Evaluation Proposal

Date: 2022-05-31

Author: Cole Brookson

To better understand how the R Manual can be optimized to assist student learning and success, it would be useful to have a plan in place for evaluating the usefulness of the program, as well as areas for improvement. Since EEB225 will be the first course that the R Manual is trialed in, that class could be a great case study for rolling out an evaluation procedure.

Here outlined are a number of options to consider for discussion surrounding an evaluation procedure:

What is Our Measure of Success?

I think to develop a good evaluation procedure, we first need a really solid set of measures for the success of this R Manual. Do we care about improvement in courses? That could be determined by mark improvements or lab improvements. Or alternatively do we only care that they FEEL like they know R? That would be defined very differently as the performance in the course might not change but their comfort in R might be higher.

Additionally, as far as I understand, this manual is not supposed to service just EEB225 but eventually a suite of EEB course offerings. It would be useful to define some success metrics for the Fall semester (the trial of the manual in 225), and then separately, some more long-term success metrics. That will give us more concrete goals to reach for, and some specifics to design our thoughts and work around.

Evaluation Options

There are two main types of evaluation that I think could be used here - a quantitative assessment and and a sociologically-styled interview-based assessment.

Quantitative Evaluation

I think there are three main ways we can solicit quantitative feedback. The first is relatively easy to implement, whereas the second two are quite a bit more difficult but still would be very helpful.

1 - Survey-based Evaluation (Scaled Questions)

We can implement a series of evaluations that the students participate in, either on a volunteer basis or for participation grades in the course (EEB225). These surveys could be a set of multiple choice questions that have categorical answers (i.e. "What has the most challenging component of navigating the R Manual been?" with possible answers such as "a) using the web interface, b) finding the

content that answers my question, c) . . . ") and a number of scaled questions (i.e. "The R Manual was easy to use. Answer of on a scale of 'Strongly Agree' to 'Strongly Disagree' ".).

These evaluations could be given at strategic points in the course such as before the course begins, before the first midterm, after quizzes, before the final, etc. After each evaluation, Cole would analyze the data and create some sort of report either in the form of a presentation, video, or write-up, to share with the students, summarizing what the survey found, and outlining the plans for improving the manual or otherwise addressing the things that were identified as weaknesses. These data would then be used to guide changes/additions to the manual, as well as included in a semester-long analysis of the usefulness of the R Manual approach.

2 - A/B Testing of the Manual in EEB225

The use of A/B testing to evaluate this manual would be incredibly useful and provide a lot of insights, but would also require Jacqueline to be willing to implement the A/B test in her class. However, if that option is not on the table, it would still be worth thinking about A/B testing the inclusion of R Manual content in BIO120/220.

Either way, it would be incredibly useful to be able to use lab sections as a way to trial the manual as being successful or not in improving student performance with regard to specific learning outcomes. This could be done on some subset of assignments such as just lab quizzes, or on the course outcomes as a whole (it would likely be useful to do both but we would have to think carefully about how to parse these results).

The specifics of how this type of evaluation would work are not worth even thinking about until we as a group decide if this is feasible/worth it.

3 - Web-based Assessment of Manual Usage

If the manual is in fact deployed in a web-based fashion, there are any number of statistics that can be pulled from web pages such as but not limited to:

- page-area usage rates - gives information on where visitors spent the most time on a page
- video watch rates
- scroll times - tells you how long a user took to scroll to the bottom of the page or where they scrolled to before leaving the page
- download counts - if skeleton scripts are embedded in the page, then it can be determined how many users actually downloaded the file
- click counters - if there are "hidden" examples, we can know how many users actually look at the answers of the hidden examples

And many more metrics. This method would be more useful in determining how many students *were* using the resource, and out of those students, what parts

did they use/not use.

Some things to consider about web-based assessment:

- The possible implementation of this is completely dependent on the domain we implement this on. Best case scenario would be that the manual is deployed (at least at first) on a UofT-hosted platform, that would allow **both** a log-in under a UTORid and a "visitor" log in. That would allow us to track (anonymously) who visited the site from the course, and cross-reference their usage with their responses to the survey-based evaluation.
  – This would almost certainly require ethics approval, but would likely be fine so long as all identifying data were wiped
- It would be likely that implementing this for ALL suggested components would be time consuming and difficult, so we would likely have to strategically choose components of the site and also metrics that we wanted to collect, that would best illustrate for us the usage of the resource in ways that can help us improve or further develop it.
- While this is the most time consuming and difficult version of the quantitative assessments, it would be likely the most helpful in terms of improving the manual for other courses/additional usage

Qualitative Evaluation

The potentially complimentary method of assessment beyond just quantitatively assessing the success of the manual is a sociological set of interviews with students in the course. This would most certainly require ethics approval. I know relatively little about interview methods, but likely there would be a set of interviews at the end of the term, and possibly at the beginning and throughout at key points as well. The goal would be to ask them targeted questions to assess the success of some pre-determined set of learning outcomes, and then go through and code the interview to get the information we want out.

This is obviously a significant undertaking, and would require commitment through the fall semester and into the winter semester to be even worth considering. However, with specific regard to the usefulness of the manual in addressing the learning outcomes of EEB225 in particular, and perhaps assessing students' preparedness for upper level classes with R requirements, this would likely be the most useful type of assessment, and one that could be translated into a study or at least an internal report to the department.

Some Further Considerations

1. If we intend to do assessment of this nature on any meaningful scale, we should likely prioritize planning and implementing this as needed. It is likely we will need ethics approval for anything that verges on a "study", and it might be worth considering if we want to turn this into a study of some kind.

2. It is unlikely that this work would be completed by the middle of the Fall 2022 term, so some conversation should be had regarding how this type of work would be passed along to the next person to take over the manual (or continued by me)
3. Most of the evaluation metrics above (especially the interview coding) would be relatively time consuming work - it would be worth considering what resources are available to help that process along (i.e. 299/397/498 students, work-study students, etc)