2. Product Description

AskMissy is a web-based software application designed to help users in grades 9-12 find resources such as books to both aid them in their school careers and to invigorate their interest in reading. AskMissy will be more efficient in its recommendations by implementing machine learning algorithms to learn from the search results of the users over time and incorporate the learned preferences into future results. AskMissy databases will connect directly to the library databases of school libraries to allow users to plan out their resource acquisition through a number of options. Students will be able to search for books in the library based on several criteria such as book title, author, isbn number, or subject matter. Teachers will be able to upload the subject matter of their SOL curriculum and receive recommendations on resources that correspond. They will then be able to directly recommend those resources to their students who are organized in groups based on their classes. Librarians will maintain their libraries inventory and receive feedback from students and teachers for specific resources, as well as see statistics to reflect the popularity of those resources.

2.1 Key Product Features and Capabilities

AskMissy will be a web-based application that connects to the existing systems of each school that participates. The software would be centered around each specific school that has registered to use it. Schools will be able to authenticate users either manually by way of an administrator or automatically connecting the registered user's information to the school database of students, teachers or librarians. Once registered a user may access AskMissy anywhere so long as they have a device with internet access.

The AskMissy database will retrieve input from two types of databases: internal and external. The internal database will be both the school's database of students, teachers and librarians, as well as the libraries inventory. The external database will consist of data taken from APIs that connect to Amazon Web Services in order to find information on each book in the libraries database.

Machine learning algorithms will be implemented to both fill in the specific data for each resource in a school library as well as find resources that match a user's interest and expand their horizons. Once AskMissy is connected to a school library database, the algorithm will go through the current inventory and connect those items based on their title, author and isbn number to the external database. Language recognition will parse through the external database's resources and draw primarily from the synopsis and reviews to find out what the subject matter of the resource is. This populates the tags for that resource which is stored in the AskMissy database. When a user searches for a resource, they will be given the chance to look for specific subject matter by entering a list of words. The algorithm will then search for resources whose tags most closely match the search criteria and return that resource to the user.

Users will be able to supply feedback to AskMissy based both on their overall impression of the resource and how well a given tag fits that resource. The feedback will take the form of a 'swipe left or right' rating, as well as the chance to submit a short (less than 200 characters) written feedback. Once this feedback is received, AskMissy's search algorithm will take this into consideration when supplying similar searches in the future, learning what resources best match any given search criteria.

A user who never registers or wants to access the software before logging in will have limited access to the AskMissy software; a basic search function of the database with limited ability to input search criteria. Users who register will have full access to the user-facing features of AskMissy, namely the machine learning search algorithms and group interactions. There are three different user types, but a single user may only ever be registered as a single type. The authentication process will ensure that a user is not placed in the wrong category, but if an error does occur then the administrator will have access to correct this.

A student user may create a profile for themselves and input information about themselves to a limited extent. This limit will be based around their school career, such as their grade or school club participation. Students will have access to the AskMissy search algorithm and can view their previous search results as well as the specific searches themselves. The specific amount of data saved for each student to view, and retrieve will be decided at a later date. They will be entered as a member of their specific class group as established by their teacher. They will have the option to view recommendations from other student users in their class and share their own recommendations. There will be limited ability to share details of their feedback that take the form of short (maximum 100 characters) reviews which other students

may observe. Students may submit longer forms of feedback that are sent directly to higher tier users they are connected to, namely their teacher or one of the librarians in their school group. This will take the form of resource requests for items like books that are missing from their school library.

Teachers will also have access to the AskMissy search algorithm. They will be able to create class groups and connect their students into the group. Once a teacher has a group that teacher can then forward resources, they have found to their students either individually or in sets of students. These recommendations will not affect the AskMissy search algorithm of any students until they submit feedback on that resource. Teachers may also submit resource requests to librarians. Teachers will have the option to create lesson plans in AskMissy based on the subject matter of each program. Teachers can then use AskMissy to provide them a list of resources that correspond to each lesson. Teachers will also be able to review the feedback that their students submit and run personalized reports on what students are searching and responding to.

Librarians will function similarly to teachers, though their groups will be school groups, and they will be connecting all registered teachers to the same group. They will also maintain the inventory of their school library. Librarians may receive the resource requests of the students and teachers. They may also directly contact administrators for bug support.

Administrators will have access to all the features that the other user types have, as well as having access to the AskMissy data servers to ensure data integrity. They will do so by maintaining the 3rd party APIs and their web-scraping functionality. They will have access to all

user profiles and user metadata. They will be responsible for authorizing librarian users and connecting the AskMissy program to the school's systems.

2.2 Major Components (Hardware / Software)

AskMissy will require the user to have access to some device capable of accessing the internet. On the backend, there will be one or more servers dedicated to each of the following: The frontend website, the main backend database (holding user information and school library data), the web scraping process, the machine learning algorithm, and the external database (up to date data from web scraping). The full view of the components necessary is listed in Figure 1.

Figure 1

Major Functional Components Diagram
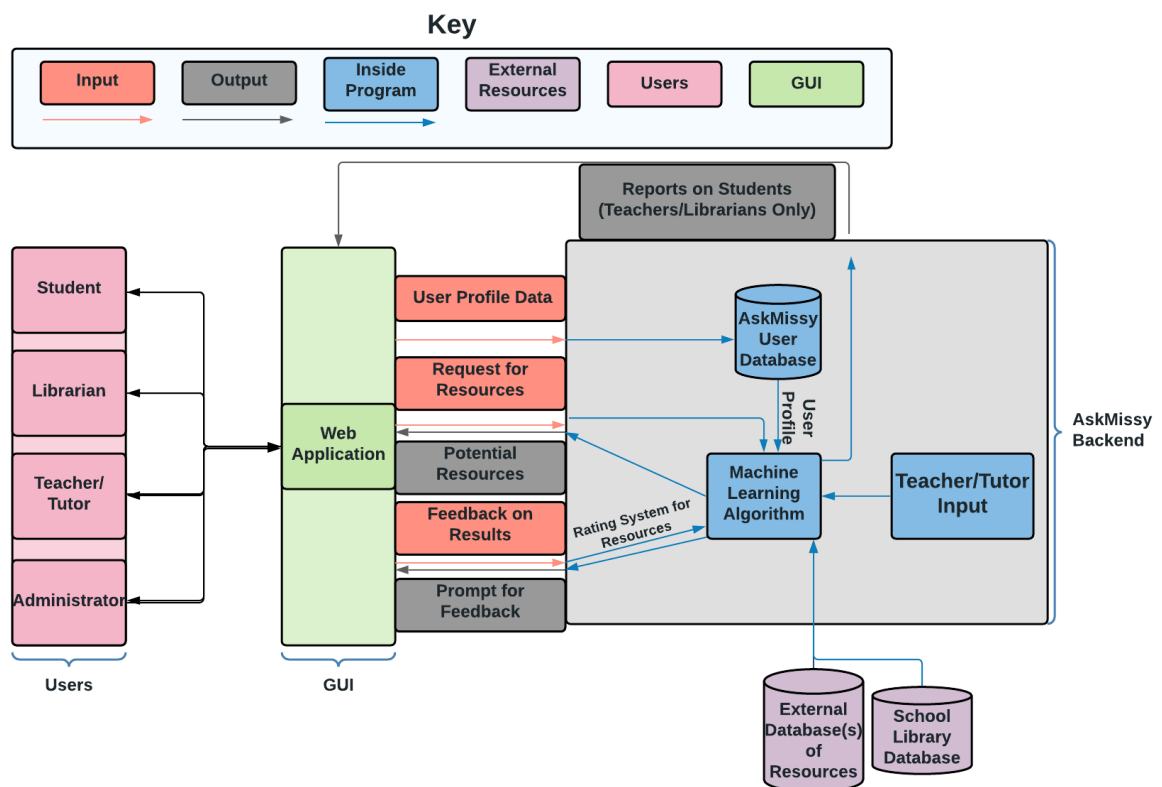
AskMissy will be centered around the Python programming language and the application server will be decided at a later date, though most likely be chosen from Django, Gunicorn or Python Paste. The data that AskMissy collects from web scraping will be stored in mySQL file format.

The frontend of the website will be written in a combination of HTML5, CSS3 and Javascript. The backend will be python. The team's programming IDE will be Microsoft Visual Studio Code.

3.   Identification of Case Study

AskMissy's main intended users are students, teachers and librarians in grades 9-12 whose school curriculum is centered around the Virginia SOL. The schools in which AskMissy is connected are some of the primary stakeholders, along with the parents of the students as secondary stakeholders.

The main goal of AskMissy is to supply a user with a tool that allows that user to obtain a list of resources tailored to their intended subject matter and interests. This goes beyond a customary search engine tool to look for a book on that user's interests; instead, AskMissy will learn how the user reacts to certain subjects and the quality of resources, then tailor future searches to zero-in on one type of resource, exclude others that offend their sensibilities, or open the user to a new world of subject matter in a way that encourages expanding horizons.

Other parties that might benefit from AskMissy are publishers of resources who get overlooked in favor of more popular businesses. Because AskMissy is made to connect users to lesser-known resources, it follows that lesser-known publishers would glean a net positive from this as well.
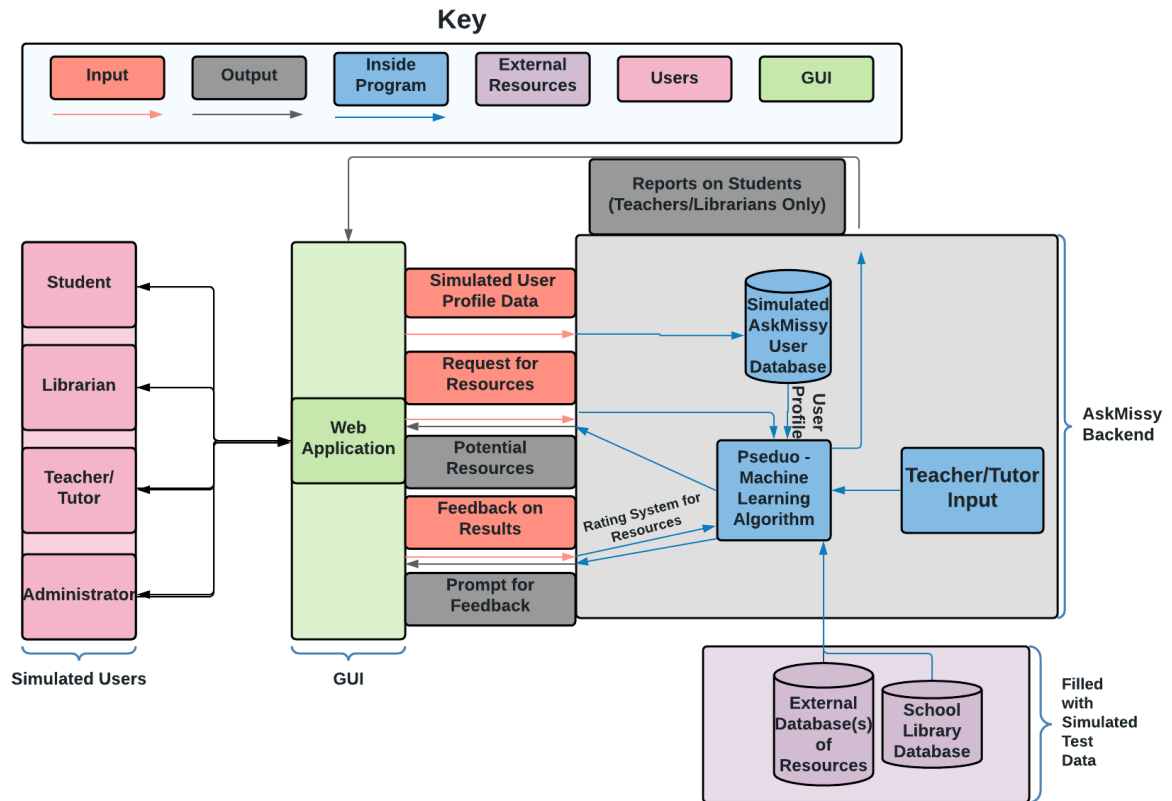
4.   AskMissy Prototype Description

The prototype contains most of AskMissy's features, either fully or partially implemented. External sources like Amazon RDS will not be implemented in the prototype due to a lack of knowledge of a workable execution. Security encryption will not be implemented in the prototype, but login credentials will. All user roles will have full functionality and testability in both the real-world product and the prototype, though the prototype will have a 6th user type in the Tester. The Tester will be responsible for verifying the accuracy of the algorithm and the accessibility of the UI.

4.1 Prototype Architecture (Hardware / Software)

The web interface will be available on desktop and laptop devices with internet connection. For users who choose to use their phones to participate in AskMissy, functionality will be prioritized over the visual aspect. The prototype and its database will be hosted on a physical server provided by ODU's Computer Science department in a containerized environment using Docker. This environment also contains the MySQL server. There would be more separation between these resources in the real-world product to prevent a single failure point. Figure 2 shows a visual representation about the prototype implementation.

Figure 2

Prototype Major Functional Components Diagram

Key

Input	Output	Inside Program	External Resources	Users	GUI

Reports on Students (Teachers/Librarians Only)

Student
Librarian
Teacher/Tutor
Administrator

Simulated Users

Web Application

GUI

Simulated User Profile Data
Request for Resources
Potential Resources
Feedback on Results
Prompt for Feedback

Simulated AskMissy User Database

Pseduo - Machine Learning Algorithm

Teacher/Tutor Input

User Profile

Rating System for Resources

AskMissy Backend

External Database(s) of Resources
School Library Database

Filled with Simulated Test Data

The AskMissy machine learning algorithm will be written in the Python language using the PyCharm Integrated Development Environment, or IDE. Our database will be a MySQL Server. Testing will be done using unittest. GitHub will be utilized to keep track of which version of software is in use. User accounts will be simulated to properly test the prototype and will be stored in the database.

4.2 Prototype Features and Capabilities

The prototype for AskMissy will keep most of the functionality as in the real-world product, but with a reduced scale. The application will utilize the AskMissy Search to personalize the recommended book to the individual user, so long as the user is logged in and is

not a guest. The tag creation algorithm will be implemented in the prototype to fill the test data

set with workable tags for the AskMissy Search to use but will not be implemented in the RWP.

Communication between the user types will be modeled to showcase the types of communication

in the RWP. Users will be organized into different classes and schools, with separated library

catalogs to simulate access to different data sets. In the prototype the simulated users will have

simple profiles, only having enough details to test the AskMissy Search. The prototype will not

have real connection to a school's database of students, teachers, and librarians- instead, it will

provide a simulation of administrators authenticating the registration of the school side users.

Users will be able to view their previous searches and statistics that AskMissy gathers about

themselves.

[This space is intentionally left blank]

Table 1

AskMissy Real World Product vs. Prototype deliverables

Category	Feature	RWP	Prototype	Reasoning
Data Retrieval	Metadata Report	Full	Partial	Limited test data
	Basic Search	Full	Full	
	AskMissy Search	Full	Full	
Data Management - Live Product	Machine Learning	Full	Partial	Limited test data
	Source Tag Creation	Full	Full	
	Source Tag Management	Full	Full	
	Lesson Plans	Full	Partial	Limited test data

	Internal Database Manipulation	Full	Full	
	External Database Manipulation	Full	Full	
Data Management - Development	Source Tag Development	None	Full	Use to develop default tags
	Machine Learning Training	None	Full	Use to develop algorithm defaults
	Simulated Data	None	Full	Use to fill database for testing
	User testing reports	None	Full	Use to develop user interface
Security	Login/Authentication	Full	Partial	Limited test data
	Data Encryption, moving	Full	None	Best practices will be used
	Data Encryption, resting	Full	None	Best practices will be used
Account Management	User Profile	Full	Partial	Limited test data
	Feedback	Full	Full	
	Group Management	Full	Partial	Limited test data
	Login/registration	Full	Full	
UI	Group Interaction	Full	Partial	Limited test data
	Bug Report	Full	Partial	Limited test data
	Basic Search	Full	Full	
	AskMissy Search	Full	Full	
	Communication	Full	Partial	Limited test data
	Personal Data Report	Full	Partial	Limited test data

4.3 Prototype Development Challenges

Prototyping the application will involve several challenges. While most will be planned

for with mitigation techniques, some will require handling as they arise. The AskMissy

development team has little prior experience with Machine Learning algorithms and as such

much of the functionality will need more rigorous testing and secondary input from the testing

team to ensure quality. To mitigate this, the data being used will be as simple as possible while covering as many different scenarios as possible that would arise in the RWP. Similarly, connecting all the different databases presents as the largest hurdle. Again, sufficient testing will mitigate many of the issues that arise. Staggering feature development so that the backend is complete before the frontend should provide more opportunity for the frontend development to go smoothly, connecting to an existing database without having to change the implementation functions more than needed.

[This space intentionally left blank]

5.   Glossary

Administrator: A user who is responsible for managing a majority of AskMissy's working data.

Agile: A set of frameworks and practices where solutions evolve through collaboration between self-organizing cross-functional teams.

AskMissy: A software application that will help users find more relevant resources.

AskMissy Search: The search function in AskMissy that implements Machine Learning algorithms to find books for the user based on their profile.

Application Programming Interface (API): A software intermediary that allows two applications to talk to each other.

Data Retention: The continued storage of an organization's data for compliance or business reasons.

Database: An organized collection of structured information, data, typically stored in a computer system.

Economically Disadvantaged: A student eligible for Free/Reduced Meals who receives Temporary Assistance for Needy Families (TANF) or is eligible for Medicaid.

Exact Match Search: A search for a single specific type of resource.

File Server: A device that controls access to separately stored files.

Guest: A user who is not a student, teacher, librarian, or administrator, who has limited access to the AskMissy program.

Integrated Development Environment (IDE): a software development environment that provides comprehensive facilities to computer programmers for software development, usually consisting of a source code editor, build automation tools and a debugger.

Librarian: A user responsible for managing the library's inventory/database, communicating with teachers and students.

MySQL: MySQL is an open-source relational database management system.

Personal Learning: An educational approach that aims to customize learning for each user's strengths, needs, skills, and interests.

Prototype: A model of the product in question designed as a proof of concept to allow the developers to showcase their intended design.

Real World Product (RWP): The fully realized version of a product that is developed after a prototype for use in the open market.

Student: A user studying at a K-12 education institution in need of reliable resources.

Tags: Labels attributed to various subjects, such as subject topics or genres, used to organize the books found in the AskMissy database.

Teacher: A user who helps K-12 students acquire knowledge. They are responsible for making plans and managing students' groups/communication.

Temporary Assistance for Needy Families (TANF): A program that provides eligible families with a monthly cash payment to meet their basic needs.

Tester: A user responsible for designing and conducting testing suites for usability testing.

Unittest: A unit testing framework in the python programming language.

Unit Test: A software testing method by which individual units of source code are tested to determine whether they are fit to use.

User: An individual using the AskMissy Interface.

Web Scraping: The process of extracting content and data from a website.

Web Server: A computer program that distributes web pages as they are requisitioned.

Windows: A series of operating systems developed by Microsoft.

6.   References

Albanese, A. (2021, May 6). *Report urges library leaders to address decline in public library usage stats*. PublishersWeekly. Retrieved December 15, 2021, from

https://www.publishersweekly.com/pw/by-topic/industry-news/libraries/article/86256-rep

ort-urges-library-leaders-to-address-decline-in-public-library-usage.html

"Allreaders.com features detailed book and movie reviews from many different genres of books!,"

Detailed Book review summaries. Retrieved September 22, 2021, from

http://allreaders.com/.

"Amazon.com: Kindle eBooks: Kindle Store: Nonfiction, Literature & Fiction, Foreign

Languages, Business & Money & More," *Amazon*, Retrieved October 01, 2021, from

https://www.amazon.com/Kindle-eBooks/

Coates, T. (2021). *Freckle report 2021: Digital or diverse?- the future for public libraries*. Tim Coates

Books.

Crain, C., & Waldman, K. (2018, June 14). *Why we don't read, revisited*. The New Yorker.

Retrieved September 22, 2021, from

https://www.newyorker.com/culture/cultural-comment/why-we-dont-read-revisited.

Ellard, C. (2020). *Covid-19 impact on Sol Assessments, verified credits, and graduation

requirements for Spring/Summer 2020*. Newport News Public Schools. Retrieved

December 14, 2021, from http://sbo.nn.k12.va.us/sol/covidimpact.html

Fuglei, M. (2019, July 22). *Why students who read for pleasure are stronger academically*.

ResilientEducator. Retrieved October 1, 2021, from

https://resilienteducator.com/classroom-resources/how-reading-for-pleasure-helps-student

s-develop-academically/.

Garcia, E., Weiss, E., & Welshans, I. (2020, October 7). *What teaching is like during the

pandemic-and a reminder that listening to teachers is critical to solving the challenges the*

coronavirus has brought to public education. Economic Policy Institute. Retrieved October 7,

2021, from

https://www.epi.org/blog/what-teaching-is-like-during-the-pandemic-and-a-reminder-that

-listening-to-teachers-is-critical-to-solving-the-challenges-the-coronavirus-has-brought-t

o -public-education/.

GeeksforGeeks. (2020, August 29). *5 most-recommended career fields in computer science*.

GeeksforGeeks. Retrieved January 25, 2022, from

https://www.geeksforgeeks.org/5-most-recommended-career-fields-in-computer-science/.

Gelles-Watnick, R., & Perrin, A. (2021, September 21). *Who doesn't read books in America?* Pew

Research Center. Retrieved September 22, 2021, from

https://www.pewresearch.org/fact-tank/2021/09/21/who-doesnt-read-books-in-america/.

Gioia, D. (n.d.). Reading at Risk. Washington D.C., Virginia ; National Endowment for the Arts.

https://www.arts.gov/sites/default/files/RaRExec_0.pdf

Henry. (2021, May 27). *The importance of Reading* Retrieved September 22, 2021, from

https://www.uopeople.edu/blog/why-its-important-to-read/

Ingraham, C. (2018, June 29). *Leisure reading in the U.S. is at an all-time low*. The Washington

Post. Retrieved September 22, 2021, from

https://www.washingtonpost.com/news/wonk/wp/2018/06/29/leisure-reading-in-the-u-s-

is-at-an-all-time-low/

"Meet your next favorite book," *Goodreads*. Retrieved October 01, 2021, from

https://www.goodreads.com/

Frequently asked questions about Sol testing. Virginia Department of Education. (2015, October).

Retrieved October 7, 2021, from

https://www.doe.virginia.gov/testing/sol_faq.pdf.

Rosalina, E. N. (2018, November 30). *The correlation between self-esteem and student's reading

comprehension.* English Language Teaching Educational Journal. Retrieved October 1, 2021,

from

https://eric.ed.gov/?id=EJ1283078

What should I read next? book recommendations from readers like you. What Should I Read Next?

(n.d.). Retrieved January 25, 2022, from https://www.whatshouldireadnext.com/

Virginia Department of Education. (n.d.). *Sol Test Pass Rates & other results*. Retrieved October 7,

2021, from

https://www.doe.virginia.gov/statistics_reports/sol-pass-rates/index.shtml.