STA 493 Data Science Education: From Tutorials to Assessment

Spring 2023 Reflection

Evan Dragich   supervised by Dr. Mine Çetinkaya-Rundel, PhD.

The following document serves as the final artifact representing my work in STA493 this Spring 2023 semester. I met with Dr. Çetinkaya-Rundel weekly on Thursdays from 2:45-3:30pm this semester, as well as with the research team from 2:00-3:00pm on alternating Tuesdays.

Thesis

The work this semester culminated in my Graduation with Distinction project: Exploring Data Science Education: From Tutorials to Assessment. The link contains my project in an easily-browsable, HTML format. The PDF version copy can be downloaded at that link, and is also stored in department archives.
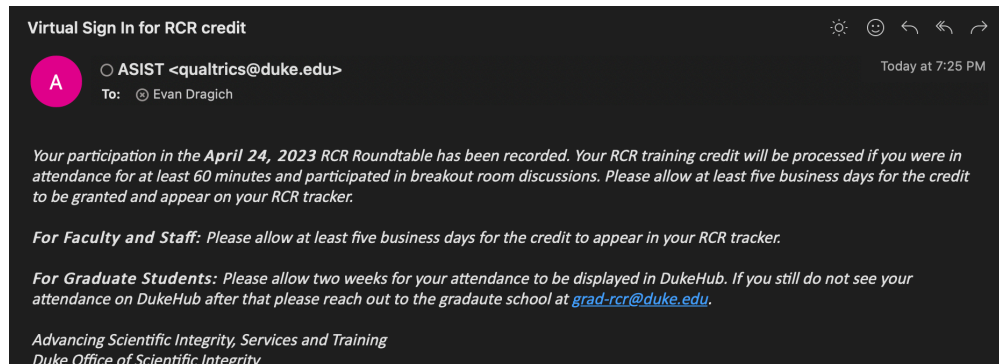
Responsible Conduct in Research Requirement

The final portion of my required work as a student in a research independent study is attendance at, and reflection upon, one Duke-sponsored Responsible Conduct in Research (RCR) event during the enrolled semester. I attended the 90 minute "Registered Reports" research town hall held on April 24th, 2023.

The RCR requirement also entails completing the three required RCR Basics courses and one additional elective. I had previously completed those for another research experience; see the attached pdf at the end of this document.

The following summarizes the roundtable discussion's topics and connections to my work:

Currently, academia is facing some issues with regards to the longstanding journal publication process. Particularly in social sciences such as psychology, issues such as "p-hacking" (exploiting statistical tests for the sole goal of achieving statistical significance), "HARKing" (hypothesizing after the results are known) are becoming more and more rampant as researchers feel funding- and tenure-driven pressure to publish at all costs. To attempt to alleviate these issues, many journals have begun accepting *registered reports* in addition to traditional manuscripts.

In a nutshell, researchers are able to design a study, identify particular hypotheses, write an introduction and analysis plan, and submit this *registered report* to journals before any data is collected. After this initial submission, journal reviewers can offer a desk rejection, acceptance with minor revisions (to the study plan), or acceptance with major revisions, as per usual. Once accepted, the researcher carries out data collection, writes up the results, and submits again. Here, the study is *published on principle*–its validity and relevance has already been confirmed by the first round of review–before undergoing a second round of review, this time focusing on whether the approved study design was actually followed.

The benefits to adopting this paradigm, the presenters explained, were many-fold. There is a reduce in publication bias, as not only significant findings would be published and researchers would feel confident examining a wider variety of research questions. Fewer resources would be wasted as insignificant findings, previously unpublished nor shared, would discourage other researchers from conducting the same study. Finally, addressing researchers' concern about "publishing or perishing," adopting the registered report paradigm allows researchers to have more CV entries, as well as a more clear timeline of their research efforts over time.

However, significantly changing such a longstanding publication process is not without pushback. Several researchers who were participating in the roundtable alongside me shared that some of their colleagues had attempted to submit registered reports, but were dissuaded by the time and effort required, and that thier registered reports had usually resulted in higher rates of rejection compared to traditional papers. As well, several researchers in fields such as physics and political science shared that they don't think registered reports will make a large impact on the status quo in their fields. In response to this hesitation, one of the roundtable leaders shared an intriguing idea: letting younger researchers in those fields complete a registered report in full, once, as a learning exercise. After that initial exposure, she explained, senior mentors can acknowledge that sometimes such a process may be infeasible in all cases in that field. However, given the close mentor-mentee relationship and this means of transfering of knowledge being key to academia, such an exercise can have a strong effect on shaping the careers of younger researchers.

Regarding my independent study project, in which I and a team of researchers have developed a concept inventory-style assessment for introductory data science, I can see clear benefits to

making use of a registered report as we continue piloting our work. Much of the work to date has been exploratory and qualitative, conducting think-aloud interviews and iterating over questions based on group discussions. However, we are transitioning to large-scale student pilots, from which we hope to begin exploring potential *subscales* for the assessment (i.e. in addition to a student's overall score, additional *subscores* could be calculated based on their responses to a set of several items clustered around one topic, e.g. data visualization).

Before doing so, I believe that submitting a registered report of the work collected so far and any *a priori* ideas for such subscales would be valuable. While some subscales may end up being the result of exploratory analysis at this stage, I believe detailing hypotheses on specific potential subscales before collecting data will corroborate the strength and validity of the assessment that we have already observed in qualitative interviews. As well, one of the largest strengths of our efforts is novelty; such concept inventories exist for statistics and computer science, but none yet for data science. By publishing our work as early as possible, we attenuate the risk of being "scooped," or watching someone else develop and publish such a concept inventory before we have finalized our results.

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)

COMPLETION REPORT - PART 1 OF 2
COURSEWORK REQUIREMENTS*

* NOTE: Scores on this <u>Requirements Report</u> reflect quiz completions at the time all requirements for the course were met. See list below for details.
See separate Transcript Report for more recent quiz scores, including those on optional (supplemental) course elements.

- **Name:** Evan Dragich (ID: 9396378)
- **Institution Affiliation:** Duke RCR (ID: 1806)
- **Institution Email:** evan.dragich@duke.edu

- **Curriculum Group:** Undergraduate Student Responsible Conduct of Research Course
- **Course Learner Group:** Undergraduate Student Responsible Conduct of Research
- **Stage:** Stage 1 - Basic Course
- **Description:**

- **Record ID:** 41049208
- **Completion Date:** 18-Feb-2021
- **Expiration Date:** 18-Feb-2024
- **Minimum Passing:** 80
- **Reported Score*:** 93

REQUIRED AND ELECTIVE MODULES ONLY	DATE COMPLETED	SCORE
Research Misconduct (RCR-Basic) (ID: 16604)	18-Feb-2021	5/5 (100%)
Data Management (RCR-Basic) (ID: 16600)	18-Feb-2021	5/5 (100%)
Duke Translational Omics Research Misconduct Case (ID: 20198)	18-Feb-2021	No Quiz
Research, Ethics, and Society (ID: 15198)	18-Feb-2021	4/5 (80%)

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

Verify at: www.citiprogram.org/verify/?kc0e65d21-1321-4d0a-b8a1-b73b035f0858-41049208

Collaborative Institutional Training Initiative (CITI Program)
Email: support@citiprogram.org
Phone: 888-529-5929
Web: https://www.citiprogram.org

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)
COMPLETION REPORT - PART 2 OF 2
COURSEWORK TRANSCRIPT**

** NOTE: Scores on this <u>Transcript Report</u> reflect the most current quiz completions, including quizzes on optional (supplemental) elements of the course. See list below for details. See separate Requirements Report for the reported scores at the time all requirements for the course were met.

- **Name:** Evan Dragich (ID: 9396378)
- **Institution Affiliation:** Duke RCR (ID: 1806)
- **Institution Email:** evan.dragich@duke.edu

- **Curriculum Group:** Undergraduate Student Responsible Conduct of Research Course
- **Course Learner Group:** Undergraduate Student Responsible Conduct of Research
- **Stage:** Stage 1 - Basic Course
- **Description:**

- **Record ID:** 41049208
- **Report Date:** 11-Jan-2022
- **Current Score****:** 93

REQUIRED, ELECTIVE, AND SUPPLEMENTAL MODULES	MOST RECENT	SCORE
Research, Ethics, and Society (ID: 15198)	18-Feb-2021	4/5 (80%)
Duke Translational Omics Research Misconduct Case (ID: 20198)	18-Feb-2021	No Quiz
Data Management (RCR-Basic) (ID: 16600)	18-Feb-2021	5/5 (100%)
Research Misconduct (RCR-Basic) (ID: 16604)	18-Feb-2021	5/5 (100%)

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

Verify at: www.citiprogram.org/verify/?kc0e65d21-1321-4d0a-b8a1-b73b035f0858-41049208

Collaborative Institutional Training Initiative (CITI Program)
Email: support@citiprogram.org
Phone: 888-529-5929
Web: https://www.citiprogram.org