

## **INSTRUCTIONS:**

- *Use this “TEMPLATE PROTOCOL (HRP-503)” to prepare a study protocol outlining your research plan.*
- *Depending on the nature of your study, some major sections might not be applicable to your research. If so, simply mark as “N/A.” For example, a simple survey might have many sections with “N/A.” For subsections (e.g., 1.x or 8.x) you can mark as “N/A” if you are certain that the subsection is not applicable.*
- *Once the IRB/HRPP approves your submission, your latest approved version of the protocol will be stored in the IRB Protocol Management online system.*
- *If your research plan changes and you need to modify the protocol, please submit an amendment to Protocol Management with the requested modifications. Download your current protocol from Protocol Management and indicate the changes/revisions using the track changes feature in order to make review of the modifications easier to follow. If you are unable to use track changes, please create a new paragraph wherever you need to make a change, and indicate “Amendment: Date” before making a change to any section. Protocol management will store the older versions of your protocol if the IRB or HRPP staff need to compare them during the review.*

## **PROTOCOL TITLE:**

*Include the full protocol title.*

Data Science Workshops for Biomedical and Health Professionals: Persona Identification and Workshop Assessment

## **PROTOCOL NUMBER:**

*Include the number assigned in Protocol Management (verify this has been added before submitting protocol to HRPP).*

20-537

## **PRINCIPAL INVESTIGATOR:**

*Full Name and Degrees:* Anne M. Brown, PhD  
*Department:* Research and Informatics, University Libraries  
*Telephone Number:* 540-231-9231  
*Email Address:* ambrown7@vt.edu

## **FUNDING:**

*Sponsor(s):* [Click here to provide a response.](#)

*Funded already or in the proposal phase?:* No funding is currently used for this work or anticipated

*Is Virginia Tech the primary awardee or the coordinating center of this grant or contract? If not, list the primary institution:* N/A

## **VERSION NUMBER/DATE:**

*Include the version number and date of this protocol. Versions should start at 1.0.*

1.0.1 2020-07-08

## **REVISION HISTORY:**

*Use this table to keep track of changes. Add more rows as needed.*

<b>Revision #</b>	<b>Version Date</b>	<b>Brief Summary of Changes (i.e., the different sections)</b>	<b>Consent Change?</b>
1.0.1	2020-07-08	Revisions for original v1.0 submission	yes
1.0.2	2020-07-14	No protocol changes. Only changes to survey	no
2.0.0	2020-09-24	Phase 2: pre and post workshop assessment	yes

## **Table of Contents**

1.0	Study Summary	4
2.0	Objectives	6
3.0	Background	7
4.0	Study Endpoints	9
5.0	Study Design and Statistical Analysis Plan	10
6.0	Setting	10
7.0	Study Intervention(s)/Investigational Agent(s)	13
8.0	Procedures Involved	14
9.0	Data and Specimen Long Term Storage and Use	18
10.0	Sharing of Results with Subjects	21
11.0	Study Timelines	22
12.0	Inclusion and Exclusion Criteria	22
13.0	Vulnerable Populations	23
14.0	Number of Subjects	24
15.0	Recruitment Methods	25
16.0	Withdrawal of Subjects	27
17.0	Risks to Subjects	27
18.0	Potential Benefits to Subjects	29
19.0	Data Management and Confidentiality	30
20.0	Provisions to Protect the Privacy Interests of Subjects	32
21.0	Provisions to Monitor the Data to Ensure the Safety of Subjects	33
22.0	Compensation for Research Related Injury	34
23.0	Economic Burden to Subjects	34
24.0	Consent Process	35
25.0	Process to Document Consent in Writing	39
26.0	Resources Available	40
27.0	Multi-Site Research	41

## 1.0 Study Summary

<b>Study Title</b>	Data Science Workshops for Biomedical and Health Professionals: Persona Identification and Workshop Assessment
<b>Study Design</b>	<p>Participants will be invited to participate in a series of surveys accompanying a data science workshop series and learning materials catered towards the biomedical sciences. There are 3 phases to the study (i.e., 3 sets of surveys): (1) Pre-workshop student self-assessment survey used to create learner personas for the biomedical field (2) pre- and post-workshop and lesson material surveys to determine success, appropriateness, and usability of the workshops or lesson materials (3) Long-term survey after the workshop to see if there workshop was useful to build on new skills. All surveys will be administered using Qualtrics. Individuals can attend the workshops and go through the lesson materials without consenting into the study and surveys.</p> <p>Phase 1 is a learner personal survey which will provide information and details of the learners that are most likely to attend the workshops. Data from the Phase 1 survey was used to develop workshop and learning materials. The Phase 2 pre/post surveys will assess the workshop and learning materials. Workshops are planned for Fall 2020-Spring 2021. The lesson materials will be posted freely online at <a href="https://ds4biomed.tech/">https://ds4biomed.tech/</a>. This amendment is to include Phase 2 research plans, which contain specific pre, and post, workshop assessment questions that were crafted based on the results of Phase 1. Phase 3 long-term assessment survey will be posted in a future amendment. No personal information will be collected in any of the surveys. For phase 2 and 3 surveys, however, there will be a separate workshop registration form where emails will be collected to coordinate workshop logistics. Anyone can sign up for workshops and the email list of attendees will be used to email Zoom information, etc about workshops and provide phase 2 and 3 study information to all potential attendees. Workshop attendees will be emailed for Phase 3 for the long-term survey before the participant email list is deleted. For individuals who wish to take the learning materials and lean on their own, they are free to do so without signing up for the workshop. We will have the same pre-post survey linked with the learning materials so we can get feedback about how the materials present themselves in a self-learning pace.</p>

20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

	<p>Workshop deliverables will also be collected during Phase 2. Students may opt-into this part of the study by using their self-created identifier as their name and this will be linked to their consent in Phase 2. If there is no identifier match from the workshop deliverables and the results of the Phase 2 pre-workshop survey, the deliverable will not be considered as a part of the research study.</p> <p>This work seeks to develop learner personas to create more effective, engaging, and useful data science workshops and learning materials for individuals in the biomedical/health field, and then measure the success of those workshops and training materials. Findings will be analyzed and written up for a dissertation, presented at conferences relevant to biomed/health education, and published in a paper. This study is a cross-sectional (Phase 1) and longitudinal study (Phase 2 and 3) using surveys and statistical/analytics procedures. There are no interventions beyond the surveys, as workshop and self-learning participation would be voluntary and the focus of the surveys is on effectiveness of teaching and content retention/use.</p>
<b>Primary Objective</b>	What are the hurdles biomedical/medical/health professionals have with regards to using data for decision making and/or research? Does the creation of learner personas before workshop creation aid in better learning outcomes and experience for learners?
<b>Secondary Objective(s)</b>	<ol style="list-style-type: none"> <li>1. Create learner personas of people who want to learn data science skills.</li> <li>2. Create lesson and workshop materials to ease data science skills to people in the medical and biomedical field.</li> </ol>
<b>Study Population</b>	Students, researchers, and practitioners in the medical and biomedical field who will be attending curated workshops or self-learn about data science. Individuals under the age of 18 will not be allowed to consent into the study. But the workshops and materials will be open to the public.
<b>Sample Size</b>	We expect a total of ~100-150 sample subjects to enroll in the phase 1 student self-assessment (persona survey) and take one of seven workshops (pre/post assessments). Two weeks after the learner persona survey is emailed out to appropriate list servers, we will close the survey regardless of sample size number. Listservs for undergrad-faculty individuals in the departments of TBMH, PHS, BIOMED will be used.
<b>Research Intervention(s)/</b>	Phase 1 - Learner persona survey to create workshop materials, Phase 2 - pre/post survey on workshop and learning material effectiveness and learning outcomes,

<b>Investigational Agent(s)</b>	workshop observations and deliverables (code/visualization produced). Phase 2 research interventions will be included as an amendment based on Phase 1 results. Phase 3 research interventions will be included as an amendment based on Phase 2 results.
<b>Study Duration for Individual Participants</b>	Each survey (learner persona, pre/post assessment) will take about 10-15 minutes to take and will be administered electronically via Qualtrics. We expect to complete this work from July 2020-April 2021.
<b>Acronyms and Definitions</b>	<p><b>API:</b> Application programming interface</p> <p><b>FBRI:</b> Fralin Biomedical Research Institute at VTC</p> <p><b>GBCB:</b> The Interdisciplinary PhD Program in Genetics, Bioinformatics, and Computational</p> <p>iTHRIV: The integrated Translational Health Research Institute of Virginia</p> <p>PHS: Population Health Sciences</p> <p><b>TBMH:</b> Translational Biology, Medicine and Health</p> <p><b>VCOM:</b> The Edward Via College of Osteopathic Medicine</p> <p><b>VetMed:</b> Virginia-Maryland College of Veterinary Medicine</p> <p>VTCSOM: Virginia Tech Carilion School of Medicine</p>

## 2.0 Objectives

### 2.1 Describe the purpose, specific aims, or objectives of this study:

The purpose of this study in Phase 1 is to identify learner personas for individuals the medical and biomedical field. This information will be used to create a set of targeted lesson materials to for the following learning objectives:

1. Name the features of a tidy/clean dataset
2. Transform data for analysis
3. Identify when spreadsheets are useful
4. Assess when a task should not be done in a spreadsheet software
5. Break down data processing into smaller individual (and more manageable) steps

6. Construct a plot and table for exploratory data analysis
7. Build a data processing pipeline that can be used in multiple programs
8. Calculate, interpret, and communicate an appropriate statistical analysis of the data

The information from this work will inform the researchers of learner personas of individuals who will take data science workshops or self-learn and create tailored materials to improve workshop participant and self-learning learning outcomes. Learning outcomes will be assessed in the Phase 2 and 3 pre, post, and long-term surveys.

## 2.2 *State the hypotheses to be tested:*

We hypothesize that

1. Creating learner personas will create better educational content because they will be more tailored to the needs of the students.
2. Learning how to program data analysis will allow learners to feel like they can do more with their data.
3. Learning basic data literacy and data science skills can empower health/biomed workers and be more proactive in making more educated decisions.
4. Learning materials with an eye towards tidy data principles will better transition students out of a spreadsheet program into programming.
5. These workshops and learning materials will help medical professionals curate better data for research.
6. Workshops and learning materials will help medical professionals work with data outside of a spreadsheet program.

## 3.0 **Background**

### 3.1 *Summarize the relevant prior research on this topic and gaps in current knowledge within the field of study:*

Evidence-based medicine is now considered the Holy Grail [1]. However, this means clinicians are required to integrate vast amounts of information from numerous sources in for their clinical practice on top of their clinical duties. This poses a challenge to both clinicians and patients [1]. Since, clinical guidelines cater towards the average treatment success rate, there is a considerable amount of uncertainty around what is best for any one patient. Probability and uncertainty is unintuitive, and humans constantly perceive them incorrectly (e.g., consider the Monty Hall problem or the base-rate fallacy) but

clinical decisions made under uncertainty are a commonplace, yet, very few doctors have any formal training in probability or decision theory [1].

At the same time, the vast influx of data allows non-clinical researchers to conduct studies without the domain expertise of a clinician. This causes resources towards statistically significant results with marginal clinical value. [1]. As a medical professional, there are many areas where they can be better integrated into the system by increasing their data science literacy [2]:

1. Improving personalized care of patients by analyzing EMR data towards personalized care plans
2. Work on better data collection techniques to limit losing data due to missing values and outliers
3. Understanding when a predictive algorithm's accuracy applies to a certain situation
4. Serve as domain experts in more research studies
5. Advocate for integration of data mining into decision making
6. Adapting decisions to errors and unpredictable incidents

By increasing data science fluency, individuals can be better prepared and conversant working in multidisciplinary teams [3]. We hope to have more clinicians join as domain experts to improve the quality of clinical care and research.

These skills can be learned through Open Educational Resources (OERs), However, simply providing training with real-world examples is not enough. Learners also need resources and support from the workplace to apply new skills [4]. By empowering clinicians with new data science skills, we hope they will be more competent advocates for their own data needs, and better communicate and integrate with other departments to improve research and clinical outcomes.

[1] Efstratios I. Charitos, Manuel Wilbring, and Hendrik Treede. Data Science Meets the Clinician: Challenges and Future Directions. Thoracic & Cardiovascular Surgeon, 66(1):7–10, January 2018.

[2] Md Saiful Islam, Md Mahmudul Hasan, Xiaoyi Wang, Hayley D. Germack, and Md Noor-EAlam. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. Healthcare (Basel, Switzerland), 6(2), May 2018.

[3] Michelle C. Dunn and Philip E. Bourne. Building the biomedical data science workforce. PLoS Biology, 15(7):1–9, July 2017.

[4] Abrar Alturkistani, Josip Car, Azeem Majeed, David Brindley, Glenn Wells, and Edward Meinert. Determining the Effectiveness of a Massive Open Online Course in Data Science for Health. International Conference on e-Learning, pages 27–34, January 2018.

### *3.2 Describe any relevant preliminary data:*



The surveys are based off research stated in the previous section (3.1).

- 3.3 *Based on the existing literature, provide the scientific or scholarly rationale for and significance of your research and how will it add to existing knowledge:*

The NIH, as a part of their Big Data Knowledge (BD2k) training program from 2013 to 2016, created the Educational Resource Discovery Index (ERuDIte) 7 to help users find educational resources for biomedical sciences. "Many of the resources gathered by ERuDIte are crowd-sourced from the immense online data science community built through the efforts of organizations such as Data and Software Carpentry and the Mozilla Foundation" [3]. At the time, the NIH had a slew of Funding Opportunity Announcements (FOAs) (See Appendix) for training biomedical data scientists, exposing biomedical scientists to data science, and holding in-person courses 8. The authors [3] were actually surprised from the low number of FOA applications at the time and encouraged more leadership and sustained commitment of resources to build multidisciplinary data science teams.

While there exists open data science curriculum for the health sciences, they have not been formally assessed in their efficacy. Additionally, no set of learner personas to identify the types of students exist so future curriculum can be better focused for the student's needs.

[3] Michelle C. Dunn and Philip E. Bourne. Building the biomedical data science workforce. PLoS Biology, 15(7):1–9, July 2017.

## 4.0 Study Endpoints

- 4.1 *Describe the primary and secondary **study** endpoints. See links below for discussion of study endpoints and how they may differ from study objectives. These are most common in clinical trials but are sometimes applicable to other types of biomedical research, as well as social, behavioral, or educational research. See link below for a discussion.*

[https://docs.google.com/document/d/1Wocz7K7a0hCQJPP0\\_khh5l1SQQjhGDDGHzcOPRHR5Tw/edit?usp=sharing](https://docs.google.com/document/d/1Wocz7K7a0hCQJPP0_khh5l1SQQjhGDDGHzcOPRHR5Tw/edit?usp=sharing)

N/A

- 4.2 *Describe any primary or secondary **safety** endpoints. These should be included for all studies that are greater than minimal risk. (Minimal risk: The probability and magnitude of harm or discomfort anticipated in the research that are not greater in and of themselves than those ordinarily encountered in*

*daily life or during the performance of routine physical or psychological examinations or tests.):*

N/A

## **5.0 Study Design and Statistical Analysis Plan**

- 5.1 Describe the basic study design/approach (e.g., qualitative study using five focus groups of first year students to describe assimilation into the university community; randomized controlled trial of a behavioral change intervention to increase dietary intake of whole grains; pre- post-test evaluation of new pedagogical techniques to improve adult literacy):*

The study will use a convenience sample of people in the medical and biomedical field through email listservs and snowball sampling. The study is separated into 3 phases. Phase 1 will use a survey to create learner personas. Those survey results will be used to create a workshop series as well as the Phase 2 pre and post workshop and learning material assessments. The results of Phase 1 and 2 will inform the researchers of the questions for the Phase 3 long-term survey.

A series of assessments, questionnaires, and deliverables will be given before, during, and after a workshop about data science. The assessments and questionnaires will ask about the participant's programming, statistics, and data management experiences. These questions will be used to formulate learner personas which will guide in the creation of tailored data science curriculum workshop content. A set of pre, post, and long-term workshop assessments will be given to evaluate the workshop to improve data literacy.

The Phase 1 pre workshop student self assessment (i.e., learner persona survey) is included as ("survey-01-pre\_workshop\_self\_assessment.pdf").

The Phase 2 pre/post working and learning material surveys are included as (survey-02-pre\_workshop.pdf and survey-03-post\_workshop.pdf).

Descriptive statistics, visualizations, and analysis will be formed using any of the various platforms, including Microsoft Excel, Python, R, or Tableau. Text based survey responses will be analyzed through text mining using R and Python.

- 5.2 Describe corresponding data analysis plan/approach (e.g., content analysis of focus group transcripts; descriptive analysis followed by linear regression modeling; nonparametric analysis of pre- and post-test measures):*

The assessment questions are written such that each participant will serve as their own internal validity check. This will be confirmed with exploratory and confirmatory factor analysis.

Participant internal consistency will be checked within the questionnaire itself as multiple questions will be asked relating to a single dimension and looking at Cronbach's alpha and intraclass correlation

The pre-workshop student self-assessment will be used to identify personas. An unsupervised clustering algorithm (e.g., K-means) will be used to identify the types and patterns of responses and the final groupings will be modified so they make sense to the real world.

Descriptive statistics will be formed using any of the various platforms, including Microsoft Excel, Python, or Tableau. Text based survey responses will be analyzed through text mining using R and Python. Finally, the results of our analyses will be visualized using Tableau software

The pre, post, and long-term assessments will be used to evaluate the workshop using a repeated measures method, latent growth curve modeling, and McNemar test.

## 6.0 Setting

6.1 *Describe the sites or locations where your research team will conduct the research. Consider each of the items listed below:*

- *Identify where your research team will identify and recruit potential subjects.*
- *Identify where the team will perform the research procedures.*
- *Describe the composition and involvement of any community advisory board(s).*
- *For research conducted in other locations, describe:*
  - *Site-specific regulations or customs affecting the research at those locations.*
  - *Local scientific and ethical review structure at those locations. Examples include work in other cultures or ethnic groups (within or outside of the U.S.) and work with churches. The HRPP will provide additional guidance for international research.*

For phase 1, our team will send an email (verbiage of email supplied in document "email-survey-01-pre\_workshop\_self\_assessment.pdf") to be forwarded to listservs for individuals working in the biomedical/health space at Virginia Tech. The email will all contain a link to the Qualtrics pre workshop student self assessment ("survey-01-pre\_workshop\_self\_assessment.pdf") providing more

## 20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

information about the study, consent, and the actual study survey. Results from this study will inform the survey creation for phase 2.

For Phase 2, this study is an evaluation of a technical hands-on workshop or on the learning materials. This can be conducted in-person in a classroom, online via webcast (e.g., Zoom), or on the student's own time.

Emails will be collected from participants to coordinate the date, time, and location of the workshop. These emails will be used to send out the pre and post workshop surveys for the workshop. The pre and post workshop surveys will also be posted along with the published learning materials so students can still participate in the study if they are self-learning.

Workshop deliverables will also be collected during Phase 2. Students may opt-into this part of the study by using their self-created identifier as their name and this will be linked to their consent in Phase 2. If there is no identifier match from the workshop deliverables and the results of the Phase 2 pre-workshop survey, the deliverable will not be considered as a part of the research study.

After the workshop is over, the emails collected will be used one more time to send out a long-term survey for Phase 3. At the end of the Phase 3, the participant list and their emails will be deleted for privacy concerns. Since no identifiable information will be collected in the survey themselves, there will not be a Phase 3 component for those who are going through the materials on their own.

The pre-workshop student self-assessment (learner persona survey) in Phase 1 will be emailed out via listservs for participants to take on their own time. The surveys for Phase 2 and 3 will be emailed to participants after they sign up for the workshop and be provided along with the published learning materials. Separate consent forms will be given to students to fill out in Phase 1 and 2. For phase 1, the consent form will be filled out during the pre-workshop student self-assessment (learner persona). The pre-workshop student self-assessment survey is included in Phase 1 of this IRB application. Results from Phase 1 survey will inform the design of the Phase 2 and 3 survey. For phase 2 and 3, an additional consent form will be filled out during the pre-workshop assessment (in case someone attends the workshop without filling out a learner persona survey and wants to participate in the study). These consent forms and surveys are and will be included as IRB amendments. The post workshop survey will be given after the workshop. E-mail address collected for the workshop will also be used to send out for the Phase 3 long-term survey. Students will be able to complete all the surveys on their own time. For the Phase 2 surveys, those who are attending the workshop will be asked to fill out the pre-workshop survey at the beginning of the workshop, and time will be provided to take the post-workshop survey at the end of the workshop. This is to maximize the number of responses from participants. Those students going through the materials on their own will have access to the survey links in the survey materials.

For in-person workshops (to be determined based on COVID-19 and at minimum online version will be offered for everybody), they will be conducted at the university and online workshops will be held using an online conferencing system such as Zoom.

All research procedures (i.e., surveys) will be conducted online using Qualtrics. And in-workshop deliverables will use an online clicker system to collect student responses.

## 7.0 Study Intervention(s)/Investigational Agent(s)

*7.1 Describe the study interventions (including behavioral interventions) and/or investigational agents (e.g., drugs or devices) to be used in this study. Consider each of the items listed below:*

- *Drug/Device Handling: If the research involves drugs or devices, describe your plans to store, handle, and administer the drugs or devices so that they will be used only on subjects, and only by authorized investigators.*
- *Describe whether any of the following will be used: microwaves, X-rays, DEXA scans, general anesthesia, or sedation*
- *If control of the drugs or devices used in this protocol will be accomplished by following an established, approved organizational SOP (e.g., Research Pharmacy SOP for the Control of Investigational Drugs, etc.), please reference the SOP in this section.*

No study interventions will be used in this study. Individuals can attend the workshops without enrolling in the study and the workshops were going to be delivered regardless of this study. The materials will also be posted freely online for anyone to go through on their own. This study seeks to more effectively create and deploy data science workshops for biomedical/health professionals.

*7.2 List the name of all drugs (including any vitamins, supplements, herbs, or nicotine) to be used in the study. Indicate whether they have FDA approval, and list any limitations for their use:*

No drugs will be used in the study.

*7.3 List all devices, how they will be used, their purpose in the study, and if they will be used in a manner consistent with their approved uses. If they will be used in ways that are not yet FDA approved, indicate whether they need an*

*IDE or a determination that they are exempt from the IDE Determination. If a determination of significant risk or non-significant risk is needed for any of the devices, include the researcher's recommendation for each of those devices:*

No devices will be provided for this study. A computer or smartphone supplied by the participant will need to be used to fill out the surveys. The students will be asked to have the necessary software to be installed on their computers to follow along the workshop.

7.4 *If the drug is investigational (has an IND) or the device has an IDE or a claim of abbreviated IDE (non-significant risk device), include the following information:*

- *Identify the holder of the IND/IDE/abbreviated IDE.*
- *Explain procedures followed to comply with sponsor requirements for FDA regulated research for the following:*

<b>FDA Regulation</b>	<b><i>Applicable to:</i></b>		
	<b><i>IND Studies</i></b>	<b><i>IDE studies</i></b>	<b><i>Abbreviated IDE studies</i></b>
<b><i>21 CFR 11</i></b>	<b><i>X</i></b>	<b><i>X</i></b>	
<b><i>21 CFR 54</i></b>	<b><i>X</i></b>	<b><i>X</i></b>	
<b><i>21 CFR 210</i></b>	<b><i>X</i></b>		
<b><i>21 CFR 211</i></b>	<b><i>X</i></b>		
<b><i>21 CFR 312</i></b>	<b><i>X</i></b>		
<b><i>21 CFR 812</i></b>		<b><i>X</i></b>	<b><i>X</i></b>
<b><i>21 CFR 820</i></b>		<b><i>X</i></b>	

No drugs will be used in this study.

## 8.0 Procedures Involved

8.1 *Describe and explain the study design:*

This study will use a longitudinal study design with 4 main survey points. Participants will take a (1) pre-workshop student self-assessment (learner persona survey), (2) pre-workshop survey, (3) post-workshop survey, and (4) long-term survey. The consent form will be given to the participant during the pre-workshop student self-assessment (Phase 1)

## 20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

and during the pre-workshop survey in Phase 2 (in the event they did not take the persona survey).

Participants will create a unique identifier that will be used to match results across surveys. To keep their personal information anonymous the unique identifier will be in the form of: Number of siblings (as numeric) + First two letters of the city you were born in (lowercase) + First three letters of your current street (lowercase).

E.g., 1pobac (Sherlock Holmes has 1 brother, Mycroft, Born in Portsmouth, lives in 221B Backer Street, London).

E-mail addresses will be collected for workshop registration, these email addresses will also be used to send out the surveys in Phase 2 and Phase 3 of the study. At the end of Phase 3, email addresses and participant list will be deleted for privacy. No e-mail addresses will be collected for learners who are self-learning through the materials (i.e, not attending the workshop).

During the workshop there will be a series of formative assessments given (i.e., questions to test what they are learning and progressing). This will also serve as an internal validity check for some of the background information in the pre-workshop surveys.

### 8.2 Provide a description of:

- *All research procedures being performed*
- *If the study has more than one procedure, session, and/or subject population, describe each procedure, session, and/or study population separately. For complex studies, you are encouraged to include a figure or chart.*

#### Phase 1: Pre-workshop student self-assessment

Participants will be contacted through VT listservs to participate in a student self-assessment survey. The listservs will be biomedical related (e.g., FBRI, FBCB, iTHRIV, PHS TBMH, VetMed, VCOM, VTCSOM, etc). The potential research participant will be taken to a Qualtrics survey that will have the research consent before the first survey research question is shown. These results will be used to create the learner personas which will inform the workshop content. The learner persona survey will be used to gather a participant's understanding of the core data science competencies as outlined in the objectives in Section 2.1. The pre-survey document titled, "survey-01-pre\_workshop\_self\_assessment.pdf" is included with this form in the submission file. A separate listserv call, consent, and surveys will be made for Phase 2 and Phase 3. The results of this survey will guide the researchers towards the pre, post, and long-term surveys in Phase 2 and 3.

#### Phase 2: Pre/post workshop and learning material surveys

A separate listserv call will be made for workshop registration. Participants who register for the workshop will be asked to partake in Phase 2 and 3 of the study. They will also be

asked to complete and consent to the Phase 1 survey if they did not participate already take the pre-workshop student self-assessment survey. The Phase 2 pre-workshop survey will have a separate consent form that will apply for Phase 2 and 3 of the study since these questions will all revolve around the workshop itself. The Phase 2 pre-workshop survey will be completed before the start of the workshop. At the end of the workshop links to the Phase 2 post-workshop survey will be given to the participants. Students have the ability to opt-in to submitting workshop deliverables at the end of the workshop. Learners who are not attending the workshop will have the option to participate in the study by accessing the links to the survey

### Phase 3: Long-term workshop surveys

Consent for Phase 3 of the study will be given during Phase 2 of the study. The attendants of the workshop in Phase 2 will be sent a Phase 3 long-term survey about 6 months from the workshop. At the end of Phase 3, the workshop participant list will be deleted for privacy concerns.

Participants will input their own unique identifier in the surveys which will link the responses longitudinally. These identifiers will be converted to an integer value for privacy concerns.

### 8.3 Describe:

- *Procedures or safeguards intended to reduce the probability and magnitude of risks. (For example: Reducing the risk of injury in a virtual reality study either by having the subjects sit during the study or by providing an obstacle-free space for walking.)*
- *Be sure to describe all drugs and devices used in the research, when they will be administered or used, and their purpose.*
- *Methods used to collect data about subjects. Please upload all data collection forms to Protocol Management. Some common examples are:*
  - *Screening questionnaires*
  - *Survey(s), including online surveys*
  - *Demographic questionnaire(s)*
  - *Interview guide(s), e.g., questions or pool of questions for semi-structured interviews*
  - *Focus group guide(s)*
  - *Other documents used to collect data*

We reduced the probability and magnitude of risks by allowing participants to cease participation in the study immediately, whenever they wish to do so. We also emphasized the vocabulary and phrasing of the surveys in a manner that



reduces the risks of negative feelings, thoughts and/or associations.

Given that this is an online research study, participants can use any personal devices necessary to complete the online surveys. These personal devices will not be administered by this research team and include, but are not limited to, computers, smartphones and tablets.

In Phase 1, only 1 online survey using the Qualtrics platform will be used to collect data about

subjects. These supplemental documents are included in the submission along with this form. Finally, there will be more separate documents, each containing the message/email being sent to invite participants to take place in the study.

This document will be the email draft.

Data collection survey filenames:

1. survey-01-pre\_workshop\_self\_assessment.pdf
2. survey-02-pre\_workshop.pdf
3. survey-03-post\_workshop.pdf

Email filename:

1. email-survey-01-pre\_workshop\_self\_assessment.docx
2. email-survey-01-pre\_workshop\_self\_assessment.docx

Website filename:

1. website-survey-02-pre\_post\_survey.docx

Survey's for Phase 2 (IRB version 2.0.0) and Phase 3 will be submitted for IRB amendment and approval after Phase 1 of the study.

*8.4 What data will you collect during the study and how you will obtain them? Please include descriptions of electronic data collection, database matching, and app-based data collection:*

Data will be collected from online electronic surveys (Qualtrics) sent to the participants. E-mail addresses will be collected during workshop registration. This list will be used to mail out the pre, post, and long-term survey for the workshop. This email list will be deleted at the end of Phase 3 for privacy concerns.

Responses from subjects are automatically integrated and condensed in the Qualtrics platform. This allows us to readily collect the electronic data and transfer it to software such as SAS JMP, R, Python, or Microsoft Excel that enables us to perform analyses.

8.5 *Who will transcribe or code audio and/or video recordings?:*

No audio or video recording will be performed.

8.6 *Include a description of any deception to be used in the study. Include justification for the use of deception (why the deception is necessary), describe the debriefing process, and describe how the study meets all the following criteria for alteration of consent (deception is considered an alteration of informed consent):*

- *The research involves no more than minimal risk to the subjects*
- *The alteration will not adversely affect the rights and welfare of the subjects*
- *The research could not practicably be carried out without the alteration/deception*
- *(Optional but encouraged in most cases) Subjects will be provided with additional pertinent information after participation (i.e., debriefing for studies involving deception)*

No deception will be used in the study.

8.7 *If the study involves long-term follow-up (once all research related procedures are complete), describe what data will be collected during the follow up period and when it will occur:*

The study does not involve long-term follow up. There is a long-term survey after the workshop is over, but this is part of the longitudinal study design.

## **9.0 Data and Specimen Long Term Storage and Use**

9.1 *If you will store data or specimens for future use, describe where you will store the data or specimens, how long they will be stored, and how and by whom the data or specimens will be accessed:*

## 20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

Data will be shared with the research team via a private, shared Virginia Tech 2FA google drive folder. Analysis will be performed using the programs above and analysis files/scripts will be uploaded to this folder. Final visualizations will be uploaded to this folder. At the completion of the study, all data (raw and analyzed) will be zipped and stored on a external hard drive of the PI in a locked office drawer. Access will be controlled by the PI to the VT 2FA google drive folder.

The code (e.g. python scripts and other computer code written for analysis of the data) and de-identified data used for analysis will be stored and shared on an open science platform (e.g., Open Science Framework (<https://osf.io/>), GitHub (<https://github.com/>), Zenodo (<https://zenodo.org/>), and/or VTechData (<https://data.lib.vt.edu/>). There will be nothing in the code that can reverse engineer the user-provided identifier once it has been converted to an integer value.

Data can be removed at the participants' request.

Attendance list from the workshop signup will be deleted at the end of the Phase 3 long-term assessments.

The data will be accessed by the principal investigator and the other study contacts that include Daniel Chen(chend@vt.edu)

### 9.2 *For specimens, list the data to be stored or associated with each specimen:*

No specimens will be kept

### 9.3 *Describe the procedures to release data or specimens outside of the research team, including the process to request a release, approvals required for release, who can obtain data or specimens, and what data will be provided with specimens:*

Survey data will be further de-identified by taking the user-provided unique ID and converting into an integer value. No other identifiers will be collected in the survey data that will be released.

E-mail addresses collected from the workshop registration will be deleted at the end of the study and will not be released. E-mail addresses will only be used for registration and contacting participants for Phase 2 and 3 of the study.

9.4 *Describe the identifiers to be included with stored data or specimens, as well as any key or code that could be used to make them identifiable. Describe where the code will be stored, who will have access to it, and when it will be destroyed:*

The participant list with email address will be kept during the workshop registration phase. This list will be used to inform the time and place of the workshop and also links to the pre, post, and long-term surveys. This list of participants and email addresses Access will be controlled by the PI to the VT 2FA google drive folder and will be deleted at the end of Phase 3. This E-mail list will not be stored on any publicly accessible system.

For survey results, participants will be asked to create a unique identifier. This will take the form of: Number of siblings (as numeric) + First two letters of the city you were born in (lowercase) + First three letters of your current street (lowercase). E.g., 1pobac (Sherlock Homes has 1 brother, Mycroft, Born in Portsmouth, lives in 221B Backer Street, London). This is the same unique identifier used by The Carpentries for their publicly released survey results. This identifier will be further de-identified into a unique number (e.g., "1pobac" becomes "001") for data analysis and sharing.

The Qualtrics data will be pulled programmatically via the Qualtrics API and the user-created identifier will be immediately processed into an integer representation to reduce the chance of accidentally storing the user-created identifier on any computer.

The code (e.g. python scripts and other computer code written for analysis of the data) and de-identified data used for analysis will be stored and shared on an open science platform (e.g., Open Science Framework (<https://osf.io/>), GitHub (<https://github.com/>), Zenodo (<https://zenodo.org/>), and/or VTechData (<https://data.lib.vt.edu/>)). There will be nothing in the code that can reverse engineer the user-provided identifier once it has been converted to an integer value.

Data can be removed at the participants' request.

9.5 *Please select the identifiers you will obtain (whether directly from participants or from another source), including but not limited to:*

<input type="checkbox"/>	<i>Name</i>
<input type="checkbox"/>	<i>Geographical subdivisions smaller than a state, including street address, city, county, precinct, zip code, and equivalent geocodes (note, the initial three digits of a zip code are not considered identifiable)</i>

20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

<input checked="" type="checkbox"/>	<i>Elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death, and single year of age over 89 and all elements of dates (including year) indicative of such age (note, such ages and elements may be aggregated into a single category of age 90+)</i>
<input type="checkbox"/>	<i>Phone numbers</i>
<input type="checkbox"/>	<i>Fax numbers</i>
<input checked="" type="checkbox"/>	<i>Electronic mail addresses (e-mail)</i>
<input type="checkbox"/>	<i>Social Security numbers</i>
<input type="checkbox"/>	<i>Medical record numbers</i>
<input type="checkbox"/>	<i>Health plan beneficiary numbers</i>
<input type="checkbox"/>	<i>Account numbers</i>
<input type="checkbox"/>	<i>Certificate/license numbers</i>
<input type="checkbox"/>	<i>Vehicle identifiers and serial numbers, including license plate numbers</i>
<input type="checkbox"/>	<i>Device identifiers and serial numbers</i>
<input checked="" type="checkbox"/>	<i>Web Universal Resource Locators (URLs)</i>
<input type="checkbox"/>	<i>Internet protocol (IP) address numbers</i>
<input type="checkbox"/>	<i>Biometric identifiers, including finger and voice prints (audio recording)</i>
<input type="checkbox"/>	<i>Full face photographic images and any comparable images (including video recording)</i>
<input type="checkbox"/>	<i>Student record number or identification number</i>
<input type="checkbox"/>	<i>User name for online or computer accounts</i>
<input checked="" type="checkbox"/>	<i>Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data): For survey results, participants will be asked to create a unique identifier. This will take the form of: Number of siblings (as numeric) + First two letters of the city you were born in (lowercase) + First three letters of your current street (lowercase). E.g., 1pobac (Sherlock Homes has 1 brother, Mycroft, Born in Portsmouth, lives in 221B Backer Street, London).</i>

## 10.0 Sharing of Results with Subjects

*10.1 Describe whether you will share results (study results or individual subject results, such as results of investigational diagnostic tests, genetic tests, or incidental findings) with subjects or others (e.g., the subject's primary care physician). If so, describe how you will share the results and include this information as part of the consent document. Upload materials you will use to explain the results to subjects:*

Data and reports will be published on an open science platform (e.g., Open Science Framework (<https://osf.io/>), GitHub (<https://github.com/>), Zenodo (<https://zenodo.org/>), and/or VTechData (<https://data.lib.vt.edu/>)). Peer

reviewed publication is planned as well as conference presentations on the results of this work.

## 11.0 Study Timelines

### *11.1 Describe:*

- *The duration of an individual subject's participation in the study (for example, 1 hour, 2-4 weeks, 3-5 years).*
- *The amount of time expected to enroll all study subjects (weeks, months, years, etc.)*
- *The amount of time expected for the investigators to complete this study including primary data analyses.*

Participants are not expected to participate in the study for longer than 1 year. This includes workshop logistics (time to set up the workshop, coordinating time and location of the workshop, and registering for the workshop) along with the questionnaires and assessments for the study, including the 6-month long-term assessment.

We are hoping to enroll 4 workshop cohorts over the course of 1 year. Where the data analysis can be performed on a rolling basis as workshops are being planned, running, and finished.

## 12.0 Inclusion and Exclusion Criteria

### *12.1 Describe how you will screen individuals for eligibility. When will screening occur and what procedures will you use? Upload any screening scripts or surveys to Protocol Management:*

Notices for the workshops will be sent out to biomedically relevant departments and programs from 2020 to 2021. This will follow a snowball sampling method to screen for participants, where workshops will serve as the basis to find more interested participants. Students who go through the materials on their own without attending a workshop may also participate in the study. Anyone who identifies as being in the biomedical community as a student, researcher, staff, or practitioner qualifies to be in the study.

### *12.2 Describe the eligibility criteria that define who will be included and who will be excluded from enrollment for each procedure of your study. Include any geographic criteria (e.g., Virginia Tech undergraduate students, a national*

*sample of adults with engineering degrees, minors aged 8-12 in the New River Valley, university faculty in Virginia and Paris, France):*

Any adult (18+ years old) who fills out the survey (phase 1) and/or attends the workshops or uses the materials on their own (phase 2/3) qualifies to be in the study.

*12.3 Indicate specifically whether you will include or exclude each of the following special populations: (You may not include members of these populations as subjects in your research unless you indicate them in the description of your subject population.)*

- *Minors, as defined by state law where the study is performed (infants, children, teenagers)*
- *Pregnant women (can be included in minimal risk studies by mentioning in section 13.1)*
- *Prisoners (including all incarcerated individuals)*
- *Adults not capable to consent on their own behalf*

Minors, prisoners, and adults not capable to consent on their own behalf are excluded from this study.

Pregnant woman who also qualify in the inclusion criteria are eligible to be in the study since the study assumes minimal risk and no procedures and/or interventions will be performed in this study.

## **13.0 Vulnerable Populations**

*13.1 If the research involves individuals who are vulnerable to coercion or undue influence, please describe additional safeguards you will include to protect their rights and welfare. Consider the applicable items listed below:*

- *If the research involves Virginia Tech students, indicate whether these are students of any of the investigators. If so, describe whether the activities will take place during class time as part of the curriculum and the steps you will take to reduce the possibility that students feel obliged to participate in order to improve their course grade. The HRPP can provide further guidance as needed. Describe whether you will request access to student records (e.g., SAT, GPA, GRE scores).*
- *If the research involves employees of Virginia Tech or the research sponsor, describe steps you will take to ensure that the employees*

*are freely participating and describe how their data will be protected from inspection by their supervisors.*

- *If the research involves Virginia Tech NCAA athletes, you must obtain approval from the athletic department.*
- *For research involving Montgomery County Public Schools, you must obtain county approval (after obtaining contingent Virginia Tech approval). Other locales have different requirements; please check on these and describe here. Approval is typically granted by the superintendent, principal, and classroom teacher (in that order). Approval by an individual teacher is insufficient. School approval, in the form of a letter or a memorandum should be uploaded as a supporting document.*
- *If the research involves pregnant women, review “CHECKLIST: Pregnant Women (HRP-412)” to ensure that you have provided sufficient information in this protocol.*
- *If the research involves prisoners, review “CHECKLIST: Prisoners (HRP-415)” to ensure that you have provided sufficient information in this protocol.*
- *If the research involves persons who have not attained the legal age for consent to treatments or procedures involved in the research (minors), review the “CHECKLIST: Minors (HRP-416)” to ensure that you have provided sufficient information in this protocol.*
- *If the research involves cognitively impaired adults, review “CHECKLIST: Cognitively Impaired Adults (HRP-417)” to ensure that you have provided sufficient information in this protocol.*

The study can include VT students, VT employees, pregnant women or NCAA athletes if they meet the inclusion criteria (age 18+).

The workshop dates will be scheduled in advance to work with participants' schedules and they are free to leave the study at any point of the study. No student records or data outside those asked in the surveys, registration, and sign-in will be requested. The study and workshops will be advertised and participants can choose to opt-in the study on their own free will. The study and accompanying workshop is low-risk and participants can leave the study at any point.

Research surveys will be posted along the materials for learners to opt-in the study if they choose to self-learn without a workshop.

## 14.0 Number of Subjects



20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

*14.1 Indicate the total number of subjects to be enrolled and how this number was determined (e.g., sample size calculation [show], number of available subjects in a finite pool, number of tests funding award would allow):*

Virginia tech software-carpentry workshops tend to be between 20 to 30 participants. This study plans to run about 4 to 7 workshops for a total of ~100-150 participants in the study based on past workshop sizes.

The survey can go to anyone at Virginia Tech (about n=~38K), but we expect to only market to biomed/health listservs and attract those interested in data science workshops. Materials will be posted freely online to make and survey links will be provided for anyone to self-learn and opt-in to the study.

*14.2 If this is a multi-site study, indicate the number of subjects to be enrolled at this site and the total to be enrolled from all sites:*

This is not a multi-site study

*14.3 If applicable, indicate the number of potential subjects you expect to screen for enrollment, and the number of subjects you will need to complete the research procedures:*

We plan to distribute the learner persona survey to all relevant biomedical/health listservs at Virginia Tech. We expect a total of 100 -150 survey participants based on known interest and request for such workshops.

*14.4 If the study has more than one procedure, indicate the total number of subjects to undergo each procedure separately:*

No procedures in this study.

## **15.0 Recruitment Methods**

*15.1 Describe when, where, and how you will recruit potential subjects:*

Subjects will be recruited through word-of-mouth snowball sampling.

We will reach out to representatives of populations pools through Virginia Tech list serves (mailing list) such as academic and organizational list serves for undergraduate students, graduate students, faculty and staff in the biomedical sciences.

Recruitment for the learner personal survey will be distributed after IRB approval and be live for 2-weeks.

*15.2 Describe the source of subjects (for example, clinic patients with specific conditions, students in the library, community members at a gathering, or members of a local gym):*

Anyone over age 18 is included in this work. This includes students, faculty, staff, practitioners, and community members.

*15.3 Describe the methods that you will use to identify potential subjects:*

We will work with biomedical departments to conduct and plan the workshops. Those who sign up for the study and workshop should meet the inclusion criteria (aged 18+). The pre-workshop student self-assessment (learner persona) has a question about what part of the biomedical field the participant is in.

*15.4 Describe materials that you will be use to recruit subjects. Attach copies of these documents with this protocol in Protocol Management and be sure to include the IRB protocol number on each document.*

- *For flyers, attach the final copy of printed flyers.*
- *For Virginia Tech News, Facebook postings and ads, newspaper ads, websites, MTurk/SONA/online survey systems, etc., attach the final wording and graphics to be used.*
- *For email recruitments, please include the subject line.*
- *For advertisements meant for audio broadcast, please submit the wording of the advertisement prior to taping (to avoid having to re-record with approved language) and submit the final recorded version for IRB review before use.*
- *Describe any compensation to subjects. Separate compensation into appropriate categories, such as: reimbursement for expenses, time and effort, and additional incentives for study participation. For each category, specify the amount (including any pro-rated amount), schedule, and method of payment.*

Emails will be used to recruit subjects. Email subject would be "Participants needed for survey on data science workshops for the biomedical sciences". The contents of the email to send to listserv managers is included as file:

"email-survey-01-pre\_workshop\_self\_assessment.docx"

Workshop attendees will be contacted after signing up for the workshop with the to take the self assessment, and be contacted before the workshop starts to take the pre-workshop assessment.

## 16.0 Withdrawal of Subjects

### *16.1 Describe circumstances under which you anticipate subjects could be withdrawn from the research without their consent:*

Subjects could be withdrawn from the study if their survey results are incomplete.

### *16.2 If applicable, describe any procedures for orderly termination (e.g., discontinuation of a study drug or debriefing after a behavioral intervention):*

There is no procedure, drug, or intervention in this study. If the participant would like to withdraw from the study they can contact the study PI or personnel and their data will be removed from the study.

### *16.3 Describe procedures that you will follow when subjects withdraw from the research, including partial withdrawal from procedures with continued data collection (e.g., participant declines to continue with regular blood draws, but continues with periodic behavioral questionnaires):*

If the study participant would like to withdraw from the study, can contact the study PI and their data will be removed from the study.

## 17.0 Risks to Subjects

### *17.1 List the reasonably foreseeable risks, discomforts, hazards, or inconveniences to the subjects related the subjects' participation in the research. Include for the IRB's consideration a description of the probability, magnitude, duration, and reversibility of the risks. Consider physical, psychological, social, legal, privacy, and economic risks. Do not indicate "No risk" or "N/A." Instead, for*

*studies with very low risk (e.g., anonymous online questionnaire on a mundane topic) indicate “The investigators are not aware of any risks from participation in this study.” or “No more than risks than are found in everyday life.” The example consent form presents a tabular method for risk information, which you can also use here. Common risk types include:*

- *Physical (e.g., potential for pain, discomfort, infection)*
- *Psychological (e.g., potential for stress, discomfort, and/or embarrassment)*
- *Social (e.g., potential for discrimination or stigmatization and disruption of personal and family relationships)*
- *Legal (e.g., potential for disclosure of illegal activity, negligence)*
- *Privacy (e.g., potential for personal information being accessed, used, or disclosed without the subjects’ knowledge or consent, breach of confidentiality/security)*
- *Economic (e.g., potential for individuals to lose access to economic services, employment, insurability)*

During the process of completing the survey, participants will be asked questions about programming experience and their thoughts and attitudes surrounding the subject of statistics and data management. If there are any questions participants would rather not answer or that they do not feel comfortable answering, they can move on to the next question.

There is minimal risk the by being a part of this study they could experience physical, psychological, privacy, legal, social, economic, or emotional distress given the subject of the survey.

This study is not meant to gather information about specific individuals, and the information you provide will be combined with that of other survey participants to gather information.

Outside this, the investigators are not aware of any other risks from participation in this study and expect no more than risks found in everyday life.

*17.2 Indicate the measures you will use to minimize risks and monitor subjects for safety. (e.g., asking a subject at regular intervals to rate how they are feeling from 1 to 10, or to slowly crouch in order to check their balance.)*

Participants can choose to skip questions they are not comfortable responding to and can leave the study at any point.

*17.3 If applicable, indicate which procedures might have risks to the subjects that are currently unforeseeable. This will be rare, and usually applicable when testing a new drug or device or a new use of an existing drug or device:*

There are no procedures or drugs in this study.

*17.4 If applicable, indicate which procedures might have risks to an embryo or fetus should the subject be or become pregnant:*

There are no procedures or drugs in this study.

*17.5 If applicable, describe risks to others who are not subjects (e.g., collection of sensitive health data that might affect sexual partners if disclosed, mandatory reporting of abuse, DNA testing that might affect family members or relationships):*

There is minimal risk to other who are not subjects in this study.

## **18.0 Potential Benefits to Subjects**

*18.1 Describe the potential benefits that individual subjects might experience from participating in the research. Include the probability, magnitude, and duration of the potential benefits, as this will be useful to the IRB's risk:benefit analysis. Do not include benefits to society or others. Do not list monetary or non-monetary compensation for participation, as this is not a benefit. These should be included in section 2 or 3 of this document:*

The study questionnaires and surveys revolve around the creation and execution of a technical workshop and its learning materials. By participating in the study the subjects are also participating in the workshop or self-learning using the free and public learning materials to develop data science skills which may benefit their professional and/or personal development.

The workshop and its training materials are of the same caliber of training workshops at conferences.

*18.2 If applicable, specify that there are no anticipated direct benefits for participants:*

N/A

## 19.0 Data Management and Confidentiality

*19.1 Describe procedures that you will use for quality control to ensure validity of collected data:*

Survey validity will be checked using factor analysis to look for construct validity. Participant internal consistency will be checked within the questionnaire itself as multiple questions will be asked relating to a single dimension and looking at Cronbach's alpha and intraclass correlation.

*19.2 Describe any existing data or biospecimens you will obtain as part of this study. Include:*

- *Variables or samples to be obtained*
- *Source of the data or specimens*
- *Your authorization to access or receive the data or biospecimens*
- *Whether the data or biospecimens are publicly available*
- *Whether the data or specimens you receive will contain identifiers*

No existing data will be used. All data collected will be from this study and the questionnaires filled out by participants.

The questionnaires themselves will be adapted from The Carpentries (<https://carpentries.org/assessment/>), RStudio learner personas (<https://rstudio-education.github.io/learner-personas/>), Teaching Tech Together (<https://teachtogether.tech/>), and How Learning Works (Ambrose, et al)

*19.3 Describe the steps that you will take to handle and secure study data during data collection, storage, use, and transmission. Include information about training of study staff, authorization of access, password protection, encryption, physical controls, certificates of confidentiality, separation of identifiers and data, etc.:*

20-537 Data science workshops for biomedical and health professionals: Persona identification and workshop assessment

The survey questionnaire results will be collected separately from workshop registration. This will inherently un-link survey responses from participant name and email address.

To link participants longitudinally, they will be asked to create a unique identifier for themselves. During data analysis, this unique identifier will be converted into an ID number that will be used during the data analysis and data dissemination portions of the study.

All hard-copy records with workshop participant names and email addresses will be locked in a cabinet in the PI's office. All participant survey responses collected contain their original participant-created unique identifier will be stored in the VT Qualtrics system and only pulled using the Qualtrics API, so the original IDs never leave Qualtrics.

*19.4 For multi-site studies, describe how data or specimens will be handled and secured for each site (e.g., central or disseminated data storage, data coordinating center):*

N/A

*19.5 Describe the plan for data disposition following the conclusion of the study (e.g., long term maintenance of data, data destruction methods).*

- *What information will be included in the long term storage of data or specimens?*
- *How long will the data or specimens be stored?*
- *Where and how data or specimens will be stored?*
- *Who will have access to the data or specimens during long term storage?*
- *Who is responsible for receipt or transmission of the data or specimens?*
- *How will data or specimens be shared or transported?*
- *When and how will personal identifiers be destroyed?*

The original participant data will have the participant-created unique identifier removed and converted into an integer value which will be used for data analysis, storage, and dissemination.

Data will be stored on an open science platform such as Open Science Framework (<https://osf.io/>), GitHub (<https://github.com/>), Zenodo (<https://zenodo.org/>), and/or VTechData (<https://data.lib.vt.edu/>).

At the conclusion of the study the de-identified version of the data will be shared, and all data containing participant names, email, and identifiers will be deleted and/or destroyed.

## 20.0 Provisions to Protect the Privacy Interests of Subjects

*20.1 Describe the steps that you will take to protect subjects' privacy interests. "Privacy interest" refers to a person's desire to place limits on with whom they interact or to whom they provide personal information (e.g., collecting the minimal amount of private information required to complete the study, protecting the data once it is obtained):*

The participants' survey responses will be kept private and de-identified. However, by participating in this study students will all be able to interact with one another during the workshop.

For online workshops, students may opt to have their cameras off and use a pseudonym when interacting with the instructor and other participants.

Aside from emails (which are needed to coordinate the workshop time, location, and send out surveys), the only other information we are requesting that can be considered personal would be demographic information (i.e., profession, ethnicity, gender) however, this information is not personally identifiable to a given respondent.

The email and workshop participant list will be deleted at the end of Phase 3.

*20.2 Describe steps that you will take to make subjects feel at ease with the research situation in terms of the questions being asked and the procedures being performed. "At ease" does not refer to physical discomfort, but the sense of intrusiveness a subject might experience in response to questions, examinations, and procedures (e.g., use of a same gender investigator to place sensors on the torso, a private changing area if clothing must be changed, sensitivity when discussing pregnancy testing with subjects, making it clear on surveys that participants can discontinue at any time, not asking questions about private or sensitive issues unless necessary for the research):*

This study does not have any procedure or physical examination. The questions asked will be given in an online survey where participants can take wherever they feel most comfortable.

If there is a question the participant feels uncomfortable answering, they are free to skip the question.



*20.3 Describe how you plan to access existing sources of information about the subjects (e.g., medical records, grades) and how you will protect participant privacy through the data security plan:*

No existing information about the participant will be collected. There will be basic demographic information that will be captured (occupation/career stage, gender, and ethnic background) but no data outside the scope of the surveys will be used or collected about the participants. Students will be asked to create their own unique identifier. This will be used to link longitudinal survey responses, but they will not leave the Qualtrics server and will be converted into an integer value for analysis and dissemination.

*20.4 Describe any required reporting that might occur as a result of your research questions, study populations, and data collection methods. Examples for Virginia and Virginia Tech include:*

- ***Any** suspicions (e.g., circumstantial, disclosed) of child abuse (physical, emotional, sexual) and neglect*
- *Sexual discrimination and/or sexual violence that involves a student*
- *Disclosure or signs of intention to harm oneself (i.e., suicidal ideation and/or plan)*
- *Disclosure or signs of desire to harm others (i.e., homicidal ideation and/or plan)*
- *Suspected abuse, neglect or exploitation of vulnerable adults (e.g., individuals with a disability, elderly persons)*

Questions in the survey will be around programming, statistics, data management, and other data science related topics. No questions will be asked where required reporting might occur.

## **21.0 Provisions to Monitor the Data to Ensure the Safety of Subjects**

*Safety monitoring is required when research involves greater than minimal risk and is sometimes appropriate for other studies.*

*21.1 Describe:*

- *The plan to periodically evaluate the data collected regarding both harms and benefits to determine whether subjects remain safe (e.g., periodic reporting to the IRB, establishing a data monitoring committee, reporting data monitoring committee findings to the IRB and the sponsor).*

- *What data you will review, including safety data, unexpected events, and data that show the ability to produce the intended results.*
- *How the safety information will be collected (e.g., with case report forms, at study visits, by telephone calls with subjects).*
- *The frequency of data collection, including when safety data collection starts.*
- *Who will review the safety data and with what frequency.*
- *The statistical tests for analyzing the safety data to determine whether harm is occurring.*
- *Any conditions that will trigger an immediate suspension of the research (e.g., a serious adverse event).*

Study does not involve greater than minimal risk.

## **22.0 Compensation for Research Related Injury**

*22.1 If the research involves more than minimal risk to subjects, describe the available compensation in the event of research-related injury, if any:*

Study does not involve more than minimal risk to subjects.

*22.2 Provide a copy of contract language, if any, relevant to compensation for research-related injury. At Virginia Tech, this is most common for sponsored research:*

Study does not involve more than minimal risk to subjects.

## **23.0 Economic Burden to Subjects**

*23.1 Describe any costs that subjects might be responsible for because of participation in the research, including any uncompensated costs for items such as transportation, missed work, and childcare:*

Participants will need to set aside time to participate in the workshop or go through the materials on their own during this study. They will need access to a computing device with the necessary software installed. The researchers will try their best to accommodate the workshop to be accessible to participants but participants will still need to set aside time

for the workshop or find time on their own to self-learn. Attending the workshop is voluntary and not related to any course/grade.

## 24.0 Consent Process

*24.1 Indicate the process by which you will obtain consent for study participation. Please upload all consent, parental permission, and assent forms, documents, and scripts referenced in this section to Protocol Management.*

*Describe the following:*

- *Where the consent process will take place (e.g., clinic waiting area, classroom, online)*
- *The time interval between sharing the consent information with the prospective subject and obtaining consent. For lab, interview, and focus group studies, the Virginia Tech IRB prefers that subjects have at least 24 hours to review the consent form and study information before the appointment where consent will be obtained. For simple online survey studies, you can typically present the consent information immediately before subjects begin participation.*
- *If applicable, processes to ensure ongoing consent or assent (e.g., for multiple sessions; for research in which a minor will turn 18 during the study; for longitudinal research with minors who will later be asked to provide or affirm their assent).*
- *Please review “SOP: Informed Consent Process for Research (HRP-090)” for recommended procedure. Describe your process, being sure to include:*
  - *The name and role of all study personnel who will be trained and certified by the PI to conduct the consent process*
  - *The time that will be devoted to the consent discussion*
  - *Steps that you will take to minimize the possibility of coercion or undue influence*
  - *Steps that you will take to gauge or ensure the subjects’ understanding*

Consent will take place during the initial online student self-assessment questionnaire (phase 1). Participants will consent to the study before being able to see the first question asked by the questionnaire.

The consent process will take place online, at the beginning of the surveys and before any part of the research study can occur.

There will be a consent page included in the Qualtrics survey form, detailing all necessary information and respondents will have to voluntarily consent before moving on to the surveys.

We will present the consent information immediately before subjects begin participation.

Ongoing consent will not be an issue, as we will require consent prior to the phase one survey and the pre-assessment survey of phase 2. This is opposed to ensuring ongoing consent from an initial subject consent.

Participants will have the ability to contact researchers if they have any concerns. Other than that, it is expected that respondents will read through the consent document and voluntarily consent to participate.

The link to the consent form, pre-workshop self-assessment, pre-workshop survey, and post-workshop survey will be provided during the workshop/study registration process. Workshops will typically be planned at least a month in advance to workout workshop logistics, participants will have up until the workshop starts to fill out the consent form. We will not be meeting with participants in person. They will participate on their own time, and under their own terms. There will be no punishment, nor reward for participating. After administering the surveys, the only role our team will play will be to close the surveys and collect responses.

We are requiring subjects to consent prior to participation. If they selected the option to not consent, then any further responses will be immediately deleted.

### ***Non-English Speaking Subjects***

- *Indicate what language(s) other than English are understood by prospective subjects or representatives.*
- *If non-English speakers will be recruited, describe the process you will use to ensure that the oral and/or written consent information provided will be in a language that they understand.*
- *If you translate consent forms and study materials, please provide a certified translation of the form as well as the certification document.*
- *Indicate the spoken language that study personnel obtaining consent will use. Describe how you will assess fluency of personnel obtaining consent to ensure that the translation is accurate.*

The workshop will only be conducted in English. The study and the consent form will only be conducted in English.

N/A

N/A

N/A

N/A (not using spoken language)

***Waiver or Alteration of Consent Process (consent will not be obtained, required information will not be disclosed, or the research involves deception)***

- *Review the “CHECKLIST: Waiver or Alteration of Consent Process (HRP-410)” to ensure you have provided sufficient information for the IRB to make these determinations (i.e., that it meets the criteria for a waiver or alteration of the consent process).*

N/A

***Subjects who are not yet adults (minors: infants, children, teenagers)***

- *Describe the criteria that you will use to determine legal age for consent to treatments or procedures involved in the research under the applicable law of the jurisdiction in which the research will be conducted (e.g., in Virginia, individuals under the age of 18 years).*
  - *For research conducted in Virginia, review “SOP: Legally Authorized Representatives, Minors, and Guardians (HRP-013)” to determine which individuals in the state meet the definition of “minor.”*
  - *For research conducted outside of the state, please describe the legal requirements for the definition of “minor.”*
- *Describe the process for obtaining parental permission.*
  - *Permission from one parent is acceptable for studies that involve no greater than minimal risk OR involve greater than minimal risk but present the prospect of direct benefit to the minor subject.*
  - *Permission from both parents is required in all other cases (unless one parent is deceased, unknown, incompetent, or not reasonably available, or when only one parent has legal responsibility for the care and custody of the minor).*
- *Describe whether you will obtain permission from individuals other than parents or Legally Authorized Representatives, and if so, who will be allowed to provide permission. Describe the process you will use to determine these individuals’ authority to consent to the minor’s general medical care.*
- *Indicate whether you will obtain assent from all, some, or none of the minors. If you will obtain assent from some minors, indicate which minors will be required to assent. Consider chronological age and intellectual capacity when determining who will be required to*

*provide assent (e.g., infants are unable to assent. However, teenagers are likely able to read and sign an assent form).*

- *When assent of minors is obtained, describe whether and how you will document it. Will minors sign an assent form or give verbal assent?*
- *Attach parental permission and minor assent forms or scripts in Protocol Management.*

Subjects who are not yet adults (minors: infants, children, teenagers).

In the state of Virginia, a minor is considered any individual under the age of 18 (so 17 years old & younger). We have decided to refrain from including anyone who is legally a minor. It is required that all participants must be 18 years of age or older. If they are not, they are not allowed to participate in consent or any part of the survey.

N/A  
N/A  
N/A  
N/A  
N/A  
N/A

#### ***Adults Unable to Consent***

- *Describe the process you will use to determine whether an individual adult is capable of consent.*
- *List the individuals from whom you will obtain permission in order of priority (e.g., durable power of attorney for health care, court appointed guardian for health care decisions, spouse, and non-minor child).*
  - *For research conducted in the Virginia, review “SOP: Legally Authorized Representatives, Minors, and Guardians (HRP-013)” to determine which individuals in the state meet the definition of “legally authorized representative.”*
  - *For research conducted outside of Virginia, please describe the legal requirements for obtaining permission from a legally authorized representative in the state where the research will occur.*
- *Describe the process for assent of the subjects.*
  - *Indicate whether you will require assent from all, some, or none of the subjects. If some, indicate which subjects will be required to assent and which will not.*
  - *If you will not obtain assent from some or all subjects, please provide justification for not obtaining assent.*
  - *Describe whether and how you will document assent.*

Prospective participants are expected to be above the age of 18. We expect any and all participants willing to participate to be capable of consent.

We expect all prospective participants to legally consent prior to participation.

No assent will be obtained. Only direct consent

## **25.0 Process to Document Consent in Writing**

*25.1 Consult “SOP: Written Documentation of Consent (HRP-091)” for recommended procedures, and describe whether and how consent of the subject will be documented in writing:*

N/A

*25.2 If the research presents no more than minimal risk of harm to subjects and involves no procedures for which written documentation of consent is normally required outside of the research context, you can request that the IRB waive the requirement to obtain written documentation of consent (e.g., consent to participate is indicated by pressing a button for an online questionnaire – after the consent information is presented and before the questionnaire begins):*

Study present no more than minimal risk of harm to subjects and involves no procedures.

*25.3 If you will document consent in writing, attach a consent document with places for signatures. If you will obtain consent, but not document consent in writing, please attach the consent script or text. Review “CHECKLIST: Waiver of Written Documentation of Consent (HRP-411)” to ensure that you have provided sufficient information. You should use “TEMPLATE CONSENT DOCUMENT (HRP-502)” to create the consent document or script:*

N/A

## 26.0 Resources Available

26.1 *Describe the resources available to conduct the research. For example, as appropriate:*

- *Describe the PI's availability to supervise the research.*
- *Justify the feasibility of recruiting the required number of suitable subjects within the agreed recruitment period. For example, how many potential subjects do you have access to? What percentage of those potential subjects do you need to recruit?*
- *Describe the time that you will devote to conducting and completing the research.*
- *Describe your facilities.*
- *Describe the availability of medical or psychological resources that subjects might need as a result of an anticipated or unanticipated consequence of participation in the research.*
- *Describe your process to ensure that all persons assisting with the research are adequately informed about the protocol, the research procedures, and their duties and functions (e.g., training plans, detailed study notebooks).*

The PI has previously performed surveys to broad university groups in needs assessments. They are familiar with the background in education research and has supervised several research studies in similar nature to assessment and surveys regarding experiential education and course assessment. They have worked with the team to review all research protocol and safety to ensure the correctness of the project in-line with human subject research.

All students, researchers, faculty, and staff at Virginia Tech in are eligible for the study. There are 38,000 potential subjects who will have access to our research study. Of the 38,000 potential subjects, we will primarily market to the biomed/health listservs and based on previous workshops, plan around 4-7 workshops for a total of 100-150 participants.

The overall timeline of conducting and completing the research will be on a rolling basis as workshops are planned during the 2020-2021 Academic Year (Summer 2020, pending IRB approval)

Despite very minimal risk involved with the study, students have access to university medical or psychological resources.

All persons assisting with the research have obtained IRB certificated of completion at Virginia Tech. They have completed and passed the certification assessment and received a certificate of approval.



## **27.0 Multi-Site Research**

*Contact the HRPP for multi-site research (involving multiple institutions) and the details required for this section will be provided. Otherwise, indicate N/A.*

N/A